

Проектиране на бизнес референтни архитектури насочени към използване на NoSQL бази от данни

Автореферат на дисертационен труд за придобиване на научна и
образователна степен "доктор"

Докторант: Мариана Бойкова Ковачева

Научен ръководител: проф. д.ик.н. В. Кисимов

Катедра „Информационни технологии и комуникации“

Факултет „Приложна Информатика и Статистика“

Университет за Национално и Световно Стопанство

София 2023

Дисертационният труд е обсъден и допуснат до публична защита на редовно заседание на катедрен съвет на катедра "Информационни Технологии и Комуникации" на Университета за Национално и Световно Стопанство - София (УНСС), включващо членове на катедрата, състояло се на 30 март 2023 г. Авторът на дисертационния труд е задочен докторант в катедрата.

Публичната защита на дисертационния труд е насрочена за **27 юни 2023 г.**, от **15:00ч.**

Материалите по защитата са публикувани на Интернет страницата на Университета за Национално и Световно Стопанство - София (<https://www.unwe.bg>) и са на разположение на интересуващите се в сектор "Научни съвети и конкурси" към отдел "Наука" на УНСС.

Съдържание

Обща характеристика на дисертационния труд.....	5
1. Актуалност	5
2. Обект, цели и задачи на дисертационния труд.....	6
3. Работни хипотези на дисертационния труд.....	7
4. Приноси на дисертационния труд.....	8
4.1. Научни	8
4.2. Научно-приложни	8
5. Обем и структура на дисертационния труд	8
Кратко изложение на дисертационния труд	10
1. Въведение.....	10
2. Глава I. Анализ на финансови услуги и генерираните от тях Големи данни. Възможности за използване на NoSQL бази от данни	10
Финансови услуги – дефиниция и видове.....	10
Големи данни – характеристики и видове.....	14
NoSQL бази от данни – дефиниция, типове, характеристики	15
Системи за Големи данни – дефиниция, преимущества, недостатъци, архитектура.....	16
Референтна архитектура – дефиниция, видове и подход за създаване	17
Заключение	19
3. Глава II. Метод за избор и проектиране на бизнес референтна архитектура за NoSQL бази от данни за финансови услуги	20
Метод за избор на NoSQL база от данни за финансови услуги	20
Финансови услуги в резултат на практически разработените към момента бизнес приложения	20
Дефиниране на метод за анализ и оценка на референтна архитектура.....	30
Проектиране на бизнес референтна архитектура за финансови услуги с NoSQL бази от данни.....	39
Подход за използване на проектирана РА водеща към създаване на ICT Архитектура.....	45
Заключение	46
4. Глава III. Инструмент за избор на NoSQL бази от данни за финансови услуги.....	47
Потребителска роля Потребител/Компания.....	47
Потребителска роля Администратор/Експерт	51

Изводи.....	56
5. Глава IV. Приложение на метод за проектиране на референтна архитектура за финансови услуги	57
Архитектура за изпълнение на финансови услуги с NoSQL бази от данни	57
Инсталация и конфигурация на необходими компоненти за реализиране на прототипа на референтната архитектура за финансови услуги	58
Приложение на типични неструктурирани данни при финансови услуги	61
Тестване на референтна архитектура с финансови данни	61
Изводи.....	65
Заключение	65
Списък на публикациите по темата на дисертационния труд	67
- Публикувани.....	67
- Очаква се публикуване	67
6. Литература	68

Обща характеристика на дисертационния труд

1. Актуалност

В настоящия момент, когато промените и дигиталната трансформация са необходимост за бизнеса, за да остане конкурентноспособен, компании, които не успяват да се справят с промяната, рискуват да останат на заден план или да прекратят своето съществуване. Технологиите, развиващи се с по-бързи темпове от обикновени телефони през сензори, събиращи данни и използващи се почти навсякъде в ежедневието ни до компютри, роботика, Изкуствен интелект, облачни технологии, Големи данни, Интернет на нещата и др. се считат за основата на дигиталната трансформация в различни сфери.

Дигитализацията и дигиталната трансформация обхващат различни видове процеси в живота ни, като прогресират с изключително бързи темпове. Адаптирането на нови технологии помага за повишаването на ефикасността и продуктивността при дигитализацията на ръчните процеси. Има изключително много ползи от дигиталната трансформация, но един от най-важните фактори е, че помага на бизнеса да стане по-ефективен и организиран. Този процес има потенциала да направи организациите по-гъвкави, ефективни и клиентско-ориентирани.

В последното десетилетие управлението на експоненциално нарастващите количества данни става все по-предизвикателна задача. С развитието на технологиите, дигитализацията на процесите и дигиталната трансформация на бизнесите води до генерирането на огромни количества данни, които трябва да бъдат съхранявани, управлявани и анализирани, чрез подходящи методи. Развитието на технологиите и генерирането на данните представляват първоначалната и най-лесна стъпка. Предизвикателствата започват със съхраняването, обработката, анализиранието и извличането на резултати от събраните данни.

В миналото, когато е поставено начало на развитието на информационните технологии, основният тип събирани данни са били структурираните. С дигитализацията на все повече и повече бизнеси и дигитално трансформиране, започна и генерирането на различни типове данни – полуструктуриране и неструктурирани. За разлика от структурираните данни, другите два типа представляват изключително голямо предизвикателство за съхранение и обработка. Обикновените релационни бази от данни, използвани за съхранението на структурирани данни не са подходящи за полуструктурираните и неструктурираните, което води до необходимостта от създаването и развитието на нови начини за съхранение на

различните видове данни. Едно от ключовите изисквания за съхранението на Големи данни е, че трябва да се справя с огромни количества данни, като възможностите за съхранение продължават да се увеличават без да бъде прекъсван работния процес.

През последните години, финансовият сектор се развива и се дигитализира все повече, което води до нуждата от промени. За да могат финансовите институции да отговорят на изискванията, нуждите и очакванията на своите клиенти е от изключителна важност да се правят промени и да се адаптират бързо към развиващите се технологии. В допълнение на това, този тип институции трябва да управляват изключително големи обеми от данни докато правят промени на техните системи, без да причинят загубата им или проблеми за клиенти си. Финансовите услуги предлагани от банки, кредитни дружества, счетоводни компании, застрахователни такива, инвестиционни фондове, стокови борси и други, се нуждаят от бази данни, които да могат да се адаптират към нуждите за автоматизиране, на които нерелационните бази могат да отговорят. NoSQL базите могат да се справят със съхранението и обработката на големи количества от данни, със своята скалируемост и предоставянето на по-добра производителност при обработката на данни.

1. Обект, цели и задачи на дисертационния труд

Обект на настоящата дисертация са бизнес референтните архитектури насочени към използването на нерелационни (NoSQL) бази от данни свързани със системи за Големи данни насочени към финансовите услуги.

Целта на дисертационния труд е проектирането на бизнес референтна архитектура насочена към използването на NoSQL бази от данни свързани със системи за Големи данни, която да бъде приложена и тествана с данни за конкретна финансова услуга и дефинирането на метод за оценка и избор на NoSQL бази от данни за финансови услуги.

Предмет на дисертационния труд е проектиране на референтна архитектура за избор на NoSQL бази от данни свързани със системи за Големи данни, насочени към финансови услуги и методът за избор и оценка на нерелационните бази от данни за финансови услуги.

Изследователският въпрос поставен в настоящия дисертационен труд е „Как се създава бизнес референтна архитектура насочена към използването на NoSQL бази от данни свързани със системи за Големи данни в сферата на финансовите услуги, по какъв начин може да се получи количествена оценка, чрез която да изберем подходяща нерелационна база от данни за референтната архитектура? “

Задачи, които се изпълняват в процеса на реализация на поставената цел на настоящия дисертационен труд са следните:

- **Изследване и анализ на референтна архитектура за финансови услуги.**
 - ✓ Обзор и анализ на актуалните проблеми, свързани със съхранението на различните видове Големи данни и референтни архитектури, свързани с използването на нерелационни бази от данни насочени към финансовите услуги.
- **Дефиниране на критерии за избор на NoSQL база от данни свързани със системи за Големи данни насочени към използването на финансови услуги**
 - ✓ Дефиниране на подход за количествена оценка за избор на нерелационна база от данни в сферата на финансовите услуги.
 - ✓ Избор на критерии за оценка на NoSQL бази в сферата на финансовите услуги.
 - ✓ Проектиране на концептуален, логически и физически модел на референтна архитектура за финансови услуги с NoSQL бази от данни
- **Разработка на инструмент за оценка и избор на NoSQL бази от данни, свързани със системи за Големи данни насочени към финансови услуги за използването ѝ в референтната архитектура**
 - ✓ Проектиране на инструмент за оценка и избор на нерелационна база от данни свързани със системи за Големи данни насочена към финансови услуги.
 - ✓ Инструмент за избор и оценка на NoSQL бази от данни за финансови услуги.
- **Практическо приложение на метода за проектиране на референтни архитектури за финансови услуги**
 - ✓ Създаване на експериментален прототип на референтна архитектура свързана със система за Големи данни насочени към финансови услуги
 - ✓ Практическо приложение на експерименталния прототип
 - ✓ Оценка на резултатите от практическото приложение

2. Работни хипотези на дисертационния труд

Хипотеза 1:

Възможно е да се създаде референтна архитектура, приложима за съхранение, обработка и работа с Големи данни свързани с финансови услуги.

Хипотеза 2:

Възможно е да се даде количествена оценка за референтна архитектура, като се създаде метод за оценка, както и инструмент за автоматизирането и прилагане.

Хипотеза 3:

Възможно е да се постигне ефективна интеграция на компонентите на референтната архитектура за финансови услуги с NoSQL бази от данни за работа с Големи данни.

3. Приноси на дисертационния труд

Научни

- Създаден е метод за количествена оценка на NoSQL бази данни за внедряване на цифрови финансови услуги
- Формулиране на индикатори и измерители за избор на NoSQL бази от данни за финансови услуги
- Проектиране на концептуален, логически и физически модел на референтна архитектура за финансови услуги с NoSQL бази от данни

Научно-приложни

- Принципи за проектиране и създаване на инструмент за избор на NoSQL бази от данни за финансови услуги
- Структурна и интеграционна трансформация на бизнес референтна архитектура към детайлна архитектура за използване на NoSQL бази от данни

4. Обем и структура на дисертационния труд

Дисертационният труд се състои от въведение, четири глави, заключение, приноси, публикации, литература, приложения, списък с фигури, списък с таблици, списък с графики и приложения. Той е в общ обем от 122 страници (110 страници без приложенията). Използваните източници са 85 броя, от тях 1 на български език, 84 на английски език и 73 интернет адреса. В изложението присъстват 45 фигури, 5 таблици и 1 графика.

Въведението представя актуалността, обекта, целите, задачите и работните хипотези на проблемите разгледани в дисертационния труд. Формулирани са три хипотези, а като негова цел е поставено „проектирането на бизнес референтна архитектура насочена към използването на NoSQL бази от данни свързани със системи

за Големи данни, която да бъде приложена и тествана с данни за конкретна финансова услуга и дефинирането на метод за оценка и избор на NoSQL бази от данни за финансови услуги.“

Първа глава е с обзорен характер. В нея се въвеждат основните понятия и дефиниции използвани в дисертационния труд. Направен е кратък анализ на Hadoop, неговите основни преимущества и недостатъци. Разгледани са типовете данни, както видовете NoSQL бази от данни и е направен анализ на референтните архитектури като цяло и е конкретизирано в видовете такива за финансови услуги.

Втора глава представя метод за оценка и избор на NoSQL бази от данни за финансови услуги. Той е базиран върху анализ на литературни източници, както и на практически реализирани решения. Изведени са критерии и е направена експертна оценка за осъществяването на избора. Проектиран е концептуален, логически и физически модел на референтна архитектура за финансови услуги

В трета глава е проектиран и създаден инструмент за оценка и избор на NoSQL бази от данни за финансови услуги, базиран върху метода за оценка и избор на нерелационни бази от данни създаден в Глава II.

Четвърта глава представлява експерименталната част на дисертационния труд. Разглежда практическата приложимост на метода за оценка и избор на NoSQL, както и реализирания физически модел на проектираната референтна архитектура, който е тестван с финансови данни.

В заключението се обобщава, че като резултат от разработките, формулираните хипотези са доказани, като се изразява намерение да се продължат изследванията и разработките в областта на нерелационните бази от данни, както развитието на инструмента за оценка и избор на NoSQL бази от данни, както и продължаване на практическата работа по реализираната бизнес референтна архитектура.

Кратко изложение на дисертационния труд

1. Въведение

Въведението представя актуалността, обекта, целите, задачите и работните хипотези на проблемите разгледани в дисертационния труд. Формулирани са три хипотези, а като негова цел е поставено „проектирането на бизнес референтна архитектура насочена към използването на NoSQL бази от данни свързани със системи за Големи данни, която да бъде приложена и тествана с данни за конкретна финансова услуга и дефинирането на метод за оценка и избор на NoSQL бази от данни за финансови услуги.“

1. Глава I. Анализ на финансови услуги и генерираните от тях Големи данни.

Възможности за използване на NoSQL бази от данни

Финансови услуги – дефиниция и видове

Финансовите услуги са икономически услуги, които са предоставяни от финансовата индустрия, която обхваща широк кръг от бизнеси, занимаващи се с управление на пари включително кредитни съюзи, банки, компании за кредитни карти, застрахователни компании, счетоводни компании, потребителско финансиране, борсови посредници, инвестиционни фондове, правителствено спонсорирани предприятия. [1]

Финансовите услуги се делят на банкови, обменни, инвестиционни, осигурителни, финансов износ и други финансови услуги. Всяка една от тези основни финансови услуги се разделя на още подуслуги.

- ✓ **Банките** - данните за анализ в банковия сектор идват от различни източници: личната информация на клиентите, детайлите на акаунтите, транзакции на клиенти, клиентски оплаквания и обслужване, информационните страници в социалните мрежи, настройки на пазара, представяне на продукти и др. С помощта на Големите данни банките могат да преодолеят големи бизнес предизвикателства, като доходност и представяне чрез анализ на Големи данни, като също помага на банките да намалят риска и разходите при придобиване на клиенти, предвиждане на риска при заеми и ипотечи. Потенциалният анализ на Големи данни в банковата индустрия би довел до предотвратяването на измами, идентифицирането и придобиването на клиенти, задържането на клиенти, подобряването на клиентското

изживяване, оптимизиране на операции, подобряване на продуктов дизайн, увеличаване на прозрачността, спазване на регулаторни изисквания. [2]

- ✓ **Обменни услуги/в чужбина/** се предлагат от банки и специални валутни брокери. Чуждестранните обменни услуги генерират изключително големи обеми от данни – структурирани, полуструктурирани и неструктурирани, като информацията, която остава скрита вътре би могла да бъде полезна за участниците на пазара. [3]
- ✓ **Инвестиционни услуги** са широк кръг от активности свързани с инвестицията в различни финансови инструменти. За да се намали риска свързан с инвестициите, инвеститорите използват различни технологични решения базирани върху Големи данни, за да могат да вземат правилните решения. Използвайки този тип инструменти, могат да взимат информирани инвестиционни решения, чрез събирането на достъпна информация, идентифицирането на тенденциите в индустрията и правилното управление на активите. Освен стандартна структурирана информация, при инвестиционните услуги може лесно да се съберат аудио файлове, които са неструктурирани данни. Този тип данни могат да помогнат за по-бързото създаване на отчети, както и за извличането на най-важната информация за заинтересованите лица. [4] [5]
- ✓ **Застрахователни услуги** представляват договор, представен от полица, между две страни, в който притежателят на полица получава финансова защита или възстановяване срещу загуби от застрахователната компания. За изключително продължителен период от време, застрахователните услуги се базират върху традиционните бизнес модели, но със случващата се дигитална трансформация и развитие на технологиите, индустрията започна да използва прогресивни технологии и дигитални платформи, които позволяват на застрахователните компании да проследяват, измерват и контролират риска. Компаниите за застраховане на пътувания, здравни застраховки и други, разчитат на статистиката, за да сегментират добре своите клиенти. Статистика за инциденти, информация за застрахованите, както и данни събрани от трети страни, помагат за групирането на хората в различни

рискови категории, за предотвратяването на измами и оптимизирането на разходи. Събраните структурирани, полуструктурирани и неструктурирани данни могат да се използват и анализират за ценовото оформление, рейтинга, маркетинга и др. [6] [7]

- ✓ **Финансовият износ** е финансова услуга предоставена от местна фирма към чуждестранна такава или индивид.
- ✓ **Други финансови услуги** [8] [9]

Финансовата индустрия използва повече системи с релационни бази данни, настъпващата дигитализация и дигитална трансформация ги принуждава да трансформират своята работа, за да посрещнат нуждите на своите клиенти. Повечето компании във финансовата индустрия съхраняват Големи данни и в момента, но не знаят как да извлекат техния потенциал, тъй като има полуструктурирани и неструктурирани такива. На първо място организациите трябва да разберат стойността на събраните данни и какво означават те за клиентите и бизнес процесите. Поради факта, че наследствените системи (legacy systems) не могат да поддържат полуструктурирани и неструктурирани данни, трябва да се търсят нови решения на този проблем. Големите данни във финансовата сфера през последните години доведоха до големи иновации в технологиите, като позволяват създаването на удобни, персонализирани и сигурни решения. Търговията на акции и облигации, инвестиции, засичането и предотвратяването на измами и анализ на риска са само част от финансовите услуги, които се възползват от Големите данни. [10]

В миналото Големите данни не бяха присъщи за финансовите институции, като по тази причина е необходим дълъг процес на промяна, който изисква както технологична, така и поведенческа промяна. Наличието на Големи данни, в частност полуструктурирани и неструктурирани, води до необходимостта за намирането на подходящ начин за съхранение, където се намесват нерелационните (NoSQL) бази от данни. [10]

Нерелационните бази от данни предлагат не само гъвкавост за приложенията, които са разработени върху тях, но и мащабируемост на ниска цена, нещо, което е невъзможно при релационните бази от данни. Необходимостта за обработката и анализ в реално време за автентикация, авторизация, засичане на опити за измама води до нуждата от NoSQL базите.

- ✓ **Подобрено предотвратяване на измами и мониторинг на престъпления във финансовата сфера.** Това е едно от най-големите предимства на нерелационните бази от данни с анализа, който може да се извърши в реално време и е от изключителна важност за финансовите услуги – банки, кредитни и дебитни карти. Всички тези институции работят с изключително чувствителна информация, която е от изключителна важност за разбирането и удовлетворяването на клиентите. Измамите във финансовите услуги се увеличават, което може да се предотврати, чрез анализ, анализиране на аномалии, класификация и клъстериране, което може да се реализира с нерелационните бази от данни. . [12]
- ✓ **Дигитална авторизация в реално време.** От изключителна важност за финансовите услуги, за да може да е сигурно, че няма да се случи измама и за реалната идентификация на хората, с които се комуникира. NoSQL базите от данни имат такива възможности поради намалената латентност на четене, като авторизацията може да се случи почти веднага. [11][12]
- ✓ **Персонализирани оферти.** За клиентите е важно да получават персонализирани оферти, които да отговарят на техните нужди. Това не е по-различно във финансовите услуги – осигурителни полици базирани на персоналния риск, лихвени проценти на база погасяване и капацитет и др. За реализирането и представянето на персонализираните оферти на клиентите е необходимо да се събере информация – поведение, ангажираност в социалните медии, бисквитки, онлайн пазаруване и др., като всички данни идват от различни източници, като за тази цел на помощ идват нерелационните бази от данни. NoSQL базите помагат за бърз анализ в реално време, за да може клиентите да бъдат снабдени с подходящи персонални оферти. [12] [13]
- ✓ **Оказване на съдействие преди търговия.** Проучванията сочат, че информацията събрана от различни източници, като данните могат да бъдат както структурирани, така и полуструктурирани и неструктурирани, а по този начин може да се направи оценка на риска и могат да бъдат взети решения свързани със закупуването на даден продукт. [11]

- ✓ **Целеви маркетинг.** Събираната информация и данни за търсене, купуване, предлагане е необходима за извършването на анализ в реално време и да предложи на клиентите правилните реклами. Целевият маркетинг също е от изключителна важност, когато става дума за финансови услуги, защото когато данните са анализирани правилно, това дава на компаниите знания за нуждите, изискванията и очакванията на клиентите, и определя кой е перфектния продукт за предлагане, който в последствие да прерасне в продажба. [11] [13]

Големи данни – характеристики и видове

Големи данни (Big Data) е терминът, който описва данните, които са толкова големи и сложни, че е трудно или в някои случаи невъзможно да се обработват с помощта на традиционни методи. Събирането на такива огромни количества данни е полезно за извършване на по-добри и по-широки анализи на дадени проблеми, които са важни за компаниите. [14] Големите данни са такива големи обеми, които не могат да бъдат обработени от традиционните методи, като релационните бази от данни. Основните характеристики, с които те разполагат са петте V [15] [16]:

- ✓ **Разнообразие (Variety)** от данни и източници, от които биват събирани. В днешно време източниците, от които могат да се събират данни са имейли, изображения, ПДФ файлове, аудио файлове, видео и др., като данните могат да бъдат структурирани, полуструктурирани и неструктурирани в зависимост от източника, от който идват.
- ✓ **Обем (Volume)** - изключително големи количества, защото всички системи, устройства, социални мрежи и процеси генерират данни в големи количества, не само ежедневни, ежечасно, но и всяка секунда.
- ✓ **Скорост (Velocity)** - генерирането на Големи данни се случва с голяма скорост. Тази характеристика дефинира скоростта, с която данните се генерират в реално време и скоростта, с която се обработват.
- ✓ **Стойност (Value)** - данните трябва да бъдат важни и ценни. За да бъдат събирани, съхранявани, обработвани и анализирани по конкретен начин е важно да бъдат полезни за бизнеса.

- ✓ **Достоверност (Veracity)** - данните трябва да бъдат точни и надеждни, като това може да бъде доказано с източника на данните.

В зависимост от източниците, от които се събират данните има 3 типа – структурирани, полуструктурирани и неструктурирани – Таблица 1. [17] [18] [19]

Видове данни	Описание	Примери
Структурирани	Предефиниран табличен модел, с връзки между редовете и колоните.	счетоводни транзакции, детайли за адреси, демографска информация, оценка от клиенти, локации от различни устройства, информация събирана от сензори и умни устройства.
Полуструктурирани	Този тип данни има дефинирани постоянни характеристики, но няма точна структура, както се очаква от релационните бази данни.	Имейли
Неструктурирани	Съхранението в табличен вид е невъзможно и не може да се създаде предефиниран модел. Поради липсата на структура търсенето, управлението и анализа на неструктурирани данни е изключително труден.	Изображения, видео, аудио, ПДФ файлове, съдържание от социални мрежи

Таблица 1 Видове данни, описание и примери

NoSQL бази от данни – дефиниция, типове, характеристики

Има четири типа NoSQL бази данни, които могат да разрешават различни проблеми въз основа на нуждите на компания. NoSQL базите от данни са нерелационни бази, които използват ключ-стойност за съхранението, достъпването и извличането на данни вместо традиционния начин използващ таблици с колони и редове. Нерелационните бази от данни не се нуждаят от предварително дефинирана схема, като по този начин могат да съхраняват различни видове данни – Таблица 2. [20]

Видове бази	Описание	Примери
Документно-ориентирани	Документно-ориентираните бази съхраняват информацията в документи. Този тип бази имат гъвкава схема, която позволява модела на данни да се развива заедно с приложението. Има възможност за хоризонтална мащабируемост. [21]	MongoDB, CouchBase, OrientDB
Ключ стойност	Ключ-стойност базите данни използват най-обикновени ключ-стойност методи за съхранение на данни. Този тип съхранение е оптимизиран за четене и записване на данни. Ключ-стойност базите имат добра хоризонтална мащабируемост. [22]	Amazon DynamoDB, Redis
Широко-колонни бази	Широко-колонните хранилища на данни организират данните по такъв начин, че позволяват по-ефективното съхранение и по-бързо изпълнение на заявки. Този тип бази се разделят лесно, което позволява разпределението на големи сетове от данни в множество възли за висока ефективност и ниска латентност. [23]	HBase, Cassandra
Граф бази	Граф базите съхраняват възли и връзки вместо таблици или документи. Данните се съхраняват ограничение в предефиниран модел позволявайки голяма гъвкавост. [24]	Neo4j, OrientDB

Таблица 2 Видове бази, дефиниция и примери

С дигитализацията на процесите, генерирането на огромни обеми от различни типове данни се разви и необходимостта от нерелационните бази от данни, тъй като релационните такива нямат необходимите параметри за съхраняването на полуструктурирани и неструктурирани данни. Основните характеристики на нерелационните бази от данни, които ще разгледаме са: скалируемост, гъвкавост, липса на необходимост от предефинирана схема, разпределена обработка и издръжливост. [25]

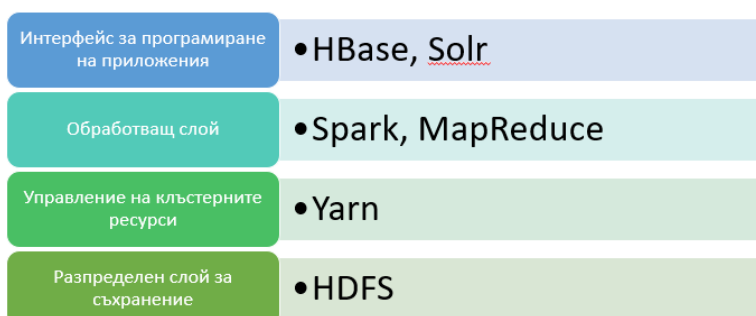
Системи за Големи данни – дефиниция, преимущества, недостатъци, архитектура

Hadoop позволява разпределена обработка на големи сетове от данни разпределени в множество клъстъри от компютри, използвайки прости програмни модели. Apache Hadoop е проектиран, за да се разширява от единични до множество сървърни машини, като всяка

една от тях предлага локални изчисления и място за съхранение. Реализиран по този начин, Hadoop представлява рентабилно решение за съхранение и обработка на големи количества структурирани, полуструктурирани и неструктурирани данни без изисквания за формат. [26] [27] Преимуществовата на Apache Hadoop са следните: цена, мащабируемост, гъвкавост, скорост, устойчивост на грешки, висока производителност и минимален трафик в мрежата. Недостатъците на Apache Hadoop са следните – проблем с малки файлове, уязвимост, сигурност, пакетна обработка, обработка на високо и ниска производителност с обработка на малки файлове.

Apache Hadoop е с топология на главен - подчинен(master-slave) и позволява съхранението на големи обеми от данни в система от възли. При този вид топология, работата на главния възел е да разпредели конкретна задача на различни подчинени възли и да разпредели ресурсите, като реалната работа и изчисления се извършват от подчинените възли. Архитектурата на Hadoop позволява паралелна обработка, която използва няколко компонента: Разпределен слой за съхранение - HDFS(Hadoop Distributed File System), Управление на клъстерни ресурси - Hadoop Yarn, Обработващ слой – MapReduce, Apache Spark и интерфейс за програмиране на приложения – HBase, Solr, показани на Фигура 1 долу. [28] [29]

Слоевете на архитектура на Hadoop



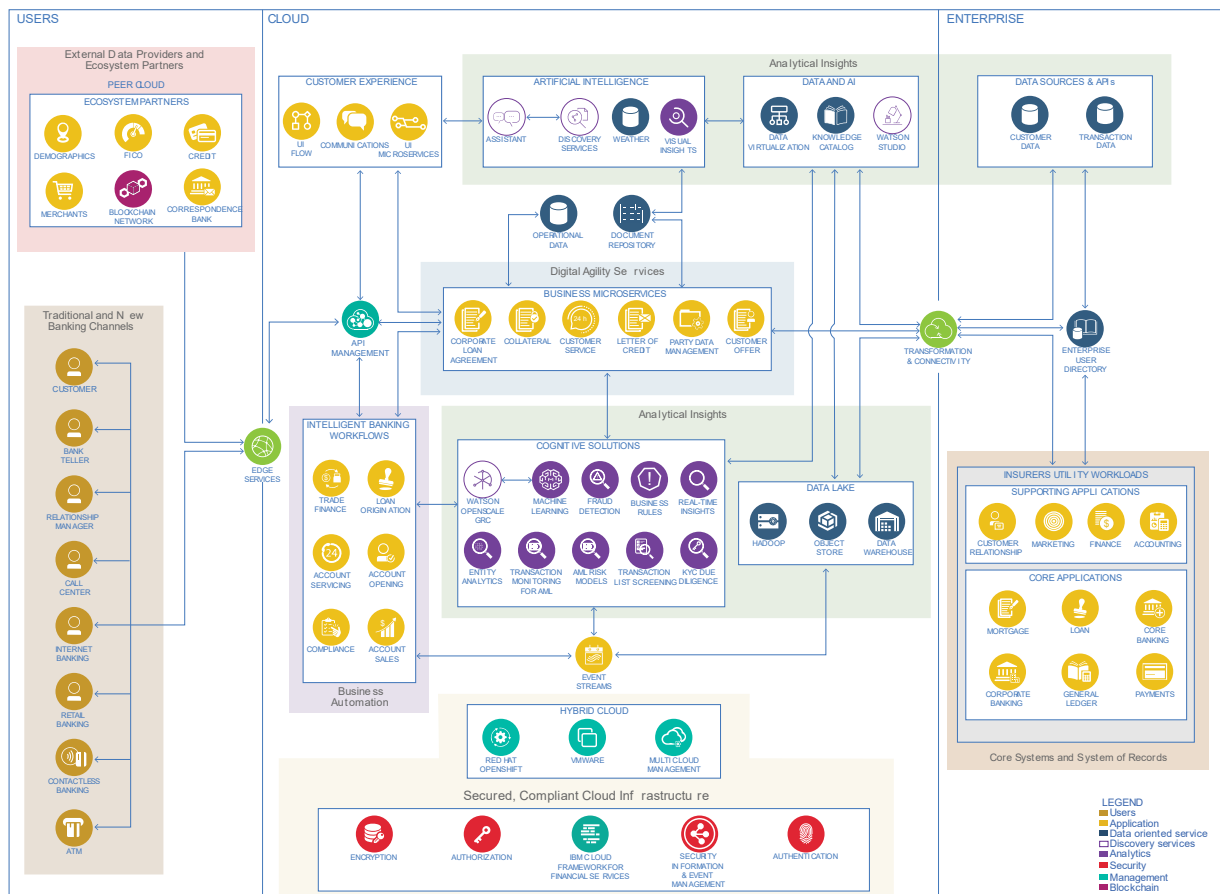
Фигура 1 Слоевете на архитектурата на Hadoop

Референтна архитектура – дефиниция, видове и подход за създаване

Референтната архитектура е документ или съвкупност от документи, които предоставят препоръчана структура и интеграции на ИТ продукти, които да формират решение. РА обединява в себе си най-добрите практики в индустрията, като обикновено предлага оптималния метод за специфични технологии. Референтната архитектура предлага най-

добрите ИТ практики в лесно разбираем формат, който направлява прилагането и използването на сложни технологични решения. [30]

Използването на добри практики при създаването на референтни архитектури подобрява ефикасността, отговаря на нормативните изисквания и намалява шанса за грешки. Референтната архитектура помага при вземането на решения за избор на най-добрия модел и начин за създаване на софтуерна архитектура, за да се срещнат бизнес целите. [31]



Фигура 2 Референтна архитектура за банкови услуги на IBM. Източник: [32]

Референтната архитектура за банковата индустрия използва методи и модели за индустриални стандарти, които да допринесат за постигане на по-голямо съответствие между ИТ и индустрията. Финансовите институции могат да следват консултативен метод, за да идентифицират възможностите за подобрене и иновации, като използват Мрежата за архитектурата на банковата индустрия (Banking Industry Architecture Network - BIAN),

Информационна рамка на IBM (IBM Information Framework – IFW) и IBM Компоненти за бизнес модели (IBM Component Business Models - CBM) за финансови услуги. Тези активи предоставят специфични за индустрията рамки, които могат да бъдат използвани за цялостната модернизация на банковата програма – Фигура 2. [32]

Създаването на референтна архитектура е изключително предизвикателна задача при липсата на предварително зададени стъпки или процес за създаването ѝ. Има пет основни стъпки, които трябва да се изпълнят: идентифициране на целта, формулиране на принципи, залагане на технически правила и стандарти, построяване на правила и стандарти, и прилагане на контекст. [31]

Заключение

Дигитализацията и дигиталната трансформация доведоха до много промени във всички сфери, а автоматизацията на процесите и използването на нови технологии като Интернет на нещата, Изкуствен интелект, сензори и други, доведоха до генерирането на огромни количества Големи данни от различни типове – структурирани, полуструктурирани и неструктурирани, а това от своя страна носи необходимостта за нови начини на съхранение и обработка, тъй като релационните бази от данни не могат да се справят с полуструктурираните и неструктурираните данни.

Финансовите услуги от своя страна са базирани основно на системи с релационни бази от данни, което носи необходимостта от промяна. Този тип услуги се делят на банкови, обменни, инвестиционни, застрахователни, финансов износ и други финансови услуги, като с навлизането на новите технологии идват и Големите данни и необходимостта да се съхраняват в подходящи места, които да могат да се справят със съхранението и последващите действия с полуструктурираните и неструктурирани данни, където се намесват NoSQL базите от данни и системите за Големи данни.

От представените по-горе финансови услуги и видове данни, които се генерират от тях, начините на съхранение към момента, както и наличните референтни архитектури за такъв тип услуги, се открива възможността за създаването на метод оценка и избор на NoSQL бази от данни за финансови услуги. Методът, на първо място може да бъде приложен практически, чрез създаването на инструмент базиран върху критериите и оценките, както

и да се създаде практическа реализация, чрез експериментален прототип на референтна архитектура свързана със система за Големи данни насочена към финансови услуги.

2. Глава II. Метод за избор и проектиране на бизнес референтна архитектура за NoSQL бази от данни за финансови услуги

Метод за избор на NoSQL база от данни за финансови услуги

Всеки вид NoSQL бази от данни има своите предимства и недостатъци за специфични ситуации. В тази част на научното изследване ще бъдат разгледани практически приложени примери на нерелационни бази от данни във финансовата сфера от реални компании, от които в последствие ще бъдат изведени критерии и оценки, които ще могат да бъдат използване за избор на NoSQL бази и за подходяща референтна архитектура за конкретните нужди на компанията.

Финансови услуги в резултат на практически разработените към момента бизнес приложения

В следващите точки ще бъдат разгледани вече съществуващите финансови услуги базирани върху различните нерелационни бази от данни, заедно с предимствата, поради, които могат да се използват тези конкретни бази, като след това тези финансови услуги ще бъдат добавени към разработения метод.

✓ Документно-ориентирани бази

a. MongoDB

MongoDB е нерелационна база от данни без схема, която предоставя висока производителност, като поддържа MapReduce, инструменти за акумулиране, хоризонтална скалируемост и вградена репликация. Базата поддържа специализиран конектор с Hadoop. MongoDB предоставя анализиране в реално време, като данните могат да идват от различни източници, което спомага за анализа на нуждите и изискванията на клиентите. MongoDB могат да се създават приложения, които са били невъзможни за създаване с релационни бази от данни. Тази база допринася с технологична основа, която позволява: [33]

- ✓ Съхраняването на данни в гъвкави JSON документи, което на практика означава, че полетата, които се съхраняват могат да варират от документ до документ, както и, че структурите от данни могат да бъдат променяни.

- ✓ Документният модел, който картографира до обектите в кода на приложението прави данните лесни за работа.
- ✓ Разпределена база с висока достъпност, хоризонтална скалируемост и географска разпределеност са вградени и лесни за използване.

