



# Обзор систем для построения аналитических хранилищ данных

Максим Стаценко  
Ведущий разработчик

# Обо мне



**Максим Стаценко**

Ведущий разработчик, Yandex

Опыт в Big Data с 2010 года в различных областях:  
энергетические компании, образование, банки,  
IT, научные исследования

# О чем и зачем этот доклад?



- › Рассказать о технологиях, о которых говорят меньше, чем о хайповых технологиях
- › Поделиться опытом, полученным в рамках PoC на большом объеме данных
- › Поделиться опытом принятия решений, как и где строить хранилище

# План

## Выбор между облаком и Bare Metal

- › Почему мне облака нравятся больше
- › Почему облака в РФ в моих глазах выглядят удобнее
- › PoC и как облако с этим поможет

## Технологии, когда не хватает SMP БД

- › MPP
- › Hadoop
- › ClickHouse
- › Greenplum
- › Teradata
- › Exasol
- › Vertica

## Заключение

- › Нужно пробовать и не бояться комбинировать

# Опрос



**Есть ли у вас аналитическое хранилище данных?**

- › Да
- › Нет

**Какую функцию вы выполняете в компании?**

- › Data Engineer
- › Data Architect
- › Data Officer
- › Analyst
- › Manager
- › Other

# Что мы строим?



- › Строим систему, в которой будут храниться данные для анализа
- › Данных много, классические SMP БД уже не справляются
- › Новые данные появляются каждый день
- › Данные доступны 24/7
- › Доступ к данным разграничен
- › К данным можно подключить BI-инструмент

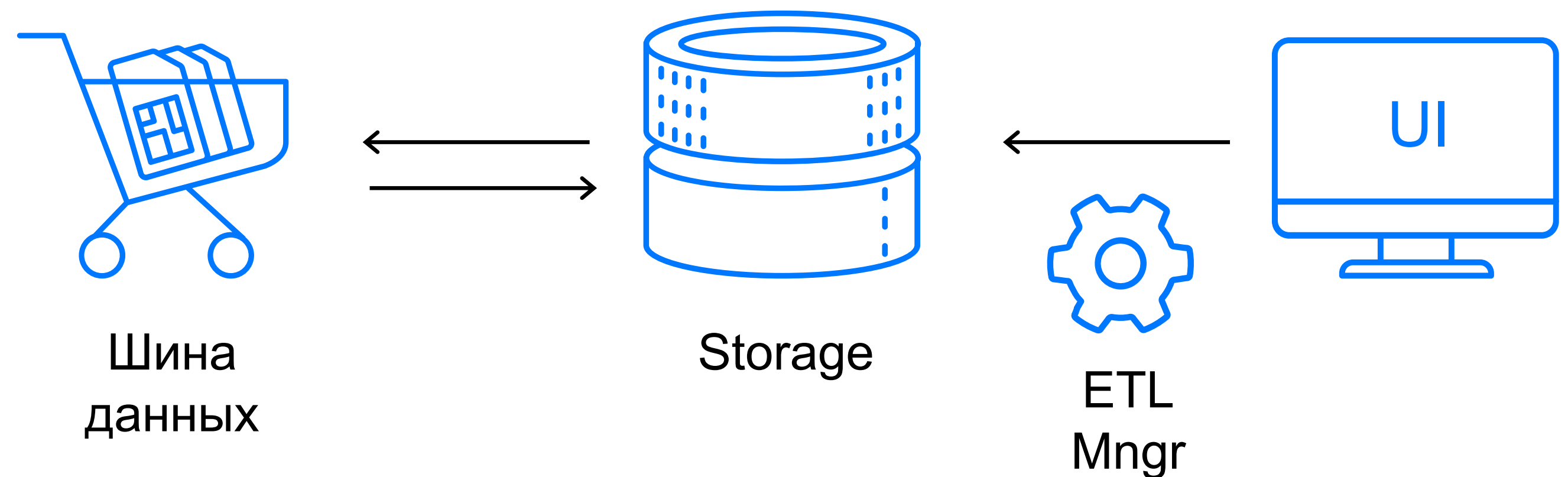
# Специфика современности



- › Количество данных растет быстрее, чем вычислительные мощности
- › Угроза безопасности чувствительных данных
- › SQL, Python/R — де-факто стандарт интерфейса доступа к данным
- › Машинное обучение на данных — это важная и большая часть анализа данных
- › Постоянно у людей появляются новые (внешние) данные, которые нужно проанализировать вместе со старыми (внутренними)
- › Fail Fast — множество дешевых экспериментов, гипотез

