



ОТКРЫТАЯ НАУКА
издательство

Международный журнал информационных технологий и
энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.62

РАЗРАБОТКА ГИБРИДНОЙ АРХИТЕКТУРЫ ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ ПОСТРОЕНИЯ ОТЧЕТНОСТИ С ПРИМЕНЕНИЕМ OLAP И OLTP СИСТЕМ

Серда И.А.

ФГБОУ ВО «МИРЭА - РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ», Москва, Россия (119454, г. Москва, Пр-т Вернадского, д. 78, стр.4), e-mail: ilya.sereda2002@mail.ru

В данной работе рассматривается разработка гибридной архитектуры информационной системы для построения отчетности, объединяющей возможности OLAP и OLTP-систем. Предложенный подход использует ClickHouse в качестве высокопроизводительного хранилища для обработки и агрегации больших объемов данных, а также OLTP-базу данных PostgreSQL для хранения предварительно рассчитанных показателей и оперативного доступа к отчетным данным.

Классическая схема аналитики предполагает выполнение всех расчетов и отчетных запросов в OLAP-хранилище, что приводит к высокой нагрузке на систему и увеличению времени отклика. В предлагаемой архитектуре высоконагруженные аналитические вычисления выполняются в OLAP, а полученные агрегированные метрики загружаются в OLTP-хранилище, откуда BI-системы могут быстро извлекать данные с минимальными задержками.

Таким образом, предложенная архитектура представляет собой эффективное решение для построения высокопроизводительных систем отчетности, объединяющее преимущества OLAP и OLTP для оптимального распределения нагрузки и ускоренного доступа к данным.

Ключевые слова: Гибридная архитектура, OLAP, OLTP, построение отчетности, хранилище данных.

DEVELOPMENT OF A HYBRID ARCHITECTURE OF INFORMATION SYSTEM FOR BUILDING REPORTS USING OLAP AND OLTP SYSTEMS

Sereda I.A.

MIREA - RUSSIAN TECHNOLOGICAL UNIVERSITY, Moscow, Russia (119454, Moscow, avenue. Vernadsky, 78, b. 4), e-mail: ilya.sereda2002@mail.ru

This paper considers the development of a hybrid architecture of information system for reporting, combining the capabilities of OLAP and OLTP-systems. The proposed approach uses ClickHouse as a high-performance storage for processing and aggregation of large amounts of data, as well as OLTP-database PostgreSQL for storing pre-calculated indicators and operational access to reporting data.

The classical analytics scheme assumes that all calculations and reporting queries are performed in OLAP-storage, which leads to high load on the system and increased response time. In the proposed architecture, heavy analytical calculations are performed in OLAP and the resulting aggregated metrics are loaded into OLTP storage from where BI systems can quickly retrieve data with minimal latency.

Thus, the proposed architecture is an efficient solution for building high-performance reporting systems, combining the advantages of OLAP and OLTP for optimal load balancing and accelerated data access.

Keywords: Hybrid architecture, OLAP, OLTP, report building, data warehouse.

В современных условиях цифровой трансформации бизнеса эффективное управление данными становится критическим фактором успеха организаций. Рост объемов информации, повышение требований к скорости её обработки и необходимость предоставления аналитических отчетов в режиме, близком к реальному времени, создают потребность в разработке новых архитектурных решений для информационных систем.

Традиционно в корпоративных информационных системах применяется разделение на транзакционные (OLTP) и аналитические (OLAP) системы. OLTP-системы оптимизированы для обработки множества коротких транзакций и обеспечения целостности данных, в то время как OLAP-системы предназначены для сложной многомерной аналитики и агрегации больших объемов данных. Обычно данные перемещаются из OLTP-систем в OLAP через процессы извлечения, преобразования и загрузки (ETL), формируя однонаправленный поток информации.

Однако такой классический подход имеет ряд ограничений. Во-первых, сложность и продолжительность ETL-процессов часто приводит к задержкам в получении актуальной аналитики. Во-вторых, предоставление доступа к OLAP-системам конечным пользователям требует дополнительных ресурсов и компетенций. В-третьих, при увеличении нагрузки на систему отчетности могут возникать проблемы с производительностью, особенно если многие пользователи одновременно выполняют ресурсоемкие аналитические запросы.

В данной статье предлагается альтернативный подход к построению архитектуры информационной системы, основанный на гибридном использовании OLAP и OLTP технологий. Представленная архитектура использует в качестве OLAP-системы СУБД ClickHouse [4] для выполнения сложных аналитических расчетов, а результаты этих расчетов сохраняет в OLTP-базу данных, реализация которой будет осуществлена с помощью СУБД PostgreSQL, в виде предварительно агрегированных витрин данных. Такой подход позволяет объединить вычислительную мощность OLAP-систем с быстродействием и доступностью OLTP-систем для конечных пользователей и инструментов визуализации.