Базирането на системи върху MongoDB като база от данни допринася за: модернизирани наследствените системи (legacy systems), защитаването на информацията на клиентите, Постигане на висока достъпност и скалируемост

MongoDB предоставя решения за следните финансови услуги – Фигура 3:



Фигура 3 Финансови услуги базирани върху MongoDB

- ✓ **Финансови институции базирани своите системи върху MongoDB**

Три големи финансови институции, които базират своите системи върху MongoDB – Royal Bank of Scotland, ANL – голяма инвестиционна компания и Bendigo and Adelaide Bank.

b. Couchbase

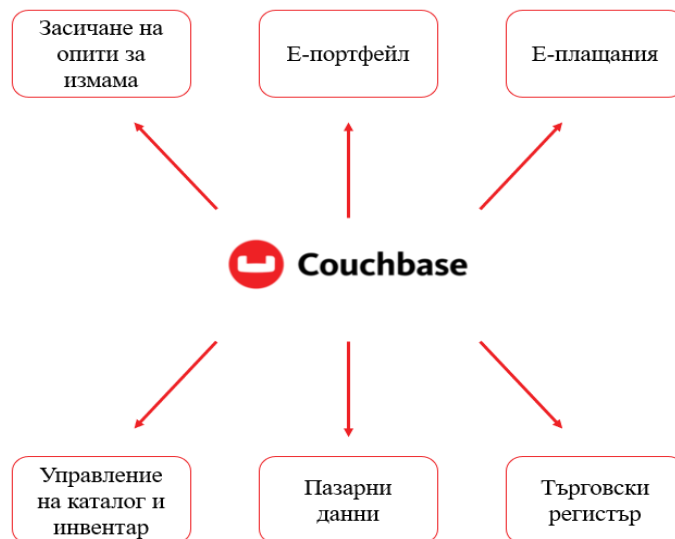
Couchbase предлага скалируемост с много измерения, което означава, че мащабирането може да се случи лесно дори, докато базата се използва. По този начин производителността на приложението се повишава, а цената се намалява. Скалируемостта с много измерения разделя, изолира и мащабира услуги на базата –

индекс, заявка и данни. Това е начинът, по който ресурсите могат да бъдат оптимизирани без да пострада работата и представянето на системата. Използването на Couchbase за анализ в реално време е изключително бърз и е важен за предотвратяването на измами, реализирането на търговски анализи и други. [34]

Базирането на финансови услуги върху Couchbase се случва поради:

- Скалируемост в много измерения, анализ в реално време, отговор на системата за милисекунди, репликация на кръстосани центрове за данни, мощен език за заявки, мобилни финансови услуги на ръба

Couchbase предоставя решения за следните финансови услуги – Фигура 4:



Фигура 4 Финансови услуги базирани върху Couchbase

✓ Финансови институции базирани своите системи върху Couchbase

Част от големите компании във финансовата сфера, базирани своите услуги върху Couchbase са PayPal, FICO, EQUIFAX и REVOLUT.

✓ Ключ-стойност бази

a. Amazon DynamoDB

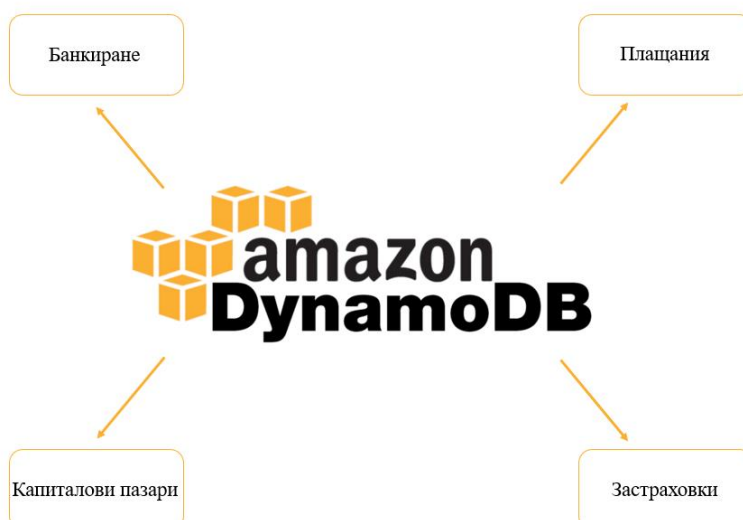
Amazon DynamoDB е нерелационна база от данни, която поддържа хоризонтална скалируемост, има гъвкава схема, което позволява лесно адаптиране, когато изискванията на приложението се променят. DynamoDB предлага

автоматична репликация и мащабируемост. Върху тази NoSQL база са базирани банкови институции, капиталови пазари и други. [35]

Базирането на финансови услуги върху Amazon DynamoDB се случва заради:

- Последователна производителност във всякакъв мащаб, което означава, че няма нужда от преработване на базата с увеличаването на обема от данни.
- DynamoDB помага за подsigуряването на данните с криптиране и непрекъснато архивиране на данните.
- Гъвкава схема, което позволява лесното преработване, когато бизнес изискванията се променят без да се налага предефиниране на цялата схема, както трябва да се случи, ако базата е релационна.

Amazon DynamoDB предоставя решения за следните финансови услуги – Фигура 5:



Фигура 5 Финансови услуги базирани върху Amazon DynamoDB

✓ Финансови институции базирани своите системи върху Amazon DynamoDB

Част от големите компании във финансовата сфера, базирани своите услуги върху Amazon DynamoDB са Raiffeisen Bank и CI Financial.

b. Redis

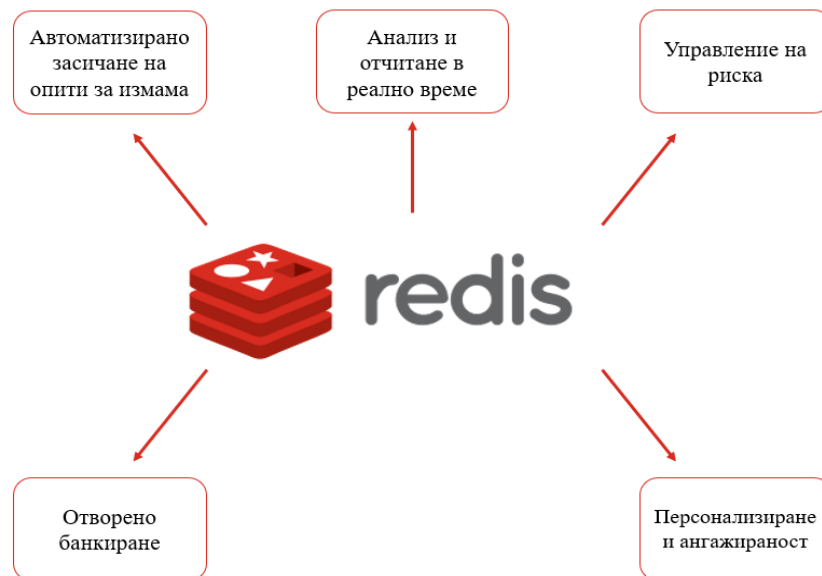
Redis предлага in-memory съхранение, което дава възможност за по-бързи отговори на системата и поддържа милиони операции. Този вид нерелационна база от данни има висока ефективност и добра скалируемост – вертикална и

хоризонтална. Redis е непроменлив по отношение на данните, което означава, че те ще останат налични, дори при повреди на сървъра, като също предлага голямо разнообразие от структури от данни, което дава възможност на разработчиците да изберат най-подходящата възможност за тяхното приложение. Много социални мрежи използват Redis за база от данни, но и много финансови институции започват да се обръщат към този вид NoSQL. [36]

Базирането на финансови услуги върху Redis се случва заради:

- Високо скоростни транзакции, които са от изключително голяма важност за финансовите институции.
- Redis предлага различни модели на данни, които могат лесно да бъдат приложени за различни услуги.
- Redis позволява на приложенията да поддържа едновременни ъпдейти на множество географски локации без да компрометират латентността или ефективността им.
- Има висока толерантност към грешки, устойчивост и висока наличност.

Redis предоставя решения за следните финансови услуги – Фигура 6:



Фигура 6 Финансови услуги базирани върху Redis

✓ **Финансови институции базирани своите системи върху Redis**

Част от големите компании във финансовата сфера, базирани своите услуги върху Redis са Ekata и Deutsche Börse Group.

✓ **Широко-колонни бази**

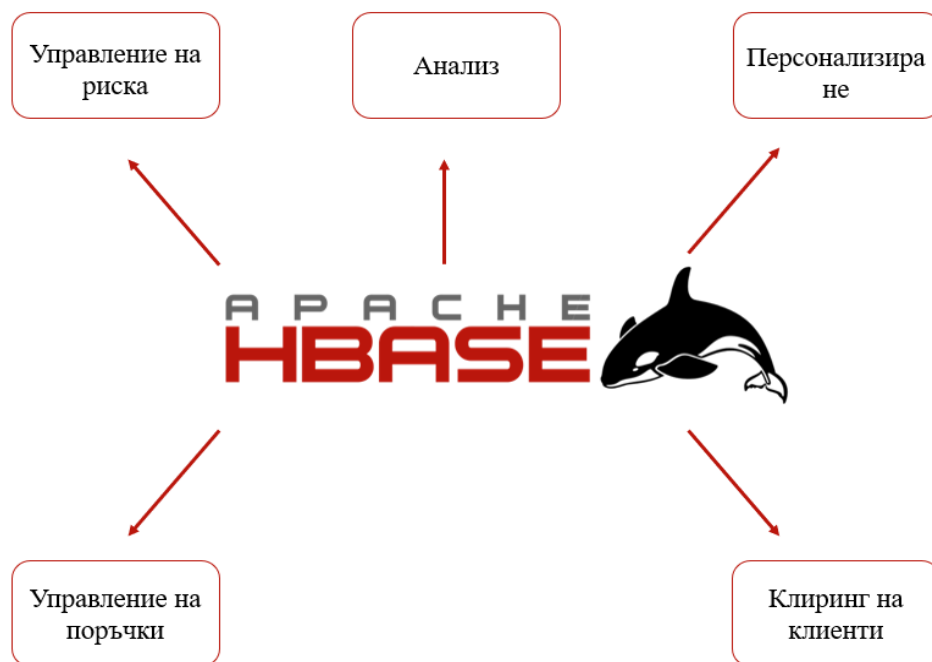
a. HBase

HBase е част от екосистемата на Hadoop. В този тип широко-колонни бази, данните могат да бъдат съхранявани в Hadoop, като най-важните функционалности са, че е скалируема, има автоматична поддръжка при проблем, интегрира се с Hadoop и има репликация на данните в клъстерите. HBase се използва от социалните мрежи, но вече има и финансови услуги, които са я адаптирали към своите системи.

Базирането на финансови услуги върху HBase се случва заради:

- Линейна и модулна мащабируемост. HBase е създаден за скалируемост през хиляди сървъри и управление на достъпа до РВ от данни.
- Бърз достъп и ниска латентност на четене и запис на данни.
- Толерантност към грешки, като данни се съхраняват на множество хостове в клъстера. [37]

HBase предоставя решения за следните финансови услуги – Фигура 7:



Фигура 7 Финансови услуги базирани върху Apache HBase

✓ **Финансови институции базирани своите системи върху Apache HBase**

Част от големите компании във финансовата сфера, базирани своите услуги върху HBase са FINRA и Celer Technologies.

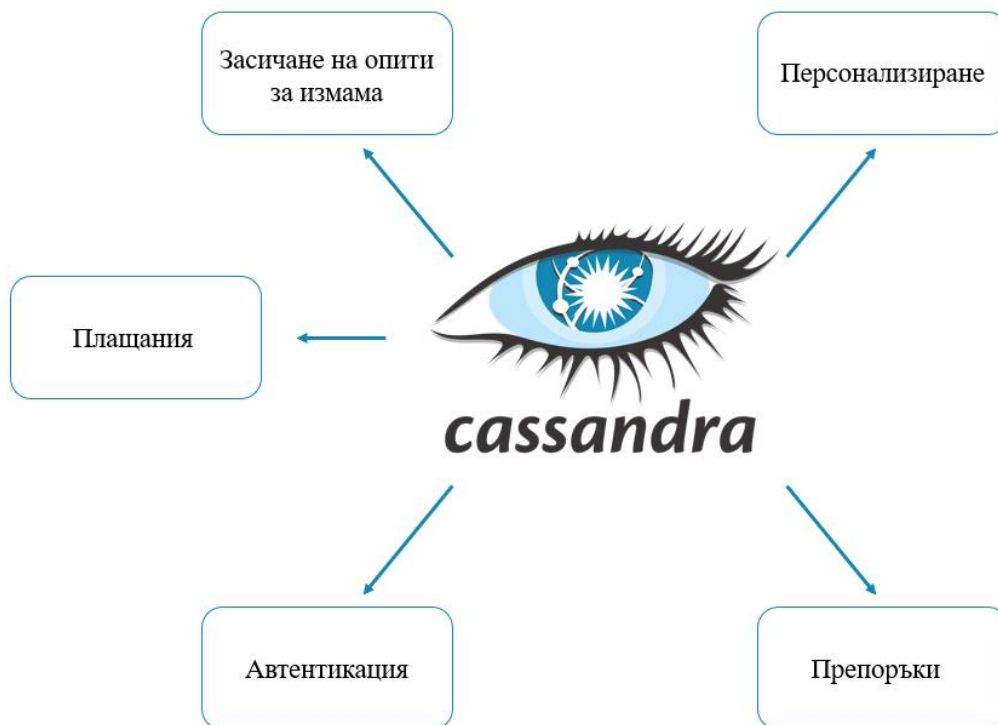
b. Cassandra

Основните функционалности на Cassandra са, че е скалируема, разпределена и изключителна толерантна към грешки, защото данните са автоматично съхранявани и репликирани. Cassandra поддържа MapReduce, както и Apache Hive и Apache Pig. Базата е хоризонтално скалируема и когато компанията има необходимост се добавят повече кълъстърри. [38] [39]

Базирането на финансови услуги върху Cassandra се случва заради:

- Cassandra се справя отлично с големи обеми от данни, които трябва да бъдат записани. Разпределението на данни вътре в кълъстъррите е бързо, записите са евтини, което прави Cassandra изключително добра с големите обеми от данни.
- Cassandra се справя с репликацията на данни, тъй като те са разпределени на множество кълъстърри и в множество центрове за данни. Поради тази причина, ако част от кълъстъррите не работят, или целия център с данни, Cassandra ще се справи със ситуацията.
- Ако се очаква нарастване на данните в реално време, Cassandra има изключително голямо предимство.

Cassandra предоставя решения за следните финансови услуги – Фигура 8:



Фигура 8 Финансови услуги базирани върху Cassandra

✓ **Финансови институции базирани своите системи върху Cassandra**

Част от големите компании във финансовата сфера, базирани своите услуги върху Cassandra са Fractal Labs и ACI Worldwide.

✓ **Граф бази**

а. Neo4J

Neo4j е нерелационна база от данни с гъвкава схема, която използва граф модел. Този модел съдържа възли, които са свързани един с друг. Neo4j поддържа ACID свойства, като също е скалируема и надеждна база. [40]

Базирането на финансови услуги върху Neo4j се случва заради:

- Изключително бързото представяне на заявките в реално време заради граф архитектурата.
- Целостта на данните се запазва благодарение на ACID свойствата, като независимо дали транзакциите са напълно успешни или провалени, данните няма да бъдат повредени.

- Гъвкавостта е още едно предимство на Neo4j. Базата няма нужда от предефинирани таблици, както и, ако бизнес изискванията се променят, моделът на данните може да се адаптира за много кратко време.

Neo4j предоставя решения за следните финансови услуги – Фигура 9:



✓ **Финансови институции базирани своите системи върху Neo4j**

Фигура 9 Финансови услуги базирани върху Neo4j

Част от големите компании във финансовата сфера, базирани своите услуги върху Neo4j са UBS и Global 50 Bank.

b. Orient DB

Основните характеристики на Orient DB са скоростта, като тази NoSQL база от данни може да съхрани до 120 000 записа за секунда. Базата е гъвкава и може да работи, на която и да е платформа без конфигурации и инсталации. Orient DB разполага с еластична линейна скалируемост, както и постепенно нарастващи записи. [41]

Базирането на финансови услуги върху Orient DB се случва заради:

- Orient DB записва и обработва данните с изключително висока скорост, като надминава в пъти релационните бази от данни, които са бавни и скъпи. Orient DB връща отговори на заявки за секунди.
- Този тип база позволява скалируемост, което води до лесно адаптиране на системите при нужда от промени при изискванията.

Orient DB предоставя решения за следните финансови услуги – Фигура 10:



Фигура 10 Финансови услуги базирани върху Orient DB

✓ Финансови институции базирани своите системи върху Orient DB

Част от големите компании във финансовата сфера, базирани своите услуги върху Orient DB е 99Bill.

След задълбочено разглеждане на научна литература, както и на практически приложени решения за финансови услуги базирани върху NoSQL бази от данни се обособиха следните услуги.

Дефиниране на метод за анализ и оценка на референтна архитектура

От практическа гледна точка на вече реализирани решения се открииха следните услуги – Фигура 11, реализирани върху различни типове нерелационни бази от данни и приложени на практика от бизнеса, които разгледахме в предишната точка.



Фигура 11 Видове финансови услуги открити на база практически опит

От теоретична гледна точка, от Глава 1 се откриха следните финансови услуги –
 Фигура 12.



Фигура 12 Видове финансови услуги открити на база теоритичен опит

Всяка една от откритите финансови услуги от теоретична и практическа гледна точка, е част от различните видове финанси. Те включват три под категории – Графика 1:



Графика 1 Под категории финанси

Най-голяма част от услугите, които бяха открити от теоретична и практическа гледна точка са от първата под категория на финансите – персоналните финанси или финансовите услуги, които са пряко свързани с клиента. Услугите от този тип генерират Големи данни, които разполагат с основните характеристики, които бяха разгледани по-рано в Глава I. От гледна точка на финансовите услуги ги разглеждаме по следния начин:

❖ **Обем (Volume)**

Финансовите институции разполагат с голям брой клиенти, което от своя страна води до голям брой операции, които се извършват за клиентите, а това от своя страна води до генерирането на големи обеми от данни.

❖ **Скорост (Velocity)**

Скоростта, с която се генерират данните в тази сфера и това, че трябва да бъдат изпълнявани операции в реално време или почти реално време, базирани върху тези данни, в противен случай може да се стигне до загуба на клиенти и отливът им към други финансови институции.

❖ **Разнообразие (Variety)**

Поради факта, че финансовите институции пряко свързани с клиентите генерират различни видове данни заради начините на обслужване – присъствено, дигитално и комбинирано. Тези типове обслужване водят до различни източници на данни

– структурирани, полуструктурирани и неструктурирани, което води и до голямото разнообразие от данни.

❖ **Стойност (Value)**

Въпреки наличието на големи обеми от данни от различни източници генерирани с голяма скорост е много важно те да бъдат стойностни и ценни. За целите на финансовите операции това е от изключителна важност, за да могат да бъдат полезни за анализирането и извличането на изводи. Например, за финансовите институции би било от огромна полза и стойност да извлекат данни за клиента за това каква застраховка могат да му предложат в зависимост от това какви са нуждите му – дали има кола, апартамент, къща, семейство и т.н., тоест стойността на генерираните, събрани и анализирани данни ще бъде от голяма полза, ако в крайна сметка това се превърне в продажба.

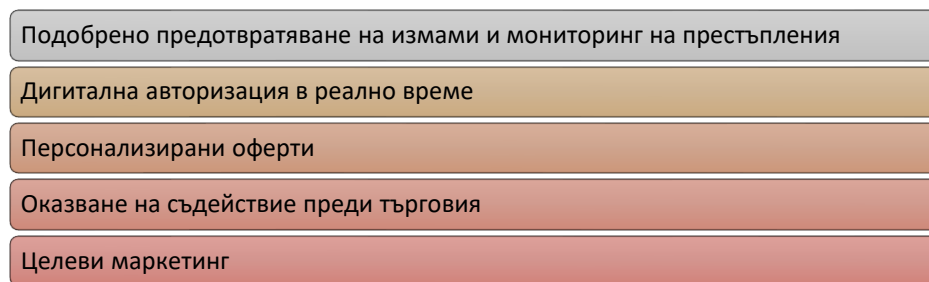
❖ **Достоверност (Veracity)**

В сферата на финансовите услуги е много важно данните да бъдат надеждни и достоверни, като това трябва да бъде доказано от източника на данните.

С развиването на технологиите, дигитализацията и дигиталната трансформация, генерирането на големи обеми от горепосочените видове данни, води до заключенията за много компании, че адаптирането на NoSQL бази от данни към техните системи би донесло само ползи. Дори финансовите услуги, които използват наследствени системи, базирани върху релационни бази от данни, осъзнават необходимостта от промяната, за да се справят с новите изисквания на клиентите.

След подробното проучване на нерелационни бази от данни за финансовите услуги – от теоретична и практическа гледна точка, не бе открит специфичен начин как да се избере подходяща NoSQL база от за финансова услуга. Има много компании и уебсайтове, които предлагат сравнение на различните характеристики на базите от данни, но като цяло нищо за конкретната сфера или услуга.

По тази причина би било целесъобразно да се разработи метод за количествена оценка на NoSQL бази от данни за конкретни финансови услуги. В Глава 1 са дефинирани предимства на нерелационните бази от данни, които са следните – Фигура 13:



Фигура 13 Предимства на NoSQL бази от данни, дефинирани като критерии

Базирано на тези предимства, ще ги дефинираме като критерии за оценка на нерелационните бази от данни. Листът с избраните критерии не е изчерпателен и могат да бъдат добавени още в зависимост от нуждите и изискванията на клиентите/компанията

Подходът, който е разработен цели да даде възможност за избор на тип нерелационна база от данни, която да се съобрази от една страна със зададените критерии, от друга страна с типа данни, които обработват както и времето за обработка. В Таблица 3 са разгледани финансови услуги, за които имаме практически приложени решения от компании, съответно са оценени с определена тежест, като в зависимост от това дали съществува решение и се дава тежест.

Критериите: Сигурност, Анализ в реално време, Персонализация, Съдействие преди търговия и Целеви маркетинг са оценявани по скалата от 1 до 5, като те са базирани на оценката на експерти с различни области на експертиза. Критериите са оценявани с различно тегло: 1 – Слабо, 2 – Незадоволително, 3 – Добро, 4 – Много добро и 5 – Отлично. Всеки от тези пет критерия се оценява на базата на експертна оценка и съответно важноста му за всяка една финансова услуга.

В таблицата са представени изведените критерии за четирите вида бази от данни – документно-ориентирани, двойка ключ-стойност, широко-колонна и граф бази с по два примера за всяка една от тях. За документно-ориентираните са разгледани Couchbase и Mongo DB, за двойка ключ-стойност – Amazon DynamoDB и Redis, за широко-колонните – HBase и Cassandra, и за граф базите – Neo4j и Orient DB.

Вторият елемент, който се оценява е дали има вече създадено решение, такова в развитие или никакво. За целта се използва скалата от 0 до 2, като 0 – означава, че няма

решение, 1 – в разработка и 2 – напълно работещо решение за финансова услуга, базирано върху NoSQL база от данни.

Прилагайки този метод за оценка на нерелационни бази от данни, всяка една от тях събира общ сбор от точки, който показва колко подходяща е цялостно за прилагане към финансови услуги. При документно-ориентираните бази Couchbase събира 190 точки, а MongoDB -126. При нерелационните бази от тип ключ-стойност Amazon Dynamo DB събира 156 точки, а Redis – 192. За широко-колонните бази HBase – 154, Cassandra – 250, а при граф базите Neo4j – 146, Orient DB – 60 точки.

На базата на създадения метод и резултатите получени при оценката на документно-ориентираните бази MongoDB с 126 и Couchbase с 190 точки, ще създадем референтна архитектура за финансови услуги.

Референтната архитектура ще бъде базирана върху MongoDB, който въпреки че получава по-малко точки според експертната оценка от документно-ориентираните бази, има вече създадени реално базирани решения, които са в сферата на основното банкиране, където са типът данни, който ще използваме за тестови – договори, фишове, репорти, имейли и други.

			Документно-ориентирани				Двойка ключ-стойност				Широко-колонни				Граф			
			Couchbase		MongoDB		Amazon DynamoDB		Redis		Hbase		Cassandra		Neo4j		OrientDB	
Финансови услуги	Критерии	СК	CP?	КС	CP?	КС	CP?	КС	CP?	КС	CP?	КС	CP?	КС	CP?	КС	CP?	КС
Засичане на опити за измама	Сигурност	5	2	10	2	10	2	10	0	0	0	0	2	10	2	10	2	10
	ДА в реално време	5		10		10		10		0		0		10		10		10
	Персонализация	3		6		6		6		0		0		6		6		6
	Съдействие преди търговия	2		4		4		4		0		0		4		4		4
	Целеви маркетинг	1		2		2		2		0		0		2		2		2
Е-портфейл	Сигурност	5	2	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	ДА в реално време	5		10		0		0		0		0		0		0		0
	Персонализация	4		8		0		0		0		0		0		0		0
	Съдействие преди търговия	2		4		0		0		0		0		0		0		0
	Целеви маркетинг	1		2		0		0		0		0		0		0		0
Е-плащания	Сигурност	5	2	10	2	10	2	10	2	10	0	0	2	10	0	0	2	10
	ДА в реално време	5		10		10		10		10		0		10		0		10
	Персонализация	2		4		4		4		4		0		4		0		4
	Съдействие преди търговия	1		2		2		2		2		0		2		0		2
	Целеви маркетинг	1		2		2		2		2		0		2		0		2
Пазарни данни	Сигурност	3	2	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	ДА в реално време	3		6		0		0		0		0		0		0		0
	Персонализация	3		6		0		0		0		0		0		0		0
	Съдействие преди търговия	5		10		0		0		0		0		0		0		0
	Целеви маркетинг	5		10		0		0		0		0		0		0		0
Trade	Сигурност	5	2	10	0	0	0	0	0	0	2	10	0	0	0	0	0	0
	ДА в реално време	5		10		0		0		0		10		0		0		0
	Персонализация	4		8		0		0		0		8		0		0		0
	Съдействие преди търговия	4		8		0		0		0		8		0		0		0
	Целеви маркетинг	3		6		0		0		0		6		0		0		0
Основно банкиране	Сигурност	5	0	0	2	10	2	10	0	0	0	0	0	0	0	0	0	0
	ДА в реално време	5		0		10		10		0		0		0		0		0
	Персонализация	2		0		4		4		0		0		0		0		0
	Съдействие преди търговия	2		0		4		4		0		0		0		0		0
	Целеви маркетинг	1		0		2		2		0		0		0		0		0
Персонализиране	Сигурност	2	0	0	2	4	0	0	2	4	2	4	2	4	2	4	0	0
	ДА в реално време	2		0		4		4		4		4		4		4		0

	Персонализация	3		0		6		0		6		6		6		6		0
	Съдействие преди търговия	1		0		2		0		2		2		2		2		0
	Целеви маркетинг	1		0		2		0		2		2		2		2		0
Анализ и отчитане в реално време	Сигурност	2	0	0	2	4	0	0	2	4	0	0	0	0	0	0	0	0
	ДА в реално време	2		0		4		0		4		0		0		0		0
	Персонализация	1		0		2		0		2		0		0		0		0
	Съдействие преди търговия	3		0		6		0		6		0		0		0		0
	Целеви маркетинг	1		0		2		0		2		0		0		0		0
Капиталови пазари	Сигурност	5	0	0	0	0	2	10	0	0	0	0	0	0	0	0	0	0
	ДА в реално време	5		0		0		10		0		0		0		0		0
	Персонализация	4		0		0		8		0		0		0		0		0
	Съдействие преди търговия	4		0		0		8		0		0		0		0		0
	Целеви маркетинг	2		0		0		4		0		0		0		0		0
Застраховки	Сигурност	4	0	0	0	0	2	8	0	0	0	0	0	0	0	0	0	0
	ДА в реално време	4		0		0		8		0		0		0		0		0
	Персонализация	3		0		0		6		0		0		0		0		0
	Съдействие преди търговия	1		0		0		2		0		0		0		0		0
	Целеви маркетинг	1		0		0		2		0		0		0		0		0
Автентикация	Сигурност	5	0	0	0	0	0	0	2	10	0	0	2	10	0	0	0	0
	ДА в реално време	5		0		0		0		10		0		10		0		0
	Персонализация	2		0		0		0		4		0		4		0		0
	Съдействие преди търговия	2		0		0		0		4		0		4		0		0
	Целеви маркетинг	1		0		0		0		2		0		2		0		0
Отворено банкиране	Сигурност	5	0	0	2	0	0	0	2	10	0	0	0	10	0	0	0	0
	ДА в реално време	2		0		0		0		4		0		4		0		0
	Персонализация	4		0		0		0		8		0		8		0		0
	Съдействие преди търговия	1		0		0		0		2		0		2		0		0
	Целеви маркетинг	3		0		0		0		6		0		6		0		0
Управление на каталог и инвентар	Сигурност	4	2	0	0	0	0	0	0	8	0	0	0	8	0	0	0	0
	ДА в реално време	5		0		0		0		10		0		10		0		0
	Персонализация	2		0		0		0		4		0		4		0		0
	Съдействие преди търговия	1		0		0		0		2		0		2		0		0
	Целеви маркетинг	1		0		0		0		2		0		2		0		0
Търговски регистър	Сигурност	4	2	8	0	0	0	0	0	8	0	0	0	8	0	0	0	0
	ДА в реално време	3		6		0		0		6		0		6		0		0
	Персонализация	1		2		0		0		2		0		2		0		0

	Съдействие преди търговия	0		0		0		0		0		0		0		0		0
	Целеви маркетинг	0		0		0		0		0		0		0		0		0
Анализ	Сигурност	4	0	0	0	0	0	0	0	2	8	0	0	0	0	0	0	0
	ДА в реално време	3		0		0		0			6		0		0		0	
	Персонализация	5		0		0		0			10		0		0		0	
	Съдействие преди търговия	3		0		0		0			6		0		0		0	
	Целеви маркетинг	4		0		0		0			8		0		0		0	
Клиринг на клиенти	Сигурност	5	0	0	0	0	0	0	0	2	10	0	0	0	0	0	0	0
	ДА в реално време	4		0		0		0			8		0		0		0	
	Персонализация	3		0		0		0			6		0		0		0	
	Съдействие преди търговия	2		0		0		0			4		0		0		0	
	Целеви маркетинг	1		0		0		0			2		0		0		0	
Управление на клиенти	Сигурност	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	ДА в реално време	4		0		0		0			0		0		0		0	
	Персонализация	4		0		0		0			0		0		0		0	
	Съдействие преди търговия	2		0		0		0			0		0		0		0	
	Целеви маркетинг	4		0		0		0			0		0		0		0	
Препоръки	Сигурност	3	0	0	0	0	0	0	0	0	0	2	6	0	0	0	0	0
	ДА в реално време	5		0		0		0			0		10		0		0	
	Персонализация	5		0		0		0			0		10		0		0	
	Съдействие преди търговия	4		0		0		0			0		8		0		0	
	Целеви маркетинг	5		0		0		0			0		10		0		0	
Управление на достъпа и самоличността	Сигурност	5	0	0	0	0	0	0	0	0	0	0	0	2	10	0	0	0
	ДА в реално време	4		0		0		0			0		0		8		0	
	Персонализация	2		0		0		0			0		0		4		0	
	Съдействие преди търговия	3		0		0		0			0		0		6		0	
	Целеви маркетинг	1		0		0		0			0		0		2		0	
Мониторинг на мрежова и ИТ инфраструктура	Сигурност	5	0	0	0	0	0	0	0	0	0	0	0	2	10	0	0	0
	ДА в реално време	4		0		0		0			0		0		8		0	
	Персонализация	1		0		0		0			0		0		2		0	
	Съдействие преди търговия	0		0		0		0			0		0		0		0	
	Целеви маркетинг	0		0		0		0			0		0		0		0	
Антпране на пари	Сигурност	5	0	0	0	0	0	0	0	0	0	0	0	2	10	0	0	0
	ДА в реално време	4		0		0		0			0		0		8		0	
	Персонализация	1		0		0		0			0		0		2		0	
	Съдействие преди търговия	0		0		0		0			0		0		0		0	

	Целеви маркетинг	0		0		0		0		0		0		0		0		0
Киберсигурност	Сигурност	5	0	0	0	0	0	0	0	0	0	0	0	0	2	10	0	0
	ДА в реално време	4		0		0		0		0		0		0		8		0
	Персонализация	1		0		0		0		0		0		0		2		0
	Съдействие преди търговия	2		0		0		0		0		0		0		4		0
	Целеви маркетинг	1		0		0		0		0		0		0		2		0
Управление на риска	Сигурност	5	0	0	0	0	0	0	2	10	2	10	2	10	0	0	0	0
	ДА в реално време	4		0		0		0		8		8		8		0		0
	Персонализация	1		0		0		0		2		2		2		0		0
	Съдействие преди търговия	2		0		0		0		4		4		4		0		0
	Целеви маркетинг	1		0		0		0		2		2		2		0		0
Общо				190		126		156		192		154		250		146		60

Таблица 3 Оценка на NoSQL бази данни спрямо дефинирани критерии

Проектиране на бизнес референтна архитектура за финансови услуги с NoSQL бази от данни

Всяка една финансова институция занимаваща се с различни видове финансови услуги генерира, събира, обработка и анализира данни от различни типове – структурирани, полуструктурирани и неструктурирани в големи обеми, което от своя страна води до различни нужди на институциите. В зависимост от типа финансова институция зависи вида данни, който бива генериран, което пък от своя страна води до различни изисквания за място за съхранение, методи за обработка и инструменти за анализи.

С внедряването на нерелационни бази от данни във финансовите институции и адаптирайки техните исторически системи към новите бази, води до развиването и нуждата на услугите в сферата на финансите да запазят своето конкурентно предимство, както и да излязат по-напред от техните конкуренти в сферата.

Целта на създаването на референтна архитектура, независимо дали за използване само в една организация или за универсалното и адаптиране в множество финансови институции, е да подсигури последователността и приложимостта на използването на дадени технологии в конкретна организация.

На база на изискванията, които могат да произлязат от финансовите услуги, заедно с направеното теоретично проучване на литература, както и на практическото такова на вече съществуващи решения на финансови услуги върху различни нерелационни бази от данни,

бе създаден модел на универсална компонентна референтна архитектура за финансови услуги.

- Концептуален модел

Концептуалният модел на референтната архитектура за финансови услуги представен на Фигура 14 представлява генералната структура и елементи, от които се състои тя, както и необходимите данни и системи, за да се постигнат бизнес изискванията, което означава поддръжката на фирмени процеси, записване на бизнес събития и проследяването на представянето. В Глава I е дефиниран подход за изграждането на референтна архитектура, който се състои от пет основни стъпки. Създаденият концептуален модел на бизнес референтната архитектура за финансови услуги е в резултат от първите две стъпки – идентифицирането на целта и формулирането на целта. Целта на бизнес референтната архитектура за финансови услуги е да адаптира модерните начини на съхранение на Големи данни за различни типове – полуструктурирани и неструктурирани данни към финансовите институции, които продължават своето бързо развитие и дигитална трансформация.

Референтната архитектура се състои от 3 слоя: Система за съхранение на данни, Интеграционен слой и Сървър за съхранение на данни, като те са свързани със системи за обработка на финансови услуги.