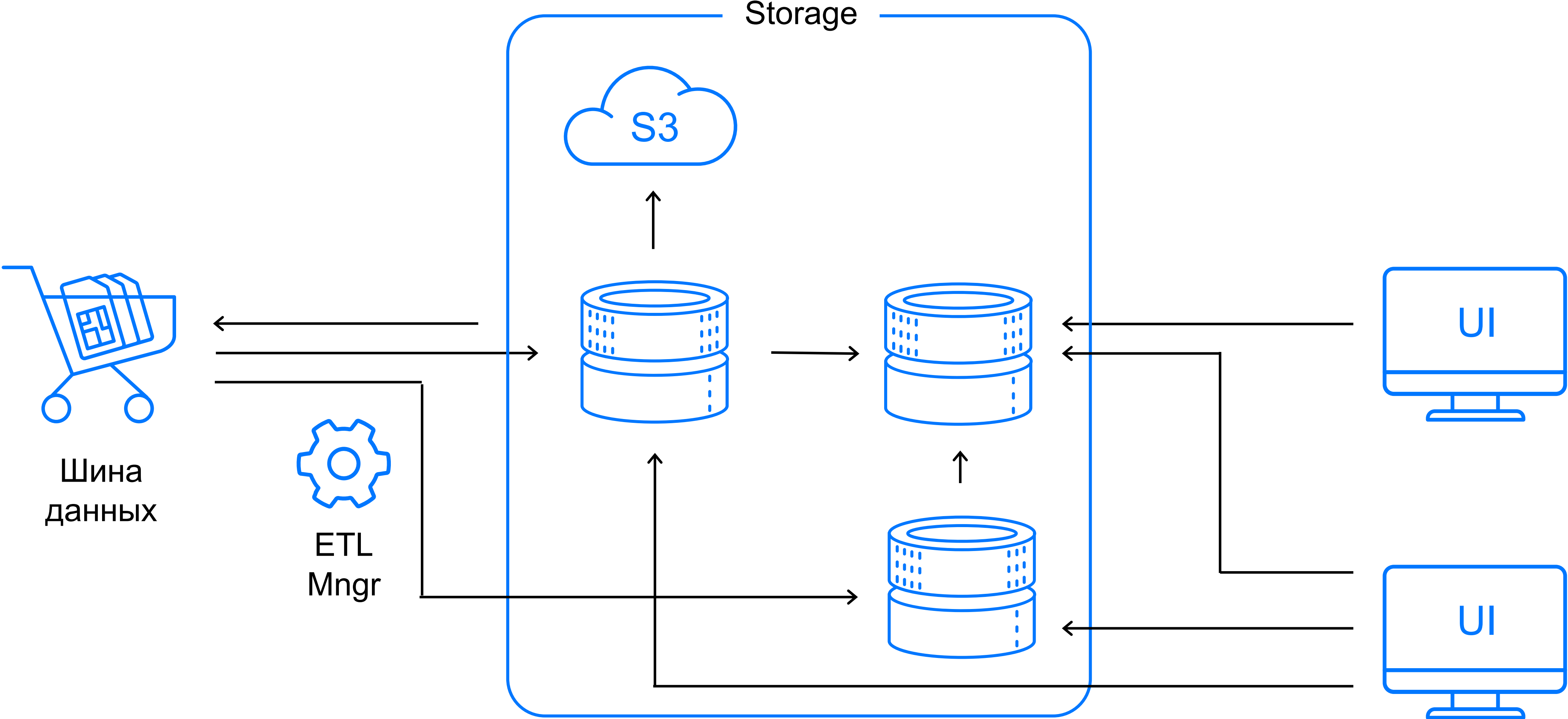
# Мы будем говорить о Storage

- › **Хранение данных**
- › Логическое хранение данных
- › Пользовательский интерфейс
- › ETL Manager
- › Шина данных





# Storage может быть сложнее:



# Варианты, где строить



- › «Сервер в кладовке»
- › Аренда стойки в ДЦ
- › Аренда сервера
- › Аренда виртуальной машины в облаке
- › Managed Services в облаке

# Опрос



**Где вы располагаете свою инфраструктуру хранилища?**

- › «Сервер в кладовке»
- › Аренда стойки в ДЦ
- › Аренда сервера
- › Аренда виртуальной машины в облаке
- › Managed Services в облаке

