МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ
МГПУ

Лекция 2.
Инструменты интеграции данных

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ
МГПУ

https://www.softwaretestinghelp.com/tools/26-best-data-integration-tools/

Какой бывает аналитика

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

- **Продуктовая аналитика.**

Нужна, чтобы улучшать продукт. Продуктовая аналитика собирает данные, которые помогают изучать поведение пользователей во время их взаимодействия с продуктом. Например, производителю важно знать, как часто пользуются его продуктом, какие проблемы при этом возникают, какую пользу от использования получает клиент.

Продуктовая аналитика: пошаговая инструкция по сбору и визуализации данных

Какой бывает аналитика

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

- **Маркетинговая аналитика.**

Нужна, чтобы оценивать эффективность маркетинговых и рекламных кампаний. Такая аналитика собирает данные из рекламных каналов и CRM. С её помощью определяют, с какой рекламной кампании пришёл пользователь, купил продукт или нет, сделал это сразу или через какое-то время и т. д. Работа с данными маркетинговой аналитики помогает понять, почему пользователи покупают или не покупают продукт, какой бюджет нужен для рекламной кампании, что нужно изменить на сайте, в работе отдела продаж или логистике.

Какой бывает аналитика

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

- **BI-аналитика (Business Intelligence-аналитика).**

Нужна, чтобы собирать, хранить, анализировать, обрабатывать и наглядно представлять все данные, которые есть в компании. BI-аналитика помогает собирать данные из разных источников, разрабатывать и подтверждать гипотезы, моделировать возможные решения. Компании, которые используют BI-аналитику, могут анализировать операционные расходы, прогнозировать доходы, сегментировать целевую аудиторию по разным признакам и т. д.

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ
МГПУ

Инструменты аналитики

Инструменты аналитики

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

- **Для сбора и хранения данных.** В любой компании есть своя база данных. В одной это могут быть таблицы **Excel**, в другой — серьёзные решения типа **Oracle** или **MySQL, Postgre**. Задача этих инструментов бизнес-анализа — хранить большие объёмы данных и быстро извлекать их.

Инструменты аналитики

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

- **Для анализа данных.** Чтобы собранные данные не лежали мёртвым грузом, а работали, их нужно доставать из базы данных и анализировать по определённым критериям с помощью различных программ. Один из самых популярных инструментов для аналитики данных — **Jupyter Notebook**.

Инструменты аналитики

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

- **Для визуализации данных.** Информацию, которую получили после анализа данных, нужно представить в удобном и понятном виде. Чтобы создавать наглядные графики и отчёты, используют программы и сервисы для **визуализации**. К простым относятся **Power Point** или **Miro**. Более сложные инструменты работы с аналитикой — **Tableau**, **Power BI**.

Инструменты аналитики

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

- **Для прогнозирования данных.** Такие инструменты нужны, чтобы на основании прошлого опыта компании могли принимать успешные решения в будущем, создавать модели поведения клиентов, составлять прогнозы ежедневного спроса определённой группы товаров и т. д. Чтобы создавать достоверные прогнозы, специалисты используют ключевые инструменты аналитиков: языки программирования **Python**, **R** и другие.

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ
МГПУ

Какими инструментами должен владеть аналитик данных

Какими инструментами должен владеть аналитик данных

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

- **Основные** инструменты аналитика помогают ему собирать, обрабатывать, анализировать и интерпретировать данные.
- **Несмотря на** большое количество сервисов и программного обеспечения, на практике специалист использует в работе **3–4 ключевых инструмента**. Их выбор зависит не только от знаний и опыта аналитика, но и от того, с чем уже работает компания.

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ
МГПУ

Онлайн обработка аналитических запросов (OLAP)

Онлайн обработка аналитических запросов (OLAP)

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

- **В обычной**, «строковой» СУБД, данные хранятся в таком порядке:

a	WatchID	JavaEnable	Title	GoodEvent	EventTime
#0	89354350662	1	Investor Relations	1	2016-05-18 05:19:20
#1	90329509958	0	Contact us	1	2016-05-18 08:10:20
#2	89953706054	1	Mission	1	2016-05-18 07:38:00
#N	…	…	…	…	…

То есть, значения, относящиеся к одной строке, физически хранятся рядом.
Примеры строковых СУБД: **MySQL, Postgres, MS SQL Server**.

Онлайн обработка аналитических запросов (OLAP)

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

- **В столбцовых** СУБД, данные хранятся в таком порядке:

Строка:	#0	#1	#2	#N
WatchID:	89354350662	90329509958	89953706054	…
JavaEnable:	1	0	1	…
Title:	Investor Relations	Contact us	Mission	…
GoodEvent:	1	1	1	…
EventTime:	2016-05-18 05:19:20	2016-05-18 08:10:20	2016-05-18 07:38:00	…

В примерах изображён только порядок расположения данных. То есть, значения из разных столбцов хранятся отдельно, а данные одного столбца - вместе.