❖ Система за съхранение на данни

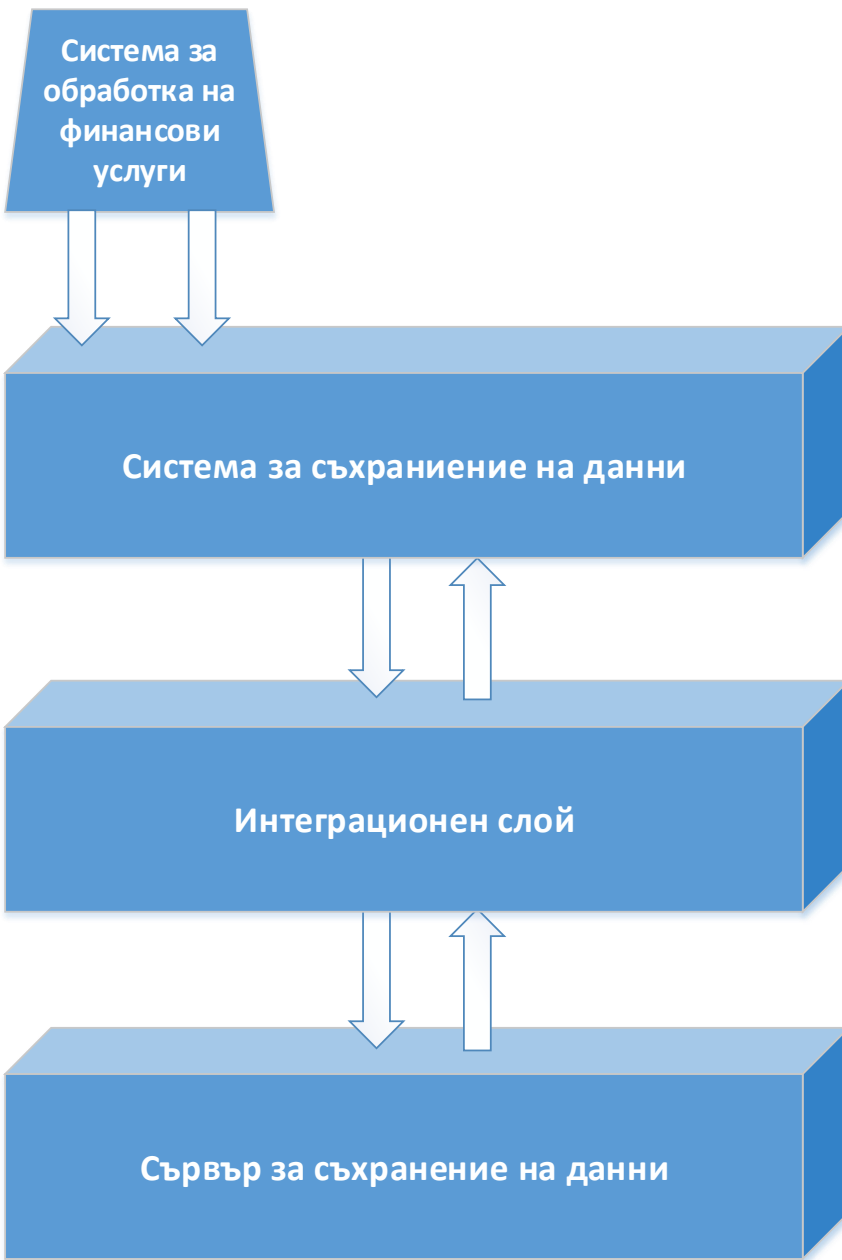
Системата за съхранение на данни е системата, която съхранява данните след като са постъпили от системите за обработка на финансови услуги.

❖ Интеграционен слой

Интеграционният слой в референтната архитектура се отнася до компонентите отговорни за интегрирането на системите за съхранение на данни и сървърите за съхранението им. Този слой предоставя централизирана платформа за управлението, движението на данни и комуникациите между различните системи. Интеграционният слой се състои от конектори за данни между системата за съхранение и сървъра.

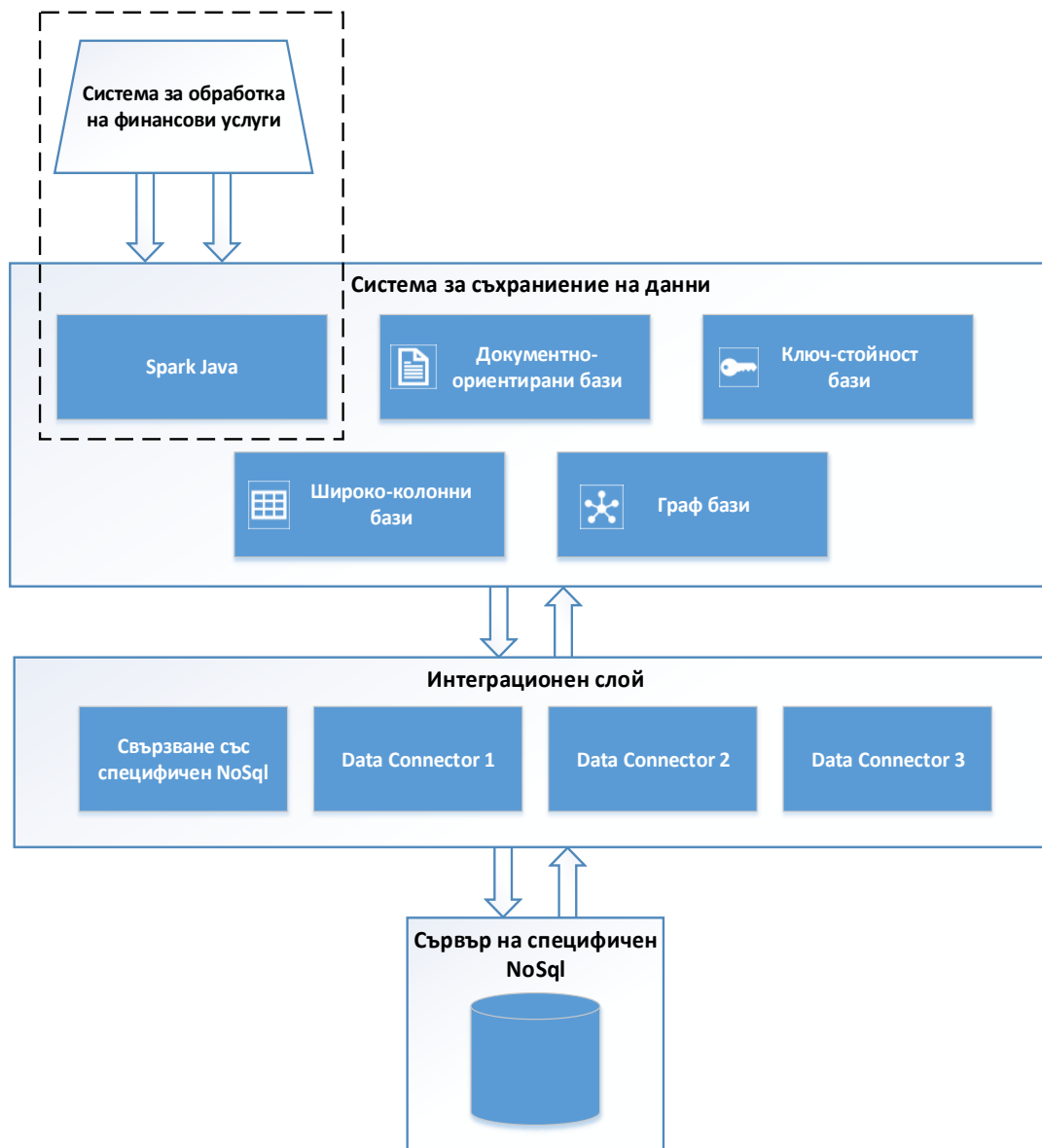
❖ Сървър за съхранение на данни

Сървърът за съхранение на данни е мястото, на което се съхраняват данните, които достигат, чрез интеграционния слой и конекторите за данни от системите за съхранение на данни, които влизат към системата за съхранение на данни от системите за обработка на финансови услуги.



Фигура 14 Концептуален модел на бизнес референтна архитектура за финансови услуги

- Логически модел



Фигура 15 Логически модел на бизнес референтната архитектура за финансови услуги

В логическият модел са заложили техническите правила и стандарти за създаването на бизнес референтната архитектура, като при него се залагат правилата, за да могат да се постигнат заложените принципи. През системата за обработка на финансови услуги влизат данните от финансовите институции, които в последствие са съхранявани, обработвани и анализирани в останалите слоеве на референтната архитектура. На Фигура 15 е представен

логическият модел на бизнес референтната архитектура за финансови услуги, като по-долу е разгледан подробно всеки един от нейните слоеве.

❖ Система за съхранение на данни

В първият слой на референтната архитектура, т.нар. система за съхранение на данни могат да бъдат разположени 1 или повече четири типа нерелационни бази от данни – документно-ориентирани, ключ-стойност, широко-колонни и граф бази. Типовете NoSQL бази от данни бяха разгледани на теоретично ниво, както какви са техните възможности в Глава I, а в настоящата глава бяха разгледани на практическо ниво – какви приложения свързани с финансовите услуги са създадени върху конкретни нерелационни бази от данни, което да спомогне изборът на конкретна база.

Изборът на такава се реализира базирано върху следните елементи, които трябва да бъдат разгледани внимателно преди пристъпването към конкретна NoSQL база:

- Налична система/системи на финансовата институция, която ще трябва да бъде мигрирана към новата база
- Тип финансови услуги, с които се занимава финансовата институция
- Типове данни, които генерира финансовата институция на база финансовите услуги, които обслужва – структурирани, полуструктурирани и неструктурирани данни

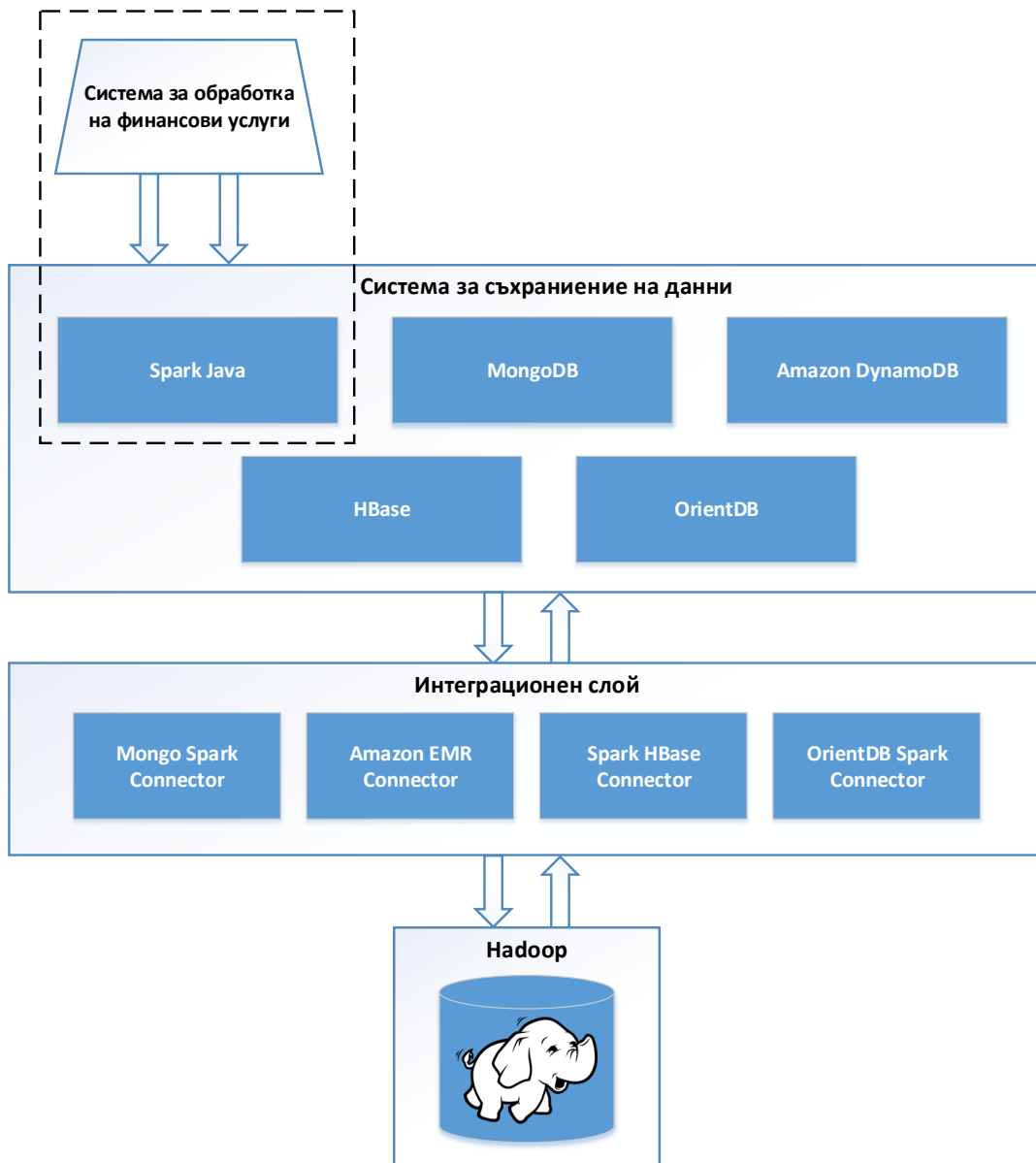
❖ Интеграционен слой

На база на нерелационната база, която е избрана в системата за съхранение на данни на база на разгледаните елементи, се избира начинът за свързване със специфичен NoSQL, който се позиционира в интеграционния слой на референтната архитектура.

❖ Сървър на специфична NoSQL база от данни

В зависимост от избраната нерелационна база от данни в системата за съхранение на данни, последвалия избор на начин за свързване на конкретна NoSQL база от данни, се достига до сървър за специфичен NoSQL, в който се съхраняват данните на финансовата институция.

- **Физически модел**



Фигура 16 Физически модел на бизнес референтната архитектура за финансови услуги

Създаденият физически модел на Фигура 16 е реализиран с конкретни примери на нерелационни бази от данни и конектори за тях към Hadoop. Всеки един от тези елементи може да бъде заменен според нуждите на финансовата институция, която иска да адаптира тази референтна архитектура с NoSQL бази от данни за собствените си нужди.

В първия слой, който е системата за съхранение на данни, могат да се разположат четирите типа нерелационни бази от данни – документно-ориентирани, ключ-стойност, широко-колонни и граф бази, като в този слой се разполага конкретната база, която е

избрана чрез метода за оценка и избор на NoSQL бази от данни за финансови услуги. Във физическия модел са разположени бази от всяка една от видовете:

- Документно-ориентирана база: **MongoDB**
- Ключ-стойност база: **Amazon DynamoDB**
- Широко-колонна база: **HBase**
- Граф база: **Orient DB**

Във втория слой, т.нар. интеграционен слой се намират свързващите елементи с различните NoSQL бази от данни, напр. NiFi, Spark Connector и други видове Data Connectors подходящи за другите типове нерелационни бази от данни. Конкретните конектори на данни от системата за съхранение на данни към сървъра на специфичната NoSQL база от данни са следните:

- За документно-ориентираната база **MongoDB** може да бъде използван **Mongo Spark Connector**
- За ключ-стойност базата **Amazon DynamoDB** може да бъде използван **Amazon EMR Connector**
- За широко-колонна база **HBase** може да бъде използван **Spark HBase Connector**
- За граф базата **Orient DB** може да бъде използван **Orient DB Spark Connector**

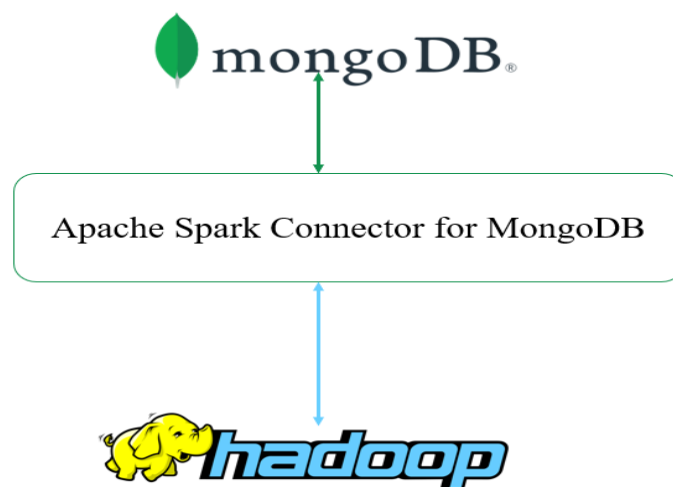
Третият слой е сървър на специфичната NoSQL база, като за реализацията на физическия модел на референтната архитектура за финансови услуги е избран Hadoop

Подход за използване на проектирана РА водеща към създаване на ICT Архитектура

ICT Архитектурата (Информационни и комуникационни технологии) осигурява концептуален модел, специфицирайки на основно ниво елементите на ICT архитектурата (приложение, бази от данни, технологични ICT елементи), както и връзките помежду им. На базата на референтната архитектура за финансови услуги с NoSQL бази от данни, която бе създадена в предишната точка и е демонстрирана на Фигура 22, ще създадем конкретна ICT архитектура, която ще е реализирана и тествана в Глава IV.

Разработеният физически модел за референтна архитектура съдържа 3 слоя – система за съхранение на данни, интеграционен слой и слой за специфичен NoSQL, като в предишната точка разгледахме подробно различните типова за нерелационни бази от данни, с различни видове конектори за данни, които могат да бъдат свързани със слоя за специфичен NoSQL.

При създаването на ICT архитектурата за система за съхранение на данни(първият слой), ще бъде използвана документно-ориентираната нерелационна база от данни MongoDB, тъй като тя ба избрана чрез създадения метод в предишната точка и отговаря на изискванията и има вече създадена финансова услуга, която използва типа данни, с който ще бъде тествана архитектурата. Вторият интеграционен слой ще използва Apache Spark Connector за MongoDB. Поради факта, че поддръжката на директния конектор на MongoDB към Hadoop е спряна, ще използваме този през Apache Spark. Третият слой остава за системата големи данни Hadoop, която ще използваме за съхранение – Фигура 17.



Фигура 17 Конкретна ICT архитектура

Заклучение

Разработен е метод за избор и оценка на NoSQL бази от данни за финансови услуги, като той е базиран на разработени и използващи се практически приложения на базите, както и на експертна оценка. Разгледани са и са анализирани практически примери от четирите вида нерелационни бази от данни. На база на тях е предложена бизнес референтна архитектура, както и подход за използването и към създаването на ICT архитектура, която е реализирана и ще бъде подробно разгледана и тествана в Глава IV.

На база на създадения метод за анализ и оценка на NoSQL бази от данни е създаден автоматизиран инструмент за избор на нерелационни бази от данни за финансови услуги, който ще бъде представен в следващата глава.

3. Глава III. Инструмент за избор на NoSQL бази от данни за финансови услуги

На база създадения метод за избор и оценка на NoSQL бази от данни за финансови услуги е създаден автоматизиран инструмент за избор на такива. За създаването му е избрана уеб технология базирана върху Microsoft .NET Framework, използвайки C# и Visual Studio.

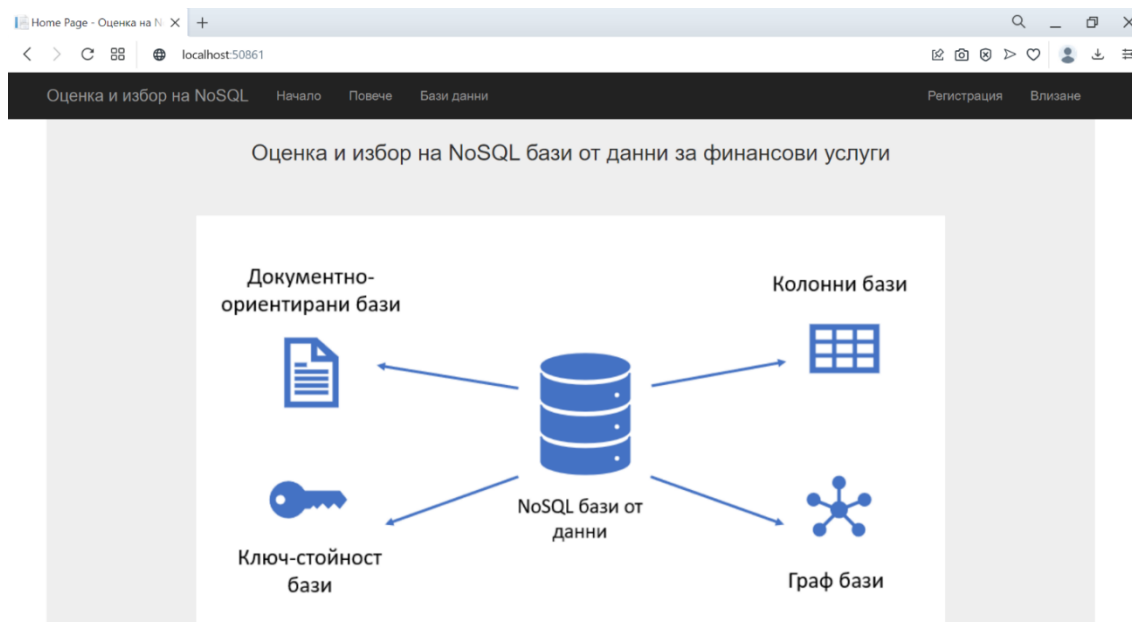
	Потребител/Компания	Администратор/Експерт
Регистрация	Да	Да
Вход	Да	Да
Оценка и избор на NoSQL според критерии	Да	Да
Добавяне на NoSQL база	Не	Да
Редакция на NoSQL база	Не	Да
Изтриване на NoSQL база	Не	Да
Добавяне на финансови услуги	Не	Да
Редактиране на финансови услуги	Не	Да
Изтриване на финансови услуги	Не	Да
Добавяне на критерии за оценка	Не	Да
Редактиране на критерии за оценка	Не	Да
Изтриване на критерии за оценка	Не	Да

Таблица 4 Потребителски роли

Инструментът за оценка и избор на NoSQL база от данни за финансови услуги има две потребителски роли – Потребител/Компания и Администратор/Експерт – Таблица 4. Системата се състои от различни страници и функционалности в зависимост от потребителската роля.

Потребителска роля Потребител/Компания

За Потребител/Компания инструментът се състои от начална страница, страница с повече информация, както и такава със списък на наличните бази от данни с допълнителна информация за тях. В дясната част на менюто се намират бутоните за Регистрация на потребители и за Влизане в системата на вече съществуващите потребители – Фигура 18.



Фигура 18 Начална страница на инструмент за Клиент/Компания

След влизане в приложението с регистрация потребителят/компанията избира тип финансова услуги, като може да маркира няколко вида от предварително дефинираните такива в инструмента в зависимост от това какви услуги са предлагани от институцията. След това потребителят заявява дали има предпочитан тип нерелационна база от данни, като това зависи от системите и софтуерите, които се ползват в дадената компания. Ако има желан тип нерелационна база от данни към, който компанията се ориентира, избира съответното поле на страницата за избор, а след неговото маркиране се появява падащото меню с четирите типа нерелационни бази, от които може да избере – документно-ориентирани, широко-колонни, ключ-стойност и граф бази – Фигура 19. Ако няма конкретен тип база, която да бъде желана от компанията, се отбелязва съответно поле, което в този случай не позволява на падащото меню с типове бази да се появи. След това се натиска бутонът напред, който отвежда потребителя към следващата страница с информация, която следва да бъде попълнена.

Избор на NoSQL база от данни за финансови услуги

Финансови услуги	<input checked="" type="checkbox"/> Банкиране
	<input type="checkbox"/> Инвестиционни услуги
	<input type="checkbox"/> Обменни услуги
	<input checked="" type="checkbox"/> Застрахователни услуги
	<input checked="" type="checkbox"/> Финансов износ
	<input type="checkbox"/> Други финансови услуги
Имате ли предпочитан тип NoSQL база?	<input checked="" type="checkbox"/> Да
	<input type="checkbox"/> Не
Вид NoSQL база	<input type="text" value="Документно-ориентирана база"/>
<input type="button" value="Напред"/>	

© 2023 - Оценка и избор на NoSQL бази

Фигура 19 Стъпка 1 за избор на NoSQL база от данни за финансови услуги

В случаят, който демонстрираме с работата на инструмента, потребителят иска нерелационна база от данни, която е подходяща за услуги в сферата на банкирането, застрахователните услуги и финансовия износ, като също има предпочитан тип нерелационна база от данни, която да бъде от документно-ориентираните. Преди да избере дали има предпочитан тип NoSQL база от данни, клиентът трябва да съобщи системите, с които работи и това дали ще трябва да интегрира и мигрира настоящите системи на институцията или ще изгражда напълно нови, които да са напълно съвместими с новата база. Случаят, при който настоящите системи трябва да се мигрират към нова база е много чувствителен и трябва да бъде внимателно преценен от техническите екипи на компанията. Другият вариант, в който се създават абсолютно нови системи за работа с новоизбраната нерелационна база от данни отнема изключително много време и финансов ресурс. Изборът при такава стъпка трябва да бъде много добре претеглен.

При следващата стъпка от оценката на нерелационните бази данни от предварително дефинираните критерии, които са изведени чрез метода от Глава II и съответно им е зададена оценка от 1 до 5 от експертите в системата, компанията може да избере само някои от тях, на които да се наблегне в тяхната система базирана върху новата NoSQL база от данни или съответно всички от тях – Фигура 20.

Избор на NoSQL база от данни за финансови услуги

Критерии

- Подобрено предотвратяване на измами и мониторинг на престъпления
- Дигитална авторизация в реално време
- Персонализирани оферти
- Оказване на съдействие преди търговия
- Целеви маркетинг

Изчисли

© 2023 - Оценка и избор на NoSQL бази

Фигура 20 Стъпка 2 от избор на NoSQL база от данни за финансови услуги

В зависимост от горепосоченото, както и на предварително дефинираната информация за това кои NoSQL бази вече имат практически приложими решения за финансови услуги, което носи допълнителна тежест към оценката и на база алгоритъма зададен в инструмента за оценка и избор на нерелационни бази от данни за финансови услуги след натискане на бутона Изчисли на екрана се появяват различни възможности за NoSQL бази, които могат да бъдат предложени при съответно избрани условия – Фигура 21.

Избор на NoSQL база от данни за финансови услуги

Резултат от оценка на NoSQL бази от данни за финансови услуги

Място	Наименование на база	Резултат от оценка
1	Couchbase	190
2	MongoDB	126
3	-	-

Започни отново

© 2023 - Оценка и избор на NoSQL бази

Фигура 21 Резултат от оценка на NoSQL бази от данни за финансови услуги

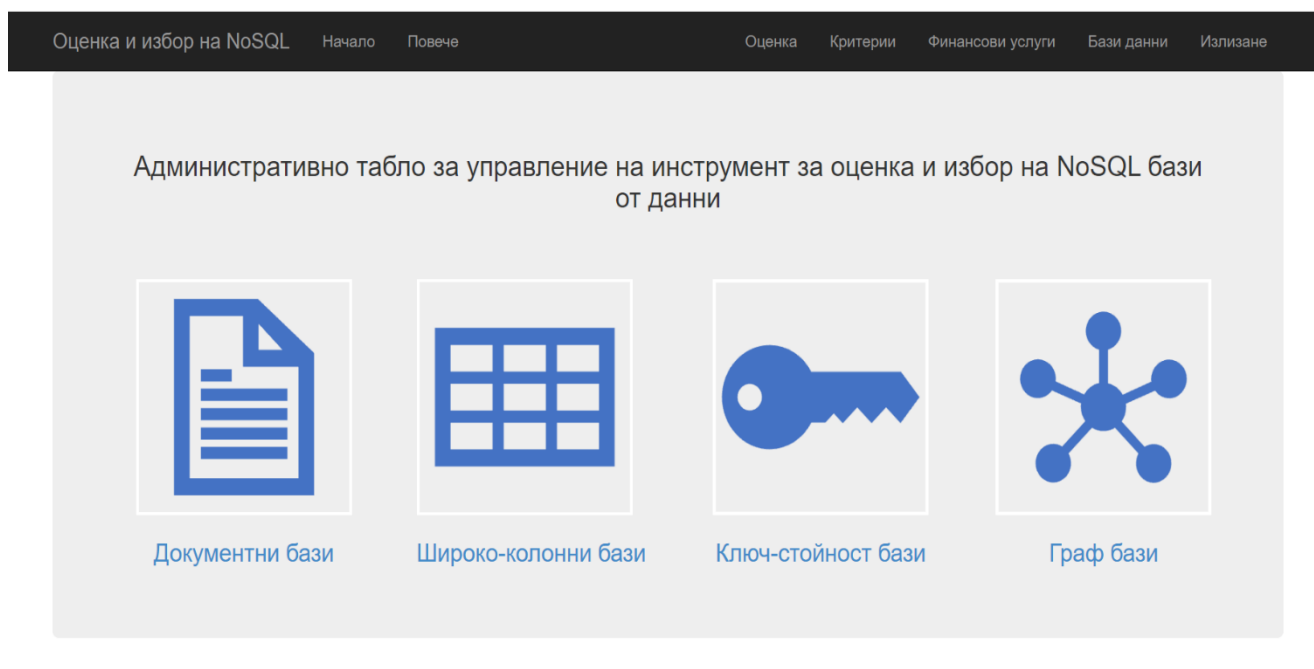
При необходимост, оценката може да бъде започната отначало, като на страницата, но която се изобразява резултата, има бутон Започни отначало, който занулява резултатите получени до момента и препраща потребителите към страницата със Стъпка 1.

Потребителска роля Администратор/Експерт

Администратор/Експерт разполага с възможността да създава, редактира и изтрива следните елементи, които са необходими за работата на инструмента:

- ✓ NoSQL бази от данни
- ✓ Финансови услуги
- ✓ Критерии

Експертът също разполага с възможността да прави оценка и избор на нерелационни бази от данни с цел тестване на правилната работа на инструмента. Административното табло за управление на инструмента за оценка и избор на нерелационни бази от данни се визуализира по следния начин – Фигура 22.



© 2023 - Оценка и избор на NoSQL бази

Фигура 22 Административно табло за управление на инструмент за оценка и избор на NoSQL бази от данни

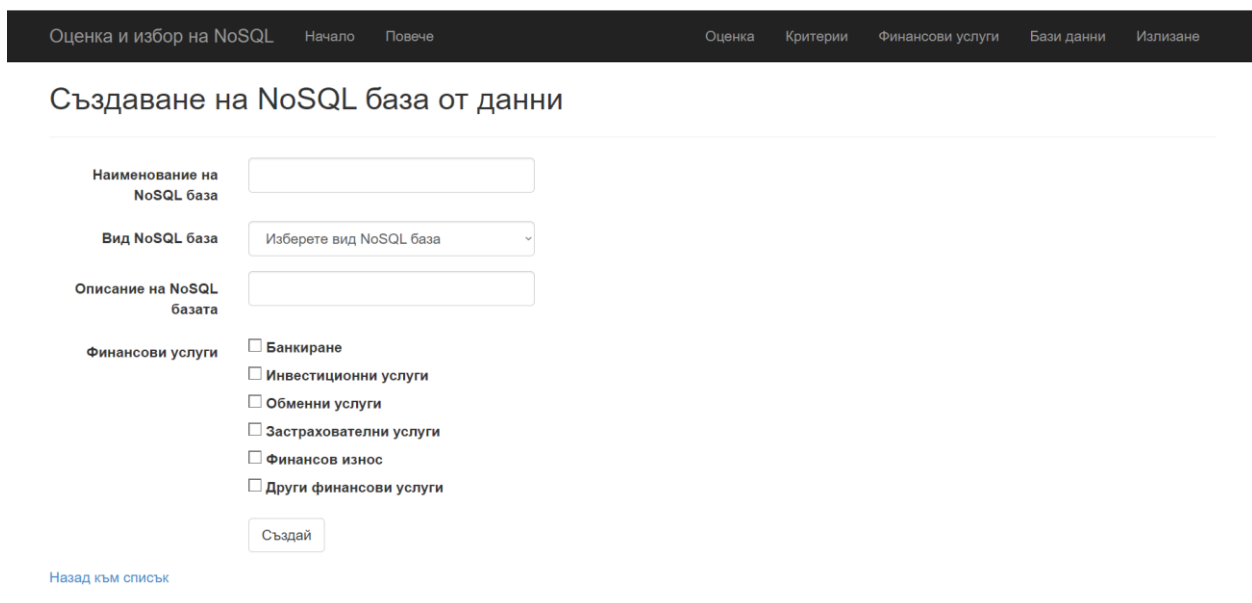
В менюто отгоре на страницата администраторът може да достигне до страниците добавяне, редакция и изтриване на критерии, финансови услуги и нерелационни бази от данни.

а. NoSQL бази от данни

При реализацията и управлението на елементите свързани с нерелационни бази от данни, инструментът предлага следните функционалности.

і. Добавяне

При добавянето на NoSQL бази от данни в инструмента се дефинират наименованието на базата, описание, от кой тип е – документно-ориентирана, широко-колонна, ключ-стойност или граф база, както и за кой тип финансови услуга вече има създадено практическо решение базирано върху тази нерелационна база от данни. Това е необходимо, за да може в последствие при изчисляване на коефициентите заедно със стойността на критериите да се получи визуално представяне колко подходяща е конкретна NoSQL база за целите на дадена организация – Фигура 23.

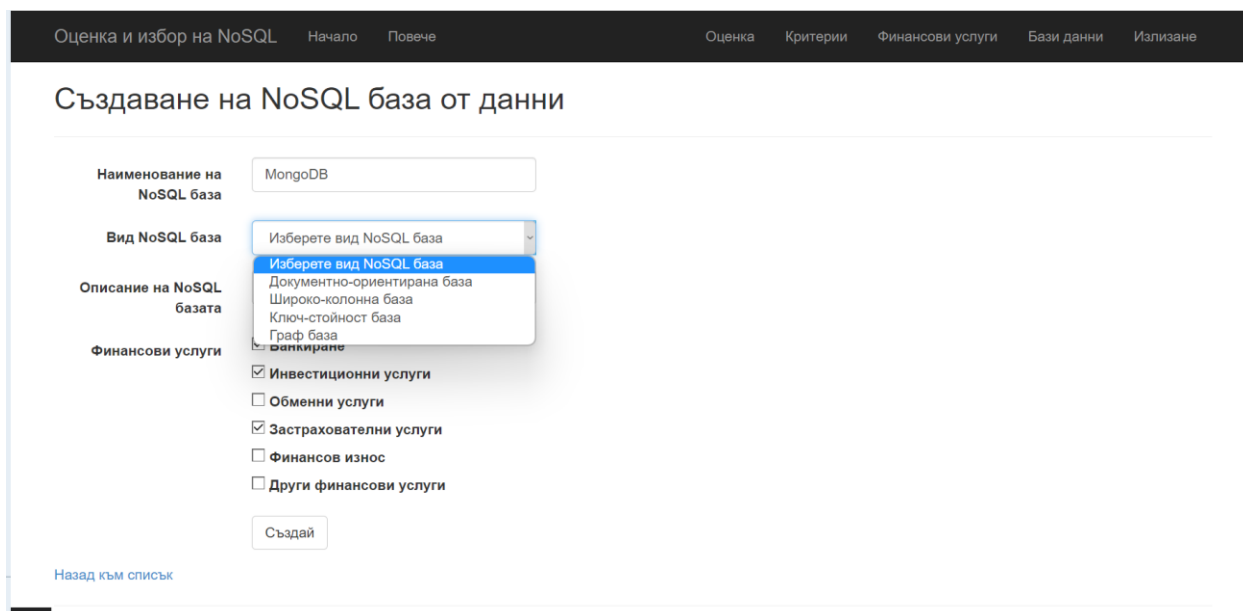


The screenshot shows a web interface for creating a NoSQL database. At the top, there is a navigation bar with the following items: 'Оценка и избор на NoSQL', 'Начало', 'Повече', 'Оценка', 'Критерии', 'Финансови услуги', 'Бази данни', and 'Излизане'. The main heading is 'Създаване на NoSQL база от данни'. The form contains the following fields and options:

- 'Наименование на NoSQL база': A text input field.
- 'Вид NoSQL база': A dropdown menu with the text 'Изберете вид NoSQL база'.
- 'Описание на NoSQL базата': A text input field.
- 'Финансови услуги': A list of checkboxes with the following labels:
 - Банкиране
 - Инвестиционни услуги
 - Обменни услуги
 - Застрахователни услуги
 - Финансов износ
 - Други финансови услуги
- 'Създай': A button to submit the form.
- 'Назад към списък': A link below the button.

Фигура 23 Създаване на NoSQL база от данни – Стъпка 1

При добавянето на нова нерелационна база от данни към инструмента, видът, от който се избира от падащо меню, където те са предварително дефинирани, за да не се получат допълнителни обърквания при работа. Финансовите услуги, от които може да се избира при дефинирането на нерелационна база от данни са онези, които администраторът е въвел предварително чрез страницата за управление на финансови услуги в инструмента, като при въвеждане на конкретна база могат да се изберат няколко възможности – Фигура 24.



