



# ТЕХНИЧЕСКИ УНИВЕРСИТЕТ-ГАБРОВО

Факултет „Електротехника и електроника“

маг. инж. Велин Сабинов Хаджиев

Моделиране на операции по структуриране, съхраняване  
и обработка на данни в Интернет

## А В Т О Р Е Ф Е Р А Т

на дисертация

за придобиване на образователна и научна степен „доктор“

Област на висше образование: 5. Технически науки

Професионално направление: 5.3. Комуникационна и  
компютърна техника

Докторска програма: Автоматизирани системи за обработка  
на информация и управление

Научен ръководител: доц. д-р инж. Алдениз Енверов Рашидов

Рецензенти: 1. проф. д.т.н. инж. Райчо Тодоров Иларионов

2. доц. д-р инж. Галина Иванова Иванова

Габрово, 2024 г.

Дисертационният труд е обсъден и насочен за официална защита на заседание на Разширен катедрен съвет на катедра „Автоматика, информационна и управляваща техника“ към факултет „Електротехника и електроника“ на Технически университет – Габрово, проведен на 11.12.2024 г.

Дисертационният труд съдържа 156 страници. Научното съдържание е представено в увод, четири глави и заключение, и включва 32 фигури и 13 таблици. Цитирани са 118 литературни източника. Номерацията на фигурите, таблиците и формулите в автореферата е в съответствие с тази в дисертацията.

Изследванията по дисертационния труд са извършени в катедра „Автоматика, информационна и управляваща техника“ към факултет „Електротехника и електроника“ на Технически университет – Габрово.

Официалната защита на дисертационния труд ще се състои на 12.03.2025 г. от 13:00 ч. в зала 3410 на Технически университет – Габрово.

Материалите по защитата са на разположение за интересуващите се в кабинет 3209, корпус №3 Ректорат на Технически университет – Габрово.

Рецензиите и становищата на членовете на научното жури и авторефератът са публикувани на сайта на университета: [www.tugab.bg](http://www.tugab.bg).

Автор: © Велин Сабинов Хаджиев  
имейл: [hadjo75@abv.bg](mailto:hadjo75@abv.bg)

Заглавие: Моделиране на операции по структуриране, съхраняване  
и обработка на данни в Интернет

Тираж: 8 бр. (на Български език)

Място на отпечатване: Университетско издателство „Васил Априлов“  
при Технически университет - Габрово

## **ОБЩА ХАРАКТЕРИСТИКА НА ДИСЕРТАЦИОННИЯ ТРУД**

### **Актуалност на проблема**

В ежедневната си дейност – научна, изследователска, бизнес процеси и др. – хората непрекъснато генерират данни, които често трябва да бъдат съхранени в определен формат и схема, за да могат по-късно да бъдат използвани за обработка и различни нужди. Съхраняването на данните в определена схема предполага те да бъдат подредени и съхранени по начин, който позволява тяхната обработка с цел извличане на зависимости и стойностна информация. Поддредането и съхраняването на данните в определена схема е свързано с извършването на операции по структуриране, съхраняване и обработка на тези данни, които винаги се извършват от специалисти.

*Настоящият дисертационен труд* е посветен на актуален научен проблем – намирането на решение, което да позволи на всеки потребител в Интернет, генериращ или обработващ данни, да описва по свое желание структурата за съхраняване на данните, както и да извършва манипулации върху тези данни (въвеждане, редактиране, споделяне и др.).

### **Цел и задачи на дисертационния труд**

**Целта на настоящия дисертационен труд** е моделиране на операции по структуриране, съхраняване и обработка на данни в Интернет.

В съответствие с поставената цел се решават следните основни задачи:

1. Модифициране и изследване на методи за структуриране, съхраняване и обработка на данни.
2. Разработване и изследване на алгоритми за структуриране, съхраняване и обработка на данни.
3. Разработване на модели на операции по структуриране, съхраняване и обработка на данни в Интернет.
4. Изследване и сравнение на предложените модели.
5. Определяне на факторите, характеризиращи информационните системи като системи за структуриране, съхраняване и обработка на данни в Интернет.
6. Приложение на разработените модели при проектиране и разработване на системи за структуриране, съхраняване и обработка на данни в Интернет.

### **Методи на изследване**

*Методите на изследването*, използвани при решаването на поставените задачи в дисертацията, са: теоретичен анализ, компютърно проектиране, моделиране и симулационни изследвания, като са приложени подобрени за целта методи за структуриране, съхраняване и обработка на данни. Реализирано е моделиране на процеси по структуриране, съхранение и обработка на данни в Интернет, включително симулация на функционирането на хибриден модел за управление на разпределени данни и неговото взаимодействие с глобалната мрежа. Извършен е анализ на ефективността и рентабилността на внедряване на подобни модели, което допринася за подобряване на качеството и достъпността на информационната инфраструктура за потребителите.

### **Приложимост**

*Приложимостта на дисертационния труд* се изразява в разработването на метод за оптимизация на операциите с данни, който интегрира добри практики и техники за тяхното структуриране, съхраняване и обработка. Ефективността на предложения

хибриден модел е доказана чрез симулации и тестове в реални условия, които оценяват неговата производителност, устойчивост и мащабируемост. На базата на този метод е създадена уеб-базирана система за управление на данни, която осигурява достъп до бази данни за широк кръг от потребители. Проведените тестове в практически сценарии потвърждават високата ѝ ефективност и приложимост в реални условия.

### **Апробация на дисертационния труд**

Основните резултати показани в дисертационния труд са представени в девет публикации на международни конференции и научни издания, като четири от тях са в чужбина и пет в България. Едната от тях е самостоятелна. Броят на тези публикации напълно покрива минималните изисквания относно разглеждания критерий. Два от трудовете са изнесени на Международна научна конференция „Автоматика и информатика, 2019“, като и двата са публикувани в списание „Списание Автоматика и информатика, 2019“, кн.2 и кн.3. Една от публикациите е изнесена на международна научна конференция ICAI, 2021, Varna, Bulgaria, три са изнесени на международна научна конференция ELECO, Bursa, Turkey през периода 2019 - 2021 и една на международна научна конференция ICCCNT 2022, Mandi, India, които са реферирани в Scopus с SJR. В публикациите са представени голяма част от извършените изследвания и са изложени основните изводи от дисертационния труд.

Постигнатите резултати от дисертационния труд са докладвани на различни научни конференции, като голяма част от изнесените доклади са в съавторство с научният ръководител. Части от този труд и дисертацията като цяло многократно са обсъждани в катедра „Автоматика, Информационна и Управляваща Техника“ на Технически университет - Габрово.

### **Структура и обем на дисертационния труд**

Дисертационният труд включва: въведение, съдържание, четири глави, класификация на приносите, публикации, свързани с дисертацията, и литература. Общият обем на труда е 156 страници, като включва 32 фигури и 13 таблици. Списъкът на използваните литературни източници съдържа 118 заглавия.

## **СЪДЪРЖАНИЕ НА ДИСЕРТАЦИОННИЯ ТРУД**

### **ГЛАВА I: ОБЗОР И АНАЛИЗ НА СЪЩЕСТВУВАЩИ МОДЕЛИ И МЕТОДИ ЗА СТРУКТУРИРАНЕ, СЪХРАНЯВАНЕ И ОБРАБОТКА НА ДАННИ В ИНТЕРНЕТ**

#### **1.1. Актуалност на темата**

В ежедневноста си дейност – научна, изследователска, бизнес процеси и др. – хората непрекъснато генерират данни, които често трябва да бъдат съхранени в определен формат и схема, за да могат по-късно да бъдат използвани за обработка и различни нужди. Съхраняването на данните в определена схема предполага те да бъдат подредени и съхранени по начин, който позволява тяхната обработка с цел извличане на зависимости и стойностна информация. Поддръждането и съхраняването на данните в определена схема е свързано с извършването на операции по структуриране, съхраняване и обработка на тези данни, които винаги се извършват от специалисти.

Настоящият дисертационен труд е посветен на актуален научен проблем – намирането на решение, което да позволи на всеки потребител в Интернет, генериращ или обработващ данни, да описва по свое желание структурата за съхраняване на данните, както и да извършва манипулации върху тези данни (въвеждане, редактиране, споделяне и др.).

## 1.2. Обзор на съществуващи модели, методи и архитектури за структуриране, съхраняване и обработка на данни в Интернет

**Метод за моделиране на корпоративни данни чрез нормализация (Normalized Data Factory Modeling):** въведен е от Бил Инмон въз основа на концепцията за нормализиране на данните [27].

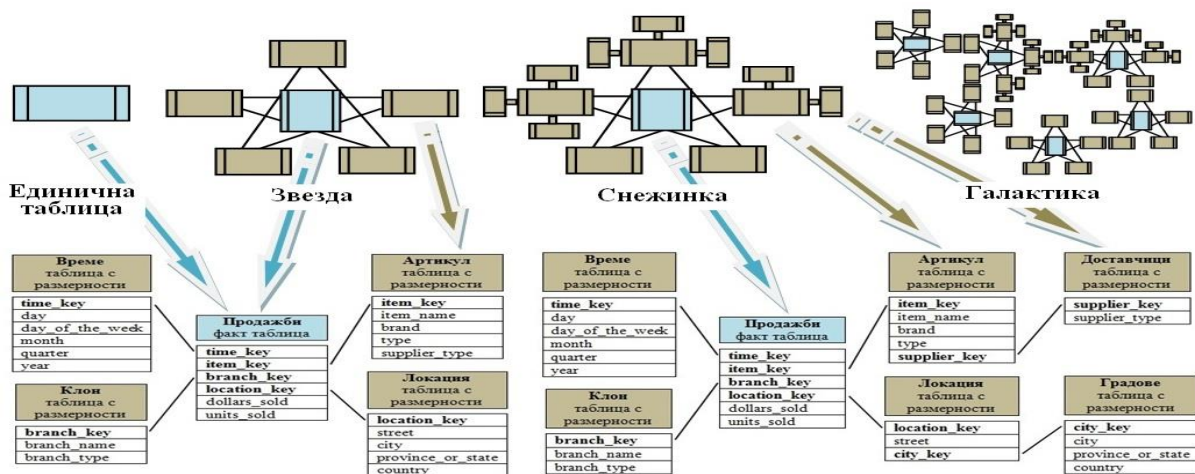
На фиг.1.1 е представен примерен процес на нормализация до трета нормална форма, който е в основата на метода, използван от Бил Инмон.



Фиг. 1.1 Примерен процес на нормализация до трета нормална форма

Процесът на нормализация има за цел да премахне излишните връзки между релациите, което води до увеличаване на ефективността и избягване на аномалиите при съхранение на данни.

**Многоизмерно моделиране (Dimensional Modeling):** Въведен от Ралф Кимбъл [27]. При многоизмерното моделиране транзакционните и референтните данни са изолирани в две основни структури: факт-таблицы (fact tables) и таблици с измерения (dimensional tables). Фигура 1.2 илюстрира схемите, използвани при изграждането на системи за съхранение на данни.



Фиг. 1.2 Базови схеми, използвани при многоизмерното моделиране

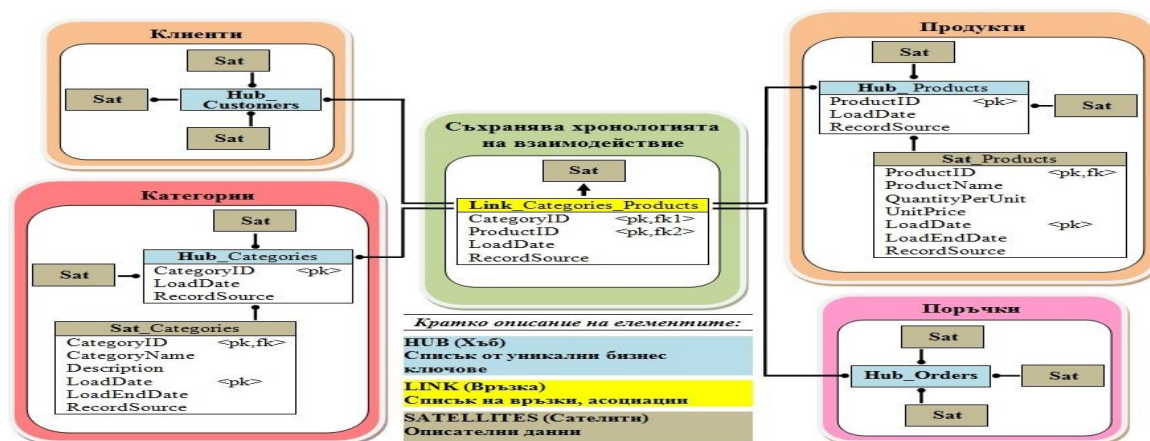
Таблица 1.1 обобщава основните разлики между схеми тип звезда и снежинка. Тези разлики показват основните предимства и недостатъци на всяка от схемите в зависимост от конкретните нужди и изисквания на даден проект или организация.

Таблица 1.1 Разлики между схема „Звезда“ и схема „Снежинка“

Характеристика	Схема „Звезда“	Схема „Снежинка“
Излишък на данни	Наличен е излишък от данни заради денормализираните измерения.	Намален излишък от данни поради нормализацията.

Характеристика	Схема „Звезда“	Схема „Снежинка“
<b>Управление на данни и съхранение</b>	По-лесно за управление и поддръжка. Необходимо е повече място за съхранение на таблиците с измерения.	По-трудно за управление и поддръжка поради сложността. Необходимо е сравнително по-малко място за съхранение на таблиците с измерения.
<b>Нормализация</b>	Съдържа денормализирани таблици с измерения, които са във втора нормална форма (2НФ).	Съдържа нормализирани таблици с измерения, които са нормализирани до трета нормална форма (3НФ).
<b>Структура</b>	Централна таблица с факти, заобиколена с денормализирани таблици с измерения.	Централна таблица с факти, заобиколена с йерархично свързани нормализирани таблици с измерения.
<b>Сложност</b>	Опростена структура, лесна за разбиране и използване. Заявките представляват директни съединения между факти и измерения за извличане на данните.	Сложна структура поради нормализацията на таблиците с измерения. Заявките представляват сложни съединения между факти и измерения за извличане на данните.
<b>Производителност</b>	Необходимо е по-малко време за изпълнение на заявките, заради по-малкото връзки между таблиците.	Необходимо е по-вече време за изпълнение на заявките поради по-много връзки между таблиците.
<b>Подход на изграждане и приложение</b>	Използва подход отгоре надолу. Подходяща за малки до средни по размер хранилища за данни.	Използва подход отдолу нагоре. Подходяща за големи хранилища за данни, където нормализацията е от съществено значение.

**Трезорно моделиране на данни (Data Vault Modelling):** Трезорното моделиране на данни (Data Vault Modelling) е метод, въведен от Дан Линстед [30][31], който се фокусира върху групиране на входящите данни в три основни типа обекти: хъб, връзка и сателит (Hub, Link, Satellite). На Фиг.1.3 са представени основните елементи, участващи в изграждането на системи за съхранение на данни, които използват методологията Data Vault, както и връзките между тях.

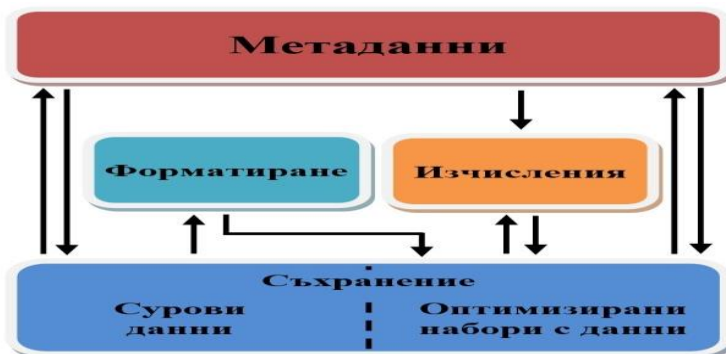


Фиг. 1.3 Основни елементи при трезорното моделиране

**Хибридно моделиране (Hybrid Modelling):** Методът представлява комбинация от многоизмерни архитектури и архитектури в 3NF, включваща оперативно хранилище за данни (ODS) с нормализирани данни и многоизмерни информационни матрици (data marts).