Примеры столбцовых СУБД: **Vertica, Paraccel (Actian Matrix, Amazon Redshift), Sybase IQ, Exasol, Infobright, InfiniDB, MonetDB (VectorWise, Actian Vector), LucidDB, SAP HANA, Google Dremel, Google PowerDrill, Druid, kdb+**

Онлайн обработка аналитических запросов (OLAP)

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

- **Сценарий работы с данными** - это то,

- какие производятся запросы, как часто и в каком соотношении;
- сколько читается данных на запросы каждого вида - строк, столбцов, байт;
- как соотносятся чтения и обновления данных;
- какой рабочий размер данных и насколько локально он используется;
- используются ли транзакции и с какой изолированностью;
- какие требования к дублированию данных и логической целостности;
- требования к задержкам на выполнение и пропускной способности запросов каждого вида.

Онлайн обработка аналитических запросов (OLAP)

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

Ключевые особенности OLAP сценария работы

- подавляющее большинство запросов - на чтение;
- данные обновляются достаточно большими пачками (> 1000 строк), а не по одной строке, или не обновляются вообще;
- данные добавляются в БД, но не изменяются;
- при чтении, вынимается достаточно большое количество строк из БД, но только небольшое подмножество столбцов;
- таблицы являются «широкими», то есть, содержат большое количество столбцов;
- запросы идут сравнительно редко (обычно не более сотни в секунду на сервер);
- при выполнении простых запросов, допустимы задержки в районе 50 мс;
- значения в столбцах достаточно мелкие - числа и небольшие строки (пример - 60 байт на URL);
- требуется высокая пропускная способность при обработке одного запроса (до миллиардов строк в секунду на один сервер);
- транзакции отсутствуют;
- низкие требования к консистентности данных;
- в запросе одна большая таблица, все таблицы кроме одной маленькие;
- результат выполнения запроса существенно меньше исходных данных - то есть, данные фильтруются или агрегируются; результат выполнения помещается в оперативку на одном сервере.

Онлайн обработка аналитических запросов (OLAP)

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

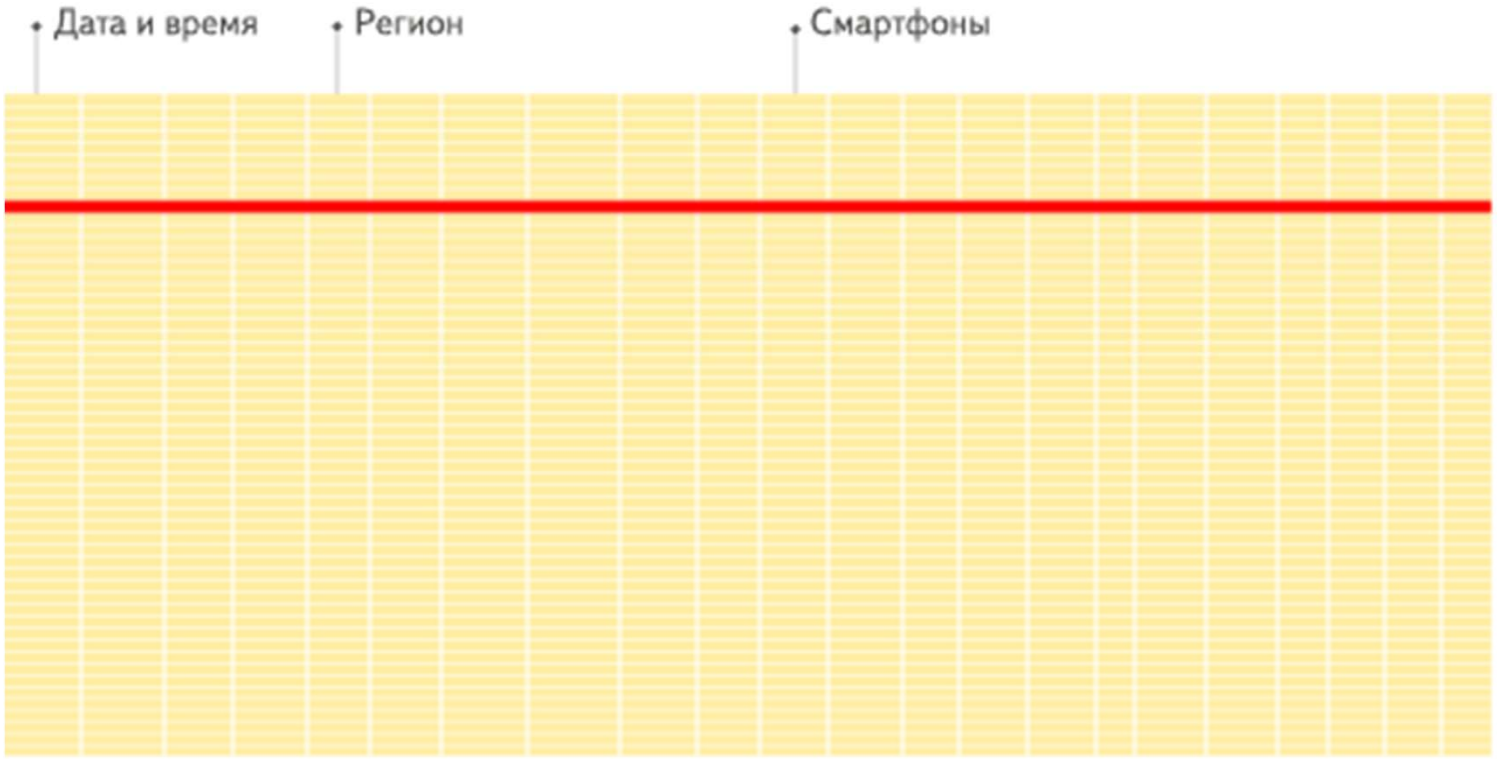
- OLAP сценарий работы существенно отличается от других распространённых сценариев работы (например, OLTP или Key-Value сценариев работы).