Фигура 24 Създаване на NoSQL бази от данни - Стъпка 2

ii. Редактиране

В инструмента се предоставя възможност за редактиране на вече създадената нерелационна база от данни, като всички нейни елементи могат да бъдат променени.

iii. Изтриване

При необходимост вече излишните бази могат да бъдат изтривани от инструмента за оценка и избор на NoSQL бази от данни.

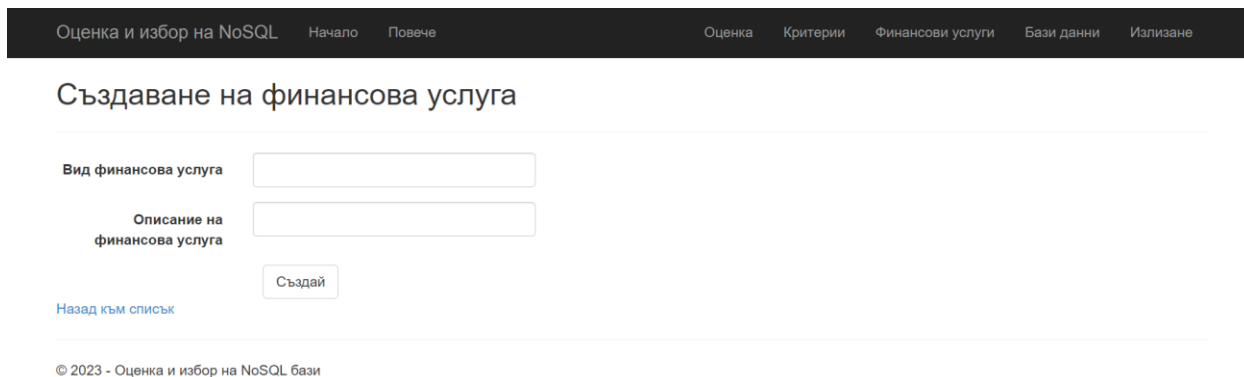
iv. Списък с NoSQL бази от данни

За улеснение на администратор/експерт всички налични нерелационни бази от данни с техния тип, описание и за кои финансови услуги работят се показват в списък.

в. Финансови услуги

i. Добавяне

Чрез създадения интерфейс в инструмента се добавят различните видове услуги, които са налични на пазара и се реализират от институциите. Наличният списък в приложението от предварително разгледаните в Глава I финансови услуги – Фигура 25.



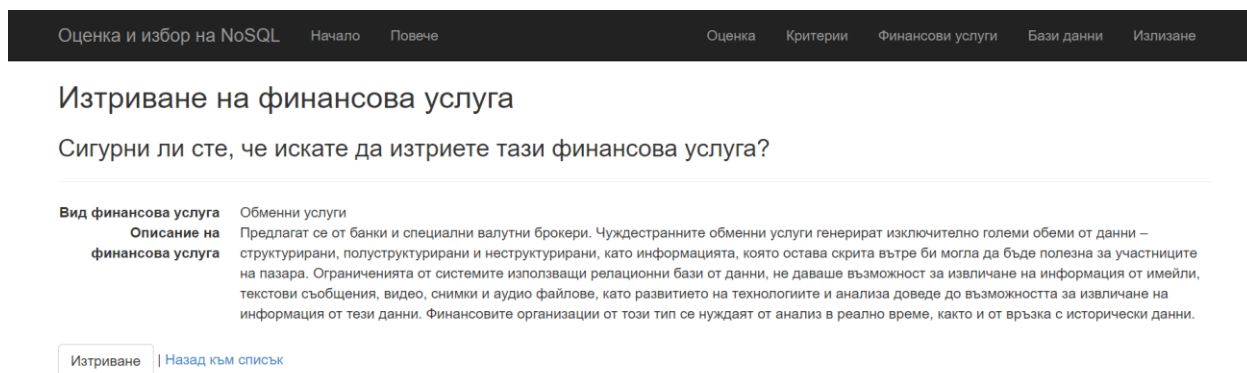
Фигура 25 Създаване на финансова услуга

ii. Редактиране

От страницата с подробно описание на финансовата услуга може да се прегледа какво точно предлага и съответно да се предприеме редакция при необходимост

iii. Изтриване

Чрез интерфейса на инструмента финансовите услуги също могат да изтриват, като съответно преди извършването на операцията се изисква потвърждение за това – Фигура 26.



Фигура 26 Изтриване на финансова услуга

iv. Списък с финансови услуги

Финансовите услуги се представят в пълен списък с тяхното описание, както и с възможностите за редакция и изтриване.

с. Критерии

Управлението на критерии се реализира, чрез създадения интерфейс на приложението, като за тях се добавя съкратено наименование, което да се визуализира на потребителите, подробно описание и тежест на самия критерий.

i. Добавяне на критерии – Фигура 27

Оценка и избор на NoSQL Начало Повече Оценка Критерии Финансови услуги Бази данни Излизане

Създаване на критерий

Съкратено наименование на критерий

Описание на критерий

Тежест на критерий

[Назад към списък](#)

© 2023 - Оценка и избор на NoSQL бази

Фигура 27 Създаване на критерии в инструмент за оценка на NoSQL бази от данни

ii. Редактиране

Всеки един от елементите на добавените критерии може да бъде променен от администратор/експерт, чрез създадения интерфейс, включително и тежестта на самия критерий – Фигура 28.

Оценка и избор на NoSQL Начало Повече Оценка Критерии Финансови услуги Бази данни Излизане

Редакция на критерии за оценка на NoSQL бази от данни

Критерий

Съкратено наименование на критерий

Описание на критерий

Тежест на критерий

[Назад към списък](#)

© 2023 - Оценка и избор на NoSQL бази

Фигура 28 Редактиране на критерии за оценка на NoSQL бази от данни



iii. Изтриване

При необходимост всеки един от критериите може да бъде изтрит.

iv. Списък с критерии

В системата за администраторското табло се визуализират и всички налични в момента критерии за оценка на нерелационни бази от данни – Фигура 29.

Съкратено наименование на критерий	Описание на критерий	Тежест на критерий	Редакция Детайли Изтриване
Предотвратяване на измами и мониторинг	Това е от изключителна важност за финансовите услуги – банки, кредитни и дебитни карти. Всички тези институции работят с изключително чувствителна информация, която е от изключителна важност за разбирането и удовлетворяването на клиентите. Имамите във финансовите услуги се увеличават, което може да се предотврати, чрез анализ, анализиране на аномалии, класификация и клъстериране, което може да се реализира с нерелационните бази от данни. Предотвратяването на измами и мониторинг на престъпления са от изключителна важност, за да се защитят клиентската лична и чувствителна информация, както и техните онлайн трансакции и данни.	5	Редакция Детайли Изтриване
Дигитална авторизация в реално време	Тази функционалност е от изключителна важност за финансовите услуги, за да може да е сигурно, че няма да се случи измама и за реалната идентификация на хората, с които се комуникира. NoSQL базите от данни имат такива възможности поради намалената латентност на четене, като авторизацията може да се случи почти веднага. Дигиталната авторизация в реално време е директно свързана с предотвратяването на измами, като по тази причина отново е с изключително висок приоритет, когато се говори за финансови услуги и е необходимо да бъдат реализирани, когато се правят онлайн трансакции.	5	Редакция Детайли Изтриване
Персонализирани оферти	За клиентите е от изключителна важност да получават персонализирани оферти, които да отговарят на техните нужди. Това не е по-различно във финансовите услуги – осигурителни полици базирани на персоналния риск, лихвени проценти на база погасяване и капацитет и др. За реализирането и представянето на персонализираните оферти на клиентите е необходимо да се събере информация – поведение, ангажираност в социалните медии, бисквитки, онлайн пазаруване и др., като всички данни идват от различни източници, като за тази цел на помощ идват	3	Редакция Детайли Изтриване

Фигура 29 Списък с критерии за оценка на NoSQL бази от данни за финансови услуги

Изводи

Създаденото приложение базирано на метода за анализ и оценка на нерелационни бази от данни цели да спомогне и улесни избора на NoSQL бази от данни за финансови услуги, след като експертната оценка е извела конкретни критерии с тежест, като е направено голямо проучване на литература и практически опит за вече съществуващите практически решения, както и за потенциалното развиване на такива. Изборът на NoSQL база от данни върху, която да бъдат мигрирани вече съществуващи системи или създадени нови, трябва да бъде направен изключително внимателно.

Избраната нерелационна база от данни, в последствие може да бъде използвана за структурна и интеграционна трансформация на бизнес референтна архитектура към детайлна архитектура.

4. Глава IV. Приложение на метод за проектиране на референтна архитектура за финансови услуги

Архитектура за изпълнение на финансови услуги с NoSQL бази от данни

С все по-голямата автоматизация и дигитализация на процесите, и навлизането на големи обеми от данни от различни типове – най-вече от полуструктурирани и неструктурирани, бизнесите се изправят пред огромното предизвикателство за събирането, съхранението и анализа на неструктурирани данни [79]. Финансовите институции и услугите, които предлагат са от бързо развиващите се бизнеси, като по тази причина те трябва да се адаптират бързо и навременно към разгръщащите се технологии и към увеличаващите се изисквания на клиентите, както и на постоянната конкуренция.

На база направеният детайлен анализ на научна литература свързана с нерелационните бази от данни и използването им за реализиране на системи за финансови услуги, както и на база разработените бизнес решения, в Глава II бе разработена бизнес референтна архитектура, състояща се от система за съхранение на данни, интеграционен слой и сървър на специфичен NoSQL, която бе развита и достигна до реализирането на физически модел. В този модел могат да бъдат разположени всеки един от четирите типа нерелационни бази от данни, както в комбинация заедно, така и по отделно.

Базирано на резултата получен от метода, отново разработен в Глава II, конкретната ICT архитектура, която ще бъде разработена е структурна и интеграционна трансформация на бизнес референтна архитектура към детайлна такава, за използването на NoSQL бази от данни за финансови услуги.

ICT архитектурата, която е тествана практически се състои от система за съхранение на данни – MongoDB, интеграционен слой – Apache Spark Connector за MongoDB и сървър на специфичен NoSQL, като в случая е използвана разпределената файлова система на Hadoop.

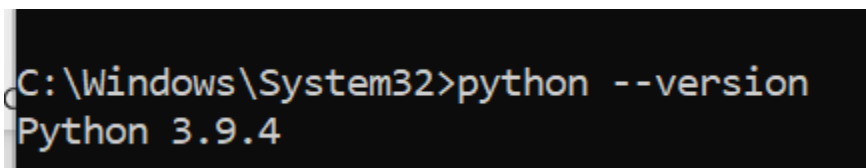
Конекторът за данни на MongoDB през Spark позволява интеграцията на документно-ориентираната база MongoDB и Apache Spark, което позволява на потребителите да реализират сложни анализи с големи сетове от данни. В допълнение на това, че конекторът позволява операции за четене, също дава възможност за запис на данни обратно към MongoDB през Spark конектора. Това дава изключителна възможност за записване на резултати от извършените обработки в Spark да бъдат записани обратно в MongoDB и след това да бъдат използвани за допълнителен анализ.

Инсталация и конфигурация на необходими компоненти за реализиране на прототипа на референтната архитектура за финансови услуги

За реализацията на физическия модел на бизнес референтната архитектура са необходими следните компоненти:

- ✓ Hadoop
- ✓ Apache Spark конектор за MongoDB
- ✓ Python

Към момента ICT архитектурата е реализирана на локално ниво - връзка на MongoDB чрез Apache Spark конекторът към Hadoop. Започва се с локална инсталация на Python, тъй като ще бъде използван PySpark. За да се инсталира и конфигурира Python се свалят необходимите файлове, които се стартират на локалната машина, чрез стандартен инсталационен софтуер, с който разполага. След завършването и, се проверява наличната версия на Python на машината, за да е сигурно, че процесът е завършил правилно и работата по следващите стъпки може да бъде продължена. В Command Prompt се изписва следната команда `python --version`, като ако инсталационния процес е преминал успешно трябва да се появи следния резултат – Фигура 30.



```
C:\Windows\System32>python --version
Python 3.9.4
```

Фигура 30 Резултат при правилна инсталация на Python

Инсталацията на Spark на локално ниво се извършва след избор на конфигурациите и сваляне на необходимите файлове на машината. При правилна инсталация на Spark и въвеждане на следната команда в Command Prompt: `C:\Spark\spark-3.1.1-bin-hadoop2.7\bin\spark-shell`, Apache Spark се стартира и се визуализира екрана по-долу – Фигура 41.

```

at org.apache.spark.SecurityManager.<init>(SecurityManager.scala:79)
at org.apache.spark.deploy.SparkSubmit.secMgr$lzycompute$1(SparkSubmit.scala:368)
at org.apache.spark.deploy.SparkSubmit.secMgr$1(SparkSubmit.scala:368)
at org.apache.spark.deploy.SparkSubmit.$anonfun$prepareSubmitEnvironment$8(SparkSubmit.scala:376)
at scala.Option.map(Option.scala:230)
at org.apache.spark.deploy.SparkSubmit.prepareSubmitEnvironment(SparkSubmit.scala:376)
at org.apache.spark.deploy.SparkSubmit.org$apache$spark$deploy$SparkSubmit$$runMain(SparkSubmit.scala:894)
at org.apache.spark.deploy.SparkSubmit.doRunMain$1(SparkSubmit.scala:180)
at org.apache.spark.deploy.SparkSubmit.submit(SparkSubmit.scala:203)
at org.apache.spark.deploy.SparkSubmit.doSubmit(SparkSubmit.scala:90)
at org.apache.spark.deploy.SparkSubmit$$anon$2.doSubmit(SparkSubmit.scala:1030)
at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:1039)
at org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to

  _____
 /_ _ _ _ _ \
| | | | | |
| |_|_|_|_|_|
 \_ _ _ _ _ /
  version 3.1.1

Using Python version 3.9.4 (tags/v3.9.4:1f2e308, Apr 6 2021 13:40:21)
Spark context Web UI available at http://LAPTOP-2FKTL05R:4040
Spark context available as 'sc' (master = local[*], app id = local-1680097593081).
SparkSession available as 'spark'.

```

Фигура 31 Резултат при правилна инсталация на Apache Spark

При правилната инсталация и стартиране на Apache Spark, освен съобщението в Command prompt, ни се предоставя потребителски интерфейс през браузъра, който може да бъде използван за наблюдението на различните обработки, които се извършват – Фигура 31, на следния адрес: <http://localhost:4040/>.

The screenshot shows the Apache Spark web UI interface. At the top, there are navigation tabs: Jobs, Stages, Storage, Environment, Executors, and SQL. The 'Stages' tab is active. Below the tabs, there's a section titled 'Stages for All Jobs' with a sub-section 'Failed Stages (1)'. A table displays the details of the failed stage:

Stage Id	Description	Submitted	Duration	Tasks: Succeeded/Total	Input	Output	Shuffle Read	Shuffle Write	Failure Reason
0	showString at NativeMethodAccessorImpl.java:0 +details	2023/03/29 16:49:47	10 s	0/1 (1 failed)					Job aborted due to stage failure: Task 0 in stage 0.0 failed 1 times, most recent failure: Lost task 0.0 in stage 0.0 (TID 0) (LAPTOP-2FKTL05R executor driver); org.apache.spark.SparkException: Python worker failed to connect back.

Фигура 32 Интерфейс на локално ниво в браузъра за наблюдение на изпълнение на задачи на Apache Spark

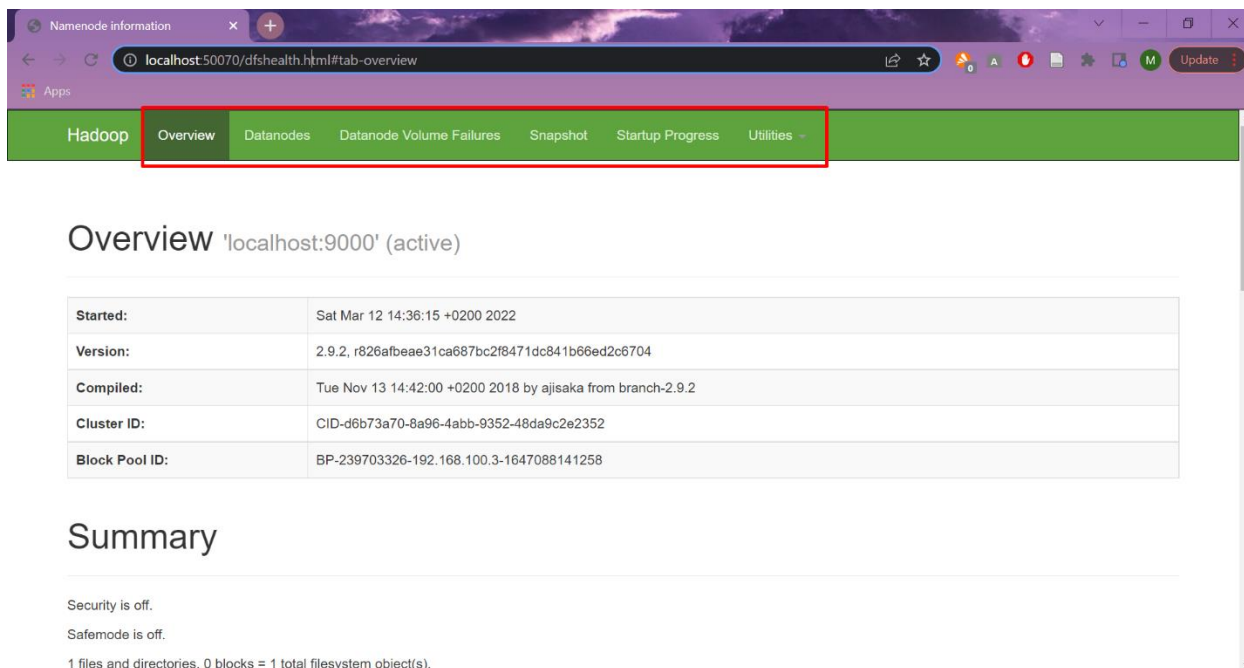
Следващата стъпка за реализацията на физическият модел на бизнес референтната архитектура, която трябва да се изпълни е инсталацията на Hadoop. Той се сваля, правят се необходимите конфигурации на машината, както се добавят и допълнителни конфигурации към някои от файловете на Hadoop, които могат да бъдат открити в приложенията на настоящия дисертационен труд. За проверка на това дали инсталацията е била успешна в

Command prompt на машината се изписва следната команда: `cd Hadoop-2.9.2\sbin` , като трябва да се появи следното нещо на Фигура 32.

```
C:\>cd Hadoop-2.9.2\sbin
C:\hadoop-2.9.2\sbin>
```

Фигура 33 Успешна инсталация на Hadoop

В браузърът на следния адрес: <http://localhost:50070/> се отваря потребителският интерфейс, чрез който при правилно стартиране на Hadoop може да се проследят клъстерите с данни и други.



Фигура 34 Интерфейс на Hadoop

Последната стъпка за реализирането на физическата архитектура е инсталацията на Apache Spark конектора за MongoDB, което се реализира чрез PySpark и следните команди `--packages`, като след това от предоставените възможности се избира `mongo-spark-connector`, който се инсталира на машината.

Стартирането на връзките се случва, чрез следния код:

```

./bin/pyspark --conf
"spark.mongodb.read.connection.uri=mongodb://localhost:27017/FinancialData.Stocks?readPref
erence=primaryPreferred"
--conf
"spark.mongodb.write.connection.uri=mongodb://localhost:27017/FinancialData.Stocks"
--packages org.mongodb.spark:mongo-spark-connector_2.12:10.1.1

```

Приложение на типични неструктурирани данни при финансови услуги

Всяка една финансова услуга работи с различни типове данни – структурирани, полуструктурирани и неструктурирани, събирани от различни източници в различни формати. Създадената референтна архитектура за финансови услуги с документно-ориентираната база MongoDB свързана чрез конектор на Apache Spark с Hadoop, работи с данни в документен вариант.

Този тип неструктурирани данни са типични за много услуги в основно банкиране, е-плащания, акции, облигации, персонализиране, е-портфейл, застраховки, клиринг на клиенти, управление на клиенти и други. По тази причина създадената референтна архитектура е тествана с данни, като договори, фишове, имейли, репорти, данни за стокови акции и облигации и други. Част от тези данни са макетни, генерирани допълнително, поради факта, че данните, с които работят финансовите услуги са изключително чувствителни и не могат да се открият такива със свободен достъп.

Тестване на референтна архитектура с финансови данни

- Източник на финансови данни – акции

Данните за и свързани с клиенти, които са генерирани, използвани и съхранявани във финансовите институции са изключително чувствителни и не бива да бъдат разпространявани свободно, като също трябва да бъдат съхранявани според изискванията за защита на личните данни. [81] [82] Финансови данни могат да бъдат открити много трудно свободно в интернет пространството, като за целите на тестването са използвани свободно достъпни данни за акции от Forbes, Nasdaq, Nyse и SP500 в json формат.

Данните Forbes, Nasdaq, Nyse и SP500 съдържат следните полета :

Полета	Стойности
--------	-----------

currency	Съответната валута
symbol	СИМВОЛ
exchangeName	Обменно име
instrumentType	Тип инструмент
firstTradeDate	Първа дата на продаване
regularMarketTime	Регулярно време на пазара
gmtoffset": -18000	
timezone	Времева зона
exchangeTimezoneName	Времева зона на обмен
regularMarketPrice	Редовна цена на пазара
chartPreviousClose	
priceHint	Подсказка за цената
currentTradingPeriod	Настоящ период за търгуване
pre	
timezone	Часова зона на предварителната продажба
start	Стартова дата
end	Крайна дата
gmtoffset	Разлика от времето по Гринуич
regular	
timezone": "EST",	Часова зона на регулярна продажба
"start": 1670855400,	Стартова дата
"end": 1670878800,	Крайна дата
gmtoffset	Разлика от времето по Гринуич
post	
timezone	Часова зона след продажба
start	
end	Крайна дата
gmtoffset	Разлика от времето по Гринуич
dataGranularity	Детайлност на данните
range	Диапазон

Таблица 5 Полета с данни от Forbes, Nasdaq, Nyse и SP500

Данните за акции от различните източници - Forbes, Nasdaq, Nyse и SP500 са над 10ГБ, които ще бъдат заредени в архитектурата, която се реализира в настоящата глава – Таблица 5.

- Зареждане на финансови данни

Създадената архитектура може да обменя данни от MongoDB през MongoDB конектора към Apache Spark и обратно. Зареждането на данни от Hadoop към Mongo се реализира, чрез PySpark.

За целта, първо е необходимо да се добавят необходимите библиотеки и да се създаде сесия на Spark със следния код:

```
spark = SparkSession.builder.FinancialTest("HadoopToMongoDB")
.config("spark.mongodb.output.uri","mongodb://localhost:27017/FinancialData.Stocks")
.getOrCreate()
```

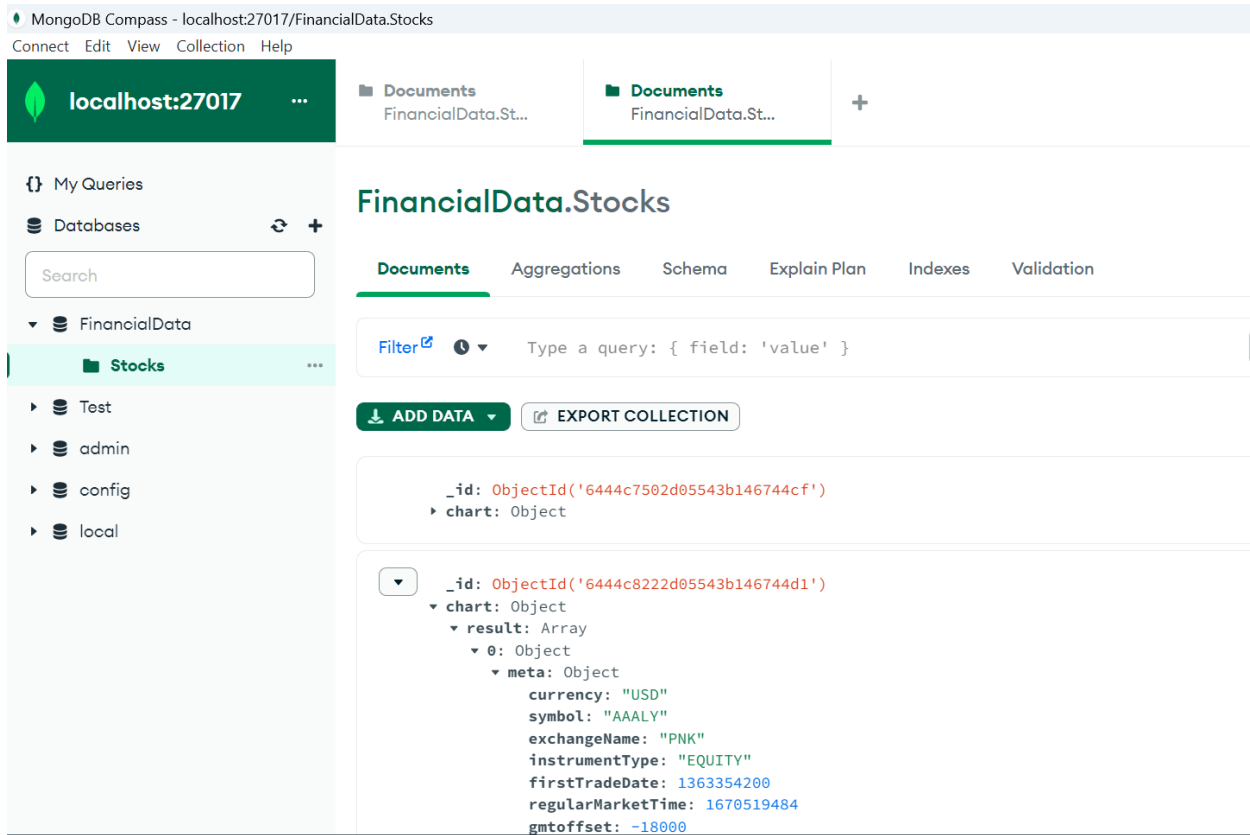
С кода по-долу зареждаме данните от Hadoop в PySpark Frame:

```
df = spark.read.format("com.mongodb.spark.sql.DefaultSource").option("uri", "mongodb://
http://localhost:50070/FinancialData.Stocks.load()
```

Като накрая записваме данните от Hadoop в MongoDB, чрез следния код:

```
df.write.format("com.mongodb.spark.sql.DefaultSource").mode("append").option("uri",
"mongodb://localhost:27017/FinancialData.Stocks") .save()
```

На Фигура 35 са визуализирани данните заредени през Apache Spark конектора на MongoDB от Hadoop.



Фигура 35 Заредени данни от Надоор към MongoDB

```

client = MongoClient()
db = client['Financial_data']
collection = db['stocks']
spark = SparkSession.builder
  .appName("Financial Stock")
  .getOrCreate()
df = spark.read.format("com.mongodb.spark.sql.DefaultSource")
  .option("database", "Financial_data")
  .option("collection", "stocks")
  .load()
df = df.filter((col("symbol") == "ACGL") & (col("firstTradeDate") >= "2017-05-02") &
(col("firstTradeDate") <= "2017-10-10"))

```



```
result = df.groupBy(year("date").alias("year"), month("date").alias("month"))
    .agg({"close": "avg"}) .orderBy("year", "month")
result.write.format("com.mongodb.spark.sql.DefaultSource").option("database",
"Financial_data") .option("collection", "stock_analysis") .mode("overwrite") .save()
client.close()
spark.stop()
```

Създадената програма на PySpark използва данните от акциите, които бяха заредени в началото на експеримента с референтната архитектура от Hadoop към MongoDB. Първата стъпка е, че се прави връзка с MongoDB, а след това се отваря сесията със Spark. Данните се филтрират по наименованието на акциите, а в последствие се търсят само онези, чиято първа дата на продажба е между зададения интервал. След това програмата групира резултата по дата и извежда усреднената стойност на цената на конкретната акция за целия период в интервала. Данните се записват в MongoDB базата и сесията се затваря.

Изводи

Проектиран е концептуален, логически и физически модел на бизнес референтна архитектура, които са създадени с помощта на подхода за създаване на референтна архитектура в Глава I. На база тези модели, в Глава IV беше реализиран успешно физическия модел на бизнес референтната архитектура за финансови услуги с NoSQL бази от данни и по-конкретно свързването на документно-ориентираната база от данни – MongoDB, чрез Apache Spark конектора за MongoDB с Hadoop. Реализираната архитектура позволява съхранението, обмена и анализа на данни с възможностите, които предлагат Apache Spark и MongoDB.

Заклучение

С развитието на технологиите, дигитализацията и дигиталната трансформация, която се случва във всички сфери, данните се генерират все по-бързо, с все по-големи обеми и по-голяма скорост. Сфери като финансовите услуги работещи със исторически системи, които са базирани върху релационни бази от данни, са принудени да се развиват и да мигрират своите системи към нерелационни бази от данни, за да успеят да се адаптират към

новите изисквания на своите клиенти и да се справят с нарастващата конкуренция, както и с бързото развитие на технологиите и новите услуги, които предлагат.

Дисертационния труд предлага разработен метод за оценка и избор на NoSQL бази от данни, базиран върху анализ на литературни източници и реални практически примери. Методът води до създаването на инструмент за оценка и избор на нерелационни бази от данни. Това става основа за проектирането на референтна архитектура на NoSQL бази от данни, която е практически приложена, чрез свързването на MongoDB с конектора на Apache Spark за Mongo и съответно чрез него свързване с Hadoop.

В началото на настоящия дисертационен труд бяха дефинирани три хипотези:

Хипотеза 1:

Възможно е да се създаде референтна архитектура, приложима за съхранение, обработка и работа с Големи данни свързани с финансови услуги.

Хипотеза 2:

Възможно е да се даде количествена оценка за референтна архитектура, като се създаде метод за оценка, както и инструмент за автоматизирането ѝ прилагане.

Хипотеза 3:

Възможно е да се постигне ефективна интеграция на компонентите на референтната архитектура за финансови услуги с NoSQL бази от данни за работа с Големи данни.

С оглед на извършената научно-изследователска работа, както и на реализирания практически прототип, може да се заключи, че хипотезите са доказани, задачите дефинирани в началото на настоящия дисертационен труд са изпълнени, което ни дава основание да считаме, че целта на дисертацията е изпълнена.

Списък на публикациите по темата на дисертационния труд

- **Публикувани**
- Mariana Kovacheva, Storing Big Data in NoSQL Databases Compared to SQL – Advantages and Problems, 10TH INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGY AND STATISTICS IN ECONOMY AND EDUCATION ICAICTSEE – 2020, November 27 – 28th, 2020, University of National and World Economy, Sofia, Bulgaria, ISSN 2367-7635 (PRINT), ISSN 2367-7643 (ONLINE), достъпна на: <https://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2020.pdf>
- Mariana Kovacheva, Ivona Velkova, Digitalization in Bulgarian Higher Education – Present and Future Opportunities, 4th International Academic Conference on Education, 2021, 10th December 2021, Barcelona, Spain, ISBN: 978-609-485-239-8), достъпна на: <https://www.dpublication.com/proceeding/4th-iaceducation/#Table-of-Contents>
- Mariana Kovacheva, Evaluation of NoSQL Databases for Digital Financial Services Implementation, AUTOMATICA and INFORMATICS, ISSN 0861-7562 (Print), ISSN 2683-1279 (Online), Year LV No. 4/2022, достъпна на: <https://sai-bg.com/wp-content/uploads/2023/01/AI-4-2022.pdf>
- Mariana Kovacheva, Creation of an automated tool for evaluation and selection of NoSQL databases intended for financial services based on predefined criteria, VIII INTERNATIONAL SCIENTIFIC CONFERENCE HIGH TECHNOLOGIES. BUSINESS.SOCIETY 2023, 06-09.03.2023, BOROSETS, BULGARIA, ISSN 2535-0005(PRINT), ISSN 2535-0013 (ONLINE), достъпна на: <http://hightechsociety.eu/sbornik/2023.pdf>
- **Очаква се публикуване**
- Mariana Kovacheva, Classification, comparison, and criteria for choosing NoSQL databases.

