

## ТРАДИЦИОННАЯ И ОБЛАЧНАЯ АРХИТЕКТУРА ХРАНИЛИЩ ДАННЫХ

Лучия НИГАЙ

Департамент Программной Инженерии и Автоматики, группа TI-196, Факультет Вычислительной Техники, Информатики и Микроэлектроники, Технический университет Молдовы, Кишинев. Молдова

Автор Корреспондент: Лучия НИГАЙ, e-mail: [nigai.lucia@isa.utm.md](mailto:nigai.lucia@isa.utm.md)

Научный руководитель: Саранчук Дориан, DISA, FCIM, UTM

**Аннотация:** В данной статье рассматриваются традиционная архитектура хранилищ данных и компоненты из которых она состоит, а также облачная архитектура хранилищ данных, ее компоненты и функционирование на примере Amazon Redshift.

**Ключевые слова:** хранилище данных, традиционная архитектура, облачные хранилища, Amazon Redshift

### Введение

Хранилище данных – это разновидность системы хранения и управления данными, предоставляющая инструменты онлайн-аналитической обработки (OLAP), и позволяющая проводить интерактивный анализ данных в многомерном пространстве, который приводит к обобщению данных и их интеллектуальному анализу.

Хранилища данных содержат в себе исторические и коммутативные данные из различных источников, поэтому их архитектура довольно непростая. В этой статье мы рассмотрим традиционную и облачную архитектуру хранилищ данных, широко используемые в различных областях, таких как обработка транзакций, аналитическая обработка, прогнозная аналитика и многое другое.

### Традиционная архитектура хранилищ данных

Традиционные хранилища данных имеют три подхода к построению: виртуальное хранилище данных (одноуровневая архитектура), киоски или витрины данных (двухуровневая архитектура) и корпоративные хранилища данных (трехуровневая архитектура). Далее рассмотрим каждый из подходов [1]:

**Виртуальное хранилище данных**, его целью является минимизация объема данных за счет удаления их избыточности. Обычно оно создается путем создания ряда представлений к базе данных или использованием специальных средств доступа.

**Киоски или витрины** данных разделяют физические источники и хранилище данных, путем построения витрин данных без создания центрального хранилища. Они способны хранить лишь ограниченное количество данных, являются нерасширяемыми, а также поддерживают лишь небольшое количество пользователей. Выделяют два типа киосков данных: независимый (данные поступают напрямую из оперативных источников данных) и зависимый (информация является подмножеством основного хранилища) [1].

**Корпоративное хранилище данных** – наиболее используемый вид хранилищ данных, состоящий из трех уровней (Рисунок 1):

- Сервер базы данных, который используется для извлечения данных из транзакционных систем и других источников данных.
- Сервер OLAP, представляющий собой абстрактное структурированное представление базы данных, подходящее для анализа и создания сложных запросов.

- Клиентский интерфейс, а именно инструменты API, которые используются для извлечения и анализа данных из хранилища.

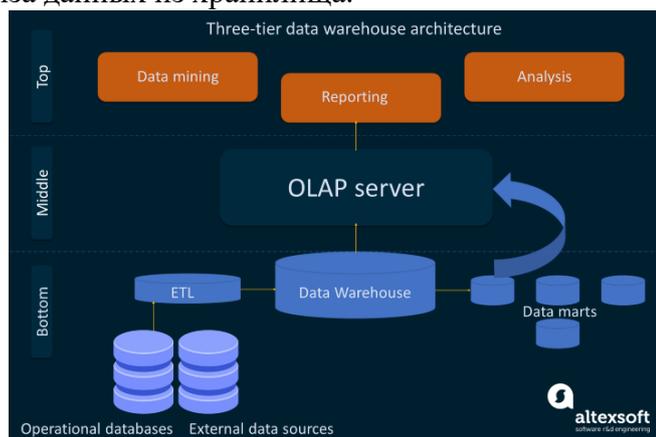


Рисунок 1. Трехуровневая архитектура

В большинстве своем хранилища данных состоят из 5 основных компонентов[2]:

- Программное обеспечение промежуточного слоя, к которому относятся сетевые и коммуникационные протоколы, драйверы и т.д.
- Транзакционные базы данных и внешние источники информации, такие как отчеты.
- Уровень доступа к данным, обеспечивающий взаимодействие с пользователем и загрузку данных из транзакционных систем.