# Таблица

\$ — скрытые дополнительные траты

|                         | Нужен админ | Нужен setup                        | Сложные обновления | Поддержка ПО | Безопасность                  | Доступность              | Мониторинги          | Реакция на аварии | Fine tuning | Локация данных     | Тип затрат | Закупка, расчет, скидка на лицензии ПО | Быстрое изменение количества ресурсов |
|-------------------------|-------------|------------------------------------|--------------------|--------------|-------------------------------|--------------------------|----------------------|-------------------|-------------|--------------------|------------|--|---------------------------------------|
| <b>Сервер целиком</b>   | Да (\$)     | Ручной (\$)                        | Руками (\$)        | На вас (\$)  | Любой кастом                  | На вас                   | Все на вас (\$)      | Ваша (\$)         | Любой       | У вас              | Сарех      | На вас (\$)                            | Закупка (\$)                          |
| <b>«Виртуалка»</b>      | Да (\$)     | OS + база стоит автоматически (\$) | Руками (\$)        | На вас (\$)  | Часть вендор определил за вас | Частично решено вендором | Часть готова за вас  | Вендор облака     | Почти любой | Локация ДЦ вендора | Орех       | На вас, иногда вендор облака (\$)      | По кнопке                             |
| <b>Managed Services</b> | Нет         | Автоматически OS + ПО              | По кнопке          | Вендор       | Стандарты вендора             | Решено вендором          | Основные уже сделаны | Вендор облака     | База        | Локация ДЦ вендора | Орех       | Вендор облака                          | По кнопке                             |

# Мое мнение

## Плюсы железа, аренды сервера

Проще адаптировать под требования ИБ, ФСБ и т. д.

Чувствительные данные в вашей инфраструктуре

Все ресурсы ваши

Пока еще привычнее

## Плюсы виртуалок

За вас применены основные best practices настройки доступа и безопасности

Простота экспериментов

За вас делают большую часть рутины (поддержка, настройка базовой части, реакция на аварии)

Гибкость

# FYI: Managed Services — МОЯ НОВАЯ ЛЮБОВЬ

**За вас всё сделали, вам только надо указать, сколько нужно RAM, CPU, DISK, и дальше само:**

- › Установится в оптимальной конфигурации и обновится нажатием кнопки
- › Новые ресурсы добавляются одним нажатием кнопки
- › Встроены мониторинги
- › Реакции на аварии по инфраструктуре не на вас
- › Основные настройки в интерфейсе
- › Баги софта чинят за вас

**Для меня, пока система не стала очень специфичной и сложной, это идеальный вариант. Каждый параметр системы не будет настраиваться кастомно, а будут использованы best practices**

# FYI: проблемы облака за пределами РФ

## Постановление Правительства Российской Федерации от 26.06.2018 № 728

- › Храним данные пользователей на территории РФ

## № 152-ФЗ

- › Нормативы, как безопасно хранить персональные данные
- › Трансграничная передача персональных данных на территории иностранных государств... может быть запрещена или ограничена

## Не везде есть русскоязычная поддержка

- › Повышаются требования и стоимость специалистов, которые будут работать с облаком

# FYI: проблемы облака за пределами РФ

## Валютные риски

- › Стоимость услуг зависит от курса рубля. Стоимость инфраструктуры может удвоиться, как это случилось, когда курс доллара вырос в два раза

## Сеть

- › Если надо заливать, выгружать большие объемы данных, то нужен хороший канал. Чем больше точек связности между двумя машинами — тем меньше вероятность, что будет хороший канал
- › Стоимость канала до европейского ЦОД может быть равна стоимости ресурсов для проекта DWH в облаках РФ



# Опрос

Какой программный продукт используете для DWH?

- › Hadoop
- › ClickHouse
- › Greenplum
- › Teradata
- › Vertica
- › Exasol
- › Другое

Какую должность предполагается заменить AI?