**Метод за моделиране на Езеро с данни (Data Lake Modelling):** Този метод е въведен от Джеймс Диксън от компанията Пентахо, който за първи път използва термина „езеро с данни“ (Data Lake) през октомври 2010 г. [36][37]. Подходът

позволява съхранение на голямо количество „сурови“ данни в естествения им формат, както и полуструктурирани и структурирани данни, с осигурен достъп до тях при поискване [38]. На фиг.1.4 са показани основните компоненти на системите за съхранение на данни, базирани на метода „езеро с данни“, и връзките между тях.



Фиг. 1.4 Основни елементи при моделирането на Езеро с данни

Предимствата и недостатъците на петте разгледани метода за структуриране и съхраняване на данни са обобщени в таблица 1.2.

Таблица 1.2 Сравнение на методите за структуриране и съхраняване на данни.

Критерии за сравнение	Методи за структуриране и съхраняване на данни				
	Normalized Data Factory Modelling	Dimensional Modelling	Data Vault Modelling	Hybrid Modelling	Data Lake Modelling
<b>Разбираемост, интерпретация и използване</b>	- Необходими са добри познания за работа с бази от данни.	+ Необходими са основни познания за работа с бази от данни.	+ Необходими са основни познания за работа с бази от данни.	Зависи от броя и сложността на методите, използвани за осигуряване на цялостно решение.	- Необходими са професионални умения, добри познания и подготовка за работа с големи обеми от данни.
<b>Наличие на дублирани данни</b>	+ Увеличава се последователността на данните, намалявайки дублирането им.	- Увеличаване на излишъкът от данни, тъй като таблиците не са йерархично разделени.	- Увеличаване на броя на таблиците до два пъти в сравнение с този при 3NF моделирането.	Зависи от броя и сложността на методите, използвани за осигуряване на цялостно решение.	- Необходим е постоянен мониторинг относно управлението и поддръжката целостта на данните. В противен случай езерото с данни може да се превърне в блато с данни.
<b>Бързодействие при търсене и запитвания</b>	+ Увеличава се бързината при търсене, поради възможността за бързо създаване на индекси. - Данните се	+ Увеличаване на производителността чрез използване прости съединения за извличане на данните. Увеличава се	- Производителността при търсене и извличане на данните не е приоритет при този метод на моделиране.	Зависи от броя и сложността на методите, използвани за осигуряване на цялостно решение.	- Производителността при търсене и извличане на данните зависи от сложността на алгоритмите, използвани за трансформа-

Критерии за сравнение	Методи за структуриране и съхраняване на данни				
	Normalized Data Factory Modelling	Dimensional Modelling	Data Vault Modelling	Hybrid Modelling	Data Lake Modelling
	съхраняват в по-вече от една таблица, което води до трудности и забавяне при изпълнение на заявки.	бързината при търсене, поради възможността за бързо създаване на индекси - Намаляване ефективността при изпълнение на заявки с увеличаване броя на таблиците с измерения.			ция на данните преди те да бъдат използвани за отчет или анализ.
<b>Приложимост в системи за онлайн обработка на данни</b>	+ Подходящ за онлайн обработка на транзакции (OLTP).	+ Подходящ за онлайн аналитична обработка (OLAP).	+ Ниска сложност на правилата за зареждане на хъбове, връзки и сателити.	Зависи от броя и сложността на методите, използвани за осигуряване на цялостно решение.	+ Висока ефективност при работа с инструменти за анализ на големи обеми от данни.
<b>Гъвкавост</b>	+ Позволява логическо групиране на подобни или свързани по същата схема данни.	- Не е подходящ за прилагане в дългосрочен план в случай, че се налагат чести промени в бизнес изискванията.	+ Отделно управление на бизнес ключовете от всички атрибути на бизнес субектите, групирайки ги в хъбове, връзки и свързани с тях сателити. Всички атрибути се управляват като бавно променящи се измерения, подобно на многоизмерното моделиране.	Зависи от броя и сложността на методите, използвани за осигуряване на цялостно решение.	+ Поради възможността за хоризонтално мащабиране могат да се съхраняват и обработват големи обеми данни Използват се технологии с отворен код, които са по-гъвкави и рентабилни от технологиите със затворен код
<b>Сложност</b>	- С нарастване броя на таблиците се увеличава и сложността на базата от данни.	- Необходимост от сложни присъединявания при изпълнение на заявки с нарастване броя на таблиците, които трябва да се добавят.	+ Всяка връзка между бизнес субекти в рамките на системата се моделира като много към много.	Зависи от броя и сложността на методите, използвани за осигуряване на цялостно решение.	- Включва големи обеми от данни, което затруднява поддръжката, управлението и използването.



**Извод:** Основната задача на дисертационния труд е да се намери решение, което да осигури на потребителите възможността да структурират, съхраняват и обработват данни според техните изисквания. Данните трябва да се съхраняват в табличен вид с определена структура и да бъдат представени с атомарните си стойности.

За проектирането и изграждането на ефективен склад за данни е необходимо да се приложи подходящият модел за съхранение на данни. Два основни модела са широко признати: моделът на Инмон и моделът на Кимбъл.

**Модел на Бил Инмон:** На фиг.1.5 е представен модел на облачна база от данни, при който за изграждането на склада за данни е използван моделът на Бил Инмон. Складът за данни се зарежда с данни от системите за онлайн обработка на транзакции (OLTP) и е в трета нормална форма (3NF). Информационните матрици са извън склада за данни, като могат да бъдат създавани нови при необходимост, които също са в 3NF. От тях се изграждат кубове с данни, които се използват за онлайн аналитичната обработка (OLAP).



Фиг. 1.5 Модел на облачна база от данни, базиран на модел на Бил Инмон

**Модел на Ралф Кимбъл:** На фиг.1.6 е представен модел на облачна база от данни, при който за изграждането на склада за данни е използван моделът на Ралф Кимбъл. Складът за данни е изграден от информационни матрици, които се зареждат с данни от OLTP системи, обикновено реляционни бази от данни в 3NF.



Фиг. 1.6 Модел на облачна база от данни, базиран на модела на Ралф Кимбъл

**Трезорен Модел:** На фиг.1.7 е представен модел на облачна база от данни, при който е използван трезорният модел на Дан Линстед. Трите нива на модела осигуряват изолиране на хранилището за необработени данни от крайните потребители и различните слоеве на извличане на данни.



Фиг. 1.7 Модел на облачна база от данни, базиран на модела на Дан Линстед

Таблица 1.3 Основни различия в моделите на Инмон, Кимбъл и Линстед

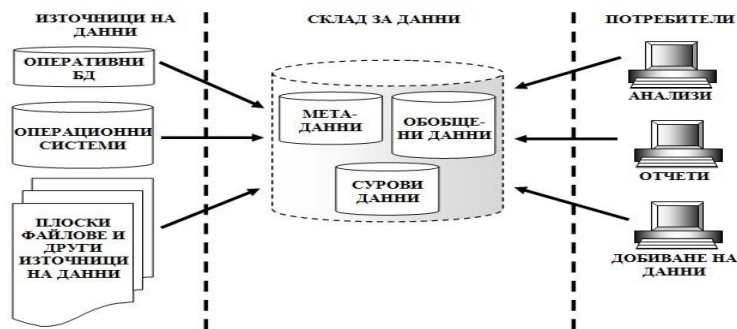
Критерий	Инмон	Кимбъл	Линстед
<b>Идеология.</b>			
<b>Основна аудитория.</b>	ИТ специалисти.	Крайни потребители.	Крайни потребители.
<b>Място в организацията.</b>	Неразделна част от корпоративната информационна система CIF (бизнес разузнаване и бизнес управление).	Трансформира и съхранява оперативни данни.	Съхранява сурови данни и данни след приложени бизнес правила.
<b>Цел.</b>	Осигуряване на стабилно техническо решение въз основа на доказани методи и технологии за структуриране, съхранение и обработка на данни.	Осигуряване на решение, което позволява на крайните потребители да инициират запитвания до базата от данни, като времето за отговор е в приемливи граници.	Осигуряване на сигурно и цялостно решение базирано на доказани методи.
<b>Методология и архитектура.</b>			
<b>Подход на изграждане.</b>	Отгоре-надолу (top-down)	Отдолу-нагоре (bottom-up)	Хибриден подход. Комбинация между 3NF и схема звезда.
<b>Предназначение.</b>	Корпоративно хранилище, съхраняващо атомарните стойности на данните. Предоставя данни за базите от данни на отделните ведомства.	Информационни матрици (Data Marts), моделиращи всеки бизнес процес по отделно. Съгласуваността на предприятието се постигва чрез шина за данни и подходящи измерения.	Трезорният склад за данни съдържа бизнес-ориентираното хранилище и хранилище за сурови данни, които предоставя данни на информационните матрици.
<b>Сложност.</b>	Доста сложен.	Доста опростен.	Доста опростен.
<b>Методология за разработка и развитие.</b>	Спирална методология за разработка и развитие.	Процес от четири стъпки за многоизмерно моделиране,	Базиран на гъвкава методология, целяща бързо разработване и внедряване.

Критерий	Инмон	Кимбъл	Линстед
		отклоняващ се от методите на релационните бази от данни.	
<b>Разходи при внедряване.</b>	Високи разходи в етапа на разработване и внедряване. По-ниски разходи за развитие и поддръжка.	Ниски разходи в етапа на разработване и внедряване. По-високи разходи за развитие и поддръжка.	По-ниски разходи в сравнение с другите два подхода.
<b>Време за внедряване.</b>	Продължителност при внедряване.	Кратко време за внедряване.	Кратко време за внедряване.
<b>Изисквания за обсъждане на физическия дизайн.</b>	Доста задълбочени.	Доста повърхностни.	Доста повърхностни.
<b>Интегриране на данните.</b>			
<b>Интегриране на множество източници.</b>	Правилата за трансформация трябва да бъдат прилагани по време на извличането и зареждането на данните.	Правилата за трансформация трябва да бъдат прилагани по време на извличането и зареждането на данните.	Разделянето на контекстните данни от бизнес ключовете намалява сложността при извличането и зареждането на данните.
<b>Сложност на процесите по извличане, трансформиране и зареждане на данните.</b>	Ниска сложност на правилата за трансформация, ако моделът на данните е подобен на моделите на източниците на данни.	Трансформациите между модела за онлайн обработка на транзакции и многоизмерния модел са сложни.	Ниска сложност на правилата за зареждане на хъбове, връзки и сателити.
<b>Моделиране на данните.</b>			
<b>Ориентация на данните.</b>	Тематична, основана на данни.	Ориентира към даден процес.	Тематична, основана на данни ориентирана към даден процес.
<b>Инструменти за реализация.</b>	Традиционни: диаграми същност - връзки и набори от данни (ERDs, DISs)	Многоизмерно моделиране, различаващо се от релационното моделиране.	Моделиране на хъбове, връзки и сателити.
<b>Достъпност от крайния потребител.</b>	Ниско.	Високо.	Високо.
<b>Управление на жизнения цикъл.</b>			
<b>Гъвкавост при промяна модела на източника на данни.</b>	Необходимост от промени в таблиците.	Честите промени в модела на източника влияят върху модела на склада за данни.	Съществуващите таблици няма да бъдат засегнати. Единствената промяна е добавянето на съответните сателити.
<b>Гъвкавост при промяна на изискванията за анализ.</b>	Необходимост от промяна на модела в случай, че необходимите данни не са налични в склада за данни.	Промените на изисквания оказват влияние върху модела на данните.	Не са необходими промени по модела. Трябва да се адаптира само доставката на данни до информационните матрици.

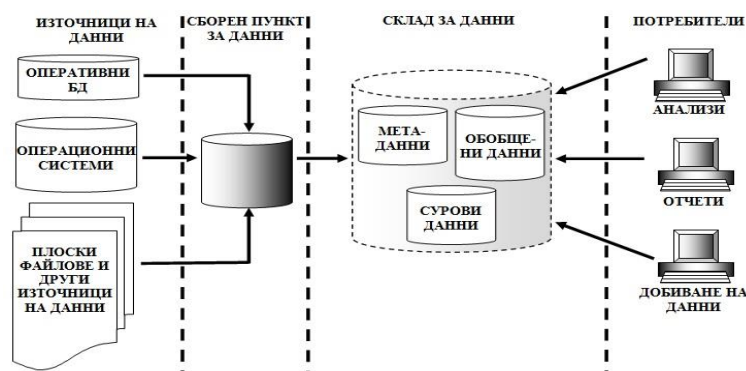
Критерий	Инмон	Кимбъл	Линстед
<b>Лесна смяна на модела.</b>	В някои случаи историческите данни трябва да бъдат мигрирани.	В някои случаи е необходимо преработването на някои таблици.	Единствената промяна е добавянето на съответните сателити.
<b>Одит и възможност за проследяване.</b>	Информацията се събира хронологично чрез вмъкване на нов запис при всяка промяна.	Използва концепцията за „бавно променящи се измерения“, за да се събира хронологично информацията.	Информацията се събира хронологично чрез вмъкване на нови връзки и сателити.
<b>Ефективност при изпълнение на заявки.</b>	Заявките се изпълняват много бавно поради нормализираната в 3 NF структура на данните.	Моделът е проектиран да бъде много ефективен при интензивни запитвания поради денормализирания характер на таблиците с измерения.	Директните запитвания са много бавни поради високата стандартизация на данните. За изготвянето на анализи и отчети са необходими многоизмерни информационни матрици.

Изборът на модел за структуриране и управление на данни трябва да бъде базиран на конкретните нужди на организацията, изискванията за данни и ресурси. Облачните складове за данни представляват съвкупност от бази от данни и механизми за достъп до данните чрез единен обект [55]. За изграждането на склад за данни могат да се използват няколко компютъра, между които данните да бъдат разпределени.

На фиг.1.8 е показана класическа архитектура на склад за данни [57][58]. В нея крайните потребители имат директен достъп до данните в склада за данни, които са получени от няколко източника.



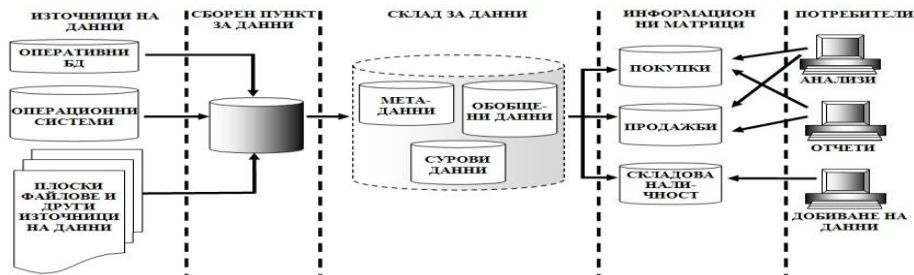
Фиг. 1.8 Класическа архитектура на склад от данни



Фиг. 1.9 Архитектура на склад за данни със сборен пункт (StagingArea)

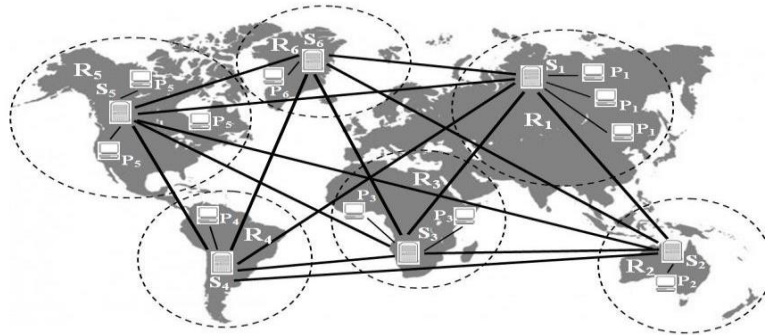
Въпреки че архитектурата на фиг.1.9 оптимизира работата на склада за данни при интеграцията на данните, това може да се окаже недостатъчно в случай, че голям брой потребители работят едновременно с тези данни.