### Загрузка и обработка

Существует два подхода: ETL и ELT.

При использовании первого подхода (ETL) данные сначала извлекаются из пула источников данных и хранятся во временной промежуточной базе данных. Затем для структурирования и преобразования данных в подходящую форму выполняются операции преобразования и лишь после этого уже структурированные данные загружаются в хранилище.

При использовании второго подхода (ELT) отсутствует промежуточная база данных и данные немедленно загружаются в единый репозиторий сразу после извлечения из исходных пулов данных. Затем они преобразуются в системе хранилища данных [3].

- Информационное хранилище, являющееся ядром всей системы и представляющее собой один или несколько серверов БД.
- Метаданные, содержащие сведения о источнике, использовании, значении и особенности хранилища данных; а также то, как данные могут быть изменены и обработаны.
- Уровень информационного доступа, обеспечивающий взаимодействие пользователя с хранилищем посредством стандартных систем анализа и манипулирования данными.
- Уровень управления, где происходит отслеживание и программирование различных процедур.

Мы рассмотрели традиционную архитектуру хранилищ данных, а также основные компоненты, из которых состоит хранилище данных.

### Облачная архитектура

Сама концепция облачных хранилищ относительно нова, и разработана на основе концепции облачных вычислений, которые являются развитием параллельной обработки, распределенных вычислений и грид-вычислений. Благодаря технологии облачных вычислений пользователи могут обрабатывать огромные объемы информации в течении нескольких секунд, за счет разбиения программы обработки на меньшие программы, которые затем обрабатываются и анализируются системой, состоящей из нескольких серверов.

Концепция облачного хранения аналогична облачным вычислениям, каждое облако состоит из нескольких узлов, которые соединены между собой быстрыми каналами передачи и управления большими объемами информации. В зависимости от требований к облачным хранениям и вычислениям существуют различные типы серверов, сетей и технологий хранения. В основе всех облачных сервисов, продуктов и решений лежат программные инструменты с тремя основными функциональными элементами: инструменты для обработки данных и запуска приложений (вычислительные серверы), перемещения данных (сеть) и хранения данных (хранилище).

Каждое хранилище имеет свою уникальную архитектуру, но несмотря на это выделяют четыре модели облачной архитектуры: общедоступная, частная, гибридная, и мультиоблачная архитектуры [4].

- Общедоступная - данная модель характеризуется тем, что вычислительные ресурсы принадлежат и управляются поставщиком и совместно используются, и перераспределяются между несколькими арендаторами через интернет.
- Частная - данная модель облака принадлежит и управляется частным образом, и может включать в себя несколько серверов. Среди ее преимуществ можно выделить настраиваемость и строгие параметры безопасности.
- Гибридная - данная модель совмещает в себе эффективность работы общедоступного облака и возможности защиты данных частного за счет использования публичных и частных облачных архитектур.
- Мультиоблачная - использует несколько общедоступных облачных сервисов. В ее преимущества входит большая гибкость в выборе и развертывании облачных сервисов, снижение зависимости от одного поставщика облачных услуг [4].