Целью работы является разработка и обоснование гибридной архитектуры информационной системы для построения отчетности, в которой OLAP-хранилище используется для выполнения тяжелых аналитических расчетов, а OLTP-система — для хранения агрегированных показателей и обеспечения быстрого доступа к данным.

OLTP (Online Transaction Processing) и OLAP (Online Analytical Processing) представляют собой два фундаментально различных подхода к хранению и обработке данных [1]. OLTP-системы оптимизированы для высокочастотных транзакций, обеспечивают атомарность операций и целостность данных. Они характеризуются нормализованной структурой хранения, малыми по объему транзакциями и ориентированы на быстрый доступ к отдельным записям. Примерами таких систем являются PostgreSQL, MySQL, Oracle.

OLAP-системы, напротив, предназначены для аналитической обработки больших объемов данных, часто используют денормализованные схемы и колоночное хранение, оптимизированы для сложных запросов, агрегации и выполнения многомерного анализа. К таким системам относятся ClickHouse, Vertica, Snowflake.

Традиционная архитектура аналитических систем [2] обычно включает в себя источники данных (OLTP-системы), ETL-процессы, хранилище данных [3] (чаще всего реализованное как OLAP) и средства визуализации. Данные движутся по направлению от операционных систем к аналитическим, формируя однонаправленный поток. Пример такой архитектуры представлен на Рисунке 1.

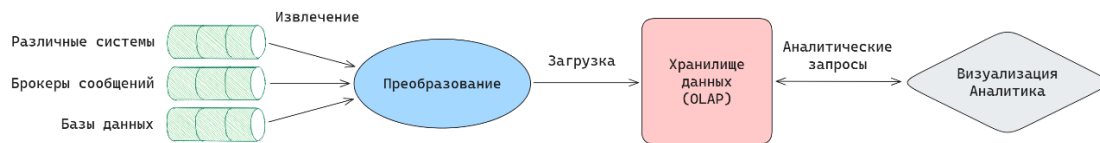


Рисунок 1 — Пример традиционной архитектуры аналитической системы

OLAP-системы, такие как ClickHouse, рассчитаны на выполнение сложных аналитических запросов, но они не оптимизированы для частых однотипных запросов. BI-системы и API-приложения могут порождать множество повторяющихся запросов, нагружая OLAP-хранилище. Если BI-инструменты напрямую запрашивают данные из OLAP-системы, нагрузка может увеличиваться, требуя горизонтального масштабирования. Если OLAP-система выходит из строя или сильно перегружена, пользователи не могут получить доступ к отчетности. В традиционной архитектуре нет альтернативного механизма быстрого доступа к предрасчитанным метрикам.

Предлагаемая модель гибридной архитектуры информационной системы представляет собой инновационный подход к организации потоков данных, отличающийся от традиционных решений. В основе данной модели лежит идея оптимального использования сильных сторон OLAP и OLTP систем при построении комплексной инфраструктуры для аналитической отчетности.

Ключевой особенностью предлагаемой архитектуры является изменение направления потока данных после их аналитической обработки. В отличие от традиционного однонаправленного движения данных (от OLTP к OLAP), в разработанной модели присутствует возвратный поток предварительно агрегированных данных из OLAP-систем обратно в специализированные OLTP-хранилища, используемые для обслуживания конечных пользователей.

Концептуально модель включает следующие основные компоненты:

1. Источники данных — традиционные OLTP-системы, генерирующие первичные транзакционные данные в процессе деятельности организации, различные системы, которые отправляют данные напрямую или брокеры сообщений.
2. ETL-процесс — процедура извлечения и загрузки данных из источников в OLAP-систему, а уже впоследствии — обработка и построение логики расчетов показателей.
3. Аналитическое ядро — OLAP-система (в нашем случае ClickHouse), выполняющая сложные аналитические вычисления, агрегацию и формирование витрин данных.
4. Витрины данных в OLTP — специализированная OLTP-система (в нашем случае PostgreSQL), хранящая предварительно рассчитанные метрики и агрегаты в структуре, оптимизированной для быстрого точечного доступа.
5. Оркестрация процессов — система управления потоками данных и регулярными расчетами (в нашем случае Apache Airflow [5;6]), координирующая работу всех компонентов обновления витрин данных.
6. Слой доступа к данным — API и инструменты визуализации (в нашем случае Apache Superset), обеспечивающие конечным пользователям доступ к аналитическим данным из витрин OLTP-хранилища.

Гибридная архитектура информационной системы для построения отчетности представлена на Рисунке 2.