Този недостатък е отстранен чрез разширяване на архитектурата [57], представена на фиг.1.9, като за целта са добавени информационни матрици (Data Marts) (фиг.1.10).

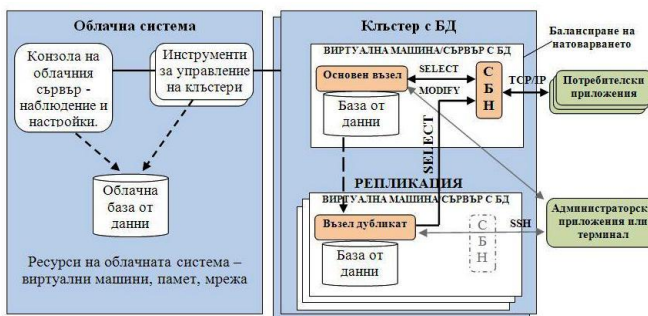


Фиг. 1.10 Архитектура на склад от данни със сборен пункт (Staging area) и модули данни (Data Marts)

На фиг.1.11 е представена архитектура на метод за структуриране, съхраняване и обработка на данни в облак от произволен потребител в Web [18]. Архитектурата поддържа неограничен брой реляционни бази от данни, разположени на сървъри (Si), които се намират на възлови места – региони (Ri) в глобалната мрежа.



Фиг. 1.11 Архитектура на метод за структуриране, съхраняване и обработка на данните в Web



Фиг. 1.12 Архитектура на облачна БД

Базите от данни, представени на фиг.1.12, са основна единица при формирането на склада за данни, който представлява реляционна база от данни, предназначена повече за изпълнение на заявки и анализи, отколкото за обработка на трансакции.

Моделите за структуриране, съхраняване и обработка на данни могат да бъдат създадени чрез използване на някои от методите: „Метод за моделиране на корпоративни данни чрез нормализация“, „Многоизмерно моделиране“, „Трезорно моделиране на данни“ и т.н. Също така, модели могат да бъдат получени чрез комбинация на някои от тези методи посредством „Хибридно моделиране“ или чрез промяна на параметрите на алгоритмите, с които се генерират моделите.

Оценката на моделите за структуриране, съхраняване и обработка на данни може да се извърши посредством SWOT анализ. Неговата задача е да обобщи силните и слабите страни, възможностите и заплахите, свързани с използването на всеки от избраните модели,

като се обърне внимание на няколко аспекта, които трябва да се имат предвид при избора на правилния модел на данни. Тези аспекти варират в зависимост от етапа на жизнения цикъл на данните, за които се проектира даден модел.

Таблица 1.4 обобщава важността на различните фактори на всеки етап от жизнения цикъл на данните.

Фактор	Създаване	Складиране	Анализ	Архив
<b>Бързина и честота за създаване и модифициране на данни</b>	Необходими са високи скорости на запис, за да се осигури по-бързо изпълнение на транзакциите. Данни, генерирани от крайни потребители или автоматизирани системи.	Необходими са умерено високи скорости на запис. Големи обеми от данни, които трябва да се съхраняват последователно.	Необходими са умерено високи скорости на запис. Може да е необходима агрегация на данните за ефективно отчитане.	Поддръжка на пониски скорости. Успешното изпълнение на процеса по архивиране е по-сигурно при пониски скорости.
<b>Скорост на извличане на данни</b>	Данните могат да бъдат извлечени веднага щом бъдат записани. Грануларността на извлечените данни може да бъде същата като вмъкнатите данни, с минимални трансформации.	Периодично извличане на данни за генериране на по-малки набори от данни. Може да изисква трансформация и агрегация на данни.	Извличането на данни е необходимо за целите на отчитането или управлението. За удовлетворяване на крайния потребител, извличането на данни трябва да бъде по-бързо, а повтарящите се трансформации и агрегации трябва да са част от съхранението и модела на данните.	Извличането на данни е необходимо само в краен случай (одити, възстановяване след бедствие и др.) Съгласуваността с действащите набори от данни е по-важна от скоростта на извличане.
<b>Свойства на атомарност, последователност, изолация, и устойчивост (ACID)</b>	Създаването на данните е част от транзакции, включващи множество стъпки. Изпълнението на изискванията за ACID е от решаващо значение за осигуряване на последователност на транзакциите.	Изпълнението на всички изисквания на ACID не е критично, но се очаква спазването на последователност на набора от данни преди и след изпълнение на операции с партида.	Изпълнението на всички изисквания на ACID не е критично, но се очаква спазването на последователността на набора от данни преди и след изпълнение на операции с партида.	Изпълнението на всички изисквания на ACID не е критично, но се очаква спазването на последователността на набора от данни преди и след изпълнение на операции с партида.
<b>Бизнес обхват</b>	Насочен към конкретна бизнес дейност.	Насочен към множество бизнес функции или цялото предприятие	Насочен към специфични изисквания за отчитане, изкуствен интелект (AI) или машинно обучение (ML)	Обхватът зависи от етапа на жизнения цикъл на данните (Създаване, Складиране, Анализ)

Фактор	Създаване	Складиране	Анализ	Архив
<b>Достъп до най-ниско ниво на данни</b>	В повечето случаи, ако една система генерира данни с най-ниската степен на детайлност, тогава се изисква и достъп до същото ниво на детайлност.	Може да се наложи данните за минали периоди да се съхраняват с различна степен на детайлност. Достъпът до най-ниското ниво на данни е важен	Може да се наложи данните за минали периоди да се съхраняват с различна степен на детайлност. Достъпът до най-ниското ниво на данни не е важен.	Може да се наложи данните за минали периоди да се съхраняват с различна степен на детайлност. Достъпът до най-ниското ниво на данни не е важен.

Най-често използваните методи за оценка на модели за структуриране, съхраняване и обработка на данни включват SWOT анализ. Целта на този анализ е да представи силните и слабите страни, възможностите и заплахите, свързани с използването на тези модели.

### SWOT анализ на модела на Инмон

Силни страни (Strengths)	Слаби страни (Weaknesses)
<ul style="list-style-type: none"> <li>• Централизирано хранилище за данни, което позволява изграждане на информационни матрици за потребителите и запазване на интегритета на данните;</li> <li>• Намаляване на аномалиите при актуализиране на данни чрез елиминиране на излишъците, което улеснява ETL процесите и ги прави по-малко податливи на грешки;</li> <li>• Лесно разбиране на бизнес процесите, тъй като логическият модел детайлно представя бизнес субектите;</li> <li>• Гъвкавост при промяна на бизнес изискванията или източниците на данни, тъй като информацията е централизирана на едно място.</li> </ul>	<ul style="list-style-type: none"> <li>• С нарастване на броя на таблиците, моделът и неговата реализация могат да станат сложни;</li> <li>• Необходими са експерти в моделирането на данни и бизнес процеси;</li> <li>• Изграждането и първоначалната настройка изискват значително време и ресурси;</li> <li>• Информационните матрици са изградени и се зареждат с данни от склада за данни;</li> <li>• Управлението и поддръжката на системата изискват голям екип от специалисти.</li> </ul>
Възможности (Opportunities)	Заплахи (Threats)
<ul style="list-style-type: none"> <li>• Ориентиран е за корпоративни нужди;</li> <li>• Организация на три нива: склад за данни, информационни матрици, информационни кубове и дублиране на данни;</li> <li>• Може да се използва за изготвяне на разнообразни отчети за нуждите на предприятието.</li> </ul>	<ul style="list-style-type: none"> <li>• С нарастването на броя на таблиците, работата по изготвяне на селекции става по-трудна;</li> <li>• Трудности при намирането на експерти в моделирането на данни и бизнес процеси;</li> <li>• Необходима е по-продължителна работа на ETL процеси, което ще доведе до необходимостта от използването на по-вече ресурси.</li> </ul>

### SWOT анализ на модела на Кимбъл

Силни страни (Strengths)	Слаби страни (Weaknesses)
<ul style="list-style-type: none"> <li>• Бързо изграждане и стартиране в първата фаза на проекта;</li> <li>• Схемата тип „звезда“ е лесно разбираема за бизнес потребителите и е подходяща за изготвяне на отчети. Много бизнес анализаторски инструменти работят добре с този тип схема;</li> <li>• Икономично използване на пространство за съхранение на данни, което улеснява управлението на системата;</li> </ul>	<ul style="list-style-type: none"> <li>• Данните се съхраняват в информационни матрици, което може да доведе до липса на интеграция;</li> <li>• Наличие на излишни данни;</li> <li>• Добавянето на колони в таблицата с факти може да доведе до проблеми с производителността. При добавяне на нови колони, размерът на факт-таблицата става по-голям, което намалява производителността. Това прави трудно модифицирането на</li> </ul>

Силни страни (Strengths)	Слаби страни (Weaknesses)
<ul style="list-style-type: none"> <li>• Модела на схема звезда е много ефективен, тъй като Системата за управление на бази от данни извършва тъй нареченото „звездно присъединяване (star join)“, при което се създава декартов продукт, използвайки всички стойности на измеренията, а факт-таблицата се поставя накрая на селектираните редове. Известно е, че това е много ефективна операция за работа с бази от данни;</li> <li>• Поддръжката на системата изисква малък екип от специалисти.</li> </ul>	<p>многоизмерния модел, тъй като се променят и бизнес изискванията;</p> <ul style="list-style-type: none"> <li>• Моделът е ориентиран към конкретни бизнес процеси, а не към предприятието като цяло;</li> <li>• Висока сложност при интегрирането на наследени данни в склада за данни.</li> </ul>

Възможности (Opportunities)	Заплахи (Threats)
<ul style="list-style-type: none"> <li>• Моделът е подходящ за отделни бизнес процеси с активно участие на потребителите;</li> <li>• Оптимизиране на процесите по извличане и управление на данни чрез организация на две нива: информационни матрици и информационни кубове;</li> <li>• Работи добре при проследяване на ключовите показатели за ефективност, тъй като информационните матрици са насочени към отчитане, съобразено с отделите или бизнес-процесите;</li> <li>• Висока ефективност при използване на инструменти за бизнес анализ, при които се налага сондиране сред множество схеми тип звезда за генериране на отчети. Това може да бъде успешно осъществено с помощта на съответстващи измерения.</li> </ul>	<ul style="list-style-type: none"> <li>• Нарушава се същността на „единен източник на информация“;</li> <li>• Поява на аномалии при актуализиране на данните с течение на времето;</li> <li>• Трудности при реализиране на специфични отчети на предприятието.</li> </ul>

**Извод:** SWOT анализът показва, че моделът на Кимбъл, базиран на подхода „отдолу нагоре“, е по-маштабируем и позволява започване с малки проекти, които впоследствие могат да се разраснат. Това осигурява по-бърза възвръщаемост на инвестициите. От друга страна, моделът на Инмон е по-структуриран и лесен за поддръжка, но изисква повече време за изграждане. Значителното предимство на модела на Инмон е, че хранилището за данни е в нормализирана форма (3NF), което значително улеснява създаването и използването на модели за извличане на данни.

#### SWOT анализ на класическа архитектура на склад за данни

Силни страни (Strengths)	Слаби страни (Weaknesses)
<ul style="list-style-type: none"> <li>• Осигурява общ модел на данните, независимо от разнообразните източници, от които те са получени.</li> </ul>	<ul style="list-style-type: none"> <li>• Оперативните данни в склада са разнородни, като някои от тях могат да бъдат излишни.</li> </ul>
Възможности (Opportunities)	Заплахи (Threats)
<ul style="list-style-type: none"> <li>• Улеснява обработката на данните и развитието на склада за данни.</li> </ul>	<ul style="list-style-type: none"> <li>• Поради липсата на механизми за премахване на излишните данни техният обем може да нарасне значително, което да затрудни работата и поддръжката;</li> <li>• Поява на аномалии при актуализиране на данните с течение на времето;</li> <li>• Трудности при реализиране на отчети.</li> </ul>

#### SWOT анализ на архитектурата на склад от данни със сборен пункт (Staging Area)

Силни страни (Strengths)	Слаби страни (Weaknesses)
<ul style="list-style-type: none"> <li>• Осигурява общ модел на данните, независимо от разнообразните източници;</li> </ul>	<ul style="list-style-type: none"> <li>• Затруднен достъп при едновременно използване на склада от голям брой</li> </ul>



Силни страни (Strengths)	Слаби страни (Weaknesses)
<ul style="list-style-type: none"> <li>Интеграция на данните посредством системите за онлайн обработка на транзакции (OLTP);</li> <li>Извличането на данни от източниците се осъществява без да се забавят операционните системи.</li> </ul>	потребители.
Възможности (Opportunities)	Заплахи (Threats)
<ul style="list-style-type: none"> <li>Предоставя възможност за улеснена обработка на данните и развитие на склада от данни;</li> <li>Възможност за установяване и отстраняване на несъответствия във формата на данните.</li> </ul>	<ul style="list-style-type: none"> <li>При едновременна работа със склада от данни на голям брой потребители времето за отговор на запитванията, генерирани от потребителите може значително да се увеличи или да доведе до невъзможност за работа.</li> </ul>

**SWOT анализ на архитектурата на склад от данни със сборен пункт (Staging Aarea) и модули данни (Data Marts)**

Силни страни (Strengths)	Слаби страни (Weaknesses)
<ul style="list-style-type: none"> <li>Осигурява общ модел на данните, независимо от разнородните източници, от които те са получени;</li> <li>Интеграция на данните посредством системите за онлайн обработка на транзакции (OLTP);</li> <li>Извличането на данни от източниците става без това да забавя работата на операционните системи;</li> <li>Балансирано натоварване при едновременно използване на склада от много потребители.</li> </ul>	<ul style="list-style-type: none"> <li>Поради сложността на архитектурата разходите за изграждане и поддръжка може да са високи.</li> </ul>
Възможности (Opportunities)	Заплахи (Threats)
<ul style="list-style-type: none"> <li>Предоставя възможност за улеснена обработка на данните и развитие на склада от данни;</li> <li>Възможност за установяване и отстраняване на несъответствия във формата на данните;</li> <li>Персонализиране на склада за данни за различни групи потребители;</li> <li>Улеснено изготвяне на отчети за тенденции, отчети за изключения и отчети, които показват действителната ефективност спрямо целите на организацията;</li> <li>Предоставя възможност за лесно добиване на данни (data mining) и извличане на информация от тях.</li> </ul>	<ul style="list-style-type: none"> <li>Възможно е дублиране на функции от склада от данни с тези на източниците на данни;</li> <li>Увеличаване на разходите с нарастване на функционалността.</li> </ul>

**Извод:** В зависимост от необходимата функционалност, тези три концепции намират широко приложение при структуриране, съхраняване и обработка на данни в уеб среда. Тези архитектури могат да включват неограничен брой бази от данни, които да бъдат разпределени между сървъри на стратегически места по целия свят.

#### 1.4. Съществуващи подходи за съхраняване и управление на данни

За основоположник на съвременната концепция за складове за данни се смята Бил Инмон, като за първи път въвежда термина „Склад за данни“ (Data Warehouse) през 1990 г.