5. Литература

- [1] “What is Financial Services?” <https://insights.btoes.com/what-is-financial-services> (отворен на Яну 19, 2022)
- [2] “Big Data in Banking – Leapfrogging into Digital Banking Era,” TechVidvan. <https://techvidvan.com/tutorials/big-data-banking/#:~:text=Banks%20generate%20various%20types%20of,added%20to%20the%20bank's%20database>. (отворен на Фев 24, 2022)
- [3] Paul Golden, “Big data comes to FX,” Euromoney, Nov. 03, 2014. <https://www.euromoney.com/article/b12kky4qp36mf4/big-data-comes-to-fx> (отворен на Яну 19, 2022)
- [4] Kate Blumberg, “7 Big Data Use Cases in Financial Services and Benefits of Data Science,” SAFEGRAPH, Nov. 04, 2021. <https://www.safegraph.com/blog/top-big-data-use-cases-financial-services> (отворен на Фев 24, 2022)
- [5] Market Trends, “Use of Big Data Analytics in Investing,” Analytix Insight, Jun. 02, 2022. <https://www.analyticsinsight.net/use-of-big-data-analytics-in-investing/>
- [6] Ritesh Pathak, “7 Uses of Big data in the insurance industry,” Analytic Steps, Jan. 09, 2021. <https://www.analyticssteps.com/blogs/7-uses-big-data-insurance-industry> (отворен на Фев 28, 2022)
- [7] JULIA KAGAN, “Insurance: Definition, How It Works, and Main Types of Policies,” Investopedia, Jul. 18, 2022. <https://www.investopedia.com/terms/i/insurance.asp#toc-what-are-the-four-major-types-of-insurance> (отворен на Фев 24, 2022)
- [8] JULIA KAGAN, “How Do Commercial Banks Work, and Why Do They Matter?,” Investopedia, Jan. 23, 2023. <https://www.investopedia.com/terms/c/commercialbank.asp#:~:text=Commercial%20banks%20provide%20basic%20banking,such%20as%20safe%20deposit%20boxes> (отворен на Фев 24, 2022)
- [9] “Financial services,” Wikipedia. https://en.wikipedia.org/wiki/Financial_services (отворен на Май 28, 2022)
- [10] “Big Data in Finance - Your Guide to Financial Data Analysis,” Talend. <https://www.talend.com/resources/big-data-finance/> (отворен на Юни 15, 2022)

- [11] MEDICI, “Why Financial Services Should Look to NoSQL.” [Online]. Available: http://pages.aerospike.com/rs/229-XUE-318/images/Aerospike_Wp_Why_Financial_Services_Should_Look_to_NoSQL.pdf (отворен на Март 11, 2023)
- [12] Kazim Hussain and Elsa Prieto, “Big Data in the Finance and Insurance Sectors,” in *New Horizons for a Data-Driven Economy*, Springer Open, p. 222. [Online]. Available: <https://library.oapen.org/bitstream/handle/20.500.12657/27764/1002241.pdf?sequence#page=222>
- [13] Sabrina Sicari, Alessandra Rizzardi, and Alberto Coen-Portisini, “Security & privacy issues and challenges in NoSQL databases,” vol. 206, Apr. 2022, doi: <https://doi.org/10.1016/j.comnet.2022.108828>.
- [14] “Big Data Insight,” SAS. https://www.sas.com/en_us/insights/big-data/what-is-big-data.html (отворен на Юни 16, 2022)
- [15] “Big Data Characteristics,” JavaTPoint. <https://www.javatpoint.com/big-data-characteristics>
- [16] Ravi Kiran, “Big Data Characteristics: Know the 5’Vs of Big Data,” Mar. 13, 2023. <https://www.edureka.co/blog/big-data-characteristics/> (отворен на Май 28, 2022)
- [17] Mark Smallcombe, “Structured vs Unstructured Data: 5 Key Differences,” Integrate.io, Feb. 16, 2023. <https://www.integrate.io/blog/structured-vs-unstructured-data-key-differences/> (отворен на Юни 18, 2022)
- [18] “Unstructured Data,” MongoDB. <https://www.mongodb.com/unstructured-data> (отворен на Юни 18, 2022)
- [19] Bernard Marr, “What’s The Difference Between Structured, Semi-Structured And Unstructured Data?,” Forbes, Oct. 18, 2019. <https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/?sh=58dbd11b2b4d> (отворен на Май 28, 2022)
- [20] “NoSQL,” TechTerms, Aug. 27, 2013. <https://techterms.com/definition/nosql> (отворен на Юни 1, 2022)
- [21] “What is a Document Database?,” MongoDB. <https://www.mongodb.com/document-databases> (отворен на Юни 15, 2022)
- [22] “What Is a Key-Value Database?,” AWS. <https://aws.amazon.com/nosql/key-value/> (отворен на Юни 15, 2022)
- [23] Margaret Rouse, “Columnar Database,” Techopedia, Aug. 16, 2013. <https://www.techopedia.com/definition/13782/columnar-database> (отворен на Юли 10, 2022)
- [24] “What Is a Graph Database?,” AWS. <https://aws.amazon.com/nosql/graph/> (отворен на Юли 10, 2022)

- [25] Mitko Radoev , "A Comparison between Characteristics of NoSQL Databases and Traditional Databases," Computer Science and Information Technology, Vol. 5, No. 5, pp. 149 - 153, 2017. DOI: 10.13189/csit.2017.050501.
- [26] "Apache Hadoop," Hadoop. <https://hadoop.apache.org/> (отворен на Август 29, 2022)
- [27] "What is Apache Hadoop?," IBM. <https://www.ibm.com/analytics/hadoop> (отворен на Август 29, 2022)
- [28] "What is Hadoop?," IntelliPaat, 09 Mar, 2023. <https://intellipaate.com/blog/what-is-hadoop/> (отворен на Август 29, 2022)
- [29] "What is Hadoop," JavaTPoint. <https://www.javatpoint.com/what-is-hadoop> (отворен на Август 29, 2022)
- [30] "Reference Architecture," HPE. <https://www.hpe.com/us/en/what-is/reference-architecture.html> (отворен на Септ 16, 2022)
- [31] Lavanya Rathnam, "What Is Reference Architecture?," TechGenix, Aug. 02, 2022. <https://techgenix.com/reference-architecture-guide/> (отворен на Септ 30, 2022)
- [32] "Banking industry architecture," IBM. <https://www.ibm.com/cloud/architecture/architectures/banking/reference-architecture> (отворен на Октомври 2, 2022)
- [33] "MongoDB for Financial Services," MongoDB. <https://www.mongodb.com/industries/financial-services> (отворен на Октомври 2, 2022)
- [34] "Financial Management System Database," Couchbase. <https://www.couchbase.com/solutions/nosql-for-financial-services/> (отворен на Октомври 15, 2022)
- [35] Max Ivashchenko, Jacky Wu, and Rohit Talluri, "DynamoDB FSI Service Spotlight," AWS, Aug. 17, 2022. <https://aws.amazon.com/blogs/industries/dynamodb-fsi-service-spotlight/> (отворен на Октомври 23, 2022)
- [36] "Redis Enterprise for financial services," Redis. <https://redis.com/solutions/industries/financial-services/> (отворен на Октомври 27, 2022)
- [37] "Apache HBase Features," Apache HBase. <https://hbase.apache.org/index.html> (отворен на Февруари 12, 2023)
- [38] Rich Edwards, "Exploring Common Apache Cassandra Use Cases," Datastax, Jun. 12, 2021. <https://www.datastax.com/blog/exploring-common-apache-cassandra-use-cases> (отворен на Декември 12, 2022)

[39] Alex Bekker, “When to use Cassandra and when to steer clear,” Towardsdatascience, Aug. 14, 2018. <https://towardsdatascience.com/when-to-use-cassandra-and-when-to-steer-clear-72b7f2cede76> (отворен на Яну 16, 2023)

[40] Nav Mathur, “Graph Technology for Financial Services How Top Financial Firms Harness Connected Data to Increase Their Bottom Line,” Neo4j. https://go.neo4j.com/rs/710-RRC-335/images/Neo4j-in-Financial%20Services-white-paper.pdf?_ga=2.90740444.799262140.1679753501-1625560426.1676516677&_gl=1*1p7o7b8*_ga*MTYyNTU2MDQyNi4xNjc2NTE2Njc3*_ga_DL38Q8KGQC*MTY3OTc1MzUwMC40LjEuMTY3OTc1Mzg2Ny42LjAuMA (отворен на Яну 16, 2023)

[41] “Payment Service Provider Fights Fraud faster and more efficiently with Orient DB Multi-Model Graph Database.,” Orient DB. https://orientdb.com/wp-content/uploads/OrientDB_99Bill_case_study.pdf (отворен на Яну 29, 2023)

Design of business reference architectures aimed for using NoSQL databases

Autoabstract of a dissertation work for the acquisition of a scientific and
educational degree "PhD"

PhD student: Mariana Boykova Kovacheva

Supervisor: Prof. DSc. Valentin Kisimov

Department of "Information Technologies and Communications"
Faculty of Applied Informatics and Statistics
University of National and World Economy

Sofia 2023

The dissertation work was discussed and admitted to public defence at a regular meeting of the departmental council of the "Information Technologies and Communications" department of the University of National and World Economy - Sofia (UNWE), including members of the department. It was held on March 30, 2023. The author of the dissertation work is a part-time PhD student at the department.

The public defence of the dissertation is scheduled for June 27, 2023, at 3:00 p.m.

The defence materials are published on the website of the University of National and World Economy - Sofia (<https://www.unwe.bg>) and are available to those interested in the "Scientific Councils and Competitions" sector of the "Science" department of the UNWE.

Contents

General characteristics of the dissertation work	76
Actuality	76
1. Object, goals, and tasks of the dissertation work	77
2. Volume and structure of the dissertation work	79
Brief presentation of the dissertation work	81
Introduction.....	81
1. Chapter I. Analysis of financial services and the Big Data generated by them.	
Opportunities to use NoSQL databases.	81
Financial services - definition and types	81
NoSQL databases – definition, types, and characteristics.....	85
Big Data Systems – definition, advantages, disadvantages, architecture.....	86
Reference architecture - definition, types, and approach to creation.....	87
Conclusion	89
2. Chapter II. A Method for Selecting and Designing a Business Reference	
Architecture for NoSQL Databases for Financial Services	89
Financial services as a result of the currently developed business applications	89
Definition of a method for analysis and evaluation of a reference architecture	98
Designing a Business Reference Architecture for Financial Services with NoSQL	
Databases.....	107
Conclusion	114
3. Chapter III. NoSQL database selection tool for financial services	114
User roles – User/Company	115
User role Administrator/Expert.....	118
Conclusion	124
4. Chapter IV. Application of a method to design a reference architecture for	
financial services.....	125
An architecture for executing financial services with NoSQL databases.	125
Installation and configuration of necessary components to implement the financial	
services reference architecture prototype.	125
Application of typical unstructured data in financial services	129
Testing the reference architecture with financial data.....	129
- Source of financial data - stocks	129
- Loading financial data.....	130

Conclusion	133
Conclusion	133
List of publications on the topic of the dissertation.....	134
- Published	134
- Expected	135
5. Literature	136

General characteristics of the dissertation work

2. Actuality

Nowadays, when change and digital transformation are a necessity for the businesses to remain competitive, companies that fail to keep up with change, risk being left behind or out of business. Technologies developing at a faster pace from simple phones to sensors collecting data and being used almost everywhere in our daily lives to computers, robotics, Artificial Intelligence, cloud technologies, Big Data, Internet of Things, and more are considered the basis of digital transformation in various spheres.

Digitization and digital transformation cover various types of processes in our lives, progressing at an extremely fast pace. The adaptation of new technologies helps to increase efficiency and productivity in the digitization of manual processes. There are countless benefits to digital transformation, but one of the most important factors is that it helps businesses become more efficient and organized. This process has the potential to make organizations more agile, efficient and customer centric.

In the last decade, managing exponentially growing amounts of data has become an increasingly challenging task. With the development of technologies, the digitization of processes and the digital transformation of businesses, it leads to the generation of huge amounts of data that must be stored, managed, and analysed through appropriate methods. Technology development and data generation represent the initial and easiest step. The challenges start with storing, processing, analysing and deriving results from the collected data.

In the past, when the development of information technology was initiated, the main type of data collected was structured. With the digitization of more and more businesses and digital transformation, the generation of different types of data - semi-structured and unstructured - has also started. Unlike structured data, the other two types are extremely challenging to store and process. The usual relational databases used for the storage of structured data are not suitable for semi-structured and unstructured, which leads to the need to create and develop new ways of storing different types of data. One of the key requirements for Big Data storage is that it must handle massive amounts of data, with storage capacity continuing to grow without disrupting the workflow.

In recent years, the financial sector has been developing and digitizing more and more, which leads to the need for changes. For financial institutions to meet the demands, needs and

expectations of their customers, it is of utmost importance to make changes and adapt quickly to evolving technologies. In addition to this, these types of institutions must manage extremely large volumes of data while making changes to their systems without causing data loss or problems for their customers. Financial services offered by banks, credit unions, accounting companies, insurance companies, investment funds, stock exchanges and others need databases that can adapt to the automation needs that non-relational databases can meet. NoSQL databases can handle the storage and processing of large amounts of data, with their scalability and providing better data processing performance.

5. Object, goals, and tasks of the dissertation work

The **subject** of this dissertation are the business reference architectures aimed for the use of non-relational (NoSQL) databases related to Big Data systems for financial services.

The **aim** of the dissertation work is the design of a business reference architecture aimed for the use of NoSQL databases connected to Big Data systems, to be applied and tested with data for a specific financial service and the definition of a method for evaluating and selecting NoSQL databases for financial services.

The **subject** of the dissertation is the design of a reference architecture for the selection of NoSQL databases related to Big Data systems aimed for financial services and the method of selection and evaluation of non-relational databases for financial services.

The **research question** posed in this dissertation is "How to create a business reference architecture aimed for the use of NoSQL databases related to Big Data systems in the field of financial services, how can a quantitative assessment to be obtained, through which we can choose an appropriate non-relational database for the reference architecture? "

Tasks that are performed in the process of realizing the set goal of this dissertation are the following:

- ✓ Research and analysis of reference architecture for financial services.
 - Overview and analysis of current problems related to the storage of different types of Big Data and reference architectures related to the use of non-relational databases aimed at financial services.
 - Defining criteria for selecting a NoSQL database related to Big Data systems aimed at the use of financial services.

- ✓ Defining a quantitative assessment approach for non-relational database selection in the financial services industry.
 - Selection of criteria for evaluation of NoSQL databases in the field of financial services.
 - Designing a conceptual, logical and physical model of a reference architecture for financial services with NoSQL databases
- ✓ Development of a tool for evaluation and selection of NoSQL databases related to Big Data systems aimed at financial services for use in the reference architecture.
 - Designing a tool for evaluation and selection of a non-relational database related to Big Data systems aimed at financial services.
 - A tool for selecting and evaluating NoSQL databases for financial services.
- ✓ Practical application of the method for designing reference architectures for financial services
 - Creation of an experimental prototype of a reference architecture related to a Big Data system aimed at financial services.
 - Practical application of the experimental prototype
 - Evaluation of the results of the practical application

2. Working hypotheses of the dissertation work

Hypothesis 1:

It is possible to create a reference architecture applicable to the storage, processing and operation of Big Data related to financial services.

Hypothesis 2:

It is possible to quantify a reference architecture by creating an evaluation method as well as a tool for automation and implementation.

Hypothesis 3:

It is possible to achieve an effective integration of the components of the reference architecture for financial services with NoSQL databases for working with Big Data.

3. Contributions of the dissertation work

- **Scientific**
 - A method for quantitative evaluation of NoSQL databases for implementation of digital financial services was created.

- Formulation of indicators and metrics for selecting NoSQL databases for financial services
- Designing a conceptual, logical, and physical model of a reference architecture for financial services with NoSQL databases
- **Scientific and applied.**
 - Principles for designing and creating a NoSQL database selection tool for financial services.
 - Structural and integration transformation of a business reference architecture to a detailed architecture for using NoSQL databases.

6. Volume and structure of the dissertation work

The dissertation consists of an introduction, four chapters, a conclusion, contributions, publications, references, appendices, a list of figures, a list of tables, a list of figures and appendices. It has a total volume of 122 pages (110 pages without the appendices). The sources used are 85, of which 1 is in Bulgarian, 84 in English and 73 internet addresses. The exhibition includes 45 figures, 5 tables and 1 graph.

The **introduction** presents the actuality, object, goals, tasks and working hypotheses of the problems considered in the dissertation work. Three hypotheses are formulated, and its goal is "the design of a business reference architecture aimed for the use of NoSQL databases connected to Big Data systems, to be applied and tested with data for a specific financial service and the definition of an evaluation method and selection of NoSQL databases for financial services.'

The **first chapter** has an overview nature. It introduces the main concepts and definitions used in the dissertation. A brief analysis of Hadoop, its main advantages and disadvantages has been made. Data types are considered, as are the types of NoSQL databases, and reference architectures are analysed in general and more specifically for financial services.

The **second chapter** presents a method for evaluating and selecting NoSQL databases for financial services. It is based on an analysis of literary sources, as well as on practically implemented solutions. Criteria were derived and an expert assessment was made for the implementation of the selection. A conceptual, logical, and physical model of a reference architecture for financial services was designed.

In the **third chapter**, a tool for evaluating and selecting NoSQL databases for financial services is designed and built, based on the method for evaluating and selecting non-relational databases created in Chapter II.

The **fourth chapter** represents the experimental part of the dissertation work. It examines the practical applicability of the NoSQL evaluation and selection method, as well as the realized physical model of the designed reference architecture, which is tested with financial data.

The **conclusion** summarizes that because of the developments, the formulated hypotheses have been proven, expressing the intention to continue research and development in the field of non-relational databases, as well as the development of the NoSQL database evaluation and selection tool, as well as continuation of the practical work on the implemented business reference architecture.

Brief presentation of the dissertation work

2. Introduction

The **introduction** presents the actuality, object, goals, tasks and working hypotheses of the problems considered in the dissertation work. Three hypotheses are formulated, and its goal is "the design of a business reference architecture aimed at the use of NoSQL databases connected to Big Data systems, to be applied and tested with data for a specific financial service and the definition of an evaluation method and selection of NoSQL databases for financial services.'

6. Chapter I. Analysis of financial services and the Big Data generated by them. Opportunities to use NoSQL databases.

Financial services - definition and types

Financial services are economic services that are provided by the financial industry, which encompasses a wide range of businesses involved in managing money including credit unions, banks, credit card companies, insurance companies, accounting firms, consumer finance, stockbrokers, investment funds, government sponsored enterprises. [1]

Financial services are divided into banking, exchange, investment, insurance, financial export and other financial services. Each of these basic financial services is divided into further sub-services.

- ✓ **Banks** - data for analysis in the banking sector comes from various sources: customer personal information, account details, customer transactions, customer complaints and service, information pages on social networks, market sentiment, product introductions, etc. With the help of Big Data, banks can overcome major business challenges such as profitability and performance through Big Data analysis, also helping banks reduce risk and cost in customer acquisition, predicting risk in loans and mortgages. The potential analysis of Big Data in the banking industry would lead to fraud prevention, customer identification and acquisition, customer retention, improved customer experience, optimized operations, improved product design, increased transparency, regulatory compliance. [2]
- ✓ **Exchange services/abroad/** are offered by banks and special currency brokers. Foreign exchange services generate extremely large volumes of data – structured, semi-structured and unstructured – and the information that remains hidden inside could be useful to market participants. [3]

- ✓ **Investment services** are a wide range of activities related to investment in various financial instruments. To reduce the risk associated with investments, investors are using various technological solutions based on Big Data to make the right decisions. Using this type of tools, they can make informed investment decisions by gathering available information, identifying industry trends and properly managing assets. In addition to standard structured information, investment services can easily collect audio files, which are unstructured data. This type of data can help generate reports faster, as well as extract the most important information for stakeholders. [4] [5]
- ✓ **Insurance services** are a contract, represented by a policy, between two parties in which the policyholder receives financial protection or reimbursement against losses from the insurance company. For an extremely long period of time, insurance services have been based on traditional business models, but with the ongoing digital transformation and technology development, the industry has started to use progressive technologies and digital platforms that allow insurance companies to track, measure and control risk. Companies for travel insurance, health insurance and more rely on statistics to segment their customers well. Accident statistics, information about the insured, as well as data collected by third parties, help group people into different risk categories, prevent fraud and optimize costs. The collected structured, semi-structured and unstructured data can be used and analysed for pricing, rating, marketing and more. [6] [7]
- ✓ **Financial export** is a financial service provided by a local company to a foreign company or individual.
- ✓ **Other financial services.** [8] [9]

The financial industry uses more systems with relational databases, the ongoing digitization and digital transformation forces them to transform their work to meet the needs of their customers. Most companies in the financial industry are storing Big Data even now, but they do not know how to extract its potential, as there are semi-structured and unstructured ones. First, organizations need to understand the value of the data collected and what it means for customers and business processes. Since legacy systems cannot support semi-structured and unstructured data, new solutions to this problem must be sought. Big data in the financial sphere has in recent

years led to major innovations in technology, enabling the creation of convenient, personalized and secure solutions. Stock and bond trading, investments, fraud detection and prevention, and risk analysis are just some of the financial services that benefit from Big Data. [10]

In the past, Big Data was not inherent to financial institutions, so a long process of change is required, requiring both technological and behavioural change. The presence of Big Data, in particular semi-structured and unstructured, leads to the need to find a suitable way of storage, where non-relational (NoSQL) databases intervene. [10]

Non-relational databases offer not only flexibility for the applications that are built on them, but also scalability at a low cost, something that is not possible with relational databases. The need for real-time processing and analysis for authentication, authorization, detection of fraud attempts leads to the need for NoSQL databases.

- **Improved fraud prevention and monitoring of financial crimes.** This is one of the biggest advantages of non-relational databases with the analysis that can be performed in real-time and is extremely important for financial services - banks, credit, and debit cards. All these institutions work with extremely sensitive information that is of utmost importance to understanding and satisfying customers. Fraud in financial services is on the rise, which can be prevented through analytics, anomaly analysis, classification, and clustering, which can be realized with non-relational databases. [12]
- **Real-time digital authorization.** Of utmost importance for financial services to be sure that fraud does not occur and for the real identification of the people with whom it is communicated. NoSQL databases have such capabilities due to reduced read latency, and authorization can happen almost immediately. [11][12]
- **Personalized offers.** It is important for customers to receive personalized offers that meet their needs. This is no different in financial services – insurance policies based on personal risk, interest rates based on repayment and capacity, etc. To implement and present personalized offers to customers, it is necessary to collect information - behaviour, social media engagement, cookies, online shopping, etc., all data coming from various sources, and for this purpose, non-relational databases come to the rescue. NoSQL databases help in fast real-time analysis so that customers can be provided with relevant personalized offers. [12] [13]

- **Provide pre-trade assistance.** Studies show that information collected from various sources, such as data can be structured, semi-structured and unstructured, and thus risk assessment can be made, and decisions related to the purchase of a product can be made. [11]
- **Target marketing.** Collected search, buy, offer information and data is necessary to perform real-time analysis and offer customers the right ads. Target marketing is also extremely important when it comes to financial services, because when the data is analyzed correctly, it gives companies knowledge about the needs, requirements and expectations of customers, and determines which is the perfect product to offer and subsequently grow on sale. [11] [13]

Big data – characteristics and types

Big Data is the term that describes data that is so large and complex that it is difficult or in some cases impossible to process using traditional methods. The collection of such huge amounts of data is useful for performing better and broader analyzes of given problems that are important to companies. [14] Big data is such large volumes that cannot be processed by traditional methods, such as relational databases. The main characteristics they have are the five V's. [15] [16]:

- ✓ **Variety** of data and sources from which it is collected. Nowadays, the sources from which data can be collected are emails, images, PDF files, audio files, video etc. and the data can be structured, semi-structured and unstructured depending on the source it comes from.
- ✓ **Volume** - extremely large quantities, because all systems, devices, social networks and processes generate data in large quantities, not only daily, hourly, but also every second.
- ✓ **Velocity** - the generation of Big Data happens at a high speed. This characteristic defines the rate at which data is generated in real time and the rate at which it is processed.
- ✓ **Value** - the data must be important and valuable. In order to be collected, stored, processed and analysed in a specific way, it is important that they are useful for the business.
- ✓ **Veracity** - the data must be accurate and reliable, and this can be proven with the source of the data.

Depending on the sources from which the data is collected, there are 3 types – structured, semi-structured and unstructured. [17] [18] [19]

Type of data	Description	Examples
Structured	A predefined tabular model, with relationships between rows and columns.	accounting transactions, address details, demographic information, customer rating, locations from various devices, information collected from sensors and smart devices.
Semi-structured	This type of data has defined persistent characteristics but does not have a precise structure as expected from relational databases.	Emails
Unstructured	Tabular storage is not possible, and a predefined model cannot be created. Due to the lack of structure, searching, managing, and analysing unstructured data is extremely difficult.	Images, video, audio, PDF files, content from social networks

Table 1 Types of data, description, and examples

NoSQL databases – definition, types, and characteristics

There are four types of NoSQL databases that can solve different problems based on a company's needs. NoSQL databases are non-relational databases that use key-value to store, access and retrieve data instead of the traditional way of using column and row tables. Non-relational databases do not need a predefined schema and thus can store different types of data. [20]

Type of database	Description	Example
Document-oriented databases	Document-oriented databases store information in documents. This type of database has a flexible schema that allows the data model to evolve along with the application. It can scale horizontally. [21]	MongoDB, CouchBase, OrientDB

Key-value databases	Key-value databases use the most common key-value methods for data storage. This type of storage is optimized for reading and writing data. Key-value databases have good horizontal scalability. [22]	Amazon DynamoDB, Redis
Wide-column databases	Wide-column databases organize data in a way that allows for more efficient storage and faster execution of queries. This type of database is easily partitioned, allowing large data sets to be distributed across multiple nodes for high performance and low latency. [23]	HBase, Cassandra
Graph databases	Graph databases store nodes and links instead of tables or documents. The data is stored within a predefined model allowing great flexibility. [24]	Neo4j, OrientDB

Table 2 Types of databases, definition, and examples

With the digitization of processes, the generation of huge volumes of different types of data, the need for non-relational databases also developed, since relational ones do not have the necessary parameters for storing semi-structured and unstructured data. The main characteristics of non-relational databases that we will consider are scalability, flexibility, no need for a predefined schema, distributed processing, and durability. [25]

Big Data Systems – definition, advantages, disadvantages, architecture

Hadoop enables distributed processing of large data sets spread across multiple clusters of computers using simple programming models. Apache Hadoop is designed to scale from single to multiple server machines, each offering local computing and storage. Implemented in this way, Hadoop is a cost-effective solution for storing and processing large amounts of structured, semi-structured and unstructured data without format requirements. [26] [27] The advantages of Apache Hadoop are the following: cost, scalability, flexibility, speed, fault tolerance, high performance, and minimal network traffic. Disadvantages of Apache Hadoop are as follows – small file problem, vulnerability, security, batch processing, high and low performance processing with small file processing.

Apache Hadoop has a master-slave topology and allows the storage of large volumes of data in a system of nodes. In this kind of topology, the master node's job is to allocate a specific task to various slave nodes and allocate the resources, with the real work and computation being done by the slave nodes. The Hadoop architecture enables parallel processing that uses several components: Distributed storage layer – HDFS (Hadoop Distributed File System), Cluster resource management - Hadoop Yarn, Processing layer - MapReduce, Apache Spark and application programming interface - HBase, Solr , shown in Figure 1 below. [28] [29]

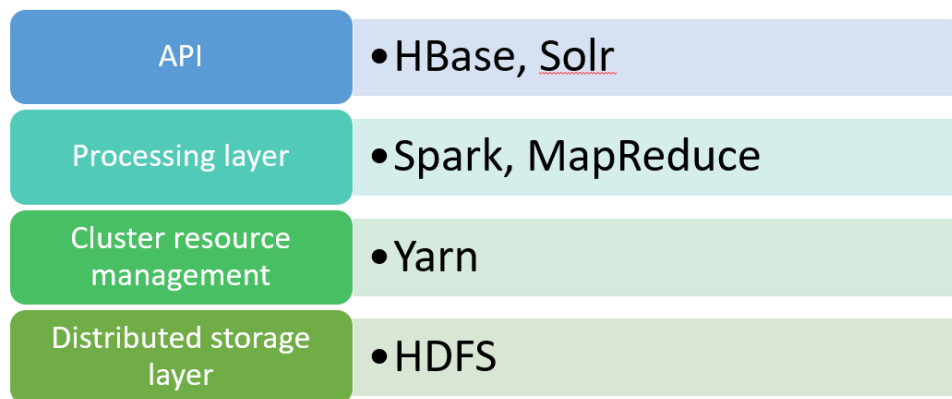


Figure 1 Layers of Hadoop architecture

Reference architecture - definition, types, and approach to creation

A reference architecture is a document or set of documents that provide a recommended structure and integrations of IT products to form a solution. RA incorporates industry best practices, usually offering the optimal method for specific technologies. The reference architecture offers IT best practices in an easy-to-understand format that guides the implementation and use of complex technology solutions. [34] Using best practices when creating reference architectures improves efficiency, meets regulatory requirements, and reduces the chance for errors. A reference architecture helps make decisions about choosing the best model and way to create a software architecture to meet business goals. [35]

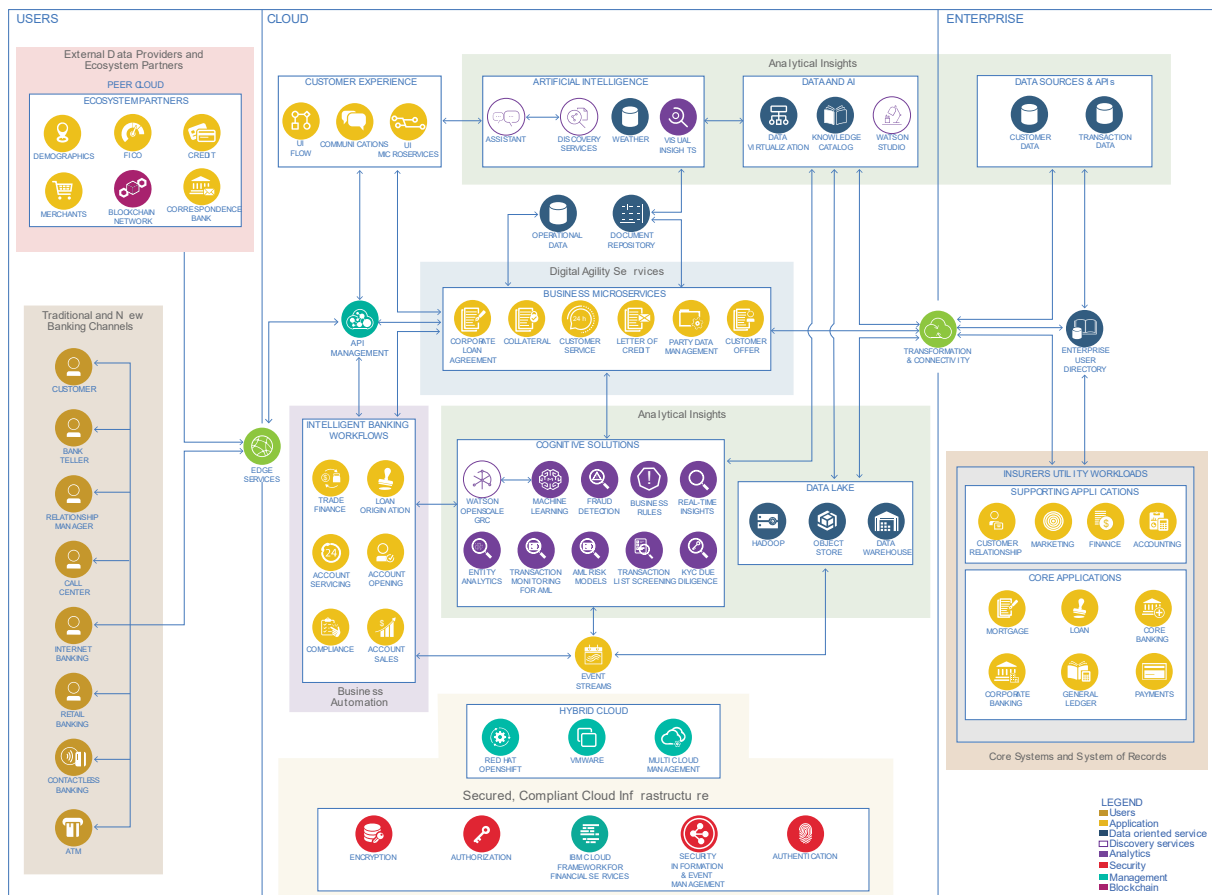


Figure 2 IBM Banking Services Reference Architecture. Source: [37]

The Banking Industry Reference Architecture uses industry standard methods and models to help achieve greater alignment between IT and industry. Financial institutions can follow a consultative method to identify opportunities for improvement and innovation using the Banking Industry Architecture Network (BIAN), IBM Information Framework (IFW) and IBM Components for Business models (IBM Component Business Models - CBM) for financial services. These assets provide industry-specific frameworks that can be used for the overall modernization of the banking program – Figure 2. [37]

Creating a reference architecture is an extremely challenging task in the absence of predefined steps or a process to create it. There are five main steps that must be taken: identifying the purpose, formulating principles, setting technical rules and standards, building rules and standards, and applying context. [35]

Conclusion

Digitization and digital transformation have led to many changes in all spheres, and the automation of processes and the use of new technologies such as the Internet of Things, Artificial Intelligence, sensors and others have led to the generation of huge amounts of Big Data of various types - structured, semi-structured and unstructured , and this in turn brings the need for new ways of storage and processing, as relational databases cannot handle semi-structured and unstructured data.

Financial services, on the other hand, are primarily based on relational database systems, which brings the need for change. This type of service is divided into banking, exchange, investment, insurance, financial export, and other financial services, and with the advent of new technologies comes Big Data and the need to store it in suitable places that can handle the storage and subsequent actions with semi-structured and unstructured data, where NoSQL databases and Big Data systems intervene.

From the financial services presented above and the types of data that are generated from them, the ways of storage now, as well as the available reference architectures for this type of services, the possibility of creating a method of evaluation and selection of NoSQL databases for financial services. The method, first, can be applied practically, by creating a tool based on the criteria and evaluations, as well as create a practical implementation, through an experimental prototype of a reference architecture related to a Big Data system aimed at financial services.

7. Chapter II. A Method for Selecting and Designing a Business Reference Architecture for NoSQL Databases for Financial Services

A Method for Selection of NoSQL Database for Financial Services

Each type of NoSQL database has its advantages and disadvantages for specific situations. In this part of the research, practical examples of non-relational databases in the financial sphere from real companies will be examined, from which criteria and evaluations will be derived, which can be used for the selection of NoSQL databases and for a suitable reference architecture for the specific needs of the company.

Financial services as a result of the currently developed business applications

In the following points, already existing financial services based on the various non-relational databases will be discussed, together with the advantages due to which these databases can be used, and then these financial services will be added to the developed method.

✓ **Document-oriented databases**

c. MongoDB

MongoDB is a non-relational, schema-free database that provides high performance by supporting MapReduce, storage tools, horizontal scalability, and built-in replication. The database supports a specialized connector with Hadoop. MongoDB provides real-time analytics, as data can come from various sources, which helps analyse customer needs and requirements. MongoDB can create applications that were impossible to create with relational databases. This base contributes a technological base that enables: [38]

- ✓ Storing data in flexible JSON documents, which in practice means that the fields that are stored can vary from document to document, and that the data structures can be changed.
- ✓ Document model that maps to objects in application code makes data easy to work with.
- ✓ A distributed base with high availability, horizontal scalability and geographic distribution are built-in and easy to use.
- ✓ Basing systems on MongoDB as a database contributes to modernizing legacy systems, protecting customer information, achieving high availability and scalability.

MongoDB provides solutions for the following financial services – Figure 3:

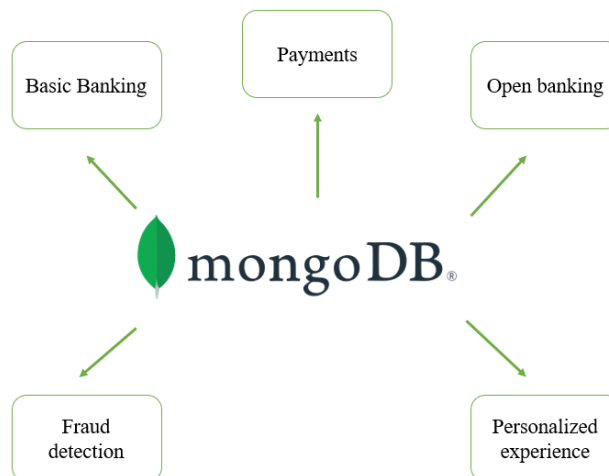


Figure 3 Financial services based on MongoDB.

✓ **Financial institutions based their systems on MongoDB.**

Three major financial institutions that base their systems on MongoDB – Royal Bank of Scotland, AHL – a major investment company and Bendigo and Adelaide Bank.

d. Couchbase

Couchbase offers multi-dimensional scalability, meaning that scaling can happen easily even while the base is in use. In this way, the performance of the application is increased, and the cost is reduced. Multidimensional scalability separates, isolates, and scales services on a per-index, query, and data basis. This is how resources can be optimized without affecting system performance and performance. Using Couchbase for real-time analytics is extremely fast and important for fraud prevention, business analytics, and more. [45]

Basing financial services on Couchbase happens because:

- ✓ Multi-dimensional scalability, real-time analytics, system response in milliseconds, cross-data centre replication, powerful query language, mobile financial services at the edge.

Couchbase provides solutions for the following financial services - Figure 4:

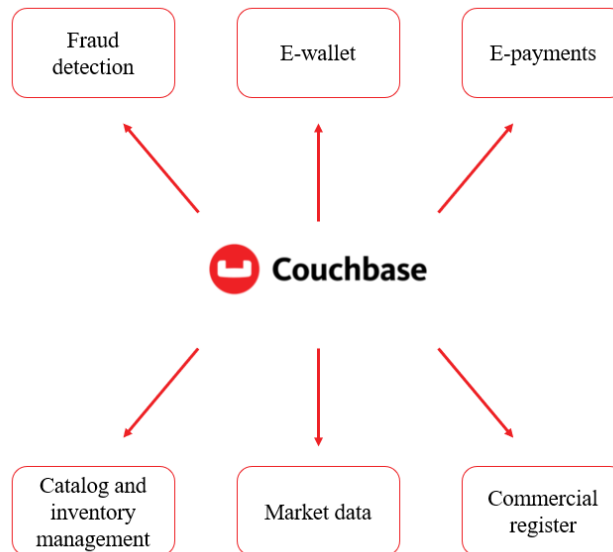


Figure 4 Financial services based on Couchbase.

✓ **Financial institutions based their systems on Couchbase.**

Some of the big companies in the financial sphere that have based their services on Couchbase are PayPal, FICO, EQUIFAX and REVOLUT.

✓ **Key-value databases**

c. Amazon DynamoDB

Amazon DynamoDB is a non-relational database that supports horizontal scalability, has a flexible schema that allows for easy adaptation as application requirements change. DynamoDB offers automatic replication and scalability. Banking institutions, capital markets and others are based on this NoSQL base. [50]

Basing financial services on Amazon DynamoDB happens because:

- ✓ Consistent performance at any scale, meaning there is no need to redo the database as data volumes increase.
- ✓ DynamoDB helps secure data with encryption and continuous data backup.
- ✓ A flexible schema, which allows for easy redesign when business requirements change without having to redefine the entire schema, as must happen if the database is relational.