Архитектура облачного хранилища включает в себя следующие основные компоненты:

- Виртуализация: облака создаются на основе виртуализации серверов, хранилищ и сетей. Виртуализированные ресурсы - это программное или виртуальное представление физического ресурса, такого как серверы или хранилище, что позволяет нескольким приложениям использовать одни и те же физические ресурсы [5].
- Инфраструктура: Облачная инфраструктура включает в себя все компоненты традиционных центров обработки данных, включая серверы, постоянное хранилище и сетевое оборудование, включая маршрутизаторы и коммутаторы.
- Промежуточное ПО: программные компоненты, такие как базы данных и коммуникационные приложения, позволяют подключенным к сети компьютерам, приложениям и программному обеспечению взаимодействовать друг с другом.
- Управление: эти инструменты обеспечивают непрерывный мониторинг производительности и емкости облачной среды.
- Программное обеспечение для автоматизации: в облачной архитектуре автоматизация используется для простого масштабирования системных ресурсов в соответствии с резким увеличением спроса на вычислительную мощность, развертывания приложений для обеспечения управления в облачной среде [5].

Рассмотрим архитектуру хранилища Amazon Redshift (Рисунок 2.). Оно использует архитектуру массивно-параллельной обработки, в которой узлы хранят данные в срезах в столбчатом формате. Иерархия архитектуры Amazon состоит из множества уровней, таких как клиентские приложения, ведущий узел, вычислительный узел, срез узла. Давайте вкратце рассмотрим функции каждого из этих слоев. Клиентские приложения используются для подключения к кластеру Amazon Redshift через драйверы JDBC или ODBC. Ведущий узел отвечает за связь с клиентским приложением и вычислительными узлами. Вычислительный узел выполняет такие функции, как загрузка данных, резервное копирование и восстановление данных. Срез узла используется для распределения данных внутри узла. Каждый раз, когда ведущий узел назначает операции срезам узла, они начинают работать параллельно, чтобы завершить операцию.

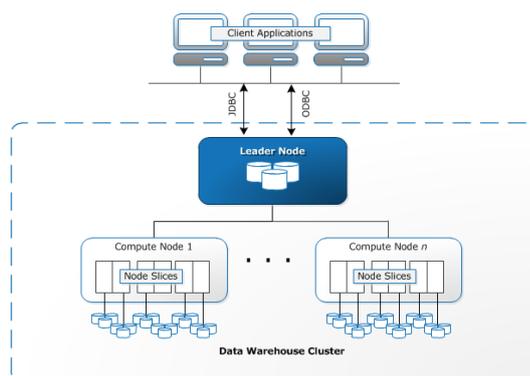


Рисунок 2. Архитектура Amazon Redshift

### Заключение

Традиционная архитектура данных все еще широко используется в основном при работе с схожими структурированными типами данных и требует больших затрат на создание и обслуживание, а также уступает в гибкости и скорости обработки данных облачным хранилищам. В свою очередь облачная архитектура отличается масштабируемостью, высокой скоростью обработки данных и гибкостью для различных форматов данных, что дает значительные преимущества в интеграции и производительности. А также благодаря массово параллельной обработке облачные хранилища способны выполнять сложные аналитические запросы в гораздо более быстром темпе.

### Библиография

1. Архитектура хранилищ данных: традиционная и облачная – [Online]- [обращения 02.03.2023], Доступно: <https://habr.com/ru/post/441538/>
2. Data warehouse system architecture – [Online], [обращения 02.03.2023] Доступно: [https://docs.aws.amazon.com/redshift/latest/dg/c\\_high\\_level\\_system\\_architecture.html](https://docs.aws.amazon.com/redshift/latest/dg/c_high_level_system_architecture.html)
3. Data Warehouse Architecture: Types, Components, & Concepts – [Online]- [обращения 02.03.2023], Доступно: <https://www.astera.com/type/blog/data-warehouse-architecture/>
4. What is Cloud Architecture? – [Online]- [обращения 02.03.2023], Доступно: <https://www.vmware.com/topics/glossary/content/cloud-architecture.html>
5. Современная архитектура хранилища данных: традиционное или облачное хранилище данных – [Online]- [обращения 02.03.2023], Доступно: <https://biconsult.ru/services/sovremennaya-arhitektura-hranilishcha-dannyh-tradicionnoe-ili-oblachnoe-hranilishche-dannyh>