След като Бил Инмон дефинира и предлага методология за изграждане на складове за данни, тази методика става основна практика в продължение на почти шест години. През 1996 г. Ралф Кимбъл, учен в областта на базите данни, разработва конкурентен модел на склада за данни. Кимбъл определя склада за данни като „копие на транзакционни данни, специално структурирано за заявки и анализ“ [66], като допълва, че целта на хранилището за данни е „да предостави информация в подкрепа на вземането на решения в една компания“ [67]. Според Кимбъл, складът за данни е база от данни, специално проектирана за извършване на анализ и вземане на решения.

Поради значителното нарастване на данните през 2010 г. се появява нова концепция, наречена „езеро от данни“ (data lake). Този термин е въведен от Джеймс Диксън от компанията Пентахо.

Стратегиите за изграждане на езеро от данни могат да включват както SQL, така и NoSQL подходи към бази от данни, възможности за онлайн аналитична обработка (OLAP) и онлайн обработка на транзакции (OLTP).

Таблица 1.5 обобщава ключовите разлики между склада за данни и езерото от данни.

Таблица 1.5 Разлики между склада за данни и езерото от данни

Критерий	Склад за данни	Езеро от данни
Данни	Структурирани данни	Структурирани, неструктурирани, полуструктурирани данни
Обработка на данните	Дефинирана структура преди зареждане Извличане, трансформиране и зареждане на данните в склада	Липса на дефинирана структура; трансформиране при нужда
Процес на ETL	Извличане, трансформиране и зареждане на данни в склада	Извличане и зареждане на данните в езерото; трансформиране при нужда
Разходи	Високи разходи	Ниски разходи (от гледна точка на софтуер и хардуер)
Зрялост на технологиите	Висока зрялост на технологиите	Технологии в процес на усъвършенстване
Потребители	Бизнес потребители, анализатори на данни	Изследователи на данни, анализатори на данни

### 1.5. Изводи

Въз основа на направения анализ на архитектури и методи за структуриране, съхраняване и обработка на данни в Интернет, могат да се направят следните изводи и заключения:

1. Анализирани са добре познати методи и модели за структуриране, съхраняване и обработка на данни, както и различни архитектури на облачни бази от данни.
2. Изследвани са класическа архитектура на склад за данни, архитектура със сборен пункт, както и архитектура със сборен пункт и модули данни, за да се идентифицират различията в техните подходи и приложения.
3. Подробно са описани спецификите на методи, модели и архитектури за структуриране, съхраняване и обработка на данни.
4. Извършен е SWOT анализ, чрез който са оценени силните и слабите страни на различни модели за структуриране, съхраняване и обработка на данни, подпомагайки избора на подходяща архитектура.
5. Всеки модел разполага със свои силни и слаби страни, което налага необходимостта от внимателен избор на подходящ метод за структуриране на данните.

6. Развитието на уеб-базираните информационни системи изисква интегриране на интелигентни елементи за обработка на данни в глобалната мрежа, за да отговори на нуждите от мащабируемост и адаптивност при големи обеми данни.
7. Приложението на хибриден подход при структурирането, съхраняването и обработката на данни минимизира въздействието на слабите страни на отделните методи.
8. Изборът на подходящ подход често се обуславя от финансови и технически фактори, като финансовите са насочени към оптимизация на разходите, а техническите фактори включват:
  - Лесна и бърза разработка;
  - Възможност за мащабиране;
  - Лесна поддръжка и управление;
  - Бързодействие при изпълнение на клиентските запитвания.
9. Широкото приложение на моделите на Инмон и Кимбъл при изграждането на решения за структуриране, съхраняване и обработка на данни води до непрекъснатото им усъвършенстване, което ги прави актуални и подходящи за задачите, обект на настоящата дисертация.
10. Посредством усъвършенстване на подходите за прилагане на моделите на Инмон и Кимбъл се постига необходимата дългосрочна перспектива за успешно решаване на задачите и реализиране на целите на настоящата дисертация.

#### **1.6. Цел и задачи на дисертационния труд**

**Целта на настоящия дисертационен труд** е моделиране на операции по структуриране, съхраняване и обработка на данни в Интернет.

В съответствие с поставената цел се решават следните основни задачи:

1. Модифициране и изследване на методи за структуриране, съхраняване и обработка на данни.
2. Разработване и изследване на алгоритми за структуриране, съхраняване и обработка на данни.
3. Разработване на модели на операции по структуриране, съхраняване и обработка на данни в Интернет.
4. Изследване и сравнение на предложените модели.
5. Определяне на факторите, характеризиращи информационните системи като системи за структуриране, съхраняване и обработка на данни в Интернет.
6. Приложение на разработените модели при проектиране и разработване на системи за структуриране, съхраняване и обработка на данни в Интернет.

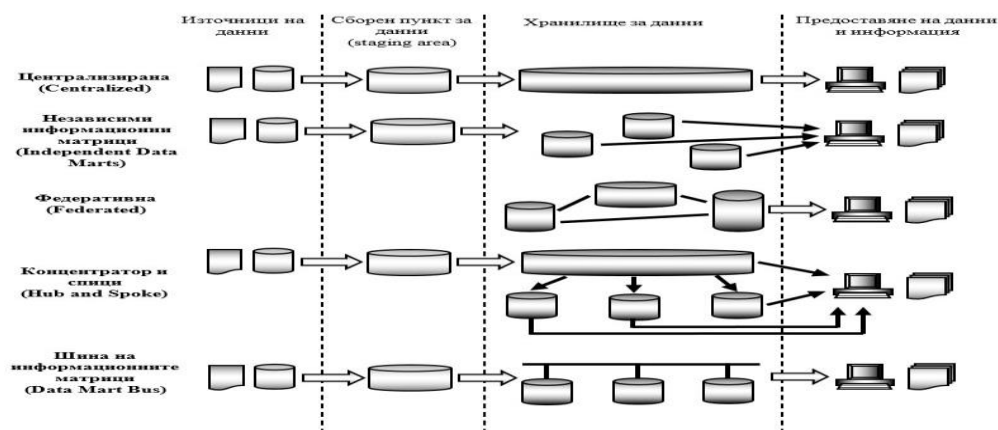
#### **ГЛАВА II: МЕТОДИКА ЗА РЕАЛИЗИРАНЕ НА ЗАДАЧАТА. СЪЗДАВАНЕ И ПРЕДСТАВЯНЕ НА МОДЕЛИ НА ОПЕРАЦИИ ПО СТРУКТУРИРАНЕ, СЪХРАНЯВАНЕ И ОБРАБОТКА НА ДАННИ**

Втора глава от дисертационния труд се състои от две части. Първата част представя методика за подбор и оценка на модели за структуриране, съхраняване и обработка на данни, която включва: избор на подход за структуриране, съхраняване и обработка на данни, определяне на подходяща концепция за изпълнение на поставените задачи и описание на конкретните дейности, които ще бъдат извършени в рамките на изследването. Във втората част са разгледани дейностите, свързани с подготовката и изследването на модели за структуриране, съхраняване и обработка на данни в Интернет.

## 2.1. Методика за подбор и оценка на модели за структуриране, съхраняване и обработка на данни

Научната задача, разглеждана в този дисертационен труд, е свързана с моделиране на операции по структуриране, съхранение и обработка на данни в Интернет. За целта е избран хибриден подход за моделиране, който използва основни методи за структуриране, съхранение и обработка на данни, подробно описани в първа глава. Този подход предоставя възможност за комбиниране на многоизмерни архитектури и архитектури в 3NF, като по този начин обхваща оперативното хранилище за данни (ODS), съдържащо нормализирани данни, както и многоизмерни информационни матрици (ИМ) (data marts). Изборът на методите за моделиране на данни, подпомагащи научните изследвания, е извършен чрез последователен анализ на основните критерии за оценка и сравнение: разбираемост, интерпретация и използваемост, наличие на дублирани данни, бързодействие при търсене и запитвания, приложимост в системи за онлайн обработка на данни, гъвкавост и сложност.

Общопризнатите архитектурни стилове [73][74][76] включват: Централизиран, Независими информационни матрици, Федеративен, Концентратор-спици и Шина на информационните матрици, както е показано на фиг.2.2.



Фиг. 2.2 Общоприети базови архитектурни стилове на складове за данни

Описанието на проведените изследвания и получените научни резултати е представено за всеки от разглежданите модели: модел на Инмон, модел на Кимбъл и Трезорен модел. Основните дейности, извършени през всеки етап на работа, са базирани на комбинация от аналитичния и експерименталния подход.

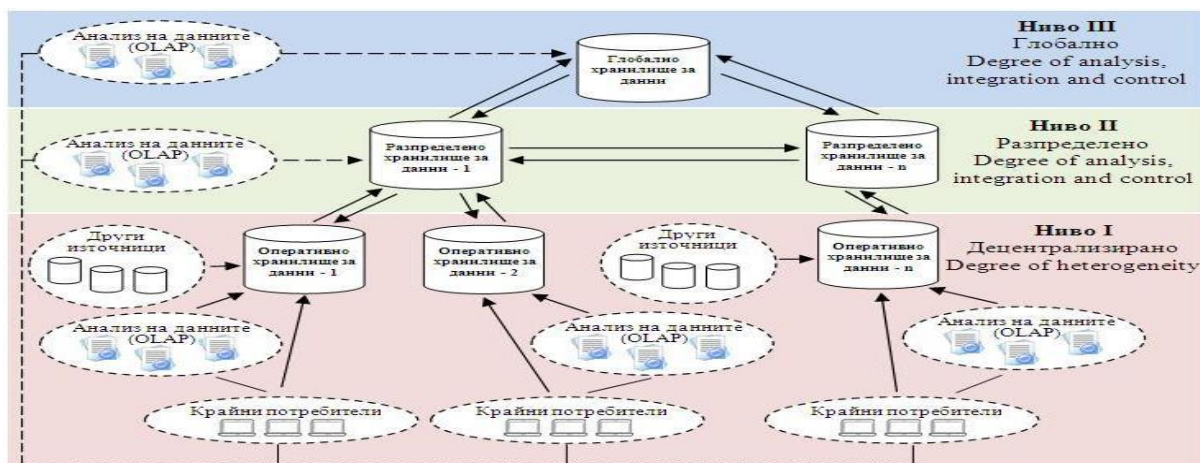
На базата на комбинация от аналитичния и експерименталния подход за моделиране са избрани необходимите стъпки в методиката на изследванията в дисертацията:

- Проучване на литературни източници, описващи различни подходи за структуриране, съхраняване и обработка на данни в областта на облачните технологии;
- Подбор на модели и данни: Изучаване на тяхната същност, произход, начин на събиране, организиране и съхранение;
- Изучаване на наличните методи и модели и предварителна подготовка;
- Визуално изследване на моделите с помощта на софтуерните средства Microsoft SQL Server, Query Analyzer и PLATINUM ERwin/ERX;
- Изследване на индивидуалното влияние на всеки модел върху жизнения цикъл на данните;
- Разработване на модели чрез прилагане на избрани методи и алгоритми за структуриране, съхраняване и обработка на данни;
- Проверка и оценка на разработените модели за структуриране, съхраняване и обработка на данни;
- Формулиране на изводи и препоръки към крайните потребители на системата за управление на данни и информация.

## 2.2. Създаване и представяне на модели за структуриране, съхраняване и обработка на данни

Научните изследвания, представени в настоящия дисертационен труд, са извършени въз основа на резултатите от Глава I, която разглежда различни методи и модели за структуриране, съхраняване и обработка на данни в Интернет.

Хибридният модел за структуриране, съхраняване и обработка на разпределени данни (фиг.2.3) включва три нива, които групират контейнери (хранилища за данни, източници и потребители) и процеси, обхващащи типични функционални групи: извличане, трансформация и зареждане (ETL), съхранение на данни, интеграция и доставка на данни. Включените компоненти представят комплексна архитектура на склад за данни, която осигурява безопасност на комуникациите, както и надзорни и аналитични функции.

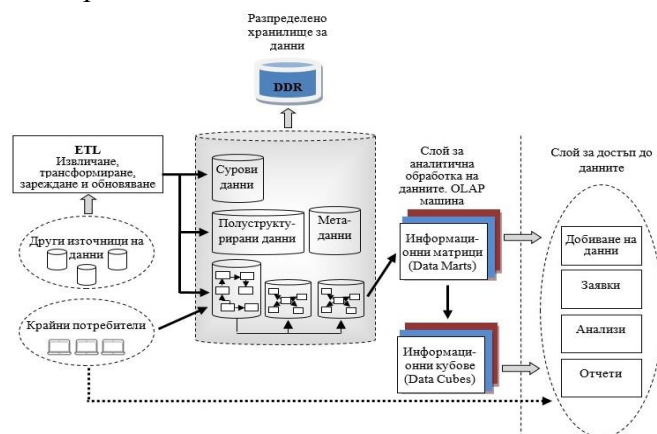


Фиг. 2.3 Хибриден модел за структуриране, съхраняване и обработка на данни

Представеният модел осигурява възможност за изграждане и внедряване на информационни системи според текущите нужди и финансови разходи. Първоначално могат да бъдат интегрирани съществуващи структурни единици за съхранение на данни, а впоследствие могат да бъдат добавяни нови, позволявайки разширяване на системата.

Архитектурата на данните и архитектурата на потока от данни са основополагащи за структурирането, съхраняването и обработката на данни. Моделирането на данните е ключов процес при създаването на архитектурата на данните, чиято цел е да определи начина на подреждане на данните във всяко хранилище, така че те да отразяват адекватно бизнес процесите. От друга страна, архитектурата на потока от данни описва как са организирани различните хранилища за данни в системите за структуриране, съхраняване и обработка на данни и как данните преминават през тези хранилища до потребителите.

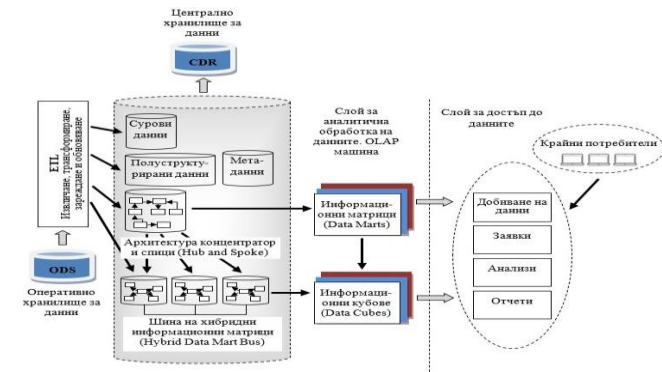
**Ниво I – Децентрализирано (локално) структуриране, съхраняване и обработка на оперативни данни**



Нивото включва децентрализирани оперативни хранилища за данни [83][84]. Оперативното хранилище на данни (виж фиг.2.4) представлява централизирана база от данни, чиято роля е да предостави обобщен изглед на най-новите данни от множество транзакционни системи за оперативно отчитане. В това хранилище данните от различни източници се комбинират и съхраняват, за да бъдат достъпни за генериране на разнообразни отчети.