- Не имеет никакого смысла пытаться использовать OLTP или Key-Value БД для обработки аналитических запросов, если вы хотите получить приличную производительность.

Причины, по которым столбцовые СУБД лучше подходят для OLAP сценария

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

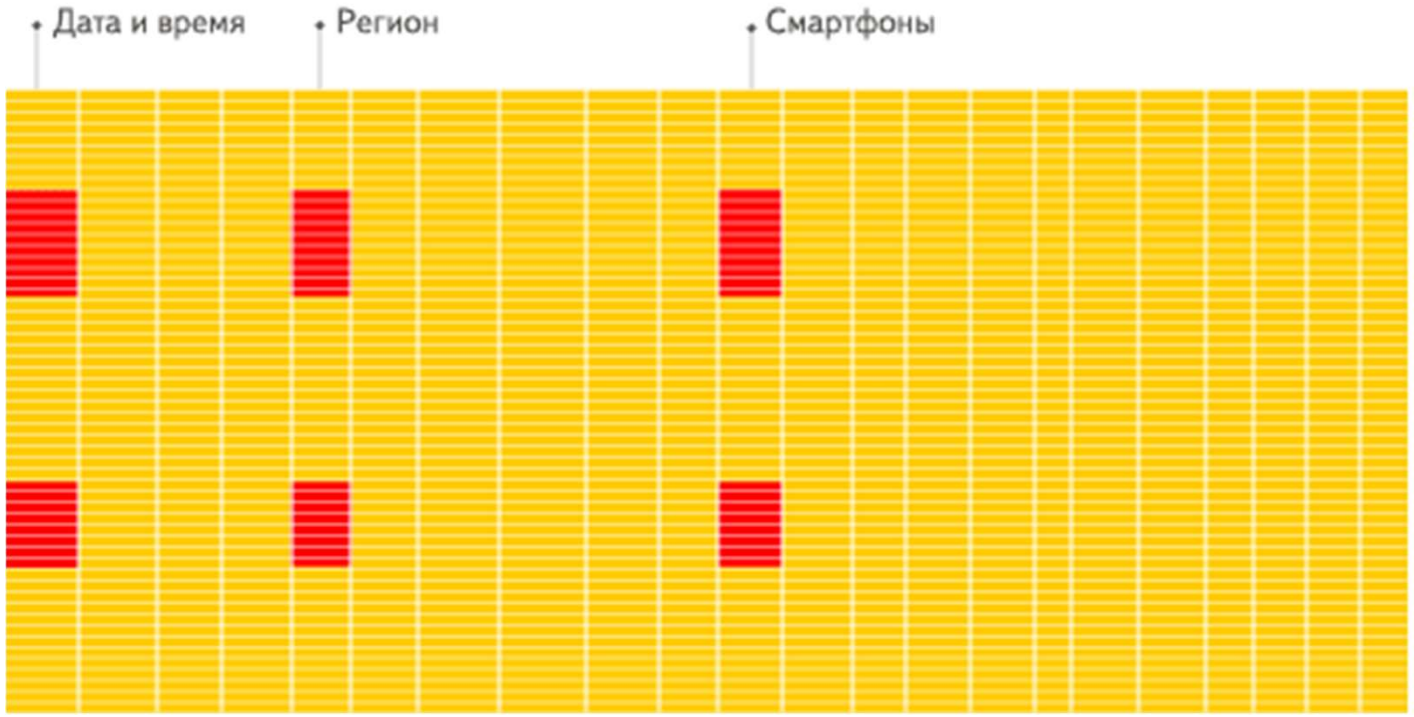
Строковые СУБД



Причины, по которым столбцовые СУБД лучше подходят для OLAP сценария

МОСКОВСКИЙ ГОРОДСКОЙ УНИВЕРСИТЕТ МГПУ

Столбцовые СУБД



МОСКОВСКИЙ
ГОРОДСКОЙ
УНИВЕРСИТЕТ
МГПУ

СПАСИБО ЗА ВНИМАНИЕ