- › Аналитик
- › Data Engineer
- › DB Admin
- › Data Architect

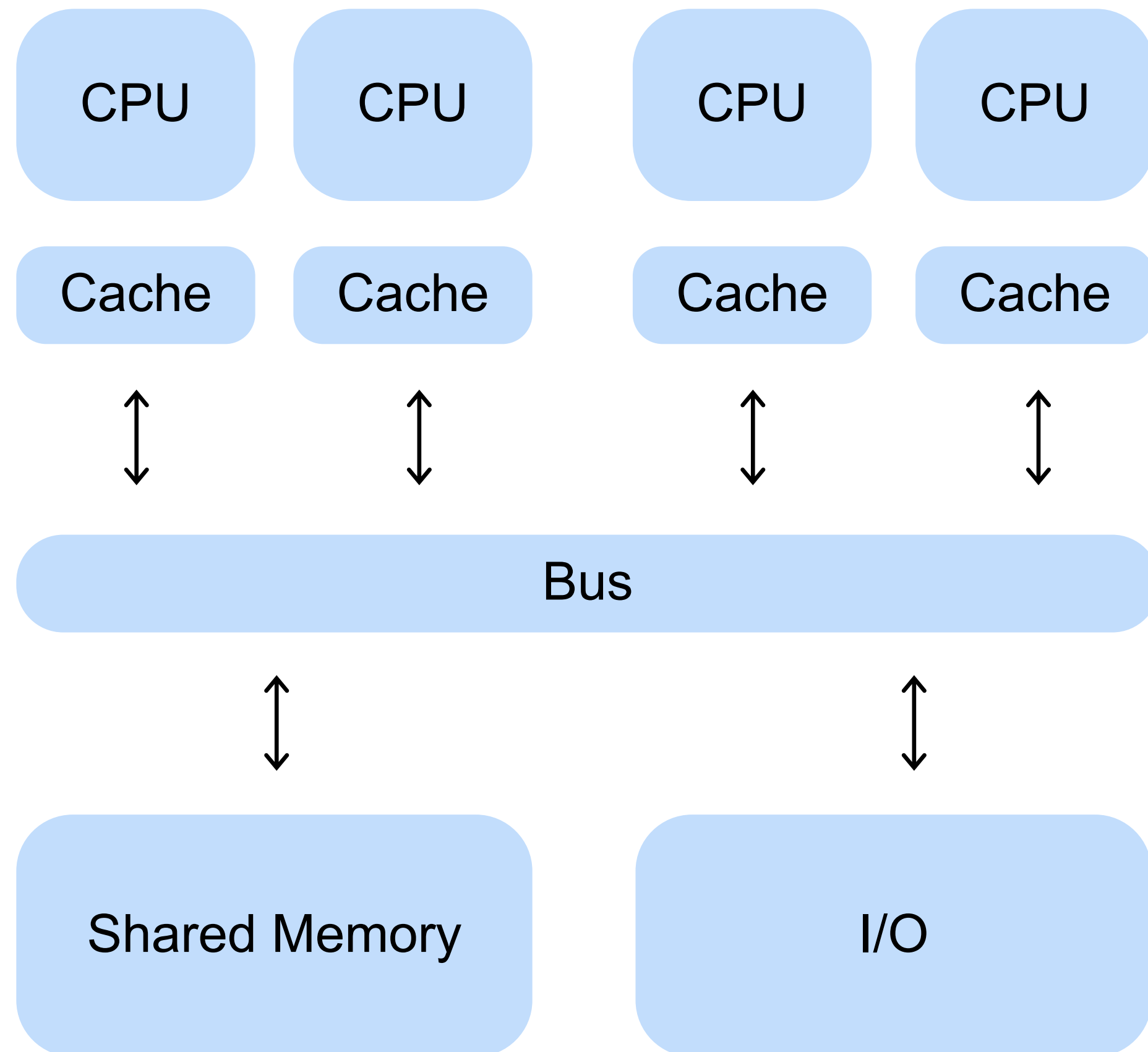
# Выбор программного продукта



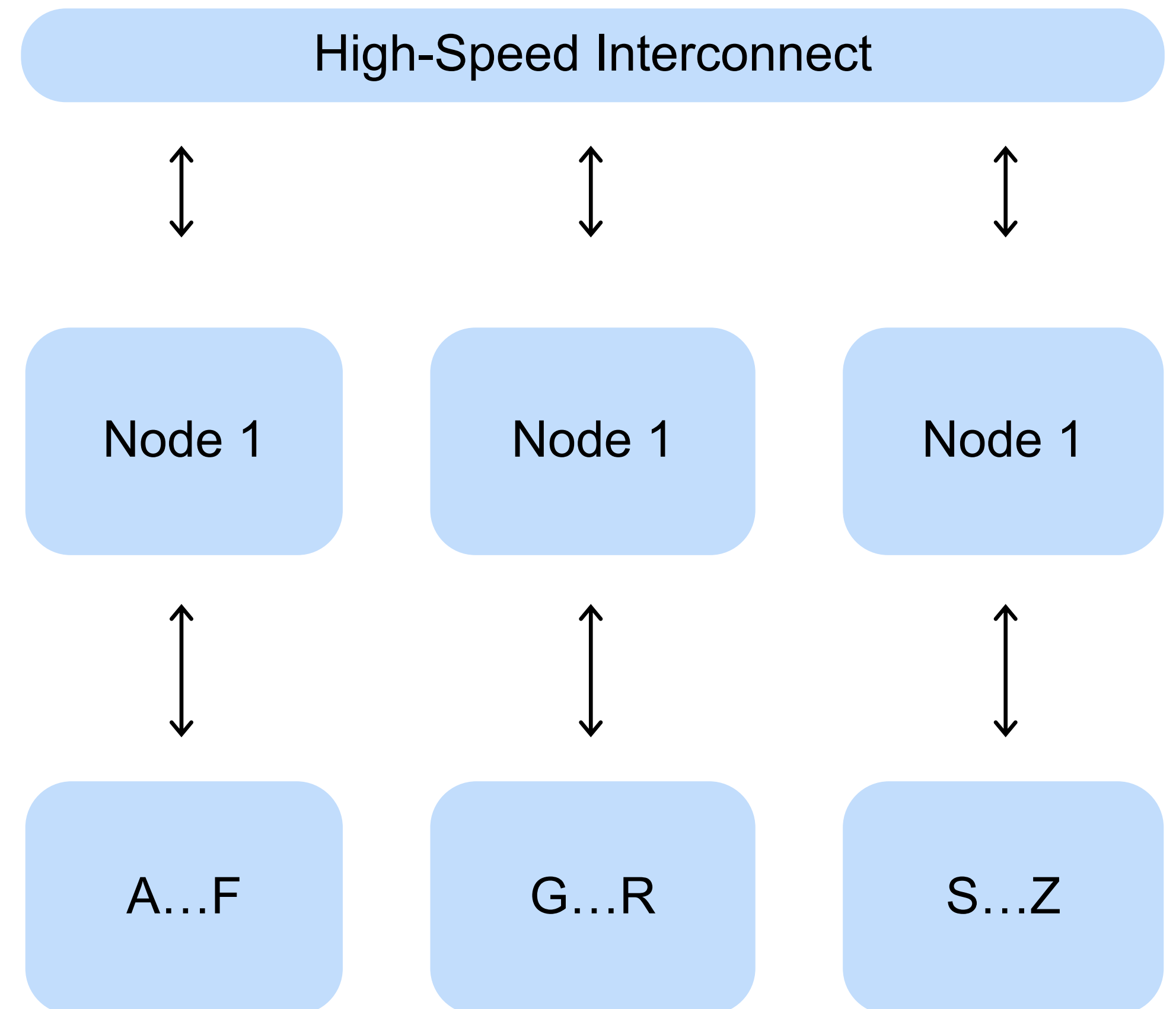
- › Нельзя ориентироваться на цену.  
У продукта могут быть большие накладные расходы для вашего сценария использования
- › Свяжитесь с представительствами и проводите PoC. Оцените, насколько это решение подходит именно вам, под ваши реалии, под ваши задачи

# MPP

## MPP vs SMP Architecture



**VS**



# На что будем смотреть у каждого решения



- › SQL — порог входа, простота для пользователей, стоимость найма специалистов/аналитиков
- › Транзакции — способность базы сохранять логическую корректность данных (пример со списанием денег со счета или зачислением на счет)
- › Все серверы (ноды) равны, или есть выделенная — устойчивость и масштабируемость. Наличие узкого места в архитектуре
- › Платное, бесплатное, есть бесплатный тест

# Hadoop и Co



- › ANSI SQL (Spark, Hive, Impala), ACID TRANSACTIONS (Impala, ~Hive), есть выделенная нода, бесплатно/платно
- › «Как хочешь» Store
- › Сложные типы
- › ML из коробки
- › UDF
- › Open Source — больно либо платно
- › Много информации
- › On-premise
- › Облака

# Disclaimer — антиреклама Hadoop

Я очень люблю Hadoop, но хорошо его приготовить сложно, **больно**, долго и дорого.

- › Безопасность = Kerberos = **Боль**
- › Vanilla Hadoop = Баги = **Боль** (например, в какой-то момент перестали работать JOIN field1 = COALESCE(field2, 0))
- › Разделение ресурсов = **Боль** (очереди YARN, отдельные инстансы Spark)
- › Удобный VI с низким RT = **Боль**
- › Работа с Hadoop + JDBC одновременно = **Боль**
- › NameNode — узкое место, ее может положить любой, и это **Боль**

# Но Hadoop хороший



- › Если осторожно, то Hadoop — хороший выбор, главное — не ждать от него многого
- › Для работы в Hadoop нужна подготовка. Аналитик с опытом MySQL не станет писать крутые запросы в Hadoop, но зато, когда обучится, он сможет обработать огромный массив данных
- › Изучив SPARK, получаешь возможность обучать модели ML распределенно без дополнительных усилий
- › Очень много информации, курсов, инструментов