Фиг. 2.4 Модел на оперативно хранилище за данни

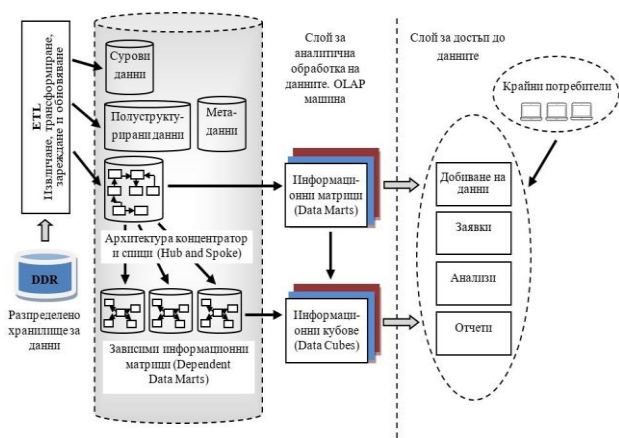
**Ниво II – Разпределено съхраняване и обработка на данни, и управление на достъпа до данни от ниво I**



Фиг. 2.5 Модел на разпределен център за данни

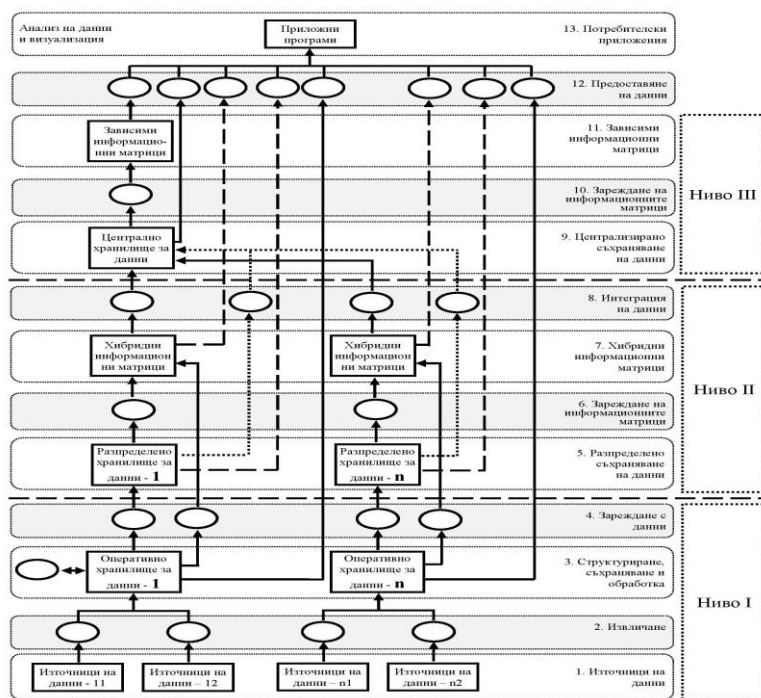
Разпределеният център за данни (фиг.2.5) представлява един възел в рамка, която използва множество разпределени възли от същия тип. Тези възли могат да бъдат контролирани с цел подобряване на производителността, повишаване на сигурността и оптимизиране на мрежовия трафик в глобалната мрежа.

**Ниво III – Централизирано съхранение на данни и управление на достъпа до данни от ниво I и II**



Фиг. 2.6 Модел на централно хранилище за данни

Това ниво представлява централно хранилище за данни, което осигурява централизирано структуриране, съхраняване на обобщени данни и управление на достъпа до данните от ниво I и II. В предложения модел, централното хранилище за данни (фиг.2.6) се състои от централизирана нормализирана база от данни, която съхранява атомарните стойности на данните и зависими информационни матрици [85]. Подобно на ниво II, тук данните също се съхраняват в нормализиран вид, обикновено в трета нормална форма.



Фиг. 2.7 Диаграма на потока от данни в хибридният модел

Диаграмата на фиг.2.7 представя архитектура за структуриране, съхранение и обработка на данни, която цели да обедини доминиращите подходи, използвани за описание на подобни архитектури. Основните елементи в диаграмата включват: процеси (абстракции на поведението) и контейнери, представляващи компоненти, способни да съхраняват, доставят или консумират данни. Тук могат да се съдържат информационни матрици, хранилища за данни, източници и други.

### 2.3. Изводи

Въз основа на анализа на предложените архитектури и методи за структуриране, съхраняване и обработка на данни са формулирани следните изводи:

1. Традиционните информационни системи, основаващи се на класически реляционни бази от данни, проявяват редица ограничения, като слаба адаптивност при динамични промени и трудност при обработка на неточно формулирани или неструктурирани запитвания. Това мотивира развитието на нови интелигентни модели, способни на по-гъвкаво и адаптивно структуриране и обработка на данни.
2. Изграждането на интелигентни интернет базирани системи е приоритетна задача за много изследователски екипи, тъй като те осигуряват по-добри възможности за динамично структуриране на данните и гъвкаво адаптиране към нови потребителски нужди.
3. Текущите решения за обработка чрез надстройка на СУБД са ограничени и се нуждаят от иновативни модели, които позволяват по-автоматизирано и адаптивно управление на данните в интернет.
4. Разработен е хибриден модел, който комбинира различни архитектури на хранилища за данни. Този модел отговаря на различни технически и бизнес изисквания, осигурявайки по-голяма гъвкавост.
5. Моделът осигурява интеграция на данни от различни източници и включва три основни нива:
  - **Децентрализирано ниво:** за локално съхранение и обработка;
  - **Разпределено ниво:** за споделяне и интегриране на данни;
  - **Глобално ниво:** за централизирано управление и обобщение на данни.
6. Архитектурата на модела осигурява плавен поток и интеграция на всички данни, постъпващи в склада. Това позволява създаването на единен източник на данни за цялата система, който включва:
  - **Централизирани таблици** (схема същност–връзки);
  - **Информационни матрици** (схема звезда) за оптимизиране на достъпа до данни.
7. Поетапна централизация на данните, позволяваща използването на предимствата на децентрализираните, разпределените и централизираните подходи за съхранение и обработка.
8. Моделът комбинира методите на Инмон и Кимбъл чрез трите типа информационни матрици: зависими, независими и хибридни. Те осигуряват компактност и гъвкавост, като разделят данните на по-малки предметно ориентирани набори за по-лесен достъп от крайните потребители.
9. Информационните матрици се създават посредством два подхода:
  - **Зависими информационни матрици:** базирани на съществуващ склад за данни, следващи подхода „отгоре надолу“ (метод на Инмон);
  - **Независими информационни матрици:** създадени от вътрешни операционни системи или външни данни, използвайки подход „отдолу нагоре“ (метод на Кимбъл).
10. Гъвкавостта и ефективността на хибридните информационни матрици е основополагаща при изграждането на разпределена облачна платформа, структурирана географски. Такава система осигурява бърз и ефективен достъп и анализ на данни, като позволява изграждането на преходни или дългосрочни клъстери от данни според аналитичните нужди.

## **ГЛАВА III: МЕТОДИКА ЗА ПРОВЕЖДАНЕ НА ИЗСЛЕДВАНЕТО. ИЗСЛЕДВАНЕ НА ОСОБЕНОСТИТЕ И ВЪЗМОЖНОСТИТЕ НА „ХИБРИДЕН МОДЕЛ ЗА СТРУКТУРИРАНЕ, СЪХРАНЯВАНЕ И ОБРАБОТКА НА ДАННИ В ИНТЕРНЕТ“**

Третата глава от дисертационния труд включва две части. В първата част е представена методиката за реализиране на изследването, която обхваща описание на основните теоретични схващания за измерване и оценка на ефективността, подходи и методи за анализ на външните и вътрешните фактори, избор на подходяща концепция за осъществяване на поставените задачи, както и описание на конкретните дейности, които ще бъдат изпълнени в рамките на изследването. Във втората част са описани извършените дейности, свързани с подготовката и изследването на модели за структуриране, съхраняване и обработка на данни в Интернет.

### **3.1. Описание на методиката на изследването**

В нашия случай оценката на ефективността се извършва чрез два набора инструменти. Първият набор включва проучвания и тестове на системи за управление на бази данни (СУБД), базирани на SQL, които представят различни подходи и метрики за оценка на производителността. Общите критерии за оценка включват скорост на изпълнение на заявки, скалируемост, надеждност и разходи. Вторият набор обхваща извършването на SWOT-анализ, който представя силните и слабите страни на модела, както и възможностите за намаляване на влиянието на слабите страни при изграждането на хибридни системи за структуриране, съхранение и обработка на данни. Резултатите от оценката на основните критерии за оценка на производителността и SWOT-анализа са представени в **точка 3.3**.

### **3.2. Подготовка за провеждане на сравнителен тест и извършване на SWOT-анализ**

За оценка на производителността на SQL Server при типични операции като четене, писане, актуализиране и изтриване се използват различни инструменти за сравнителен анализ като YCSB [90] и Sysbench [91], често използвани за оценка на производителността на базата от данни при работни натоварвания, които имитират приложение в реалния свят.

За целите на SWOT-анализа се генерират максимален брой характеристики на силните и слабите страни, възможностите и заплахите в четирите категории на анализа. Общо са идентифицирани 53 характеристики, които се въвеждат в SWOT матрица, която ги групира по съответните категории (Strengths, Weaknesses, Opportunities, Threats), като те са разпределени по следния начин:

- **Силни страни (Strengths)** – 16 характеристики;
- **Слаби страни (Weaknesses)** – 11 характеристики;
- **Възможности (Opportunities)** – 19 характеристики;
- **Заплахи (Threats)** – 7 характеристики.

Разпределението на характеристиките по категории е представено в таблица 3.5, като основните акценти са:

- Около 30% от генерираните характеристики представляват силните страни на модела;
- Около 21% от характеристиките са свързани със слабите страни на модела;
- Около 36% от характеристиките отразяват възможностите, които моделът предоставя;
- Около 13% от характеристиките се отнасят до заплахите, свързани с внедряването, сигурността и управлението на информационните системи, изградени на базата на този модел.

В заключение, може да се отбележи, че повече от половината (66%) от общо генерираните характеристики са в полза на предимствата, които моделът предоставя, докато останалите 34% описват неговите недостатъци. Това съотношение ни дава основание да продължим изследването си, използвайки метода на SWOT-анализ.



### 3.3. Резултати от проведен сравнителен тест и извършен SWOT-анализ

В нашия случай използвахме инструмента за бенчмаркинг SQLIO, разработен от Microsoft. Съществуват и алтернативи, като IOMeter – софтуер с отворен код, който има версии за Linux, Solaris и дори NetWare. Въпреки това, понеже SQL Server работи на Windows, SQLIO е по-подходящият избор в нашия случай.

След редактирането на конфигурационния файл, стартираме SQLIO веднъж, за да се създаде **testfile.dat**. Стартирането се извършва чрез командния интерпретатор със следната команда:

**sqlio -kW -s5 -fsequential -o4 -b64 -Fparam.txt**

#### *Резултати от теста:*

За да тестваме устройство D, създаваме серия от тестове в команден файл (test.cmd) и ги оставяме да се изпълнят. Това, което се търси, е максималният брой I/O операции, които устройство D може да обработи.

#### *Тест за произволни записи*

Стойностите за **входно/изходни операции в секунда (IOPS)** и **мегабайти в секунда (MB/sec)** са систематизирани в таблица 3.1.

Таблица 3.1 Резултати от теста за произволни записи

Изчакващи операции за вход/изход	Входно/изходни операции за секунда	Мегабайти за секунда
1	100.05	6.25
2	98.85	6.17
4	99.92	6.24

#### *Тест за произволно четене*

Резултатите от теста за входно/изходни операции в секунда (IOPS) и мегабайти в секунда (MB/sec) са обобщени в таблица 3.2.

Таблица 3.2 Резултати от теста за произволни четения

Изчакващи операции за вход/изход	Входно/изходни операции за секунда	Мегабайти за секунда
1	92.24	5.76
2	197.54	12.34
4	217.85	13.61
8	262.88	16.43
16	313.39	19.58
32	353.11	22.06
64	393.25	24.57
128	384.83	24.05
256	390.36	24.39
512	385.95	24.12
1024	395.28	24.70

#### *Тест за последователно четене*

Резултатите от теста за входно/изходни операции в секунда (IOPS) и мегабайти в секунда (MB/sec) са обобщени в таблица 3.3.

Таблица 3.3 Резултати от теста за последователно четене

Изчакващи операции за вход/изход	Входно/изходни операции за секунда	Мегабайти за секунда
1	2227.39	139.21
2	2233.41	139.58

Изчакващи операции за вход/изход	Входно/изходни операции за секунда	Мегабайти за секунда
4	2190.09	136.88
8	2217.39	138.58
16	2236.75	139.79
32	2262.47	141.40
64	2300.28	143.76
128	2305.87	144.11
256	2286.90	142.93
512	2288.37	143.02
1024	2295.21	143.45

### **Тест за последователно записване**

Резултатите от теста за входно/изходни операции в секунда (IOPS) и мегабайти в секунда (MB/sec) са обобщени в таблица 3.4.

Таблица 3.4 Резултати от теста за последователно четене

Изчакващи операции за вход/изход	Входно/изходни операции за секунда	Мегабайти за секунда
1	6813.03	53.22
2	6339.24	49.52
4	10323.31	80.65
8	7739.09	60.46
16	7645.11	59.72
32	8604.67	67.22
64	9164.68	71.59
128	9044.44	70.65
256	9409.32	73.51
512	9365.35	73.16
1024	9455.53	73.87

SQL Server демонстрира впечатляваща производителност при стандартни операции за работа с данни, особено когато се прилагат стратегии за оптимизация и се използват адекватни хардуерни ресурси. Разумното управление на ресурсите и прилагането на групови операции са ключови фактори за ефективността на системата, особено в контекста на приложения с висока производителност. Сравнителните анализи подчертават необходимостта от внимателно планиране и настройка, за да се гарантира оптимално функциониране на SQL Server в реални сценарии.

Въз основа на базовата матрица е създадена детайлна SWOT-матрица, която групира факторите на модела по четири аспекта. Таблица 3.5 представя обобщаваща SWOT-матрица на хибридният модел за структуриране, съхранение и обработка на разпределени данни в Интернет.

Таблица 3.5 Обобщаваща SWOT матрицата на хибридният модел

Силни страни (Strengths)	Слаби страни (Weaknesses)
<b>Ниво I – Децентрализирано (локално) структуриране, съхраняване и обработка на оперативни данни.</b>	
Локално структуриране, съхраняване и обработка на оперативни данни.	Увеличаването на възлите затруднява гарантирането на еднакъв изглед на базата от данни за потребителите.
Лесно изграждане на мрежа от системи с локални бази от данни и добавяне на допълнителни възли.	Управлението на едновременността и целостта на данните не е напълно гарантирано.
Ниски разходи при първоначално внедряване.	Възможен излишък на данни поради разпределеното съхранение на различни места.

Силни страни (Strengths)	Слаби страни (Weaknesses)
Устойчивост на натоварвания, благодарение на локалната обработка на данни.	Необходимост от сложен софтуер за управление на комуникациите между възлите.
Отказоустойчивост при излизане на възли извън строя, благодарение на софтуера за управление на комуникациите.	Трудности при намирането на специалисти с необходимия опит в управлението и поддръжката на разпределени системи.
Крайните потребители могат да структурират, съхраняват и обработват данни.	
<b>Ниво II – Разпределено съхраняване и обработка на данни, и управление на достъпа до данните от ниво I.</b>	
Локално натоварване при обработка и осигуряване на достъп до данни.	Трудност при гарантирането на еднакъв изглед на базата от данни с увеличаване броя на възлите.
Планиране на процесите по репликация и дублиране в зависимост от натоварването на системата.	Трудности при гарантиране на едновременността и целостта на данните.
Възможност за централизирано съхранение и обработка на данни в определени ситуации, когато потребителите и техните данни са в един и същи регион.	Възможен е излишък на данни, тъй като те се съхраняват на различни места.
Отказоустойчивост при излизане на възли извън строя, благодарение на софтуера за управление на комуникациите., като останалата част от системата може да продължи работата си.	Трудности при осигуряване на сигурността на данните поради децентрализирания характер и сложността на инфраструктурата, свързваща всички възли.
Групиране и съхранение на данните по тематично ориентирани области (отдели в дадена организация), като по този начин се организира йерархичен достъп.	Необходимост от сложен софтуер за управление на комуникациите между възлите.
	Възможни несъответствия при репликация на данни.
	Трудности при намирането на специалисти с необходимия опит в управлението и поддръжката на разпределени системи.
<b>Ниво III – Централизирано съхранение и обработка на данни, и управление на достъпа до данните от ниво I и ниво II.</b>	
Гарантирана е целостта на данните и излишъкът е минимизиран, поради нормализирания характер на данните.	Силна зависимост от мрежовата свързаност. Колкото по-бавна е интернет връзката, толкова повече време ще е необходимо за достъп до базата от данни.
Повишена сигурност на данните, съхранявани на едно място, тъй като се съхраняват на едно място.	Риск от срив при работа с голям брой потребители едновременно.
Данните, които се съхраняват на едно и също място, са по-лесни за обновяване, реорганизация или анализ.	Минималният или никакъв излишък на данни е предпоставка за частична загуба на данни.
Данните са достъпни едновременно от множество места.	Ограничен достъп на повече от един потребител до същия набор от данни, което може да намали ефективността на системата.
Актуализираните данни са достъпни незабавно за всеки потребител.	
Ниски разходи за труд, захранване и поддръжка.	
По-лесно за използване от крайния потребител поради простотата на дизайна.	