Amazon DynamoDB provides solutions for the following financial services - Figure 5:

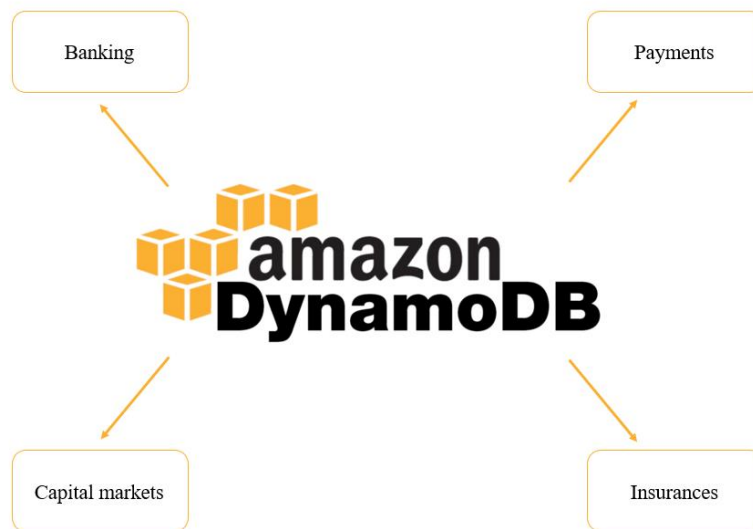


Figure 5 Financial services based on Amazon Dynamo DB.

✓ **Financial institutions based their systems on Amazon DynamoDB**

Some of the big companies in the financial sphere that have based their services on Amazon DynamoDB are Raiffeisen Bank and CI Financial.

d. Redis

Redis offers in-memory storage that enables faster system responses and supports millions of operations. This kind of non-relational database has high efficiency and good scalability - vertical and horizontal. Redis is data immutable, meaning it will remain available even in the event of server failures, and it also offers a wide variety of data structures, enabling developers to choose the most appropriate option for their application. Many social networks use Redis for their database, but many financial institutions are also starting to turn to this kind of NoSQL. [57]

Basing financial services on Redis is happening because:

- ✓ High-speed transactions, which are extremely important for financial institutions.
- ✓ Redis offers different data models that can be easily implemented for different services.
- ✓ Redis allows applications to support simultaneous updates across multiple geographic locations without compromising latency or performance.
- ✓ Has high fault tolerance, resilience, and high availability.

Redis provides solutions for the following financial services - Figure 6:

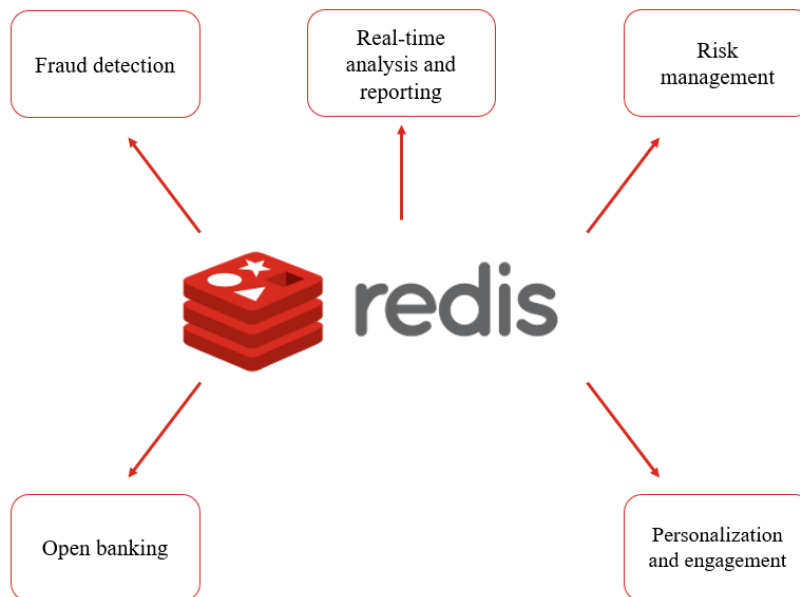


Figure 6 Financial services based on Redis.

- ✓ **Financial institutions based their systems on Redis.**

Some of the big companies in the financial sector that have based their services on Redis are Ekata and Deutsche Börse Group.

✓ **Wide-column databases**

c. HBase

HBase is part of the Hadoop ecosystem. In this type of wide-column databases, the data can be stored in Hadoop, the most important functionalities are that it is scalable, has automatic support in case of a problem, integrates with Hadoop, and has data replication in clusters. HBase is used by social networks, but there are already financial services that have adapted it to their systems.

Basing financial services on HBase happens because:

- ✓ Linear and modular scalability. HBase is built to scale across thousands of servers and manage access to PB of data.
- ✓ Fast access and low latency of reading and writing data.
- ✓ Fault tolerance as data is stored on multiple hosts in the cluster. [73]

HBase provides solutions for the following financial services - Figure 7:

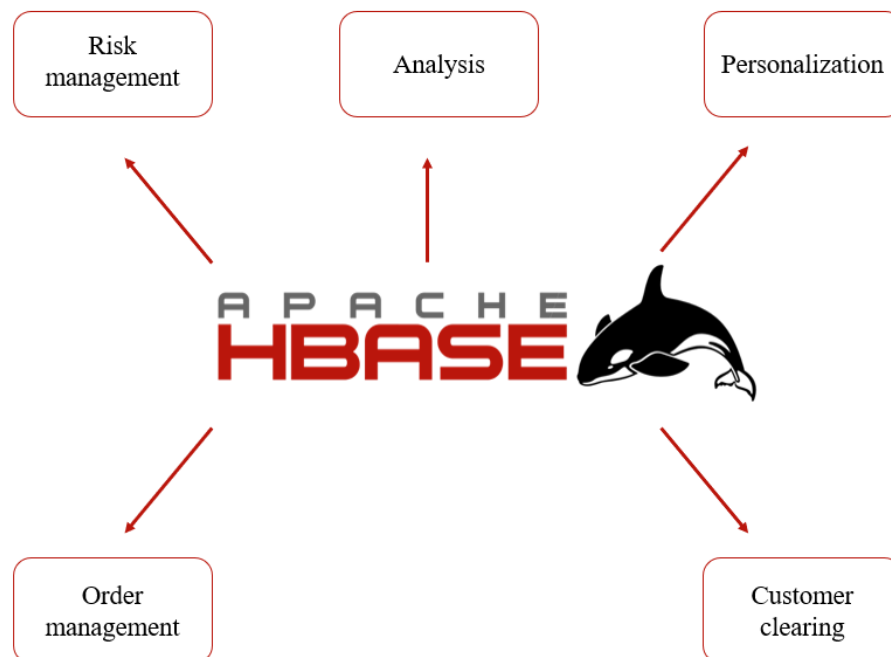


Figure 7 Financial services based on Apache HBase.

✓ **Financial institutions based their systems on Apache HBase**

Some of the big companies in the financial sphere that have based their services on HBase are FINRA and Celer Technologies.

d. Cassandra

The main functionalities of Cassandra are that it is scalable, distributed and extremely fault tolerant because data is automatically stored and replicated. Cassandra supports MapReduce as well as Apache Hive and Apache Pig. The base is horizontally scalable and when the company has a need, more clusters are added. [63] [64]

Basing financial services on Cassandra happens because:

- ✓ Cassandra excels at large volumes of data that need to be written. Data distribution inside clusters is fast, writes are cheap, making Cassandra extremely good with large volumes of data.
- ✓ Cassandra handles data replication as it is distributed across multiple clusters and across multiple data centers. For this reason, if some of the clusters are down, or the entire data center, Cassandra will handle the situation.
- ✓ If real-time data growth is expected, Cassandra has a huge advantage.

Cassandra provides solutions for the following financial services - Figure 8:

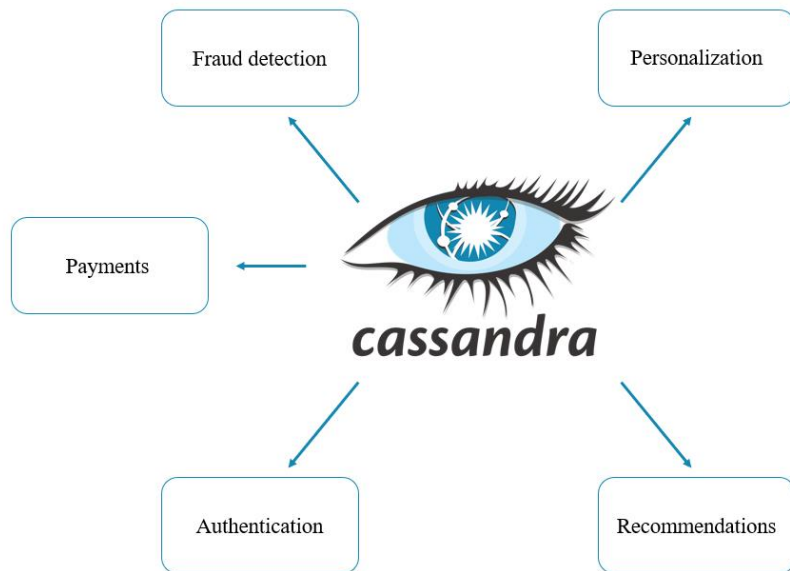


Figure 8 Financial services based on Cassandra.

- ✓ **Financial institutions based their systems on Cassandra.**

Some of the big companies in the financial sphere that have based their services on Cassandra are Fractal Labs and ACI Worldwide.

✓ **Graph databases**

c. Neo4J

Neo4j is a flexible schema non-relational database that uses a graph model. This model contains nodes that are connected to each other. Neo4j supports ACID properties and is also a scalable and reliable base.

Basing financial services on Neo4j happens because:

- ✓ Extremely fast rendering of requests in real time due to the graph architecture.
- ✓ Data integrity is preserved thanks to ACID properties, so whether transactions are completely successful or fail, the data will not be corrupted.
- ✓ Flexibility is another advantage of Neo4j. The database does not need predefined tables, and if business requirements change, the data model can be adapted in a very short time.

Neo4j provides solutions for the following financial services - Figure 9:

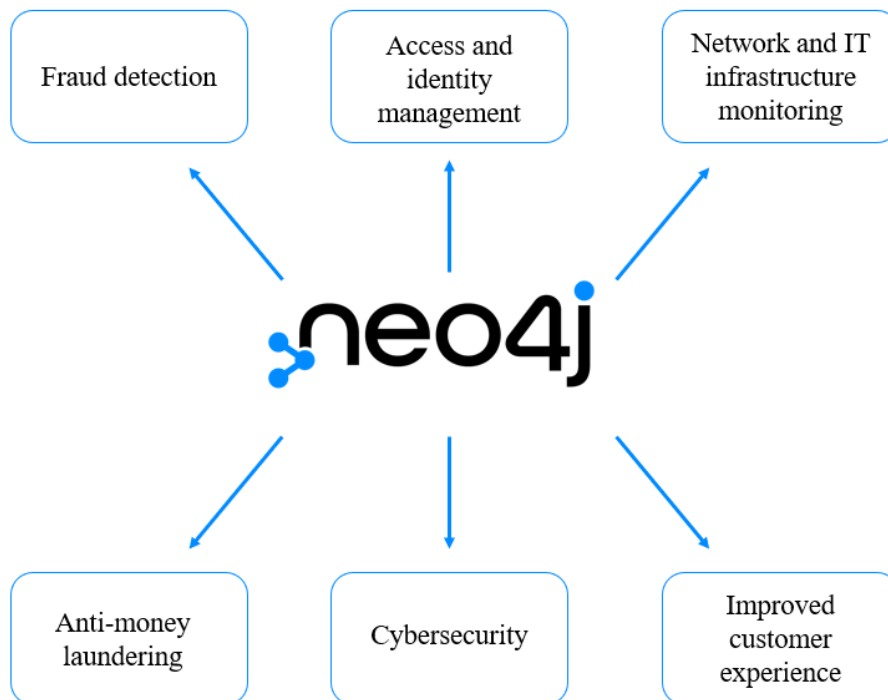


Figure 9 Financial services based on Neo4j.

- ✓ **Financial institutions based their systems on Neo4j.**

Some of the big companies in the financial sphere that have based their services on Neo4j are UBS and Global 50 Bank.

d. Orient DB

Orient DB's main features are speed, as this NoSQL database can store up to 120,000 records per second. The base is flexible and can work on any platform without configurations and installations. Orient DB features elastic linear scalability as well as incremental records. [70]

Basing financial services on Orient DB is happening because of:

- ✓ Orient DB records and processes data at extremely high speed, far surpassing relational databases, which are slow and expensive. Orient DB returns responses to queries in seconds.
- ✓ This type of base allows for scalability, resulting in easy adaptation of systems as requirements change.

Orient DB provides solutions for the following financial services - Figure 10:



Figure 10 Financial services based on OrientDB.

✓ Financial institutions based their systems on Orient DB

Part of the big companies in the financial sphere that have based their services on Orient DB is 99Bill.

След задълбочено разглеждане на научна литература, както и на практически приложени решения за финансови услуги базирани върху NoSQL бази от данни се обособиха следните услуги.

Definition of a method for analysis and evaluation of a reference architecture

From the practical point of view of already implemented solutions, the following services stood out - Figure 11, implemented on different types of non-relational databases and applied in practice by the business, which we considered in the previous point.

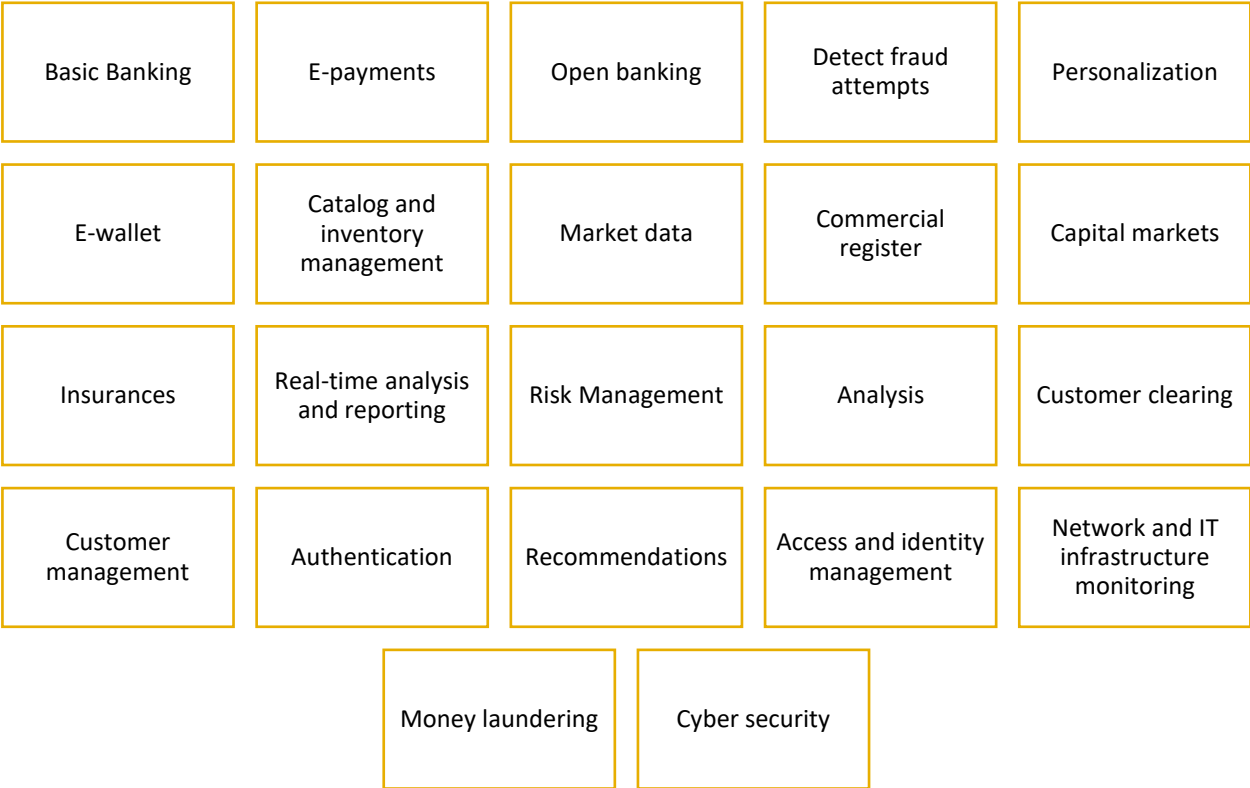


Figure 11 Types of financial services highlighted based on practical experience.

From a theoretical point of view, the following financial services stood out from Chapter 1 - Figure 12.

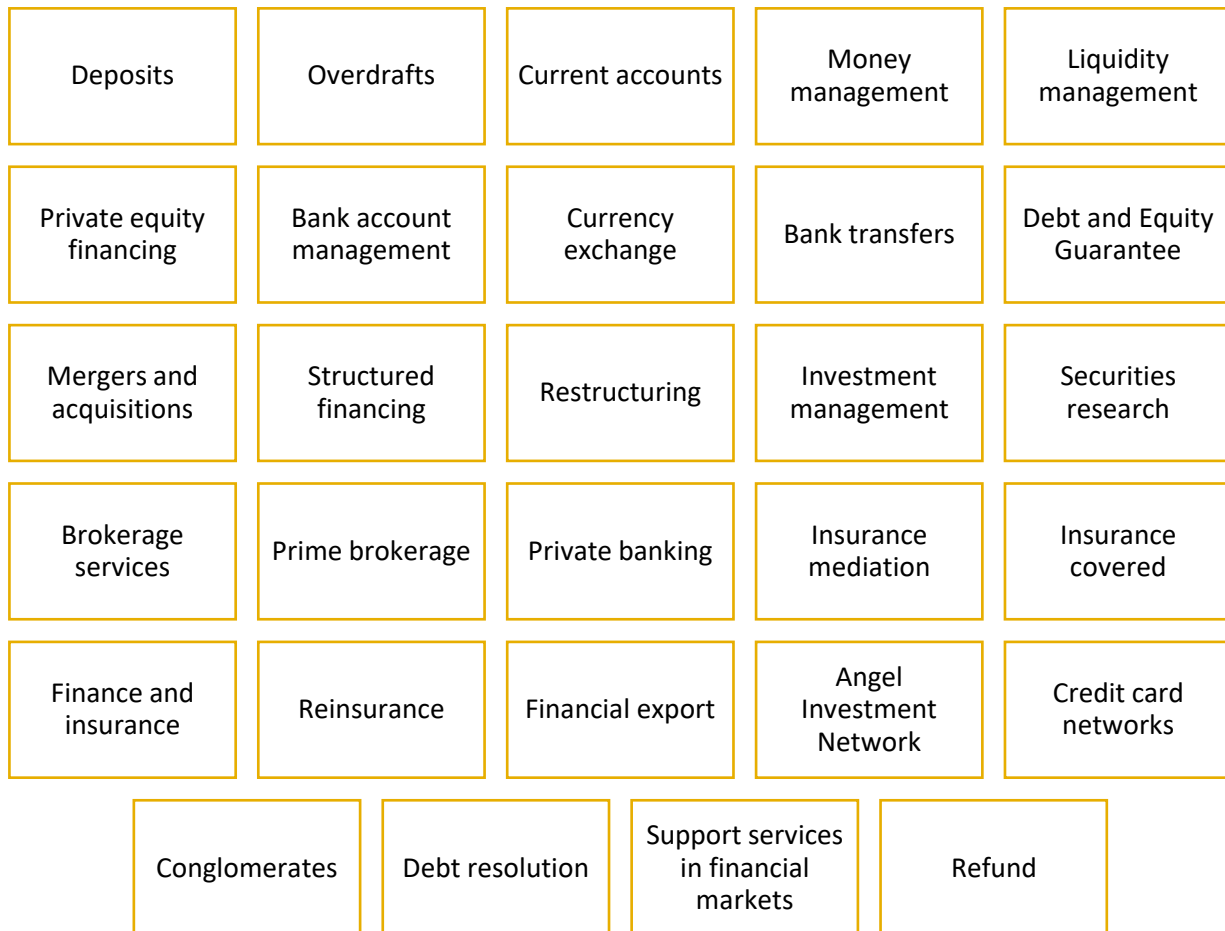


Figure 12 Types of financial services highlighted based on theoretical research.

Each of the highlighted financial services, from a theoretical and practical point of view, is part of the different types of finance. These include three sub-categories – Chart 1:

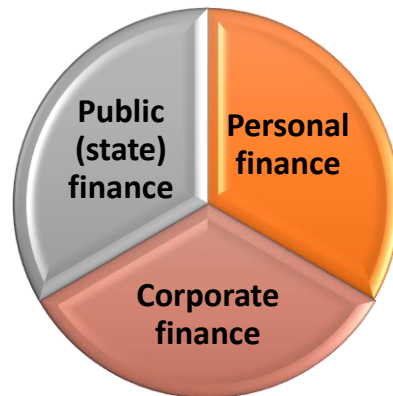


Chart 1 Finance sub-categories

Most of the services that were highlighted from a theoretical and practical point of view are from the first sub-category of finance - personal finance or financial services that are directly related to the client. Services of this type generate Big Data that has the main characteristics discussed earlier in Chapter I. From the perspective of financial services, we consider it as follows:

❖ **Volume**

Financial institutions have many customers, which in turn leads to many transactions being carried out for customers, which in turn leads to the generation of large volumes of data.

❖ **Velocity**

The speed at which data is generated in this area and the fact that real-time or near-real-time operations must be performed based on this data, otherwise customers may be lost and flow to other financial institutions.

❖ **Variety**

Since financial institutions directly related to customers generate different types of data due to the ways of service – physical, digital and combined. These types of service lead to different sources of data – structured, semi-structured and unstructured, which also leads to the wide variety of data.

❖ **Value**

Despite the availability of large volumes of data from various sources generated at high speed, it is very important that it is valuable and valuable. For the purposes of financial operations, this is of utmost importance so that they can be useful for analysing and drawing conclusions. For example, it would be of great benefit and value to financial institutions to extract customer data about what insurance they can offer them based on what their needs are – whether they have a car, apartment, house, family, etc., that is the value of the data generated, collected, and analysed will be of great benefit if this ultimately translates into a sale.

❖ **Veracity**

In the field of financial services, it is very important that the data is reliable and credible, and this must be proven by the source of the data.

With the development of technologies, digitalization and digital transformation, the generation of large volumes of the above types of data, leads to the conclusions for many companies that adapting NoSQL databases to their systems would only bring benefits. Even financial services that use legacy systems based on relational databases are realizing the need to change to meet new customer demands.

After the detailed study of non-relational databases for financial services - from a theoretical and practical point of view, no specific way was found how to choose a suitable NoSQL database for a financial service. There are many companies and websites that offer a comparison of the various features of databases, but generally nothing about the specific field or service.

For this reason, it would be appropriate to develop a method for quantitative evaluation of NoSQL databases for specific financial services. In Chapter 1, advantages of non-relational databases are defined, which are as follows - Figure 13:

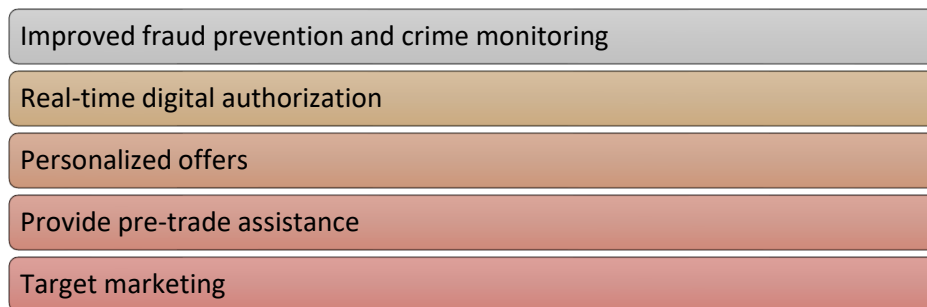


Figure 13 Advantages of NoSQL databases defined as criteria.

Based on these advantages, we will define them as evaluation criteria for non-relational databases. The list of selected criteria is not exhaustive, and more can be added depending on the needs and requirements of the customers/company.

The approach that has been developed aims to enable the selection of a type of non-relational database that will comply on the one hand with the set criteria, on the other hand with the type of data they process as well as the processing time. Table 3 examines financial services for which we have practically implemented solutions from companies, respectively, they are evaluated with a certain weight, depending on whether a solution exists, and a weight is given.

The criteria: Security, Real-time Analysis, Personalization, Pre-Trade Assistance and Target Marketing are rated on a scale of 1 to 5, based on the assessment of experts with different areas of expertise. The criteria are evaluated with different weights: 1 – Weak, 2 – Unsatisfactory, 3 – Good, 4 – Very good and 5 – Excellent. Each of these five criteria is evaluated based on expert judgment and, accordingly, its importance for each financial service.

The table presents the derived criteria for the four types of databases – document-oriented, key-value pair, wide-column, and graph databases with two examples for each of them. For document-oriented, Couchbase and Mongo DB are considered, for key-value pairs - Amazon DynamoDB and Redis, for wide-column - HBase and Cassandra, and for graph databases - Neo4j and Orient DB.

The second element that is evaluated is whether there is a solution already created, one in development, or none. A scale of 0 to 2 is used for this, with 0 meaning no solution, 1 being in development, and 2 being a fully operational financial service solution based on a NoSQL database.

Applying this method to evaluate non-relational databases, each one collects a total score that indicates how well suited it is overall for application to financial services. For document-oriented databases, Couchbase collects 190 points, and MongoDB -126. For non-relational key-value databases, Amazon Dynamo DB collects 156 points, and Redis - 192. For wide-column databases, HBase - 154, Cassandra - 250, and for graph databases Neo4j - 146, Orient DB - 60 points.

Based on the created method and the results obtained in the evaluation of the document-oriented databases MongoDB with 126 and Couchbase with 190 points, we will create a reference architecture for financial services.

The reference architecture will be based on MongoDB, which, although scores less on expert opinion than the other document-oriented databases, has already built real-world solutions that are in the realm of core banking, which is the type of data we will use for testing – contracts, slips, reports, emails and more.

			Document-oriented				Key-value store				Wide-column				Graph			
			Couchbase		MongoDB		Amazon DynamoDB		Redis		Hbase		Cassandra		Neo4j		OrientDB	
Financial services	Criteria	CK	CP?	KC	CP?	KC	CP?	KC	CP?	KC	CP?	KC	CP?	KC	CP?	KC	CP?	KC
Fraud detection	Security	5	2	10	2	10	2	10	0	0	0	0	2	10	2	10	2	10
	Real-time DA	5		10		10		10		0		0		10		10		10
	Personalization	3		6		6		6		0		0		6		6		6
	Pre-trade assistance	2		4		4		4		0		0		4		4		4
	Targeted marketing	1		2		2		2		0		0		2		2		2
E-wallet	Security	5	2	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Real-time DA	5		10		0		0		0		0		0		0		0
	Personalization	4		8		0		0		0		0		0		0		0
	Pre-trade assistance	2		4		0		0		0		0		0		0		0
	Targeted marketing	1		2		0		0		0		0		0		0		0
E-payments	Security	5	2	10	2	10	2	10	2	10	0	0	2	10	0	0	2	10
	Real-time DA	5		10		10		10		10		0		10		0		10
	Personalization	2		4		4		4		4		0		4		0		4
	Pre-trade assistance	1		2		2		2		2		0		2		0		2
	Targeted marketing	1		2		2		2		2		0		2		0		2
Market data	Security	3	2	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Real-time DA	3		6		0		0		0		0		0		0		0
	Personalization	3		6		0		0		0		0		0		0		0
	Pre-trade assistance	5		10		0		0		0		0		0		0		0
	Targeted marketing	5		10		0		0		0		0		0		0		0
Trade	Security	5	2	10	0	0	0	0	0	0	2	10	0	0	0	0	0	0
	Real-time DA	5		10		0		0		0		10		0		0		0
	Personalization	4		8		0		0		0		8		0		0		0
	Pre-trade assistance	4		8		0		0		0		8		0		0		0
	Targeted marketing	3		6		0		0		0		6		0		0		0
Core banking	Security	5	0	0	2	10	2	10	0	0	0	0	0	0	0	0	0	0
	Real-time DA	5		0		10		10		0		0		0		0		0
	Personalization	2		0		4		4		0		0		0		0		0
	Pre-trade assistance	2		0		4		4		0		0		0		0		0
	Targeted marketing	1		0		2		2		0		0		0		0		0
Personalization	Security	2	0	0	2	4	0	0	2	4	2	4	2	4	2	4	0	0
	Real-time DA	2		0		4		0		4		4		4		4		0
	Personalization	3		0		6		0		6		6		6		6		0
	Pre-trade assistance	1		0		2		0		2		2		2		2		0

	Targeted marketing	1		0		2		0		2		2		2		2		0
Real-time analysis and reporting	Security	2	0	0	2	4	0	0	2	4	0	0	0	0	0	0	0	0
	Real-time DA	2		0		4		0		4		0		0		0		0
	Personalization	1		0		2		0		2		0		0		0		0
	Pre-trade assistance	3		0		6		0		6		0		0		0		0
	Targeted marketing	1		0		2		0		2		0		0		0		0
Capital markets	Security	5	0	0	0	0	2	10	0	0	0	0	0	0	0	0	0	0
	Real-time DA	5		0		0		10		0		0		0		0		0
	Personalization	4		0		0		8		0		0		0		0		0
	Pre-trade assistance	4		0		0		8		0		0		0		0		0
	Targeted marketing	2		0		0		4		0		0		0		0		0
Insurances	Security	4	0	0	0	0	2	8	0	0	0	0	0	0	0	0	0	0
	Real-time DA	4		0		0		8		0		0		0		0		0
	Personalization	3		0		0		6		0		0		0		0		0
	Pre-trade assistance	1		0		0		2		0		0		0		0		0
	Targeted marketing	1		0		0		2		0		0		0		0		0
Authentication	Security	5	0	0	0	0	0	0	2	10	0	0	2	10	0	0	0	0
	Real-time DA	5		0		0		0		10		0		10		0		0
	Personalization	2		0		0		0		4		0		4		0		0
	Pre-trade assistance	2		0		0		0		4		0		4		0		0
	Targeted marketing	1		0		0		0		2		0		2		0		0
Open banking	Security	5	0	0	2	0	0	0	2	10	0	0	0	10	0	0	0	0
	Real-time DA	2		0		0		0		4		0		4		0		0
	Personalization	4		0		0		0		8		0		8		0		0
	Pre-trade assistance	1		0		0		0		2		0		2		0		0
	Targeted marketing	3		0		0		0		6		0		6		0		0
Catalog and inventory management	Security	4	2	0	0	0	0	0	0	8	0	0	0	8	0	0	0	0
	Real-time DA	5		0		0		0		10		0		10		0		0
	Personalization	2		0		0		0		4		0		4		0		0
	Pre-trade assistance	1		0		0		0		2		0		2		0		0
	Targeted marketing	1		0		0		0		2		0		2		0		0
Commercial register	Security	4	2	8	0	0	0	0	0	8	0	0	0	8	0	0	0	0
	Real-time DA	3		6		0		0		6		0		6		0		0
	Personalization	1		2		0		0		2		0		2		0		0
	Pre-trade assistance	0		0		0		0		0		0		0		0		0
	Targeted marketing	0		0		0		0		0		0		0		0		0
Analysis	Security	4	0	0	0	0	0	0	0	0	2	8	0	0	0	0	0	0
	Real-time DA	3		0		0		0		0		6		0		0		0
	Personalization	5		0		0		0		0		10		0		0		0

	Pre-trade assistance	3		0		0		0		0		6		0		0		0
	Targeted marketing	4		0		0		0		0		8		0		0		0
Customer clearing	Security	5	0	0	0	0	0	0	0	0	2	10	0	0	0	0	0	0
	Real-time DA	4		0		0		0		0		8		0		0		0
	Personalization	3		0		0		0		0		6		0		0		0
	Pre-trade assistance	2		0		0		0		0		4		0		0		0
	Targeted marketing	1		0		0		0		0		2		0		0		0
Client management	Security	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Real-time DA	4		0		0		0		0		0		0		0		0
	Personalization	4		0		0		0		0		0		0		0		0
	Pre-trade assistance	2		0		0		0		0		0		0		0		0
	Targeted marketing	4		0		0		0		0		0		0		0		0
Recommendations	Security	3	0	0	0	0	0	0	0	0	0	0	2	6	0	0	0	0
	Real-time DA	5		0		0		0		0		0		10		0		0
	Personalization	5		0		0		0		0		0		10		0		0
	Pre-trade assistance	4		0		0		0		0		0		8		0		0
	Targeted marketing	5		0		0		0		0		0		10		0		0
Access and identity management	Security	5	0	0	0	0	0	0	0	0	0	0	0	0	2	10	0	0
	Real-time DA	4		0		0		0		0		0		0		8		0
	Personalization	2		0		0		0		0		0		0		4		0
	Pre-trade assistance	3		0		0		0		0		0		0		6		0
	Targeted marketing	1		0		0		0		0		0		0		2		0
Network and IT infrastructure monitoring	Security	5	0	0	0	0	0	0	0	0	0	0	0	0	2	10	0	0
	Real-time DA	4		0		0		0		0		0		0		8		0
	Personalization	1		0		0		0		0		0		0		2		0
	Pre-trade assistance	0		0		0		0		0		0		0		0		0
	Targeted marketing	0		0		0		0		0		0		0		0		0
Anti-Money Laundering	Security	5	0	0	0	0	0	0	0	0	0	0	0	0	2	10	0	0
	Real-time DA	4		0		0		0		0		0		0		8		0
	Personalization	1		0		0		0		0		0		0		2		0
	Pre-trade assistance	0		0		0		0		0		0		0		0		0
	Targeted marketing	0		0		0		0		0		0		0		0		0
Cybersecurity	Security	5	0	0	0	0	0	0	0	0	0	0	0	0	2	10	0	0
	Real-time DA	4		0		0		0		0		0		0		8		0
	Personalization	1		0		0		0		0		0		0		2		0
	Pre-trade assistance	2		0		0		0		0		0		0		4		0
	Targeted marketing	1		0		0		0		0		0		0		2		0
Risk Management	Security	5	0	0	0	0	0	2	10	2	10	2	10	0	0	0	0	0

	Real-time DA	4	0	0	0	8	8	8	0	0
	Personalization	1	0	0	0	2	2	2	0	0
	Pre-trade assistance	2	0	0	0	4	4	4	0	0
	Targeted marketing	1	0	0	0	2	2	2	0	0
Total			190	126	156	192	154	250	146	60

Table 3 Financial services evaluation

Designing a Business Reference Architecture for Financial Services with NoSQL Databases

Every single financial institution dealing with different types of financial services generates, collects, processes, and analyses data of different types – structured, semi-structured and unstructured in large volumes, which in turn leads to different needs of the institutions. Depending on the type of financial institution depends on the type of data that is generated, which in turn leads to different requirements for storage space, processing methods and analysis tools.

With the implementation of non-relational databases in financial institutions and adapting their historical systems to the new databases, it leads to the development and need of services in the field of finance to maintain their competitive advantage, as well as to get ahead of their competitors in the field.

The purpose of creating a reference architecture, whether for use only in one organization or for universal and adaptation across multiple financial institutions, is to ensure consistency and applicability of the use of given technologies in a particular organization.

Based on the requirements that may arise from financial services, together with the theoretical literature review, as well as the practical one of existing financial services solutions on various non-relational databases, a model of a universal component reference architecture for financial services was created. services.

- Conceptual model

The conceptual model of the reference architecture for financial services presented in Figure 14 represents the general structure and elements of which it consists, as well as the necessary data and systems to achieve the business requirements, which means the maintenance of company processes, recording of business events and performance tracking. In Chapter I, an approach to building a reference architecture is defined, which consists of five main steps. The created conceptual model of the business reference architecture for financial services is the result of the first two steps - the identification of the goal and the formulation of the goal. The purpose of the Business Reference Architecture for Financial Services is to adapt the modern ways of

storing Big Data for different types – semi-structured and unstructured data – to financial institutions that continue their rapid development and digital transformation.