# Сводная таблица того, что я пробовал

| Vendor     | SQL         | Транзакции | Все ноды равны | Коннекторы к другим БД | Storage Type | Compression | UDF                  | Complex Data Types |
|------------|-------------|------------|----------------|------------------------|--------------|-------------|----------------------|--------------------|
| Hadoop     | ANSI+       | +/-        | -              | -/+                    | Any          | +           | Java, Scala, Python  | +                  |
| ClickHouse | Расширенный | -          | +              | +                      | Column       | +           | -                    | +                  |
| Greenplum  | ANSI+       | +          | -              | +                      | Hybrid       | +           | PL/SQL, Python, R    | -                  |
| Teradata   | ANSI+       | +          | +              | +                      | Hybrid       | +           | SQL, C, C++, Java    | +                  |
| Exasol     | ANSI+       | +          | +              | +                      | Column       | +           | Lua, Java, Python, R | -                  |
| Vertica    | ANSI+       | +          | +              | +                      | Column       | +           | Python, R, C++, Java | -                  |



# Сводная таблица того, что я пробовал

| Vendor     | ML in ecosystem | Внутреннее хранение в Object Storage (S3) | Развертывание on-premise | Развертывание в облаке | Подключение к инструментам визуализации данных | Развертывание в облаках российских вендоров | Managed Service в российских вендорах | Cost                          |
|------------|-----------------|---|--------------------------|------------------------|--|---|---------------------------------------|-------------------------------|
| Hadoop     | +               | –   | +                        | +                      | +  | +   | +                                     | Free or \$ Open Source        |
| ClickHouse | +               | +   | +                        | +                      | +  | +   | +                                     | Free Open Source              |
| Greenplum  | ?               | –   | +                        | +                      | +  | +   | +                                     | Free Open Source or \$        |
| Teradata   | +               | +   | + (VMware)               | +                      | +  | –   | –                                     | \$                            |
| Exasol     | –               | +   | + (своя OS)              | +                      | +  | –   | –                                     | \$ Free (1 node, 200 GB data) |
| Vertica    | +               | +   | +                        | +                      | +  | –   | –                                     | \$ Free (3 node, 1 TB data)   |

# Следующие восемь слайдов — мои ассоциации с продуктами и лирика о них



- › Показать в часовой презентации все продукты, которые я упомянул, я не могу
- › Я попробую описать то, что ассоциируется у меня в голове с каждым из этих продуктов
- › Информацию о пользователях каждой системы я взял из открытых источников — Habr, Medium и т. д.

# ClickHouse

Первая публичная версия — 2016 год



## Пользователи:

- › Яндекс.Метрика [clickhouse.tech/docs/en/introduction/history/](https://clickhouse.tech/docs/en/introduction/history/)
- › ВКонтакте [highload.ru/siberia/2018/abstracts/3614](https://highload.ru/siberia/2018/abstracts/3614)
- › Cloudflare [blog.cloudflare.com/http-analytics-for-6m-requests-per-second-using-clickhouse/](https://blog.cloudflare.com/http-analytics-for-6m-requests-per-second-using-clickhouse/)

# Си-плюс-плюсно — быстро и векторно



- › Код запросов для ClickHouse — ограниченный ANSI SQL с большим количеством дополнительных возможностей (например, lambda-функции над массивами: **SELECT** arrayMap(x -> (x + 2), [1, 2, 3]) **as** res)
- › Запросы написать плохо сложнее, а оптимизатору проще максимально распараллеливать и векторизовать запросы
- › Реализовано на C++ опытными разработчиками, как результат — я знаю людей, которые начали использовать ClickHouse потому, что он очень быстро парсит JSON

# Greenplum

В Open Source с 2015 года.  
Предыдущие версии с 2005 года



## Пользователи:

- › Тинькофф Банк [habr.com/ru/company/tinkoff/blog/267733](https://habr.com/ru/company/tinkoff/blog/267733)
- › Сбер [habr.com/ru/company/sberbank/blog/509936/](https://habr.com/ru/company/sberbank/blog/509936/)
- › Leroy Merlin [tech.leroymerlin.ru/stack/](https://tech.leroymerlin.ru/stack/)

# Greenplum — много больших PostgreSQL



- › Greenplum основан на PostgreSQL
- › В ядре каждой ноды лежит PostgreSQL, будь то Master, будь то DataNode
- › Следствие — почти все, что работает с PostgreSQL, будет работать и с Greenplum. Это касается и многих надстроек
- › Многие сложности хорошо известны, так же как и известны их решения

# Teradata

На рынке МРР-систем с 1984 года

The Teradata logo consists of the word "teradata" in a lowercase, sans-serif font. The letter "o" is replaced by a solid orange circle.