Възможности (Opportunities)	Заплахи (Threats)
<b>Ниво I – Децентрализирано (локално) структуриране, съхраняване и обработка на оперативни данни.</b>	
Лесен достъп до оперативни данни и обработка на опростени заявки.	Усложнения при интегрирането съпоставянето и консолидирането на данни от различни системи.
Поддръжка на оперативни и тактически решения, базирани на актуални данни. Данните могат да се използват веднага, като се отчитат разликите във времето.	Трудности при гарантиране качеството на трансформация и интеграция на данни. Степента и видът на трансформация на данните зависят от честотата на доставката им. Колкото по-кратък е периодът за доставка, толкова по-ниско е качеството на трансформацията и интеграцията. Един от начините за решаване на този проблем е използването на инструмента Extract Transform Move Load (ETML), който позволява обработка на големи обеми актуализации и работа с широк спектър от структури и формати на данни.
Интегриране на данни от нови и съществуващи системи, извършвайки централизирано съхраняване на данни. Вместо ограниченото отчитане, предлагано от системите източници. Това позволява създаването на по-изчерпателни оперативни отчети.	Нарастване на разходите за управление с увеличаване на обема на данните.
Осигурява пълен изглед, улеснен мониторинг и диагностика на проблеми при работата на информационната система.	
Изпълнява функциите на междинна станция преди данните да бъдат преместени в разпределените центрове за данни, като извършване на операции за подобряване качеството им.	
Предоставя на крайните потребители възможност за структуриране, съхраняване и обработка на данни по тяхно желание.	
<b>Ниво II – Разпределено съхраняване и обработка на данни, и управление на достъпа до данните от ниво I.</b>	
Интегриране на данни от децентрализирани оперативни хранилища (ОпХД), посредством централизирано съхранение на данни, организирано на регионален принцип. Това позволява създаването на отчети и извършването на обобщен анализ за всеки регион.	При увеличаване броя на таблиците, E-R моделът и изпълнението на заявки могат да станат сложни поради включването на повече таблици и обединения.
Лесен достъп до обобщени исторически данни, предоставени от ОпХД на ниво I.	Необходими са експерти по моделиране на данни и бизнес процеси.
Предоставяне на информация за анализ и вземане на решения, базирани на исторически данни.	Данните, съхранявани в информационните матрици (ИМ), не са напълно интегрирани и може да се появи излишък от данни.
Намаляване на излишъка от данни чрез нормализация и избягване на аномалии при актуализации.	Добавянето на колони в таблицата с факти може да доведе до намаляване на производителността, тъй като размерът на факт-таблицата нараства. Това затруднява изменянето на многоизмерния модел при промяна на бизнес изискванията.
Възможност за комбиниране на данни от хранилището и други източници на данни.	
Повишаване на бързодействието чрез използване на денормализирани предметно ориентирани многоизмерни хибридни информационни матрици (ХИМ).	

Възможности (Opportunities)	Заплахи (Threats)
Възможност за крайните потребители да използват специализиран софтуер за анализ на данни.	
Осигуряване на пълен изглед и мониторинг на съхраняваните данни, което улеснява идентифицирането и диагностицирането на проблеми в работата на информационната система.	
Изпълнение на функциите на междинна станция преди данните да бъдат преместени в централното хранилище, включително и операции за подобряване на качеството им.	
<b>Ниво III – Централизирано съхранение и обработка на данни, и управление на достъпа до данните от ниво I и ниво II.</b>	
Интегриране на данни от регионално разпределените центрове за данни, което позволява изготвянето на разнообразни отчети и извършването на глобален анализ.	При нарастване на броя на таблиците, E-R моделът и изпълнението на заявки могат да станат сложни, тъй като включват повече таблици и обединения.
Лесен достъп до обобщени исторически данни, събрани от разпределените центрове за данни (РЦД), разположени на Ниво II.	Необходими са експерти в моделирането на данни и бизнес процеси.
Предоставяне на информация за анализ и вземане на решения, основани на исторически данни.	Данните, съхранявани в информационните матрици (ИМ), не са напълно интегрирани, което може да доведе до излишък от данни.
Намаляване на излишъка от данни чрез нормализация и предотвратяване на аномалии при извършване на актуализации.	Добавянето на колони в таблицата с факти може да доведе до намаляване на производителността, тъй като размерът на факт-таблицата нараства. Това затруднява изменението на многоизмерния модел при промяна на бизнес изискванията.
Хранилището за данни е единственото място за съхранение на данни, което гарантира тяхната цялост и последователност.	
Повишаване на производителността чрез използване на денормализирани предметно ориентирани многоизмерни зависими информационни матрици (ЗИМ).	
Възможност за крайните потребители да използват специализиран софтуер за анализ на данни.	
Осигуряване на пълен изглед и мониторинг на съхраняваните данни, което улеснява идентифицирането и диагностицирането на проблеми в информационната система.	
Изпълнение на функциите на глобално хранилище за данни, включително извършване на операции за подобряване на тяхното качество.	

• Силни страни (Strengths)

Хибридният модел предлага няколко значителни предимства, включително възможността за постепенно изграждане, ниските разходи при първоначално внедряване, устойчивостта на натоварвания и лесното управление на съхранението, обработката и достъпа до данни. Сигурността на данните и бързодействието също са силни страни на модела. Въпреки това, тези предимства изискват квалифицирани специалисти за поддръжка и безпроблемна работа на всички компоненти на склада за данни. Допълнително, изготвянето на разнообразни отчети е сред предимствата на хибридният модел.

• Слаби страни (Weaknesses)

Основните слабости на модела включват гарантирането на едновременността и целостта на данните в разпределеното ниво. С увеличаването на броя на възлите, става

все по-трудно да се осигури консистентен изглед на данните за потребителите, а също така може да се появи излишък от данни поради едновременното им съхранение на различни места. Друг недостатък е необходимостта от сложен софтуер за управление на комуникацията между разпределените възли и моделирането на данните, което затруднява намирането на специалисти с необходимия опит. Също така, изготвянето на отчети, свързано с използването на сложни заявки, може да се затрудни с увеличаването на броя на таблиците и промяната на бизнес правилата.

- Възможности (Opportunities)

Поетапната централизация на съхранението предоставя възможности за извличане на информация, необходима за съставяне на по-подробни отчети и вземане на оперативни и тактически решения, базирани на исторически, текущи или почти реални данни [95]. Потребителите могат да структурират, съхраняват и обработват данни по свое желание и да използват специализиран софтуер за анализ на данни. Комбинирането на данни от хранилището и други източници чрез хибридни информационни матрици предлага бърза интеграция на множество бази от данни, съчетавайки предимствата на независимите и зависимите информационни матрици [96]. Осигуряването на пълен изглед и мониторинг на съхраняваните данни улеснява идентифицирането и диагностицирането на проблеми в информационната система [97].

- Заплахи (Threats)

Интегрирането и консолидирането на данни може да се усложни при съпоставяне на данни от различни системи. Липсата на достатъчно време за извличане и зареждане може да повлияе негативно на качеството, трансформацията и интеграцията на данните. Усложняването на модела поради увеличаване на броя на таблиците може да затрудни изпълнението на заявки и изготвянето на отчети, като също така затруднява адаптирането на модела към променящи се бизнес изисквания и увеличава разходите за управление и поддръжка поради факта, че са необходими експерти в моделирането на данни и бизнес процеси.. Високата цена за управление и поддръжка е значителна заплаха, която може да бъде смекчена чрез обучение на собствени служители или наемане на специалисти за тази цел.

### 3.5. Изводи

В резултат на извършения тест на производителността на SQL Server и SWOT-анализ на „Хибриден модел за структуриране, съхраняване и обработка на разпределени данни“ могат да се направят следните изводи и заключения:

1. SQL Server е способен да обработва значителен обем заявки при оптимизирани хардуерни настройки.
2. Използването на пакетни актуализации значително подобрява времето за изпълнение и намалява ресурсното натоварване.
3. Груповите операции, като обновяване на данните и изтриване позволяват стратегическо разделяне на задачите, което води до по-добра производителност.
4. Правилната конфигурация на SQL Server е от съществено значение за постигане на висока производителност в приложения с интензивно натоварване.
5. Моделът комбинира различни архитектури на хранилища, които отговарят както на технически, така и на бизнес изисквания. Тази комбинация осигурява по-гъвкава и приспособима среда за управление на данните.
6. Предложени са усъвършенствани методи за структуриране и обработка на данни, с акцент върху ефективното споделяне и интегриране на данни от различни източници. Това помага да се адресират предизвикателствата при обединяване на разнородни информационни потоци.
7. Всяка архитектура, използвана в отделните нива на модела, има своите уникални силни и слаби страни. Затова е от съществено значение да се подбере подходящ метод за моделиране според конкретната система.

8. Благодарение на своята адаптивност, моделът позволява компромиси между различни архитектури и възможности, като същевременно дава възможност за поетапно внедряване.
9. Хибридният модел осигурява бързо внедряване, лесна поддръжка, мащабируемост и висока производителност при клиентски запитвания. Лесната разработка и гъвкавото разширяване гарантират, че системата може да отговори на променящите се изисквания и натоварвания.
10. Интегриране на съществуващи структури и добавяне на нови според нуждите и ресурсите, с което се постига по-добра управляемост и контрол върху разходите.

## **ГЛАВА IV: СОФТУЕРНО РЕШЕНИЕ ЗА ИЗВЪРШВАНЕ НА ОПЕРАЦИИ ПО СТРУКТУРИРАНЕ, СЪХРАНЯВАНЕ И ОБРАБОТКА НА ДАННИ В ИНТЕРНЕТ**

В тази глава е представено внедряването на система за управление на данни и информация, базирана на хибридният модел за структуриране, съхранение и обработка на разпределени данни в Интернет, описан в Глава II. Дефинирани са изискванията към системата, нейната област на приложение, изборът на технологии и инструменти, както и архитектурата, организацията на работния процес и получените резултати.

### **4.1. Добри практики в управлението на данните**

Управлението на данните включва действия по тяхното структуриране, съхранение и обработка, независимо дали те са използвани или генерирани в резултат на различни дейности, като изследователска работа [98], бизнес отношения или данни, генерирани от крайни устройства. Общоприетата практика изисква данните да се съхраняват поне 10 години [99][100], което осигурява проследимост, дългосрочна интерпретируемост и устойчиво архивиране. Дефинирането на основни принципи за управление на данни може значително да подобри възможността за намиране, достъпност, оперативна съвместимост и повторна употреба на информацията.

### **4.2. Структура и формат на данните**

Структурата на генерираните данни зависи до голяма степен от използвания софтуер. Данните могат да бъдат представени в бинарен формат, в общоприети формати като ASCII или XML, или в специфични формати, определени от приложенията. Необходимостта от структуриране на данните, генерирани от потребителите, и съхраняването им в определена схема, както и изискванията за съвместно използване на тези данни от други потребители, са ключови теми на това изследване.

### **4.3. Управление на данни от работни процеси**

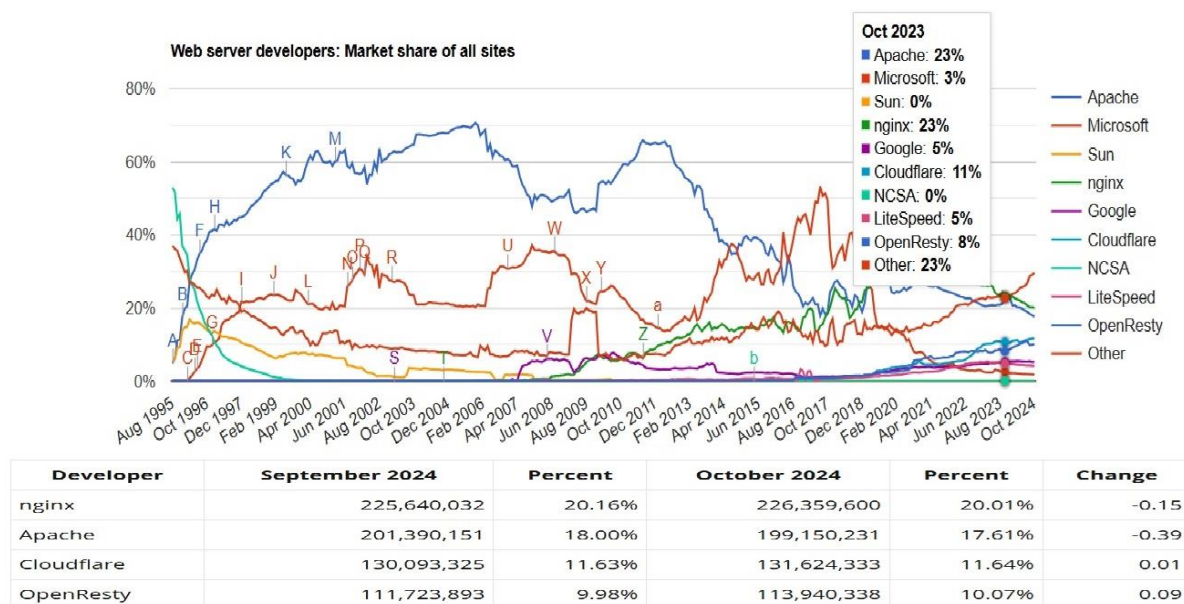
В своята дейност работните групи създават работни потоци, генерирайки и обработвайки данни, които са специфични за съответния процес. Това затруднява установяването на унифицирано управление на тези данни. Регламентирането на достъпа до съхранените данни е основна задача при изграждането на всяка локална или уеб-базирана информационна система, като по този начин се гарантират тяхната достоверност и защита от неправомерен достъп.

### **4.4. Изисквания към системата**

Предложената методология за управление на данни и информация изисква използването на подходящ инструментариум за провеждане на научни експерименти. Системата ще служи за научно-експериментална работа, предоставяйки на потребителите възможност да описват структурата за съхраняване на данни по свой избор, както и да извършват манипулации с тях (въвеждане, редактиране, споделяне и др.). Чрез тези функции системата значително ще улесни достъпа до базите от данни за широк кръг потребители.

#### 4.5. Избор на средства и технологии за реализация

На фиг.4.1 е показано разпределението на основните уеб платформи, използвани от разработчиците на уебсайтове през последните години.



Фиг. 4.1 Разпределение на пазарния дял на активните сайтове според Netcraft

От графиката на фиг.4.1 се вижда, че с времето популярността на NGINX намалява и постепенно се доближава до тази на Apache HTTP Server.

