



## СОДЕРЖАНИЕ

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ .....	3
ВВЕДЕНИЕ .....	4
1 Исследование предметной области .....	5
1.1 Исследование особенностей систем агрегирования данных .....	5
1.1.1 Определение и назначение систем агрегирования данных .....	5
1.1.2 Классификация систем агрегирования данных .....	5
1.1.3 Архитектурные принципы систем агрегирования данных .....	5
1.2 Исследование методов выполнения аналитических запросов .....	6
1.2.1 Введение в аналитические запросы .....	6
1.2.2 Основные методы выполнения аналитических запросов .....	6
1.2.2.1 OLAP (Online Analytical Processing) .....	6
1.2.2.2 MapReduce .....	6
1.2.2.3 SQL и NoSQL-подходы .....	6
1.2.2.4 Индексы и материализованные представления .....	6
1.3 Исследование особенностей работы оперативной аналитической обработки .....	7
1.3.1 Определение оперативной аналитической обработки (OLAP) .....	7
1.3.2 Основные принципы OLAP .....	7
1.3.3 Технологии оперативной аналитической обработки .....	7
1.3.4 Преимущества и недостатки OLAP .....	7

## ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящем отчете применяют следующие термины, сокращения и определения с соответствующими определениями.

OLAP	
API	
ETL	
BI-система	
БД	
NoSQL-хранилищами	
batch processing	
stream processing	
OLAP	
MOLAP	
ROLAP	
HOLAP	
<b>MapReduce</b>	
<b>SQL</b>	
PostgreSQL	
MySQL	
Oracle	
СУБД	

## **ВВЕДЕНИЕ**

В современном мире финансовая аналитика требует обработки больших объемов данных, поступающих из различных источников. Для эффективного анализа и принятия решений необходимо агрегировать данные, сводя их к структурированным наборам, удобным для последующей обработки. В данной главе рассматриваются особенности систем агрегирования данных, их архитектурные принципы и функциональные возможности.

## **1 Исследование предметной области**

### **1.1 Исследование особенностей систем агрегирования данных**

#### **1.1.1 Определение и назначение систем агрегирования данных**

Системы агрегирования данных представляют собой программные и аппаратные комплексы, предназначенные для сбора, обработки, обобщения и хранения информации из разнородных источников. Эти системы применяются для консолидации данных и их подготовки к аналитической обработке.

Основные задачи агрегирования данных:

- Объединение данных из различных источников (базы данных, API, файловые хранилища);
- Очистка и трансформация данных;
- Поддержка процессов ETL (Extract, Transform, Load);
- Обеспечение оперативного доступа к агрегированным данным;
- Подготовка данных для аналитических и BI-систем.

#### **1.1.2 Классификация систем агрегирования данных**

Системы агрегирования данных можно классифицировать по различным критериям:

- По типу источников данных: работающие с реляционными БД, NoSQL-хранилищами, потоковыми данными.
- По способу обработки: пакетная обработка (batch processing), потоковая обработка (stream processing), гибридные подходы.
- По архитектуре: централизованные системы, распределенные системы.

#### **1.1.3 Архитектурные принципы систем агрегирования данных**

Современные системы агрегирования данных строятся на основе следующих архитектурных подходов:

- Многоуровневые архитектуры, включающие уровни сбора, обработки и хранения данных.
- Микросервисный подход, обеспечивающий гибкость и масштабируемость.
- Использование облачных технологий, позволяющее динамически изменять ресурсы под нагрузку.

## **1.2 Исследование методов выполнения аналитических запросов**

### **1.2.1 Введение в аналитические запросы**

Аналитические запросы предназначены для обработки больших объемов данных с целью выявления закономерностей, трендов и аномалий. Они широко используются в финансовой аналитике для расчета премий, оценки рисков и прогнозирования.

### **1.2.2 Основные методы выполнения аналитических запросов**

#### **1.2.2.1 OLAP (Online Analytical Processing)**

OLAP-технология предназначена для многомерного анализа данных и позволяет выполнять сложные аналитические запросы. Основные типы OLAP:

- MOLAP (Multidimensional OLAP) — хранение данных в многомерных кубах.
- ROLAP (Relational OLAP) — хранение данных в реляционных таблицах, обработка с помощью SQL-запросов.
- HOLAP (Hybrid OLAP) — гибридный подход, сочетающий MOLAP и ROLAP.

#### **1.2.2.2 MapReduce**

Метод MapReduce позволяет обрабатывать большие объемы данных параллельно на распределенных системах. Этот метод эффективен для работы с неструктурированными и полуструктурированными данными.

#### **1.2.2.3 SQL и NoSQL-подходы**

- SQL-методы: Используются в традиционных реляционных БД (PostgreSQL, MySQL, Oracle) для аналитических запросов с агрегацией (SUM, AVG, COUNT, GROUP BY).
- NoSQL-методы: Используются в документоориентированных, графовых и других БД (MongoDB, Cassandra) для обработки данных в реальном времени.

#### **1.2.2.4 Индексы и материализованные представления**

Для ускорения аналитических запросов применяются индексы (B-деревья, Bitmap-индексы) и материализованные представления, хранящие предварительно рассчитанные результаты запросов.

## **1.3 Исследование особенностей работы оперативной аналитической обработки**

### **1.3.1 Определение оперативной аналитической обработки (OLAP)**

Оперативная аналитическая обработка (OLAP) представляет собой технологию, обеспечивающую быстрый доступ к агрегированным данным в многомерных структурах. OLAP используется в системах финансовой аналитики для мгновенного расчета показателей, таких как премии, рентабельность и финансовые риски.

### **1.3.2 Основные принципы OLAP**

- Многомерность данных: Данные организованы в виде кубов с различными измерениями (время, категория, география).
- Агрегация: Данные сводятся к обобщенным показателям, что снижает объем вычислений.
- Оптимизация хранения: Используются специальные структуры данных для быстрого доступа.

### **1.3.3 Технологии оперативной аналитической обработки**

Среди наиболее распространенных технологий OLAP можно выделить:

- Apache Druid: Высокопроизводительная аналитическая база данных для работы с потоковыми и историческими данными.
- ClickHouse: Колонночная СУБД с высокой скоростью выполнения аналитических запросов.
- Microsoft Analysis Services: Инструмент для работы с OLAP-кубами в экосистеме Microsoft.

### **1.3.4 Преимущества и недостатки OLAP**

Преимущества:

- Высокая скорость выполнения запросов за счет предварительной агрегации данных.
- Возможность многомерного анализа данных.
- Поддержка сложных аналитических вычислений.

Недостатки:

- Высокие затраты на вычислительные ресурсы.
- Ограниченная гибкость по сравнению с транзакционными базами данных.
- Сложность настройки и поддержки.