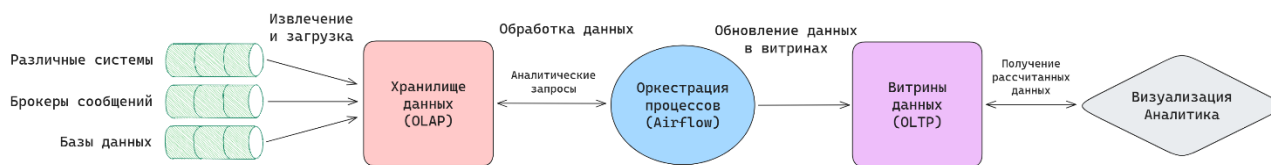


Рисунок 2 — Гибридная архитектура информационной системы для построения отчетности

При такой архитектуре OLAP-система используется по своему прямому назначению — для эффективного выполнения сложных аналитических расчетов над большими объемами данных, но результаты этих расчетов не предоставляются пользователям напрямую, а передаются в OLTP-систему. Это позволяет изолировать ресурсоемкие аналитические процессы от пользовательских запросов и обеспечить высокую производительность при обращении к уже рассчитанным показателям.

Использование ClickHouse в качестве вычислительного ядра позволяет эффективно выполнять сложные аналитические расчеты над большими объемами данных. Хранение предварительно рассчитанных витрин данных в PostgreSQL обеспечивает высокую скорость доступа к аналитическим показателям при обращении конечных пользователей.

Таким образом, предлагаемая гибридная архитектура сочетает вычислительную мощь OLAP-систем с быстродействием и масштабируемостью OLTP-систем, позволяя эффективно решать задачи построения корпоративной отчетности в условиях высоких требований к скорости доступа и значительного количества одновременных запросов:

1. OLAP используется только для сложных расчетов, а OLTP — для хранения готовых агрегированных данных.
2. BI-инструменты и API обращаются не к OLAP, а к OLTP, что снижает нагрузку на аналитическое хранилище.
3. Быстрый отклик отчетов за счет оптимизированных индексов в OLTP.
4. Снижение требований к ресурсам OLAP, так как большинство пользовательских запросов идет в OLTP.
5. Более частое обновление данных, так как в OLTP загружаются уже готовые значения, а не сырые данные, требующие перерасчета.

Список литературы

1. OLTP vs OLAP. // Engineering Resources: [сайт]. — URL: <https://clickhouse.com/engineering-resources/oltp-vs-olap> (дата обращения: 12.02.2025). — Текст: электронный.
2. Марц Н., Уоррен Дж. Большие данные: принципы и практика построения масштабируемых систем обработки данных в реальном времени. — М.: Вильямс, 2016. — 356 с. — Текст: непосредственный.
3. Data Warehouse Architecture. // Geeks for geeks: [сайт]. — URL: <https://www.geeksforgeeks.org/data-warehouse-architecture/> (дата обращения: 31.01.2025). — Текст: электронный.
4. Schulze, R., Schreiber, T., Yatsishin, I., Dahimene, R., & Milovidov, A. ClickHouse-Lightning Fast Analytics for Everyone // 50th International Conference on Very Large Databases. —

Guangzhou, China - August 26-30, 2024. – URL: <https://www.vldb.org/pvldb/vol17/p3731-schulze.pdf> (дата обращения: 29.01.2025). — Текст: электронный.

5. Apache Airflow: официальный сайт – URL: <https://airflow.apache.org/docs/> (дата обращения: 05.02.2025). — Текст: электронный.
6. Харенслак Б., де Руйтер Дж. Apache Airflow и конвейеры обработки данных / пер. с англ. Д. А. Беликова. – М.: ДМК Пресс, 2021. – 502 с. ил. ISBN 978-5-97060-970-5 — Текст: непосредственный.

References

1. OLTP vs OLAP. // Resources Engineering: [website]. — URL: <https://clickhouse.com/engineering-resources/oltp-vs-olap> (date of application: 12.02.2025). — Text: electronic.
 2. N. Martz, D. Stupidly. Big data: a trowel in the practice of not really processing the principle of data system construction in time. Moscow: Williams, 2016. p. 356 — Text: direct.
 3. Data Warehouse Architecture. // Geeks for Geeks: [website]. — URL: <https://www.geeksforgeeks.org/data-warehouse-architecture/> (date of request: 31.01.2025). — Text: electronic.
 4. Schulze, R., Schreiber, T., Yatshishin, I., Dahimene, R., & Milovidov, A. ClickHouse-Lightning Fast Analytics for Everyone // International Conference on Very Large 50th Databases. – Guangzhou, China - August 26-30, 2024. – URL: <https://www.vldb.org/pvldb/vol17/p3731-schulze.pdf> (date of application: 29.01.2025). — Text: electronic.
 5. Apache Airflow: Official website – URL: <https://airflow.apache.org/docs/> (date of request: 02/05/2025). — Text: electronic.
 6. Harenslak B. and D. Ruitier. Non-data processing of the Apache Airflow pipeline / translated from English by D. A. Belikov. Moscow: DMK Press, 2021. pp. 502. ill. ISBN 978-5-97060-970-5 — Text: direct.
-