В таблица 4.1 са представени основните характеристики на двете уеб платформи. Обобщаването и систематизирането на тези характеристики осигурява структурираност на информацията, необходима за вземане на решение, и улеснява избора на подходящ уеб сървър.

Таблица 4.1 Обобщаваща таблицата с основните характеристики на Apache и NGINX.

Apache	NGINX
Основната му функция е да бъде уеб сървър.	Има две функции: уеб сървър и реверсен прокси сървър
Не може да обработва няколко заявки едновременно при натоварен уеб трафик.	Може да обработва множество заявки едновременно с ограничени ресурси.
Използва многонишков подход при обработка на заявките.	Използва подход, управляван от събития при обработка на заявките.
Поддържа динамично зареждане и разтоварване на функционалните модули, което го прави по-гъвкав.	Функционалните модули не могат да се зареждат динамично. Те трябва да се комбинират в основния софтуер.
Обработва динамичното съдържание в самия уеб сървър.	По начало не може да обработва динамично съдържание.

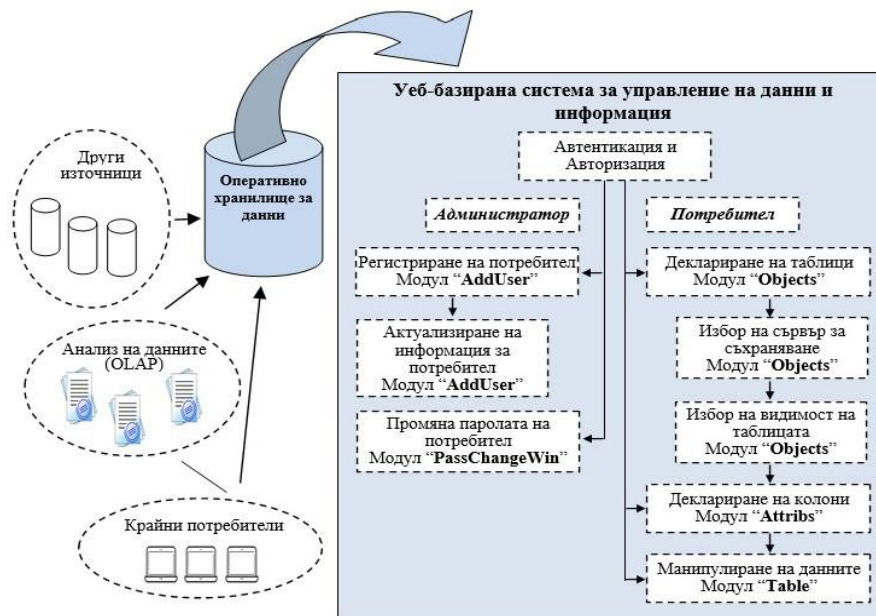
За реализирането на софтуерно решение, което включва структуриране, съхранение и обработка на данни в интернет, е избран Apache HTTP Server.

#### 4.6. Система за управление на данни и информация

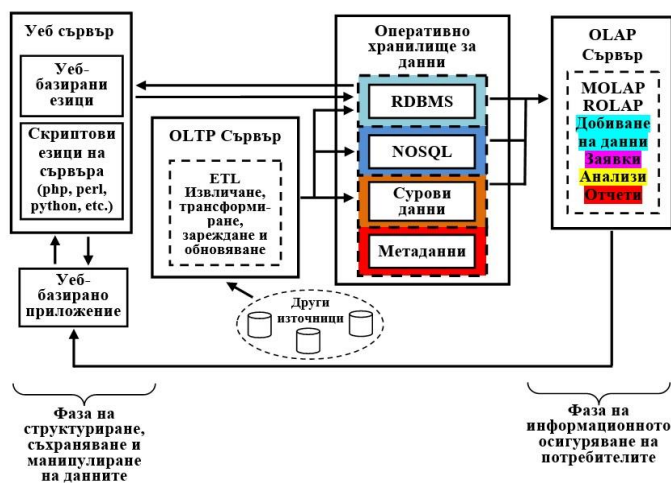
Системата се основава на методите за структуриране, съхранение и обработка на данни в Интернет, разгледани в глава първа, както и на хибриден модел за тези процеси, представен във втора глава.



На фиг.4.3 е показан обобщен изглед, който илюстрира подхода за внедряване на системата в оперативно хранилище на данни. Това хранилище е част от ниво I на хибридният модел за структуриране, съхранение и обработка на данни в интернет, като осигурява децентрализирано (локално) структуриране, съхранение и обработка на оперативни данни.

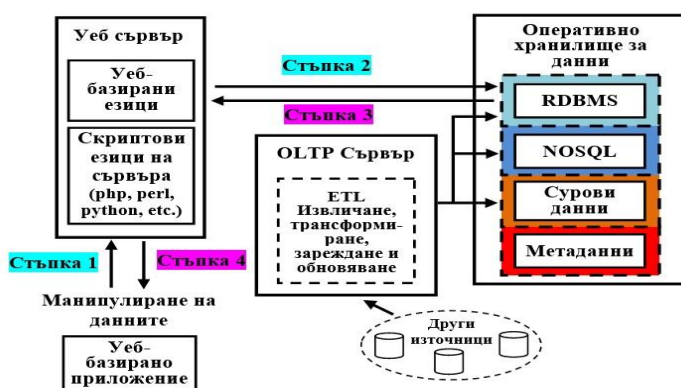


Фиг. 4.3 Обобщен изглед за внедряване на системата



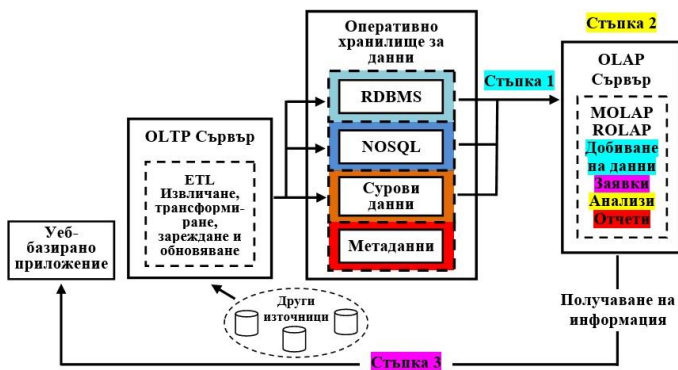
Фиг. 4.4 Блокова диаграма на архитектурата на системата

Фигура 4.4 показва хардуерните и софтуерните изисквания на компонентите, изграждащи системата, както и начина, по който те взаимодействат. Необходимите хардуерни и софтуерни компоненти за осъществяване на внедряването са избрани въз основа на три критерия: цена, наличност и леснота на използване, както и възможност за отворено програмиране.

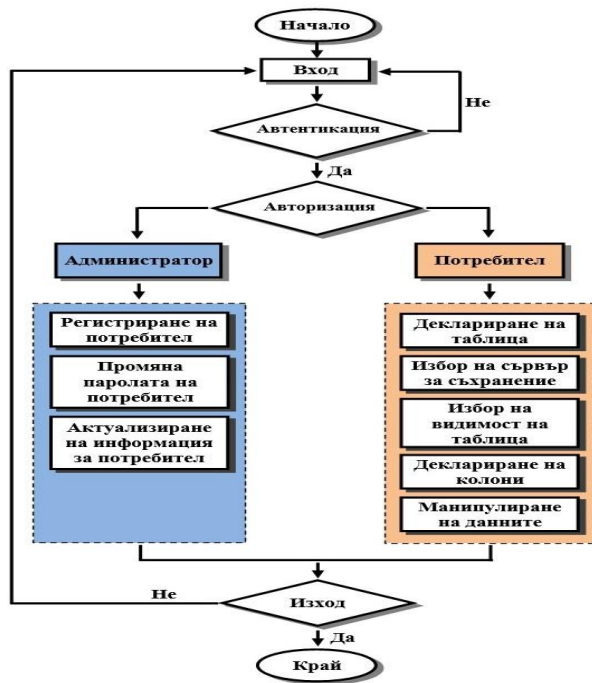


Фиг. 4.5 Процедурни стъпки при фазата на манипулиране на данните

Фигура 4.5 илюстрира процесите, които потребителите извършват при структуриране, съхраняване и манипулиране на данните в системата, както и всички стъпки и процедури, свързани с тях.

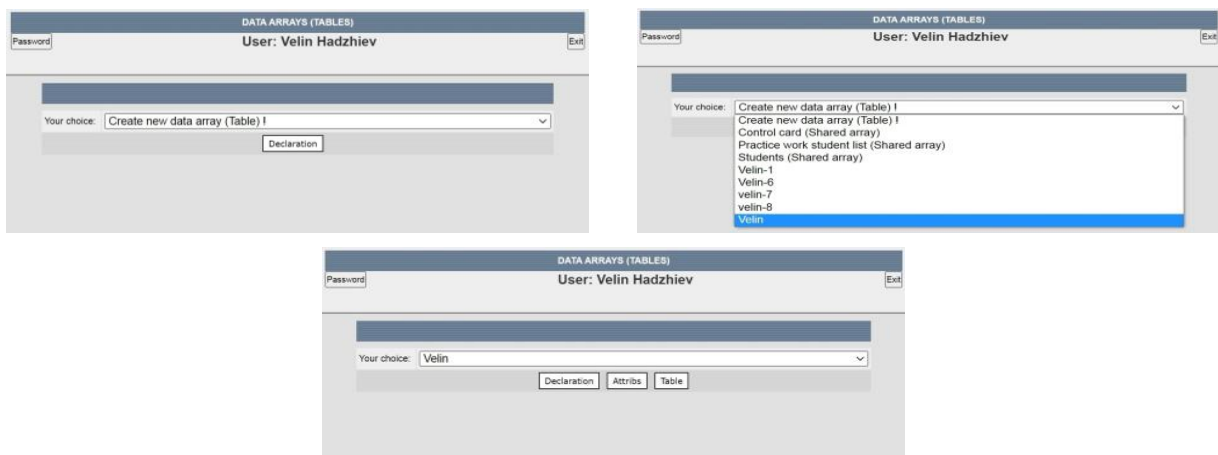


Фиг. 4.6 Процедурни стъпки при фазата на информационното осигуряване



Фиг. 4.7 Блок-схема на алгоритъма за работа на системата

Основният графичен потребителски интерфейс (фиг.4.8), който се визуализира след идентификация, съдържа поле с падащ списък, предоставящ възможност за деклариране на нова таблица или за избор на вече декларирана таблица, с която потребителят желае да работи.



Фиг. 4.8 Основен графичен интерфейс на системата

Фигура 4.6 илюстрира всички стъпки и процедури, свързани с осигуряването на информационно обслужване за потребителите.

Фигура 4.7 представя модулите на системата, разделени на две основни части – администраторска и потребителска.

- Администраторската част предоставя възможност за добавяне на потребители в базата от данни, промяна на потребителски пароли и актуализиране на информацията за потребителите;

- Потребителската част позволява извършване на следните операции: деклариране на таблица, избор на място за съхранение, определяне на достъп до таблицата, деклариране на колони и манипулиране на данните.

Чрез функционалния модул „New table“ потребителят може да декларира нова таблица или да редактира параметрите на вече декларирана таблица (фиг.4.9).

Фиг. 4.9 Параметри на таблица

Чрез функционалния модул „Table Structure“ (фиг.4.10) потребителят има възможност да добавя нови колони или да редактира параметрите на съществуващи КОЛОНИ.

Column name	Accessibility	Data type	Unique	Delete
Age	NO LIMIT	NUMBER	<input type="checkbox"/>	<input type="radio"/>
First name	NO LIMIT	STRING	<input type="checkbox"/>	<input type="radio"/>
ID number	NO LIMIT	NUMBER	<input checked="" type="checkbox"/>	<input type="radio"/>
Last name	NO LIMIT	STRING	<input type="checkbox"/>	<input type="radio"/>
	NO LIMIT	STRING		<input type="radio"/>

Фиг. 4.10 Добавяне или редактиране на колони

Чрез функционалния модул „Table Content“ (фиг.4.11) потребителят има възможност да добавя, изтрива или редактира данни в таблицата.

First name	Last name	ID number	Age	Delete
Velin	Hadzhiev	1	38	<input type="radio"/>
Ivan	Ivanov	2	45	<input type="radio"/>
Yana	Ivanova	3	23	<input type="radio"/>
Inna	Berg	4	33	<input type="radio"/>
				<input type="radio"/>

Фиг. 4.11 Манипулиране на данните

#### 4.7. Предимства и недостатъци

Целта на SWOT матрицата, показана в таблица 4.3, е да групира факторите на системата в четири категории: силни страни, слаби страни, възможности и заплахи. Обособяването на тези фактори в отделни групи предоставя подробна информация за предимствата и недостатъците на системата, което може да послужи като основа за намаляване на влиянието на слабите страни и насока за бъдещата работа по подобряване ефективността на системата.

Таблица 4.3 SWOT матрица на системата

Силни страни (Strengths)	Слаби страни (Weakness)
Интуитивен подход за структуриране, съхранение и обработка на данни от потребителите.	С увеличаването на броя на таблиците, моделът и реализацията могат да станат сложни.
Структуриране и съхраняване на неограничено количество данни в Интернет.	Силна зависимост от мрежовата свързаност: колкото по-бавна е интернет връзката, толкова по-дълго време е необходимо за достъп до базата от данни.
Съхраняване на периодично обновяваните данни в отделно пространство, различно от пространството, предназначено само за настройка на информация.	Податливост на сринове при работа с голям брой потребители едновременно.
Архивиране на данните в случай на техническа повреда.	
Възможности (Opportunities)	Заплахи (Threats)
Възможност за свързване на отделните системи с локални бази от данни в мрежа и добавяне на допълнителни възли.	Намаляване ефективността на системата с увеличаване обема от данни.
Разполагане на често използваните данни близо до потребителите с цел намаляване на уеб трафика.	
Възможност за предоставяне на достъп до данни на голям брой потребители едновременно.	
Предоставяне на данни от различни източници, включително бази от данни, директории, имейл и др.	

#### 4.8. Изводи

В заключение разработеният метод за извършване на основни операции по структуриране, съхраняване и обработка на данни, приложен в уеб базирана система за обработка и управление на данни, има следните характеристики и възможности:

1. Гъвкаво управление на таблици и атрибути, осигуряващо висока степен на персонализация и контрол над данните.
2. Създаване на неограничен брой таблици, настройка на видимостта, дефиниране и конфигуриране на атрибутите на таблиците и въвеждане на данни в таблиците в зависимост от правата на потребителите.
3. Управление както на индивидуални така и на съвместни работни процеси, независимо от областта на приложение, което я прави универсална за широк кръг от потребители.
4. Лесно адаптиране към различни научни и работни процеси, без да се налага промяна на основната ѝ структура или функционалност. Това я прави подходяща за бързо внедряване в нови приложения и домейни.

5. Системата е разработена за работа с релационни бази от данни, като за първоначалното ѝ внедряване и поддръжка са необходими специалисти, което води до по-високи разходи в началната фаза на проекта.
6. Усложняване на модела на базата от данни с нарастване броя на таблиците и намаляване ефективността на системата с увеличаване на обема на данните.
7. Представената система предоставя висока степен на персонализация, универсалност и адаптивност към различни работни процеси. За преодоляване на ограниченията, свързани с разходите и сложността при управление на големи обеми от данни, системата позволява внедряването ѝ като облачна услуга (DBaaS) или използването на NoSQL бази от данни. Тези решения могат значително да подобрят мащабируемостта и ефективността на системата при нарастващ обем от данни и разнообразие от приложения.