The reference architecture consists of 3 layers: Data Storage System, Integration Layer and Data Storage Server, which are connected to financial services processing systems.

❖ **Data storage system**

The data storage system is the system that stores the data after it has been received from the financial services processing systems.

❖ **Integration layer**

The integration layer in the reference architecture refers to the components responsible for the integration of data storage systems and their storage servers. This layer provides a centralized platform for management, data movement and communications between different systems. The integration layer consists of data connectors between the storage system and the server.

❖ **Data storage server**

The data warehousing server is where the data that arrives through the integration layer and the data connectors from the data warehousing systems that enter the data warehousing system from the financial services processing systems is stored.

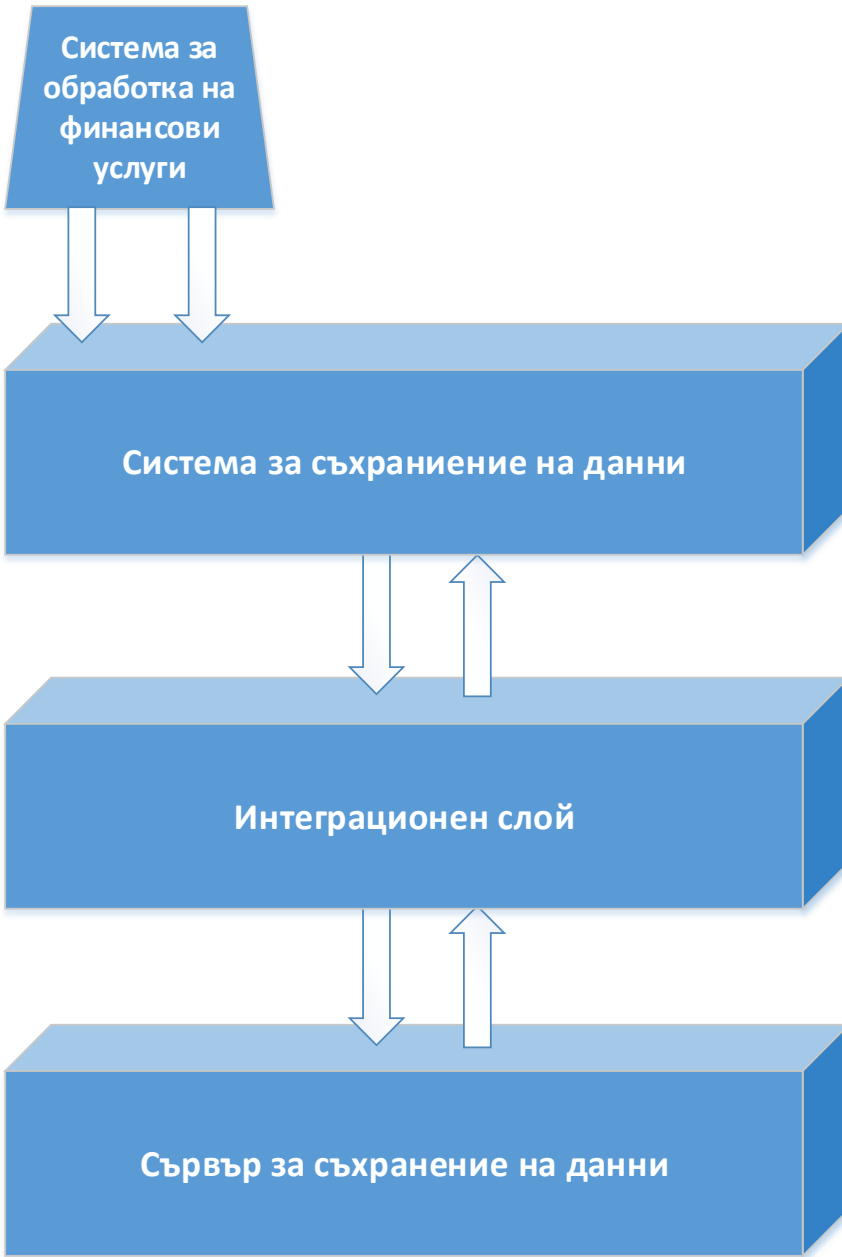


Figure 14 Conceptual model

- **Logical model**

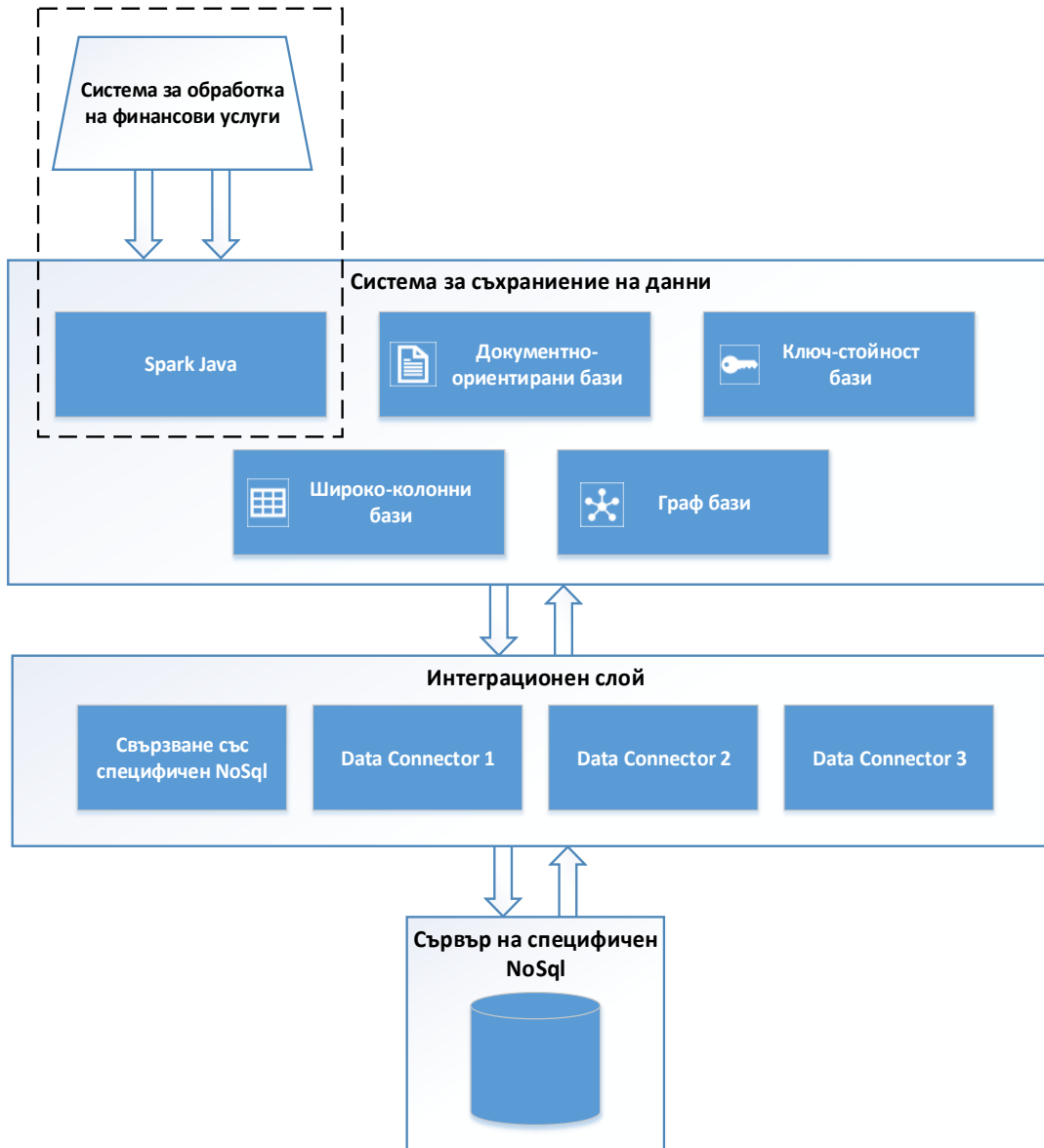


Figure 15 Logical model

The logical model lays down the technical rules and standards for the creation of the business reference architecture, and it lays down the rules so that the set principles can be achieved. Through the financial services processing system, data from financial institutions enters, which are subsequently stored, processed, and analysed in the remaining layers of the reference architecture. Figure 15 presents the logical model of the financial services business reference architecture, with each of its layers discussed in detail below.

❖ **Data storage system**

In the first layer of the reference architecture, the so-called A data storage system can host 1 or more of the four types of non-relational databases – document-oriented, key-value, wide-column, and graph databases. The types of NoSQL databases were discussed at a theoretical level, such as what they are capable of in Chapter I, and in this chapter, they were discussed at a practical level - what applications related to financial services are built on specific non-relational databases, to help the selection of specific base.

The choice of such is realized based on the following elements, which should be considered carefully before proceeding to a specific NoSQL database:

- The financial institution's existing system(s) that will need to be migrated to the new base.
- Type of financial services in which the financial institution deals
- Types of data that the financial institution generates based on the financial services it provides – structured, semi-structured and unstructured data.

❖ **Integration layer**

Based on the non-relational base that is selected in the data storage system based on the considered elements, the way to connect to a specific NoSQL is selected, which is positioned in the integration layer of the reference architecture.

❖ **Dedicated NoSQL database server**

Depending on the selected non-relational database in the data storage system, the subsequent selection of a way to connect a specific NoSQL database, a specific NoSQL server is reached in which the data of the financial institution is stored.

- **Physical model**

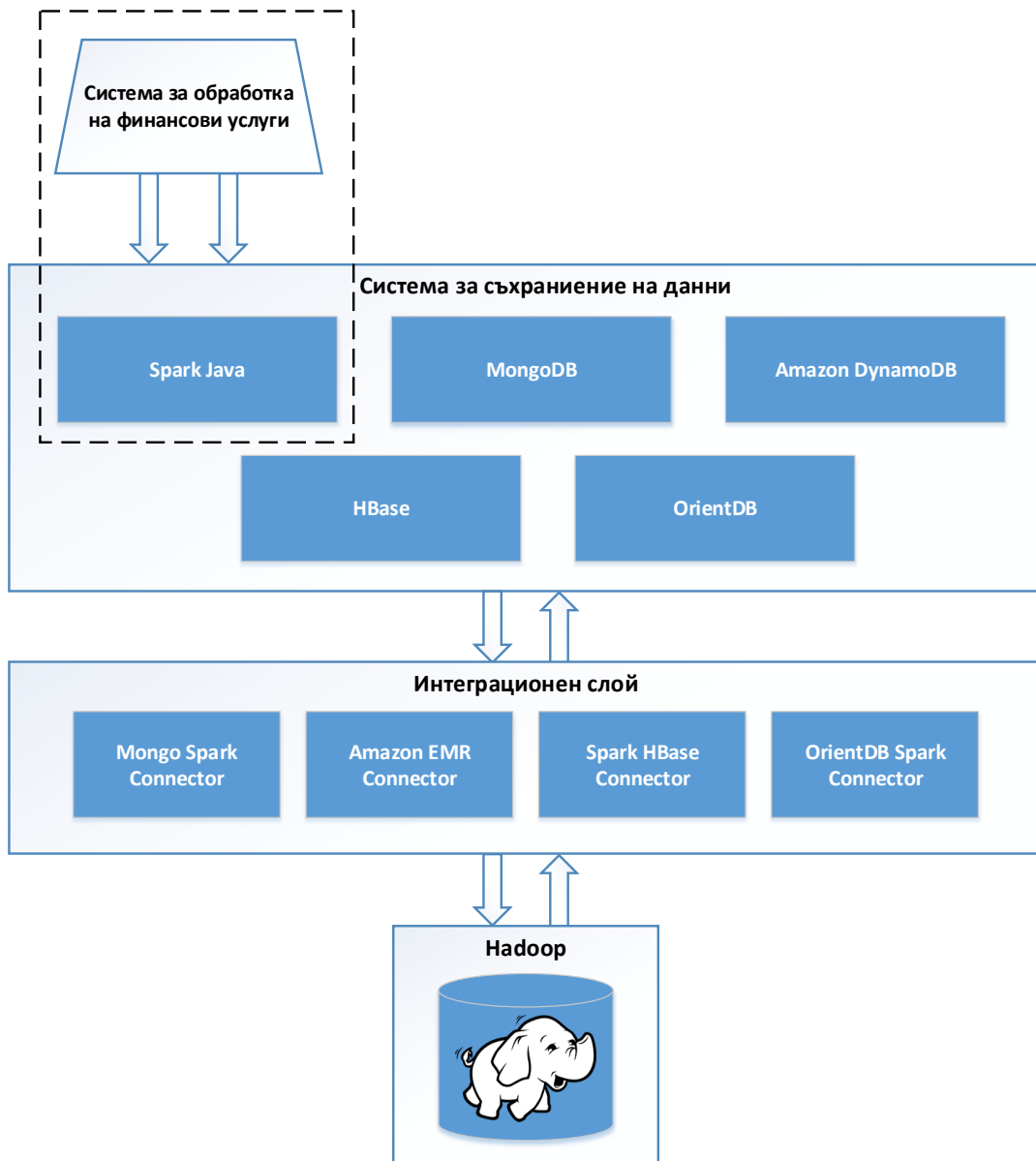


Figure 16 Physical model

The created physical model in Figure 16 is implemented with concrete examples of non-relational databases and their connectors to Hadoop. Each of these elements can be replaced according to the needs of the financial institution that wants to adapt this reference architecture with NoSQL databases for its own needs.

In the first layer, which is the data storage system, the four types of non-relational databases - document-oriented, key-value, wide-column and graph databases - can be located, and in this layer the particular database that is selected is located. through the evaluation and selection

method of NoSQL databases for financial services. Bases of each type are located in the physical model:

- **Document-oriented database:** MongoDB
- **Key-value database:** Amazon DynamoDB
- **Wide-column database:** HBase
- **Graph:** Orient DB

In the second layer, the so-called integration layer is where the connecting elements with the different NoSQL databases are located, e.g., NiFi, Spark Connector and other types of Data Connectors suitable for the other types of non-relational databases. The specific data connectors from the data storage system to the server of the specific NoSQL database are as follows:

- For the document-oriented MongoDB database, the Mongo Spark Connector can be used.
- Amazon DynamoDB key-value database can be used Amazon EMR Connector
- For wide-column HBase, the Spark HBase Connector can be used.
- For the Orient DB graph database, the Orient DB Spark Connector can be used.

The third layer is a server of the specific NoSQL database, and for the implementation of the physical model of the reference architecture for financial services, Hadoop has been chosen.

An approach for using a designed RA leading to the creation of an ICT Architecture

The ICT Architecture (Information and Communication Technologies) provides a conceptual model, specifying at a basic level the elements of the ICT architecture (application, databases, technological ICT elements), as well as the relationships between them. Based on the reference architecture for financial services with NoSQL databases that was created in the previous point and demonstrated in Figure 22, we will create a specific ICT architecture that will be implemented and tested in Chapter IV.

The developed physical model for the reference architecture contains 3 layers - a data storage system, an integration layer, and a NoSQL-specific layer, and in the previous point we discussed in detail the different types of non-relational databases, with different types of data connectors that can relate to the NoSQL-specific layer. In the creation of the ICT architecture for the data storage system (the first layer), the document-oriented non-relational database MongoDB will be used, since it has been selected through the method created in the previous point and meets the requirements and has already created a financial service. which uses the data type with which

the architecture will be tested. The second integration layer will use the Apache Spark Connector for MongoDB. Since support for MongoDB's direct connector to Hadoop is deprecated, we will use this one through Apache Spark. The third layer remains for the Hadoop big data system that we will use for storage – Figure 17.

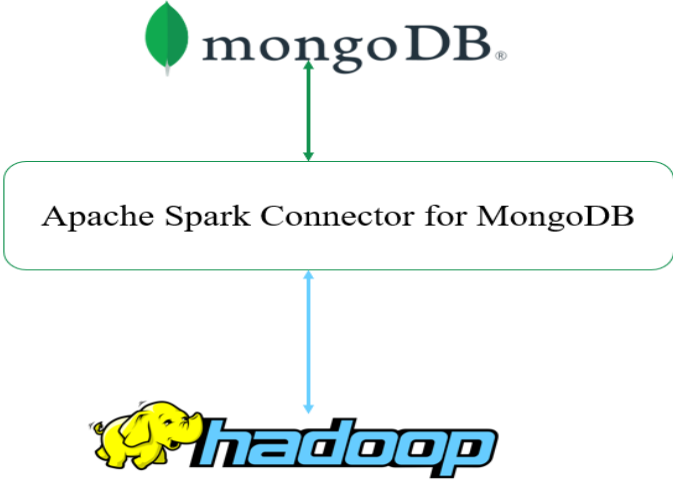


Figure 17 ICT Architecture

Conclusion

A method for selection and evaluation of NoSQL databases for financial services has been developed, and it is based on developed and used practical applications of the databases, as well as on expert evaluation. Practical examples of the four types of non-relational databases are reviewed and analysed. Based on them, a business reference architecture has been proposed, as well as an approach to the use and creation of an ICT architecture, which has been implemented and will be examined and tested in detail in Chapter IV.

Based on the created method for analysis and evaluation of NoSQL databases, an automated tool for selecting non-relational databases for financial services was created, which will be presented in the next chapter.

8. Chapter III. NoSQL database selection tool for financial services

Based on the created method for selecting and evaluating NoSQL databases for financial services, an automated tool for selecting such databases was created. Web technology based on Microsoft .NET Framework using C# and Visual Studio was chosen for its creation.

	User/Company	Administrator/Expert
Registration	Yes	Yes
Login	Да	Yes
Evaluation and selection of NoSQL according to criteria	Да	Yes
Add new NoSQL database	No	Yes
Editing of a NoSQL database	No	Yes
Delete of a NoSQL database	No	Yes
Add financial service	No	Yes
Editing of a financial service	No	Yes
Delete of a financial service	No	Yes
Add criterion for evaluation	No	Yes
Editing of a criterion for evaluation	No	Yes
Delete criterion for evaluation	No	Yes

Table 4 User Roles

The tool for evaluation and selection of NoSQL database has two user roles – User/Company and Admin/Expert. The system consists of different pages and functionalities depending on the user role.

User roles – User/Company

For a User/Company, the tool consists of a home page, a more information page, and a list of available databases with additional information about them. In the right part of the menu, there are

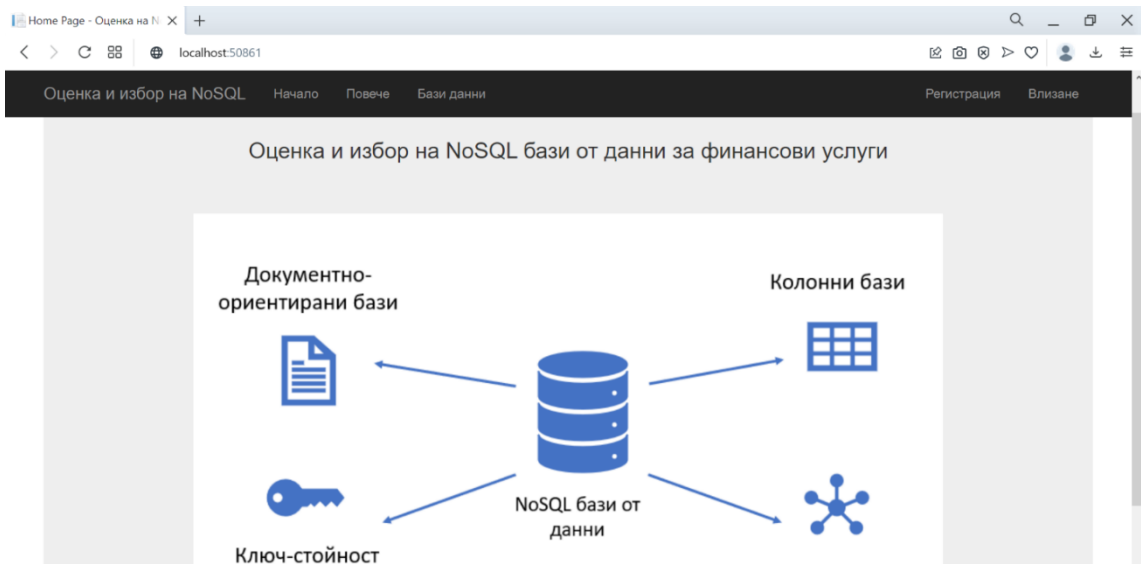


Figure 18 Homepage

the buttons for User Registration and for Login to the system of already existing users - Figure 18.

After entering the application with registration, the user/company selects the type of financial services, being able to mark several types of predefined ones in the tool depending on what services are offered by the institution. The user then states whether they have a preferred type of non-relational database, depending on the systems and software used in the company. If there is a desired type of non-relational database that the company is targeting, it selects the appropriate box on the selection page, and after marking it, a drop-down menu appears with the four types of non-relational databases from which it can choose - document-oriented, wide-column, key-value and graph bases - Figure 19. If there is no specific type of base that is desired by the company, a corresponding box is checked, which in this case does not allow the drop-down menu with types of bases to appear. The forward button is then pressed, which takes the user to the next page of information to be filled out.

Оценка и избор на NoSQL Начало Повече Критерии Финансови услуги Оценка Излизане

Избор на NoSQL база от данни за финансови услуги

Финансови услуги

- Банкиране
- Инвестиционни услуги
- Обменни услуги
- Застрахователни услуги
- Финансов износ
- Други финансови услуги

Имате ли предпочитан тип NoSQL база?

- Да
- Не

Вид NoSQL база

Документно-ориентирана база

Напред

© 2023 - Оценка и избор на NoSQL бази

Figure 19 Step 1 for selecting NoSQL database.

In the case we demonstrate with the tool working, the user wants a non-relational database that is suitable for banking, insurance, and financial export services, and also has a preferred type of non-relational database that is document oriented. Before choosing whether there is a preferred type of NoSQL database, a client should consider the systems they are working with and whether

they will need to integrate and migrate an institution's current systems or build completely new ones that are fully compatible with the new database.

The case where current systems need to be migrated to a new base is very sensitive and should be carefully considered by the company's technical teams. The other option, in which completely new systems are created to work with the newly selected non-relational database, takes an extremely large amount of time and financial resources. The choice in such a step must be very well weighed.

In the next step of evaluating the non-relational databases from the predefined criteria, which are derived by the method of Chapter II and are accordingly assigned a rating from 1 to 5 by the experts in the system, the company can choose only some of them to emphasized in their system based on the new NoSQL database, or rather all of them - Figure 20.

Оценка и избор на NoSQL Начало Повече Критерии Финансови услуги Оценка Излизане

Избор на NoSQL база от данни за финансови услуги

Критерии

- Подобрено предотвратяване на измами и мониторинг на престъпления
- Дигитална авторизация в реално време
- Персонализирани оферти
- Оказване на съдействие преди търговия
- Целеви маркетинг

Изчисли

© 2023 - Оценка и избор на NoSQL бази

Figure 20 Step 2 for selecting NoSQL databases.

Depending on the above, as well as the predefined information, therefore, which NoSQL databases already have practical solutions for financial services, which brings additional weight to the evaluation and based on the algorithm set in the tool for evaluating and selecting non-relational databases for financial services after pressing the Calculate button, various options for NoSQL databases appear on the screen, which can be offered under the correspondingly selected conditions - Figure 21.

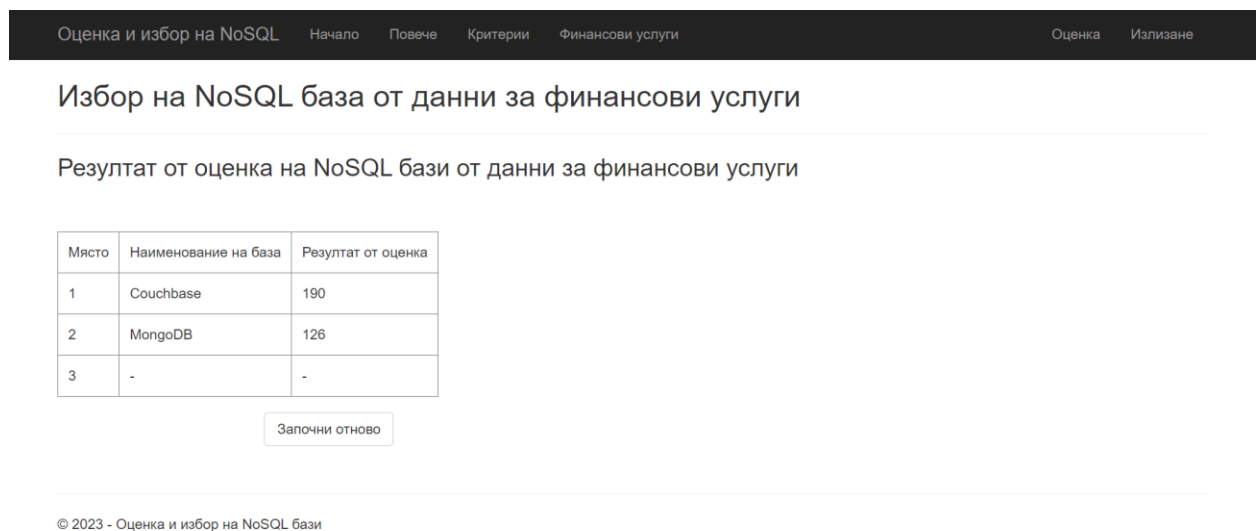


Figure 21 Result of evaluation of NoSQL databases

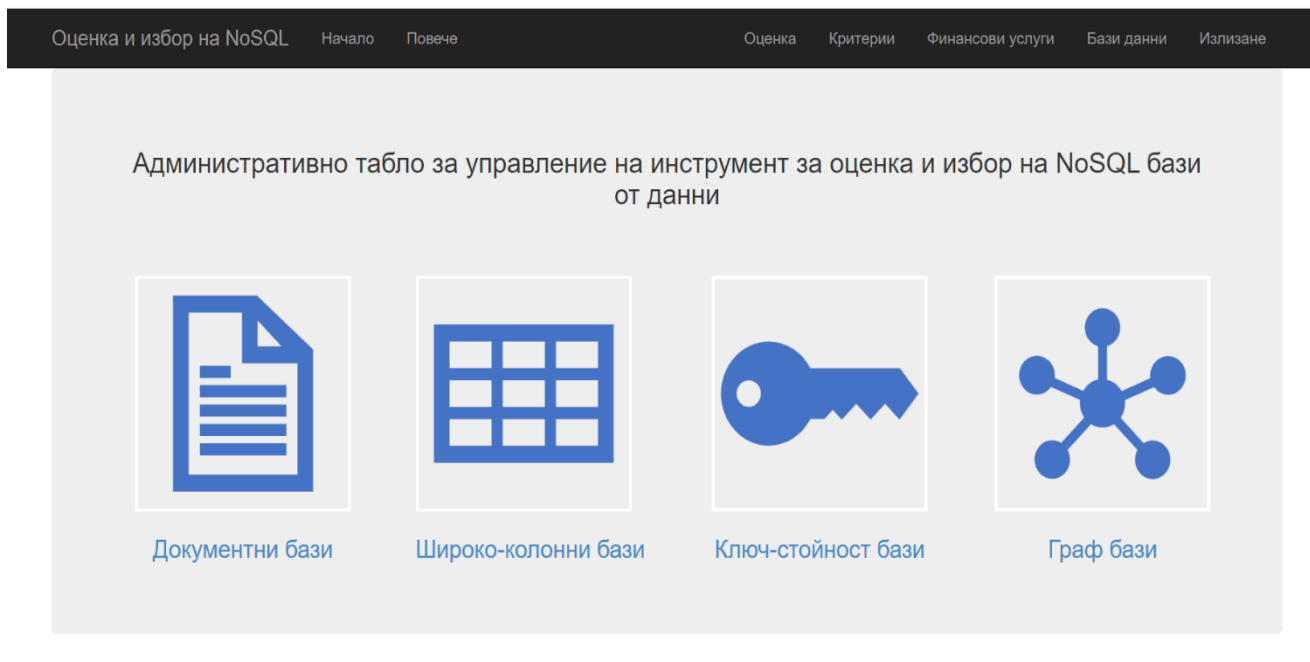
If necessary, the evaluation can be started from the beginning, as on the page that displays the result, there is a button Start over, which resets the results obtained so far and redirects the users to the page with Step 1.

User role Administrator/Expert

An Administrator/Expert can create, edit and delete the following items that are necessary for the tool to work:

- ✓ NoSQL databases
- ✓ Financial services
- ✓ Criteria

The expert also can evaluate and select non-relational databases in order to test the correct operation of the tool. The administrative dashboard of the non-relational database evaluation and selection tool is visualized as follows – Figure 22.



© 2023 - Оценка и избор на NoSQL бази

Figure 22 Administration Homepage

In the menu at the top of the page, the administrator can access the add, edit and delete criteria, financial services and non-relational database pages.

d. NoSQL databases

In the implementation and management of elements related to non-relational databases, the tool offers the following functionalities.

i. Adding

When adding NoSQL databases in the tool, the name of the database, a description, what type it is - document-oriented, wide-column, key-value, or graph database, as well as for which type of financial service a practical solution has already been created based on this non-relational database. This is necessary to subsequently calculate the coefficients together with the value of the criteria to get a visual representation of how suitable a particular NoSQL database is for the purposes of a given organization - Figure 23.

Създаване на NoSQL база от данни

Наименование на NoSQL база	<input type="text"/>
Вид NoSQL база	<input type="text" value="Изберете вид NoSQL база"/>
Описание на NoSQL базата	<input type="text"/>
Финансови услуги	<input type="checkbox"/> Банкиране <input type="checkbox"/> Инвестиционни услуги <input type="checkbox"/> Обменни услуги <input type="checkbox"/> Застрахователни услуги <input type="checkbox"/> Финансов износ <input type="checkbox"/> Други финансови услуги
	<input type="button" value="Създай"/>

[Назад към списък](#)

Figure 23 NoSQL database create – Step 1

When adding a new non-relational database to the tool, the type is selected from a drop-down menu, where they are predefined, so as not to cause additional confusion when working. The financial services that can be selected from when defining a non-relational database are those that the administrator has previously entered through the financial services management page in the tool, and when entering a specific database, several options can be selected - Figure 24.

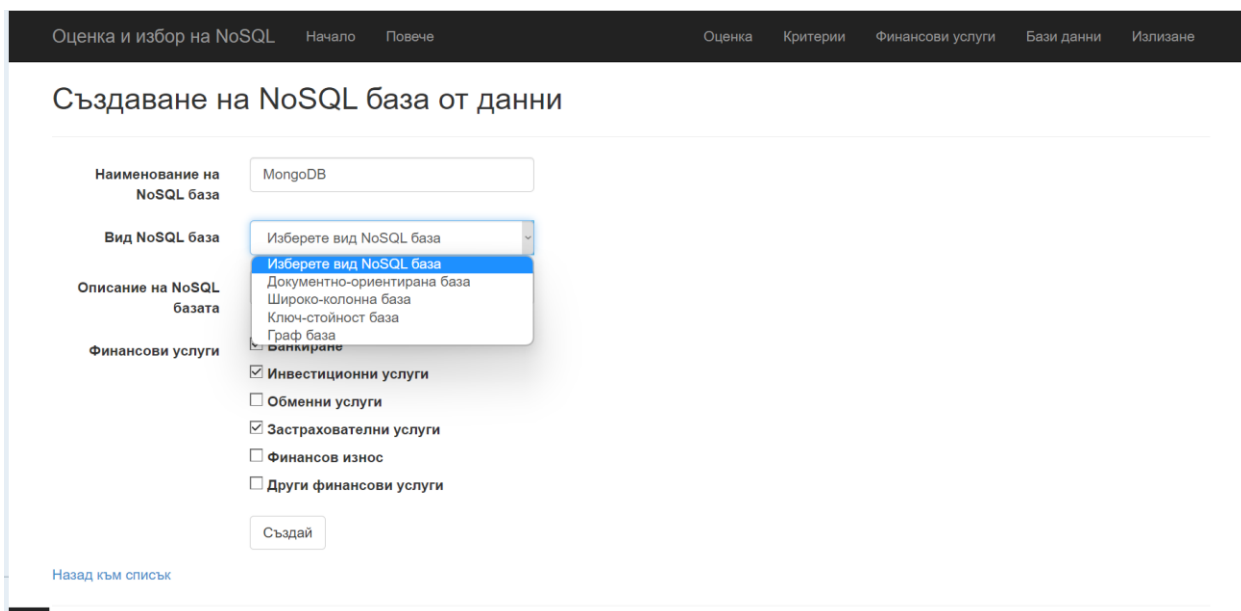


Figure 24 – Create NoSQL database – Step 2

ii. Editing

The tool provides the ability to edit the already created non-relational database, and all its elements can be changed.

iii. Delete

If necessary, already redundant databases can be deleted from the NoSQL database evaluation and selection tool.

iv. List with NoSQL databases

For the convenience of an administrator/expert, all available non-relational databases with their type, description and which financial services they work for are displayed in a list.

e. Financial services

i. Adding

Through the created interface, the different types of services that are available on the market and implemented by the institutions are added to the tool. The available list in the appendix of the previously discussed in Chapter I financial services - Figure 25.

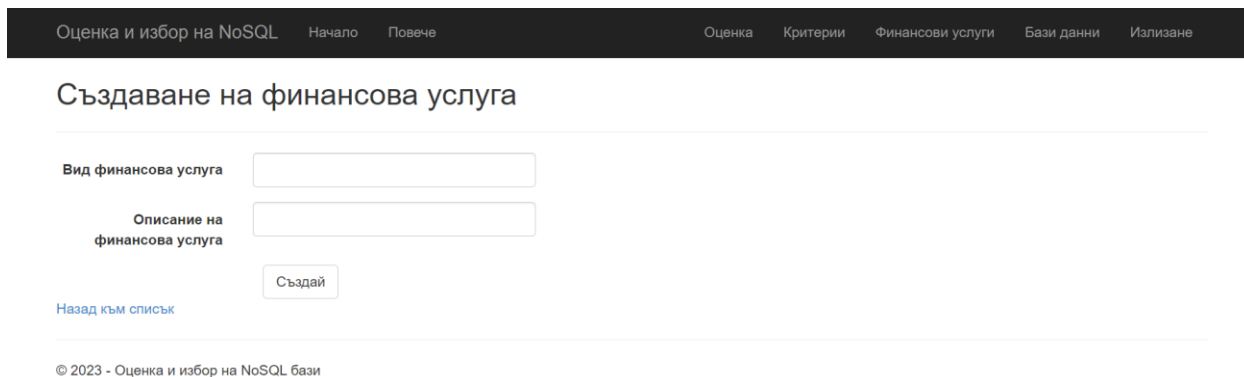


Figure 25 Create financial service

ii. Editing

From the page with a detailed description of the financial service, it is possible to review exactly what it offers and to undertake editing accordingly if necessary.

iii. Delete

Through the interface of the tool, financial services can also delete, and accordingly, before the operation is carried out, a confirmation of this is required - Figure 26.