Пользователи:

- › ВТБ [habr.com/ru/article/348534/](https://habr.com/ru/article/348534/)
- › Сбер [habr.com/ru/company/sberbank/blog/427437/](https://habr.com/ru/company/sberbank/blog/427437/)
- › BMW [forbes.com/sites/bernardmarr/2017/08/01/how-bmw-uses-artificial-intelligence-and-big-data-to-design-and-build-cars-of-tomorrow/](https://forbes.com/sites/bernardmarr/2017/08/01/how-bmw-uses-artificial-intelligence-and-big-data-to-design-and-build-cars-of-tomorrow/)

# Teradata — экосистема



- › SQL Engine — храним классические таблицы в формате Row или Column
- › TimeSeries Engine — храним и обрабатываем данные TimeSeries
- › ML Engine — набор функций, которые позволяют делать базовое ML, не покидая клиент базы
- › Graph Engine — набор функций для работы с графами
- › Коннекторы к самым разным источникам и оркестратор



# Exasol

На рынке с 2000 года

The logo for Exasol, featuring the word "Exasol" in a dark blue, sans-serif font. The letter "x" is highlighted in a light green color.

**Пользователи:**

- › Ситимобил [tadviser.ru/index.php/Проект:Ситимобил \(Exasol\)](http://tadviser.ru/index.php/Проект:Ситимобил_(Exasol))
- › Badoo [habr.com/ru/company/badoo/blog/271753/](http://habr.com/ru/company/badoo/blog/271753/)

# Магия Exasol



- › Правильные индексы — залог быстрой работы базы данных
- › Какие индексы ускорят работу базы — зависит от запросов к базе
- › Создать все индексы нельзя — будет медленная вставка, займет много места
- › Запросы к базе меняются со временем
- › Алгоритмы (или AI, как сейчас модно говорить) строят новые и удаляют старые индексы за вас

# Vertica

С 2005 года на рынке

VERTICA

Клиенты:

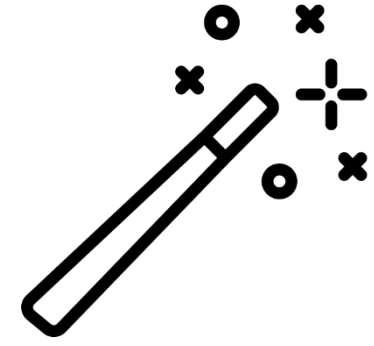
- › Avito [habr.com/ru/company/avito/blog/322510](https://habr.com/ru/company/avito/blog/322510)
- › СИБУР [habr.com/ru/company/sibur\\_official/blog/436632/](https://habr.com/ru/company/sibur_official/blog/436632/)
- › Uber [eng.uber.com/uber-big-data-platform/](https://eng.uber.com/uber-big-data-platform/)

# Vertica — надежно и прозрачно

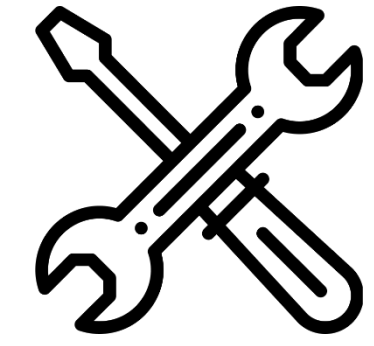
The logo for Vertica, featuring the word "VERTICA" in a bold, black, sans-serif font. The letter "V" is stylized with a diagonal slash through it.

- › Прозрачные и хорошо освещенные способы хранения и оптимизации: ROS Containers, Проекции, Merge Join, Bloom Filters
- › Понятная логика оптимизатора запросов
- › В системе можно настроить все процессы, и, если что-то буксует, понятно, что подкрутить
- › Ощущение не магии, а надежных механических швейцарских часов

Exasol



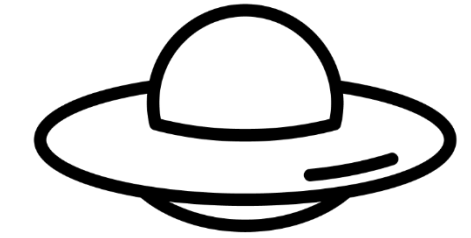
teradata.



ClickHouse



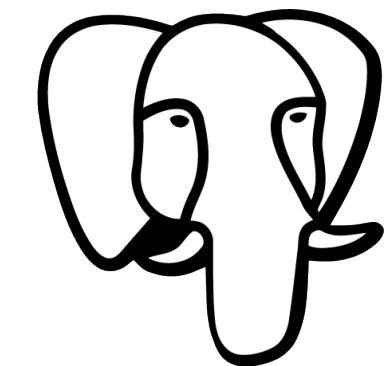
 **hadoop**



VERTICA



 **GREENPLUM  
DATABASE®**



# Есть еще решения



Я рассказал о тех частях, с которыми я работал

Еще есть:

- › MongoDB
- › Exadata
- › SAP HANA
- › Druid
- › MemSQL
- › Tarantool
- › Qlik

# Общаясь с представителем, вы можете...



- › Получить скидку
- › Получить триальную версию для PoC
- › Получить best practices и ответы на вопросы
- › Узнать других пользователей системы и поговорить об их опыте
- › Узнать о слабых сторонах системы

# Обсуждаем результаты опроса



# Логическое хранение данных — важно



- › Адаптируйте архитектуру под продукт и задачи
- › Для каждого продукта подходят разные схемы хранения: широкие таблицы, звезда/снежинка, Data Vault, Anchor, 3НФ и т. д.