## **ОБЩИ ИЗВОДИ/ЗАКЛЮЧЕНИЕ**

В резултат на извършения анализ на методи и модели за структуриране, съхраняване и обработка на данни в Интернет могат да се направят следните изводи и заключения:

1. Традиционните методи за структуриране, съхраняване и обработка на данни често не позволяват лесно адаптиране към промените в потребностите на потребителите или динамичните изисквания на интернет средата. Ограничената им способност да разрешават неточно формулирани задачи също възпрепятства гъвкавостта. Затова съвременните модели за управление на данни в интернет все повече намират приложение в интелигентни информационни системи (ИИС), които да се справят с тези предизвикателства.
2. В повечето корпоративни СУБД запитванията използват строго фиксиран синтаксис и семантика, което затруднява процеса на структуриране и достъп до данни. Ограниченията на тези методи често водят до трудности при обработка на запитвания, зададени на естествен език, и затова са необходими иновативни модели за съхраняване и обработка на данни, които да улесняват работата с разнообразни езикови структури.
3. Един от ефективните методи за структуриране на данни е създаването на системи, при които описанието на структурата на данните е представено по начин, разбираем за всички потребители, като така се улеснява проектирането и изграждането на бази от данни, дори от хора без специализирани технически знания. Тези модели значително подобряват ефективността на системите за управление на данни в интернет.
4. С цел подобряване на методите за съхраняване и обработка на данни, в съвременните СУБД се разработват технологии, които позволяват запитванията и анализът на данни да бъдат съобразени с човешкия начин на мислене и изразяване. Това включва използването на изкуствен интелект и машинно обучение, които позволяват на моделите да съхраняват и анализират данни, базирани на контекста и нуждите на потребителя.
5. Един от най-съществените методи за обработка на данни в интернет е създаването на модели, които могат да извличат информация и да генерират отговори, основани на натрупаните знания. Тези модели се развиват така, че да се справят с аналитични и предсказателни задачи, базирани на опита, и предоставят релевантни отговори в реално време.
6. Възможностите за отдалечен достъп до данни изискват методи за обработка, които позволяват гъвкавост и мащабируемост. Интегрирането на информационни системи, базирани на тези методи в интернет мрежата предоставя на потребителите възможности за структуриране и съхраняване на данни от различни източници, съобразявайки се с техните специфични нужди.

7. С напредъка на технологиите става все по-важно ясно да се дефинират и разбират основни термини като „структуриране“, „съхраняване“ и „обработка“ на данни. Изясняването на тези понятия е ключово за създаването на ефективни модели за управление на данни, които да отговарят на динамичните изисквания на интернет средата и да осигуряват надеждно и оптимизирано управление на информацията.

В съответствие с поставените цели в дисертационния труд са разработени, изследвани и експериментирани различни методи и модели за обработка на данни. В резултат на това могат да се направят следните изводи и заключения:

1. Разгледани са различни методи за структуриране, съхраняване и обработка на данни, включително както класически, така и хибридни архитектури на облачни бази от данни.
2. Извършен е анализ на всеки от разгледаните методи и архитектури, като са посочени техните силни и слаби страни. Това предоставя детайлна информация за избор на подход за проектиране на информационни системи, който позволява известни компромиси.
3. Разгледани са съществуващи подходи за съхраняване и управление на данни, като склад за данни и езеро от данни. Описани са основните им характеристики и функции.
4. Предложен е хибриден модел, който комбинира различни архитектури на хранилища за данни, осигурявайки гъвкавост и ефективност при управлението на данни от различни източници. Моделът използва три нива на организация: децентрализирано, разпределено и глобално, което позволява поэтапна централизация на данните.
5. Предложен е референтен модел [73], който включва добре познати архитектурни стилове и предлага опции за адаптиране, за да отговори на специфични цели при структурирането, съхраняването и обработката на данни. Конкретната архитектура на модела може да се конфигурира чрез добавяне или премахване на слоеве, възли и компоненти.
6. Извършен е SWOT анализ на предложения модел, от който става ясно, че хибридният подход предлага решения, подходящи за различни бизнес нужди, но всяка архитектура има свои ограничения, които изискват компромиси при проектиране на информационни системи. Гъвкавостта на модела позволява постепенно внедряване и мащабиране според нуждите.
7. Разработено е интуитивно уеб-базирано приложение за структуриране и съхраняване на данни, което може да се използва дори от неспециалисти.

Глава 4 на дисертацията представя лесен начин за управление на данни чрез уеб-базирана информационна система, която позволява структуриране, съхраняване и обработка на данни от потребителите в Интернет. Описанието на структурата на данните е представено по начин, разбираем за всички потребители, което улеснява проектирането и изграждането на бази от данни дори от непрофесионалисти. Системата също така дава възможност за споделяне на данни между потребителите, което насърчава съвместната работа и анализ на споделени данни.

Потребителите могат да създават таблици, да определят видимостта на тези таблици за други потребители, да дефинират атрибути и да въвеждат данни в собствени или споделени таблици. Системата е подходяща както за съхранение и управление на собствени работни процеси, така и за съвместна работа на няколко потребители, независимо от областта на приложение. Тя може лесно да бъде адаптирана към различни научни или работни процеси, които генерират или обработват данни, без да се налага промяна на нейната структура или функционалност. Посредством изброените функции системата реализира до голяма степен идеята за отваряне на базите от данни за широк кръг от потребители.

## **ПРИНОСИ НА ДИСЕРТАЦИОННИЯ ТРУД**

В резултат на работата по дисертационния труд, приносите могат да се обобщят както следва:

### **Научно-приложни приноси**

1. Извършен е задълбочен обзор на съвременната литература в областта на моделирането, структурирането, съхраняването и обработката на данни в Интернет.
2. Извършен е задълбочен анализ на архитектури за структуриране, съхраняване и обработка на данни в облачна среда, който служи като основа за разработването на устойчиви и мащабируеми системи, които отговарят на изискванията за достъпност на базите от данни за широк кръг потребители.
3. Разработена е методика за подбор и оценка на модели за структуриране, съхраняване и обработка на данни, осигуряваща систематичен подход за адаптация на модели към конкретни изисквания за структуриране, съхраняване и обработка на данни.
4. Оценени и анализирани са модели за структуриране, съхраняване и обработка на данни с акцент върху приложението на разработената методика за тяхната оценка. За анализа са използвани софтуерни средства, които позволяват обективна оценка на критични аспекти като ефективност, устойчивост и мащабируемост на моделите.
5. Създаден е хибриден модел, който комбинира определени функционалности на избраните модели, отговаряйки на конкретни изисквания за структуриране, съхраняване и обработка на данни и осигуряване на достъп до базите от данни от широк кръг потребители. Разработената диаграма на потока от данни подчертава ефективността на модела в различни практически сценарии.
6. Извършен е подробен SWOT анализ на хибридният модел, който потвърждава възможностите за интеграция на хибридният модел в реални системи и подчертава неговата гъвкавост и устойчивост за различни решения.

### **Приложни приноси**

1. Разработен е метод за оптимизация на операциите с данни, който интегрира добри практики и доказани техники за структуриране, съхранение и обработка на данни. Неговата приложимост е демонстрирана чрез симулации и тестове в реални условия.
2. Чрез емпирични тестове е доказана ефективността на предложения хибриден модел. Тестовите включват оценка на производителността, устойчивостта и мащабируемостта.
3. На базата на метода за оптимизация на операциите с данни е разработена уеб-базирана система за структуриране, съхраняване и обработка на данни, която осигурява възможности за предоставяне на достъп до базите от данни за широк кръг от потребители. Проведените тестове в реални сценарии потвърждават практическата ѝ ефективност.

## СПИСЪК НА ПУБЛИКАЦИИТЕ ПО ДИСЕРТАЦИОННИЯ ТРУД

- [1] Хаджиев В., Рашидов А., Обзор и анализ на методи и модели за структуриране, съхраняване и обработка на данни в Интернет, Сборник доклади от международна научна конференция Автоматика и информатика '19, София, 2019, том 1, стр. 215-218, ISSN1313-1850.
- [2] Хаджиев В., Рашидов А., Обзор и анализ на архитектури за структуриране, съхраняване и обработка на данни в Облак, Сборник доклади от международна научна конференция Автоматика и информатика '19, София, 2019, том 1, стр. 219-222, ISSN1313-1850.
- [3] Хаджиев В., Рашидов А. Обзор и анализ на методи и модели за структуриране, съхраняване и обработка на данни в интернет. Списание Автоматика и информатика. 2019, кн. 2, 27-32, ISSN 0861-7562
- [4] Хаджиев В., Рашидов А. Обзор и анализ на архитектури за структуриране, съхраняване и обработка на данни в облак. Списание Автоматика и информатика. 2019, кн. 3, 12-16, ISSN 0861-7562
- [5] V. Hadzhiev, A. Rashidov, Overview and analysis of methods and models for data structuring, storage and processing in the Internet, 2019 11th International Conference on Electrical and Electronics Engineering (ELECO), Bursa, Turkey, 2019, pp. 492-495, doi: 10.23919/ELECO47770.2019.8990416 - Научна публикация - Доклад от международна конференция в чужбина.
- [6] Velin Hadzhiev and Aldeniz Rashidov, A Hybrid Model for Structuring, Storing and Processing Distributed Data on the Internet, 2021 International Conference Automatics and Informatics (ICAI), 2021, pp. 82-85, doi: 10.1109/ICAI52893.2021.9639240 - Научна публикация - Доклад от международна конференция в България.
- [7] Velin Hadzhiev and Aldeniz Rashidov, An Organization of the Storage and Data Flow in a Hybrid Model for Structuring, Storing and Processing Distributed Data on the Internet, 2021, 13th International Conference on Electrical and Electronics Engineering ELECO, 2021, pp. 589-593, doi: 10.23919/ELECO54474.2021.9677648 - Научна публикация - Доклад от международна конференция в чужбина.
- [8] Velin Hadzhiev, SWOT Analysis of a Hybrid Model for Structuring, Storing and Processing Distributed Data on the Internet, 2021, 13th International Conference on Electrical and Electronics Engineering ELECO, 2021, pp.585-588, doi: 10.23919/ELECO54474.2021.9677789 - Научна публикация - Доклад от международна конференция в чужбина.
- [9] Velin Hadzhiev, Aldeniz Rashidov, Implementation of a Data and Information Management System Based on a Hybrid Model for Structuring, Storing, and Processing Distributed Data on the Internet, 13th International Conference on Computing Communication and Networking Technologies, Online (Technical Assistance - ИТ Mandi), India, October 3-5, 2022, doi: 10.1109/ICCCNT54827.2022.9984384 - Научна публикация - Доклад от международна конференция в чужбина.



## ЦИТИРАНИЯ НА ПУБЛИКАЦИИ ПО ДИСЕРТАЦИОННИЯ ТРУД

Открити са цитирания от списъка на публикациите, свързани с дисертацията, както следва:

1. **Едно цитиране на труд № 7** в списание с импакт фактор 7.94 за 2023 - 2024 г.

Цитирана статия: Velin Hadzhiev and Aldeniz Rashidov, An Organization of the Storage and Data Flow in a Hybrid Model for Structuring, Storing and Processing Distributed Data on the Internet, 2021, 13th International Conference on Electrical and Electronics Engineering ELECO, 2021, pp. 589-593, doi: 10.23919/ELECO54474.2021.9677648 - Научна публикация - Доклад от международна конференция в чужбина.

Списание: Computers & Industrial Engineering, Volume 190, 2024, 110082, ISSN 0360-8352

Цитиращ автор: Samuel Harno, Hing Kai Chan, Min Guo

Заглавие на статията: Enhancing value creation of operational management for small to medium manufacturer: A conceptual data-driven analytical system

Линк: <https://doi.org/10.1016/j.cie.2024.110082>

(<https://www.sciencedirect.com/science/article/pii/S0360835224002031>)

Импакт фактор: 8.78

Линк за импакт фактора: <https://www.resurchify.com/impact/details/18164>

2. **Две цитирания на труд № 8** в списания с импакт фактори 0.71 и 1.03 за 2023 - 2024 г.

Цитирана статия: Velin Hadzhiev, SWOT Analysis of a Hybrid Model for Structuring, Storing and Processing Distributed Data on the Internet, 2021, 13th International Conference on Electrical and Electronics Engineering ELECO, 2021, pp. 585-588, doi: 10.23919/ELECO54474.2021.9677789 - Научна публикация - Доклад от международна конференция в чужбина.

Списание 1: International Journal on Recent and Innovation Trends in Computing and Communication, 11(7s), 01–07, E-ISSN:2321-8169 (Scopus, SJR 0,109)

Цитиращ автор: Selvi, S. E.

Заглавие на статията: Geo-Distance Based 2-Replica Maintaining Algorithm for Ensuring the Reliability Forever Even During the Natural Disaster on Cloud Storage System

Линк: <https://doi.org/10.17762/ijritcc.v11i7s.6971>

Импакт фактор: 0.71

Линк за импакт фактора: <https://www.resurchify.com/impact/details/21101089961>

Списание 2: International Journal of Intelligent Systems and Applications in Engineering, 11(3), 769–774, ISSN: 2147-6799 (Scopus, SJR 0,209)

Цитиращ автор: Selvi, S. A. E.

Заглавие на статията: A Neoteric Geo-Distance Based 2-Replica Placing Algorithms on Cloud Storage System

Линк: <https://ijisae.org/index.php/IJISAE/article/view/3283>

Импакт фактор: 1.03

Линк за импакт фактора: <https://www.resurchify.com/impact/details/21101021990>

## **ВНЕДРЯВАНИЯ**

Хибридният модел за структуриране, съхранение и обработка на разпределени данни в Интернет е имплементиран в експериментална уеб-базирана система за управление на данни и информация. Тя позволява на потребителите да организират и управляват данните според собствените си предпочитания, дори ако нямат опит в областта на базите от данни. Основните възможности на системата включват:

- **Гъвкаво управление на таблици и атрибути** – предоставя висока степен на персонализация и контрол над данните;
- **Създаване на неограничен брой таблици** – системата позволява конфигуриране на видимостта, дефиниране на атрибути и въвеждане на данни в таблиците, като достъпът се контролира спрямо правата на потребителите;
- **Управление на индивидуални и съвместни работни процеси** – подходяща е за разнообразни приложения, което я прави универсална и полезна за широк кръг потребители;
- **Лесно адаптиране към различни научни и работни процеси** – не изисква промяна на основната структура или функционалност при интеграция в нови проекти;
- **Системата е проектирана да бъде високо персонализируема**, универсална и адаптивна към различни работни процеси, отговаряйки на нуждите на потребителите.

# **TITLE: MODELLING OF DATA STRUCTURING, STORAGE, AND PROCESSING OPERATIONS ON THE INTERNET**

**Author: M. Eng. Velin Sabinov Hadzhiev**

## **ABSTRACT:**

The paper examines methods for modelling data structuring, storage, and processing operations, which form the foundation for developing a model that implements a hybrid approach to data management on the Internet.

The proposed model combines various data warehouse architectures, tailored to meet specific technical and business requirements. The data architecture and flow organization are depicted through IDFO diagrams, which illustrate the step-by-step centralization of data structuring, storage, and processing.

The outcomes of the development process include a classification of methods and models, the creation of a generalized model, and activities related to its preparation and analysis to identify its strengths and weaknesses, along with measures to address them.

An experimental web-based system has been developed to implement this model. The system enables users to organize and manage data based on their specific needs and preferences, even without prior experience in database development.

**Key words:** Data Processing, Data Structure, Data Storage, Data Integration, Data Model, Hybrid Model, Information System, Data Architecture, Data Centre, Data Warehouse, Relational Database, Data Marts, Data Cube, Big Data, Bottom-up Approach, Top-Down Approach, Central Repository, Cloud Computing, Cloud Database.