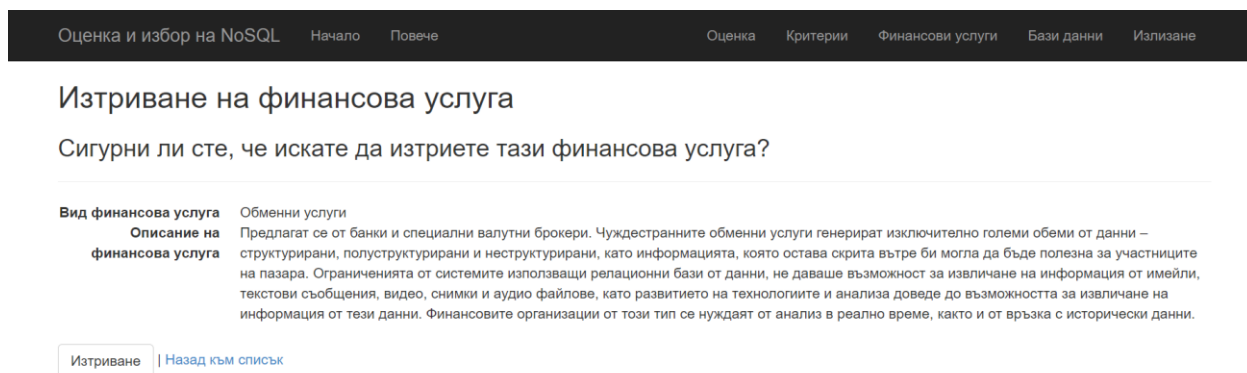


Figure 26 Financial service delete

iv. List with financial services

The financial services are presented in a complete list with their description, as well as with the options for editing and deletion.

f. Criteria

The management of criteria is realized through the created interface of the application, adding for them a shortened name to be visualized to the users, a detailed description and weight of the criterion itself.

i. Adding criteria – Фигура 27

Оценка и избор на NoSQL Начало Повече Оценка Критерии Финансови услуги Бази данни Излизане

Създаване на критерий

Съкратено наименование на критерий

Описание на критерий

Тежест на критерий

Създай

[Назад към списък](#)

© 2023 - Оценка и избор на NoSQL бази

Figure 27 – Creation of criteria

ii. Editing

Each of the elements of the added criteria can be changed by an administrator/expert through the created interface, including the weight of the criterion itself - Figure 28.

Оценка и избор на NoSQL Начало Повече Оценка Критерии Финансови услуги Бази данни Излизане

Редакция на критерии за оценка на NoSQL бази от данни

Критерий

Съкратено наименование на критерий

Описание на критерий

Тежест на критерий

Запази

[Назад към списък](#)

© 2023 - Оценка и избор на NoSQL бази

Figure 28 Editing of criteria



iii. Delete

When needed every criterion can be deleted.

iv. List with criteria

All currently available non-relational database evaluation criteria are also visualized in the admin dashboard system - Figure 29.

Съкратено наименование на критерий	Описание на критерий	Тежест на критерий
Предотвратяване на измами и мониторинг	Това е от изключителна важност за финансовите услуги – банки, кредитни и дебитни карти. Всички тези институции работят с изключително чувствителна информация, която е от изключителна важност за разбирането и удовлетворяването на клиентите. Измамите във финансовите услуги се увеличават, което може да се предотврати, чрез анализ, анализиране на аномалии, класификация и клъстеризиране, което може да се реализира с нерелационните бази от данни. Предотвратяването на измами и мониторинг на престъпления са от изключителна важност, за да се защитят клиентската лична и чувствителна информация, както и техните онлайн трансакции и данни.	5
Дигитална авторизация в реално време	Тази функционалност е от изключителна важност за финансовите услуги, за да може да е сигурно, че няма да се случи измама и за реалната идентификация на хората, с които се комуникира. NoSQL базите от данни имат такива възможности поради намалената латентност на четене, като авторизацията може да се случи почти веднага. Дигиталната авторизация в реално време е директно свързана с предотвратяването на измами, като по тази причина отново е с изключително висок приоритет, когато се говори за финансови услуги и е необходимо да бъдат реализирани, когато се правят онлайн трансакции.	5
Персонализирани оферти	За клиентите е от изключителна важност да получават персонализирани оферти, които да отговарят на техните нужди. Това не е по-различно във финансовите услуги – осигурителни полици базирани на персоналния риск, лихвени проценти на база погасяване и капацитет и др. За реализирането и представянето на персонализираните оферти на клиентите е необходимо да се събере информация – поведение, ангажираност в социалните медии, бисквитки, онлайн пазаруване и др., като всички данни идват от различни източници, като за тази цел на помощ идват	3

Figure 29 List with criteria for evaluation

Conclusion

The created application based on the method of analysis and evaluation of non-relational databases aims to help and facilitate the selection of NoSQL databases for financial services, after the expert evaluation has derived specific weighting criteria, and extensive literature research and practical experience have been done for already existing practical solutions, as well as for the potential development of such solutions. Choosing a NoSQL database on which to migrate existing systems or create new ones should be done very carefully.

The selected non-relational database can subsequently be used for structural and integration transformation of a business reference architecture to a detailed architecture.

9. Chapter IV. Application of a method to design a reference architecture for financial services.

An architecture for executing financial services with NoSQL databases.

With the increasing automation and digitization of processes, and the entry of larger volumes of data of various types – mostly from semi-structured and unstructured, businesses face the enormous challenge of collecting, storing and analysing unstructured data [79]. Financial institutions and the services they offer are fast-growing businesses, and for this reason they must adapt quickly and in a timely manner to the unfolding technologies and to the increasing demands of customers, as well as to constant competition.

Based on the detailed analysis of scientific literature related to non-relational databases and their use for the implementation of financial services systems, as well as on the basis of the developed business solutions, in Chapter II a business reference architecture consisting of a data storage system was developed, an integration layer and server of a specific NoSQL that was developed and reached the realization of a physical model. Each of the four types of non-relational databases can be deployed in this model, either in combination together or separately.

Based on the result obtained from the method again developed in Chapter II, the specific ICT architecture that will be developed is a structural and integration transformation of a business reference architecture to a detailed one, for the use of NoSQL databases for financial services.

The ICT architecture that was tested practically consists of a data storage system - MongoDB, an integration layer - Apache Spark Connector for MongoDB and a specific NoSQL server, in this case the Hadoop distributed file system was used.

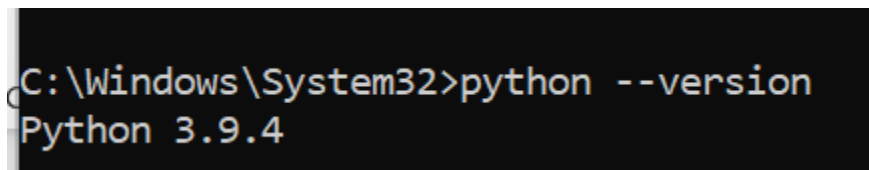
The MongoDB Data Connector over Spark enables the integration of the document-oriented database MongoDB and Apache Spark, allowing users to implement complex analyzes with large data sets. In addition to enabling read operations, the connector also enables data to be written back to MongoDB through the Spark connector. This provides an outstanding opportunity to record results from the processing performed in Spark to be written back to MongoDB and then used for further analysis.

Installation and configuration of necessary components to implement the financial services reference architecture prototype.

The following components are required for the implementation of the physical model of the business reference architecture:

- ✓ Hadoop
- ✓ Apache Spark connector for MongoDB
- ✓ Python

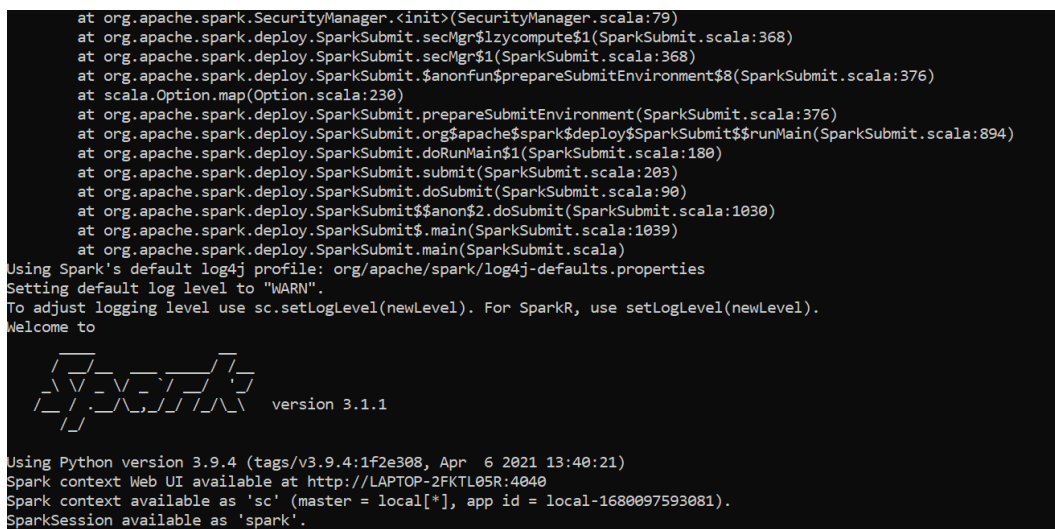
Now, the ICT architecture is implemented at the local level - a MongoDB connection through the Apache Spark connector to Hadoop. It starts with a local Python installation, as PySpark will be used. To install and configure Python, the necessary files are downloaded, which are run on the local machine, using standard installation software available. After the completion of and, the available version of Python on the machine is checked to ensure that the process completed correctly and work on the next steps can be continued. In the Command Prompt, type the following `python --version` command, and if the installation process was successful, the following result should appear - Figure 30.



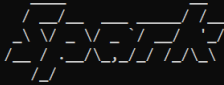
```
C:\Windows\System32>python --version
Python 3.9.4
```

Figure 30

Installing Spark locally is done after selecting the configurations and downloading the necessary files to the machine. Upon proper installation of Spark and entering the following command in the Command Prompt: `C:\Spark\spark-3.1.1-bin-hadoop2.7\bin\spark-shell`, Apache Spark starts and displays the screen below – Figure 31.



```
at org.apache.spark.SecurityManager.<init>(SecurityManager.scala:79)
at org.apache.spark.deploy.SparkSubmit.secMgr$lzycompute$1(SparkSubmit.scala:368)
at org.apache.spark.deploy.SparkSubmit.secMgr$1(SparkSubmit.scala:368)
at org.apache.spark.deploy.SparkSubmit.$anonfun$prepareSubmitEnvironment$8(SparkSubmit.scala:376)
at scala.Option.map(Option.scala:230)
at org.apache.spark.deploy.SparkSubmit.prepareSubmitEnvironment(SparkSubmit.scala:376)
at org.apache.spark.deploy.SparkSubmit.org$apache$spark$deploy$SparkSubmit$$runMain(SparkSubmit.scala:894)
at org.apache.spark.deploy.SparkSubmit.doRunMain$1(SparkSubmit.scala:180)
at org.apache.spark.deploy.SparkSubmit.submit(SparkSubmit.scala:203)
at org.apache.spark.deploy.SparkSubmit.doSubmit(SparkSubmit.scala:90)
at org.apache.spark.deploy.SparkSubmit$$anon$2.doSubmit(SparkSubmit.scala:1030)
at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:1039)
at org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to

 version 3.1.1

Using Python version 3.9.4 (tags/v3.9.4:1f2e308, Apr 6 2021 13:40:21)
Spark context Web UI available at http://LAPTOP-2FKTL05R:4040
Spark context available as 'sc' (master = local[*], app id = local-1680097593081).
SparkSession available as 'spark'.
```

Figure 31

When Apache Spark is properly installed and started, in addition to the Command prompt message, we are provided with a browser-based user interface that can be used to monitor the various processes taking place - Figure 32, at the following address: <http://localhost:4040/>.

The screenshot shows the Apache Spark web UI interface. At the top, there is a navigation bar with the Spark logo (version 3.1.1) and tabs for Jobs, Stages, Storage, Environment, Executors, and SQL. The current page is titled 'Stages for All Jobs' and shows 'Failed Stages: 1'. Below this, there is a section for 'Failed Stages (1)' with a table listing the failed stage. The table has columns for Stage Id, Description, Submitted, Duration, Tasks (Succeeded/Total), Input, Output, Shuffle Read, Shuffle Write, and Failure Reason. The first row shows a failed stage with the following details:

Stage Id	Description	Submitted	Duration	Tasks: Succeeded/Total	Input	Output	Shuffle Read	Shuffle Write	Failure Reason
0	showString at NativeMethodAccessorImpl.java:0 +details	2023/03/29 16:49:47	10 s	0/1 (1 failed)					Job aborted due to stage failure: Task 0 in stage 0.0 failed 1 times, most recent failure: Lost task 0.0 in stage 0.0 (TID 0) (LAPTOP-2FKTL05R executor driver): org.apache.spark.SparkException: Python worker failed to connect back. +details

Figure 32

The next step to implement the physical model of the business reference architecture that needs to be done is the Hadoop installation. It is downloaded, the necessary machine configurations are made, as well as additional configurations are added to some of the Hadoop files that can be found in the appendices of this dissertation. To verify that the installation was successful, type the following command in the machine's Command prompt: `cd Hadoop-2.9.2\sbin`, and the following should appear in Figure 33.

```
C:\>cd Hadoop-2.9.2\sbin
C:\hadoop-2.9.2\sbin>
```

Figure 33

In the browser, the following address: `http://localhost:50070/` opens the user interface through which, when Hadoop is started correctly, the data clusters and others can be traced – Figure 34.

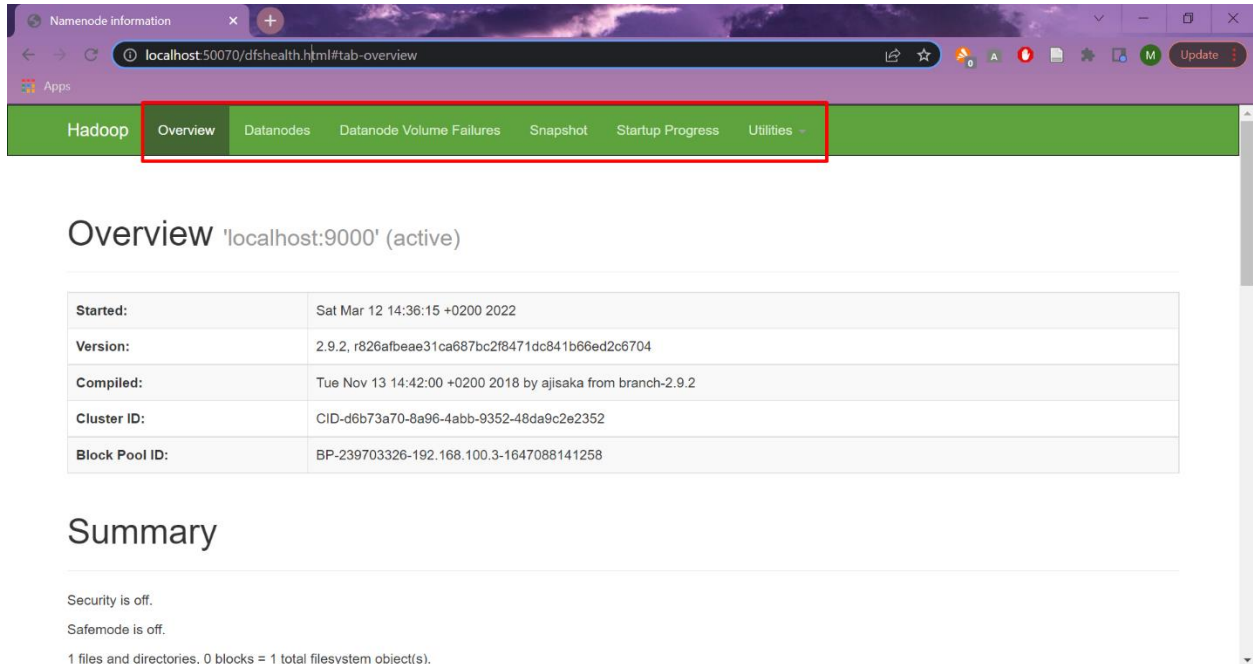


Figure 34

The last step to implement the physical architecture is the installation of the Apache Spark connector for MongoDB, which is implemented using Pyspark and the following `--packages` commands, and then from the options provided, `mongo-spark-connector` is selected, which is installed on the machine.

The initialization of the connections happens through the following code:

```
./bin/pyspark --conf "spark.mongodb.read.connection.uri=mongodb://localhost:27017/FinancialData.Stocks?readPreference=primaryPreferred" --conf "spark.mongodb.write.connection.uri=mongodb://localhost:27017/FinancialData.Stocks" --packages org.mongodb.spark:mongo-spark-connector_2.12:10.1.1
```


Application of typical unstructured data in financial services

Every single financial service works with different types of data – structured, semi-structured and unstructured, collected from different sources in different formats. The established reference architecture for financial services with the document-oriented database MongoDB connected through an Apache Spark connector to Hadoop works with data in a document version.

This type of unstructured data is typical of many services in core banking, e-payments, stocks, bonds, personalization, e-wallet, insurance, customer clearing, customer management and others. For this reason, the reference architecture created is tested with data such as contracts, slips, emails, reports, stock, and bond data, and more. Part of this data is mock-up, generated additionally since the data with which financial services work is extremely sensitive and cannot be found with free access.

Testing the reference architecture with financial data

- Source of financial data - stocks

Data about and related to customers that is generated, used, and stored in financial institutions is extremely sensitive and should not be freely distributed, and should also be stored according to the requirements for personal data protection. [81] [82] Financial data is very difficult to find freely on the Internet, and freely available stock data from Forbes, Nasdaq, Nyse and SP500 in json format was used for testing purposes – Table 5.

Forbes, Nasdaq, Nyse and SP500 data contain the following fields:

Полега	Стойности
currency	СЪОТВЕТНАТА ВАЛУТА
symbol	СИМВОЛ
exchangeName	ОБМЕННО ИМЕ
instrumentType	ТИП ИНСТРУМЕНТ
firstTradeDate	ПЪРВА ДАТА НА ПРОДАВАНЕ
regularMarketTime	РЕГУЛЯРНО ВРЕМЕ НА ПАЗАРА
gmtoffset": -18000	
timezone	ВРЕМЕНА ЗОНА
exchangeTimezoneName	ВРЕМЕНА ЗОНА НА ОБМЕН
regularMarketPrice	РЕДОВНА ЦЕНА НА ПАЗАРА

chartPreviousClose	
priceHint	Подсказка за цената
currentTradingPeriod	Настоящ период за търгуване
pre	
timezone	Часова зона на предварителната продажба
start	Стартова дата
end	Крайна дата
gmtoffset	Разлика от времето по Гринуич
regular	
timezone": "EST",	Часова зона на регулярна продажба
"start": 1670855400,	Стартова дата
"end": 1670878800,	Крайна дата
gmtoffset	Разлика от времето по Гринуич
post	
timezone	Часова зона след продажба
start	
end	Крайна дата
gmtoffset	Разлика от времето по Гринуич
dataGranularity	Детайлност на данните
range	Диапазон
validRanges	Валидни времеви диапазони

Table 5 Test data fields

Stock data from the various sources - Forbes, Nasdaq, Nyse and SP500 is over 10GB, which will be loaded into the architecture that is implemented in this chapter.

- Loading financial data

The created architecture can exchange data from MongoDB through the MongoDB connector to Apache Spark and vice versa. Loading data from Hadoop to Mongo is implemented through PySpark.

To do this, it is first necessary to add the necessary libraries and create a Spark session with the following code:

```
spark = SparkSession.builder.FinacialTest("HadoopToMongoDB")  
.config("spark.mongodb.output.uri","mongodb://localhost:27017/FinancialData.Stocks")  
.getOrCreate()
```

The code below loads the data from Hadoop into PySpark Frame:

```
df = spark.read.format("com.mongodb.spark.sql.DefaultSource").option("uri", "mongodb://  
http://localhost:50070/FinancialData.Stocks.load()
```

Finally, we write the data from Hadoop to MongoDB, using the following code::

```
df.write.format("com.mongodb.spark.sql.DefaultSource").mode("append").option("uri",  
"mongodb://localhost:27017/FinancialData.Stocks") .save()
```

Figure 35 visualizes the data loaded through the Apache Spark connector to MongoDB from Hadoop.

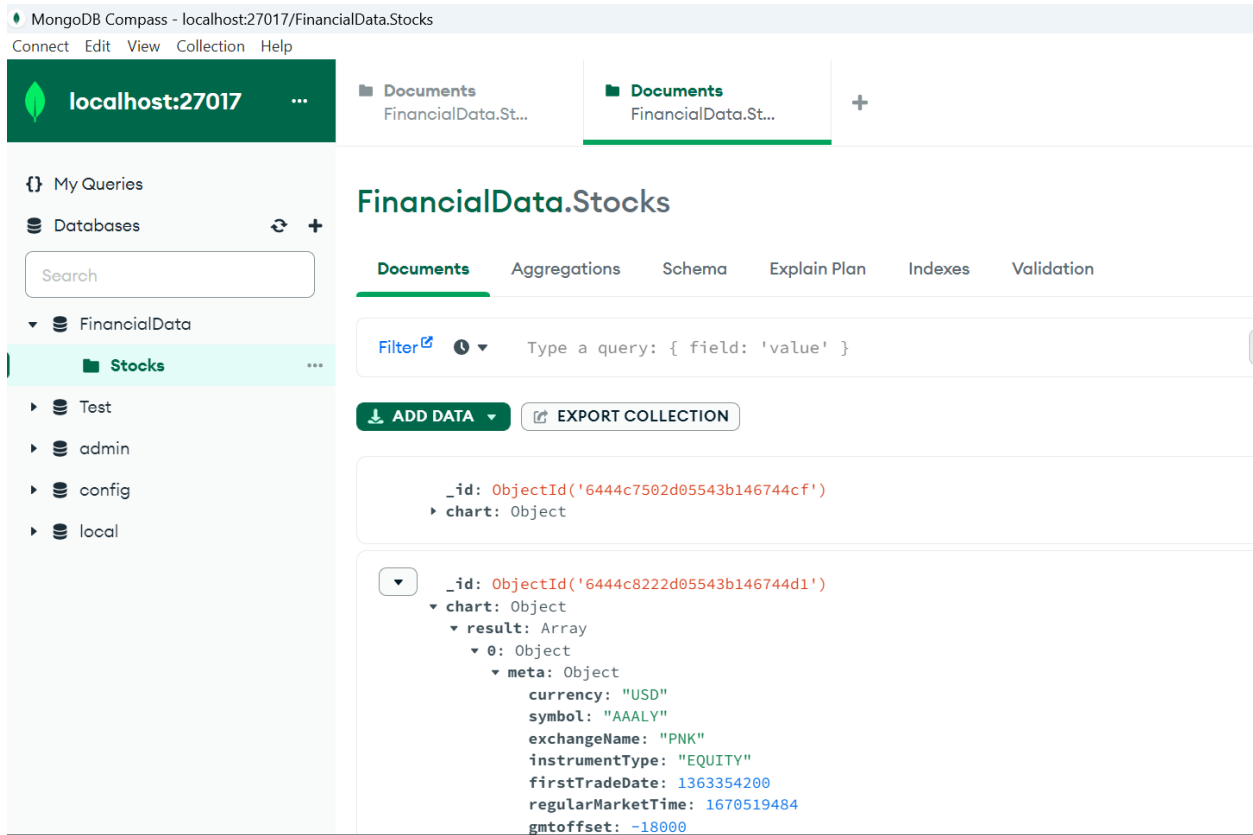


Figure 35

```

client = MongoClient()
db = client['Financial_data']
collection = db['stocks']
spark = SparkSession.builder
  .appName("Financial Stock")
  .getOrCreate()
df = spark.read.format("com.mongodb.spark.sql.DefaultSource")
  .option("database", "Financial_data")
  .option("collection", "stocks")
  .load()
df = df.filter((col("symbol") == "ACGL") & (col("firstTradeDate") >= "2017-05-02") &
(col("firstTradeDate") <= "2017-10-10"))

```

```
result = df.groupBy(year("date").alias("year"), month("date").alias("month"))
    .agg({"close": "avg"}) .orderBy("year", "month")
result.write.format("com.mongodb.spark.sql.DefaultSource").option("database",
"Financial_data") .option("collection", "stock_analysis") .mode("overwrite") .save()
client.close()
spark.stop()
```

The generated PySpark program uses the stock data that was loaded at the start of the reference architecture experiment from Hadoop to MongoDB. The first step is to connect to MongoDB and then open the Spark session. The data is filtered by the name of the shares, and subsequently only those whose first sale date is between the set interval are searched. The program then groups the result by date and outputs the average value of the price of the stock for the entire period in the interval. The data is saved to the MongoDB database and the session is closed.

Conclusion

A conceptual, logical, and physical model of a business reference architecture was designed, which were created using the reference architecture creation approach in Chapter I. Based on these models, the physical model of the business reference architecture for financial services was successfully implemented in Chapter IV with NoSQL databases, and specifically the connection of the document-oriented database - MongoDB, through the Apache Spark connector for MongoDB with Hadoop. The implemented architecture enables the storage, exchange and analysis of data with the capabilities offered by Apache Spark and MongoDB.

Conclusion

With the development of technology, digitization and digital transformation that is happening in all spheres, data is generated faster and faster, with ever greater volumes and greater speed. Fields such as financial services operating with historical systems that are based on relational databases are forced to evolve and migrate their systems to non-relational databases in order to be able to adapt to the new demands of their customers and cope with with increasing competition, as well as with the rapid development of technology and the new services they offer. The dissertation offers a developed method for evaluation and selection of NoSQL databases, based on an analysis of literary sources and real practical examples. The method leads to the

creation of a tool for evaluating and selecting non-relational databases. This becomes the basis for the design of a reference NoSQL database architecture, which is practically implemented, by connecting MongoDB to the Apache Spark connector for Mongo and, through it, connecting to Hadoop.

At the beginning of this dissertation, three hypotheses were defined:

Hypothesis 1:

It is possible to create a reference architecture applicable to the storage, processing and operation of Big Data related to financial services.

Hypothesis 2:

It is possible to quantify a reference architecture by creating an evaluation method as well as a tool to automate its implementation.

Hypothesis 3:

It is possible to achieve an efficient integration of the components of the Financial Services Reference Architecture with NoSQL databases for Big Data work.

In view of the scientific and research work carried out, as well as the realized practical prototype, it can be concluded that the hypotheses have been proven, the tasks defined at the beginning of this dissertation work have been fulfilled, which gives us reason to consider that the purpose of the dissertation has been fulfilled.

List of publications on the topic of the dissertation

- **Published**
- Mariana Kovacheva, Storing Big Data in NoSQL Databases Compared to SQL – Advantages and Problems, 10TH INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGY AND STATISTICS IN ECONOMY AND EDUCATION ICAICTSEE – 2020, November 27 – 28th, 2020, University of National and World Economy, Sofia, Bulgaria, ISSN 2367-7635

(PRINT),ISSN 2367-7643 (ONLINE), достъпна на: <https://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2020.pdf>

- Mariana Kovacheva, Ivona Velkova, Digitalization in Bulgarian Higher Education – Present and Future Opportunities, 4th International Academic Conference on Education, 2021, 10th December 2021, Barcelona, Spain, ISBN: 978-609-485-239-8), достъпна на: <https://www.dpublication.com/proceeding/4th-iaceducation/#Table-of-Contents>
- Mariana Kovacheva, Evaluation of NoSQL Databases for Digital Financial Services Implementation, AUTOMATICA and INFORMATICS, ISSN 0861-7562 (Print), ISSN 2683-1279 (Online), Year LV No. 4/2022, достъпна на: <https://sai-bg.com/wp-content/uploads/2023/01/AI-4-2022.pdf>
- Mariana Kovacheva, Creation of an automated tool for evaluation and selection of NoSQL databases intended for financial services based on predefined criteria, VIII INTERNATIONAL SCIENTIFIC CONFERENCE HIGH TECHNOLOGIES. BUSINESS.SOCIETY 2023, 06-09.03.2023, BOROSETS, BULGARIA, ISSN 2535-0005(PRINT), ISSN 2535-0013 (ONLINE), достъпна на: <http://hightechsociety.eu/sbornik/2023.pdf>
- **Expected**
- Mariana Kovacheva, Classification, comparison, and criteria for choosing NoSQL databases.

10. Literature

- [1] “What is Financial Services?” <https://insights.btoes.com/what-is-financial-services> (отворен на Яну 19, 2022)
- [2] “Big Data in Banking – Leapfrogging into Digital Banking Era,” TechVidvan. <https://techvidvan.com/tutorials/big-data-banking/#:~:text=Banks%20generate%20various%20types%20of,added%20to%20the%20bank's%20database>. (отворен на Фев 24, 2022)
- [3] Paul Golden, “Big data comes to FX,” Euromoney, Nov. 03, 2014. <https://www.euromoney.com/article/b12kky4qp36mf4/big-data-comes-to-fx> (отворен на Яну 19, 2022)
- [4] Kate Blumberg, “7 Big Data Use Cases in Financial Services and Benefits of Data Science,” SAFEGRAPH, Nov. 04, 2021. <https://www.safegraph.com/blog/top-big-data-use-cases-financial-services> (отворен на Фев 24, 2022)
- [5] Market Trends, “Use of Big Data Analytics in Investing,” Analytix Insight, Jun. 02, 2022. <https://www.analyticsinsight.net/use-of-big-data-analytics-in-investing/>
- [6] Ritesh Pathak, “7 Uses of Big data in the insurance industry,” Analytic Steps, Jan. 09, 2021. <https://www.analyticssteps.com/blogs/7-uses-big-data-insurance-industry> (отворен на Фев 28, 2022)
- [7] JULIA KAGAN, “Insurance: Definition, How It Works, and Main Types of Policies,” Investopedia, Jul. 18, 2022. <https://www.investopedia.com/terms/i/insurance.asp#toc-what-are-the-four-major-types-of-insurance> (отворен на Фев 24, 2022)
- [8] JULIA KAGAN, “How Do Commercial Banks Work, and Why Do They Matter?,” Investopedia, Jan. 23, 2023. <https://www.investopedia.com/terms/c/commercialbank.asp#:~:text=Commercial%20banks%20provide%20basic%20banking,such%20as%20safe%20deposit%20boxes> (отворен на Фев 24, 2022)
- [9] “Financial services,” Wikipedia. https://en.wikipedia.org/wiki/Financial_services (отворен на Май 28, 2022)
- [10] “Big Data in Finance - Your Guide to Financial Data Analysis,” Talend. <https://www.talend.com/resources/big-data-finance/> (отворен на Юни 15, 2022)

- [11] MEDICI, “Why Financial Services Should Look to NoSQL.” [Online]. Available: http://pages.aerospike.com/rs/229-XUE-318/images/Aerospike_Wp_Why_Financial_Services_Should_Look_to_NoSQL.pdf (отворен на Март 11, 2023)
- [12] Kazim Hussain and Elsa Prieto, “Big Data in the Finance and Insurance Sectors,” in *New Horizons for a Data-Driven Economy*, Springer Open, p. 222. [Online]. Available: <https://library.oapen.org/bitstream/handle/20.500.12657/27764/1002241.pdf?sequence#page=222>
- [13] Sabrina Sicari, Alessandra Rizzardi, and Alberto Coen-Portisini, “Security & privacy issues and challenges in NoSQL databases,” vol. 206, Apr. 2022, doi: <https://doi.org/10.1016/j.comnet.2022.108828>.
- [14] “Big Data Insight,” SAS. https://www.sas.com/en_us/insights/big-data/what-is-big-data.html (отворен на Юни 16, 2022)
- [15] “Big Data Characteristics,” JavaTPoint. <https://www.javatpoint.com/big-data-characteristics>
- [16] Ravi Kiran, “Big Data Characteristics: Know the 5’Vs of Big Data,” Mar. 13, 2023. <https://www.edureka.co/blog/big-data-characteristics/> (отворен на Май 28, 2022)
- [17] Mark Smallcombe, “Structured vs Unstructured Data: 5 Key Differences,” Integrate.io, Feb. 16, 2023. <https://www.integrate.io/blog/structured-vs-unstructured-data-key-differences/> (отворен на Юни 18, 2022)
- [18] “Unstructured Data,” MongoDB. <https://www.mongodb.com/unstructured-data> (отворен на Юни 18, 2022)
- [19] Bernard Marr, “What’s The Difference Between Structured, Semi-Structured And Unstructured Data?,” Forbes, Oct. 18, 2019. <https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/?sh=58dbd11b2b4d> (отворен на Май 28, 2022)
- [20] “NoSQL,” TechTerms, Aug. 27, 2013. <https://techterms.com/definition/nosql> (отворен на Юни 1, 2022)
- [21] “What is a Document Database?,” MongoDB. <https://www.mongodb.com/document-databases> (отворен на Юни 15, 2022)
- [22] “What Is a Key-Value Database?,” AWS. <https://aws.amazon.com/nosql/key-value/> (отворен на Юни 15, 2022)
- [23] Margaret Rouse, “Columnar Database,” Techopedia, Aug. 16, 2013. <https://www.techopedia.com/definition/13782/columnar-database> (отворен на Юли 10, 2022)
- [24] “What Is a Graph Database?,” AWS. <https://aws.amazon.com/nosql/graph/> (отворен на Юли 10, 2022)

- [25] Mitko Radoev , "A Comparison between Characteristics of NoSQL Databases and Traditional Databases," Computer Science and Information Technology, Vol. 5, No. 5, pp. 149 - 153, 2017. DOI: 10.13189/csit.2017.050501.
- [26] "Apache Hadoop," Hadoop. <https://hadoop.apache.org/> (отворен на Август 29, 2022)
- [27] "What is Apache Hadoop?," IBM. <https://www.ibm.com/analytics/hadoop> (отворен на Август 29, 2022)
- [28] "What is Hadoop?," IntelliPaat, 09 Mar, 2023. <https://intellipaate.com/blog/what-is-hadoop/> (отворен на Август 29, 2022)
- [29] "What is Hadoop," JavaTPoint. <https://www.javatpoint.com/what-is-hadoop> (отворен на Август 29, 2022)
- [30] "Reference Architecture," HPE. <https://www.hpe.com/us/en/what-is/reference-architecture.html> (отворен на Септември 16, 2022)
- [31] Lavanya Rathnam, "What Is Reference Architecture?," TechGenix, Aug. 02, 2022. <https://techgenix.com/reference-architecture-guide/> (отворен на Септември 30, 2022)
- [32] "Banking industry architecture," IBM. <https://www.ibm.com/cloud/architecture/architectures/banking/reference-architecture> (отворен на Октомври 2, 2022)
- [33] "MongoDB for Financial Services," MongoDB. <https://www.mongodb.com/industries/financial-services> (отворен на Октомври 2, 2022)
- [34] "Financial Management System Database," Couchbase. <https://www.couchbase.com/solutions/nosql-for-financial-services/> (отворен на Октомври 15, 2022)
- [35] Max Ivashchenko, Jacky Wu, and Rohit Talluri, "DynamoDB FSI Service Spotlight," AWS, Aug. 17, 2022. <https://aws.amazon.com/blogs/industries/dynamodb-fsi-service-spotlight/> (отворен на Октомври 23, 2022)
- [36] "Redis Enterprise for financial services," Redis. <https://redis.com/solutions/industries/financial-services/> (отворен на Октомври 27, 2022)
- [37] "Apache HBase Features," Apache HBase. <https://hbase.apache.org/index.html> (отворен на Февруари 12, 2023)
- [38] Rich Edwards, "Exploring Common Apache Cassandra Use Cases," Datastax, Jun. 12, 2021. <https://www.datastax.com/blog/exploring-common-apache-cassandra-use-cases> (отворен на Декември 12, 2022)

[39] Alex Bekker, “When to use Cassandra and when to steer clear,” Towardsdatascience, Aug. 14, 2018. <https://towardsdatascience.com/when-to-use-cassandra-and-when-to-steer-clear-72b7f2cede76> (отворен на Яну 16, 2023)

[40] Nav Mathur, “Graph Technology for Financial Services How Top Financial Firms Harness Connected Data to Increase Their Bottom Line,” Neo4j. https://go.neo4j.com/rs/710-RRC-335/images/Neo4j-in-Financial%20Services-white-paper.pdf?_ga=2.90740444.799262140.1679753501-1625560426.1676516677&_gl=1*1p7o7b8*_ga*MTYyNTU2MDQyNi4xNjc2NTE2Njc3*_ga_DL38Q8KGQC*MTY3OTc1MzUwMC40LjEuMTY3OTc1Mzg2Ny42LjAuMA (отворен на Яну 16, 2023)

[41] “Payment Service Provider Fights Fraud faster and more efficiently with Orient DB Multi-Model Graph Database.,” Orient DB. https://orientdb.com/wp-content/uploads/OrientDB_99Bill_case_study.pdf (отворен на Яну 29, 2023)