## Мои success story:

- › Oracle, PostgreSQL — 3НФ
- › ClickHouse — ши-и-ирокие таблицы
- › Vertica, Exasol — Data Vault
- › Hadoop, Teradata — денормализованные таблицы (но близко к 3НФ)

# Мы выбрали, где хранить. Что теперь?



## Выбрать Spare Parts

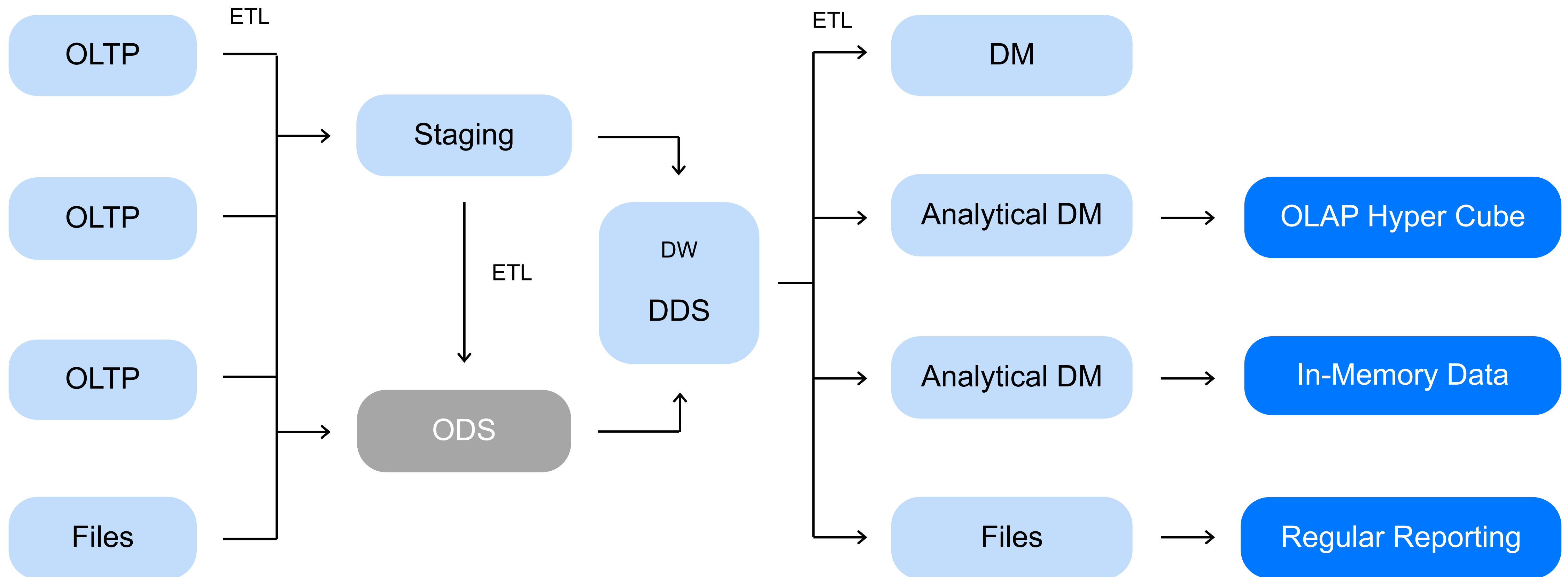
1. Выбрать шину данных (часть хранилища данных, предназначенная для загрузки и выгрузки данных в хранилище). Кажется, что сейчас это лишнее, но потом вы скажете себе спасибо
2. Выбрать UI
3. Выбрать ETL Manager

# Не бойтесь совмещать технологии. Это тренд

MDM, CDC, Data Governance, Common Metamodels and Metadata

Operational Data

Reporting Data



# Мораль



**Если у вас нет DWH, то простой MVP можно создать в облаке, на любом из описанных продуктов, «в пару кликов». И это уже будет лучше, чем совсем без аналитического хранилища**

**Если вас не удовлетворяет ваше DWH, попробуйте посмотреть вокруг. Может, появилось что-то, что улучшит пользовательский опыт**



# Спасибо

**Максим Стаценко**

Ведущий разработчик

 [maxstatsenko@yandex-team.ru](mailto:maxstatsenko@yandex-team.ru)

 @архумар