



# Аналитическая платформа Postgres Pro AXE

Андрей Кулага

Руководитель разработки

# План

1. Предпосылки

2. Цели продукта

3. Аналитическая платформа

4. Тесты производительности  
и примеры использования

5. Дорожная карта

# 1. Предпосылки



# Хранилища данных

Большие распределённые системы с медленно меняющимися данными и длинными запросами (Postgres не подходит)

2000

Классические  
shared-nothing MPP системы  
(greenplum, vertica)

Проще в использовании, но

- Закрытые
- Масштабирование ограничено



2020

Конструкторы Lakehouse

Сложные, но открытые:

- S3 и Parquet для хранения данных
- Каталог (Iceberg, Delta) для представления набора файлов в виде таблиц
- Использование вычислительных движков (Spark, Trino) для SQL запросов
- Неограниченная масштабируемость и отсутствие vendor-lock

# Вопросы

Сложно в поддержке,  
настройке, использовании

Большие задержки и  
накладные расходы при  
изменении данных

# Lakehouse в экосистеме Postgres

## Сейчас

- Большие задержки
- Накладные расходы
- Дополнительные инструменты



Postgres  
OLTP

ETL



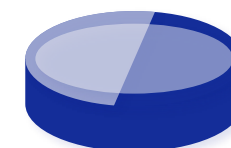
Хранилище  
данных

## ХОТИМ

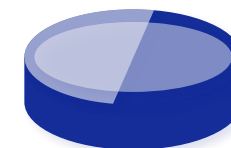
- Низкие задержки и накладные расходы
- Быстрая аналитика над данными реального времени
- Один продукт и экосистема Postgres



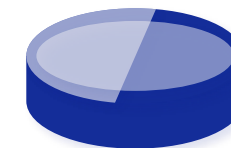
Postgres  
OLTP + OLAP



NVME



NFS

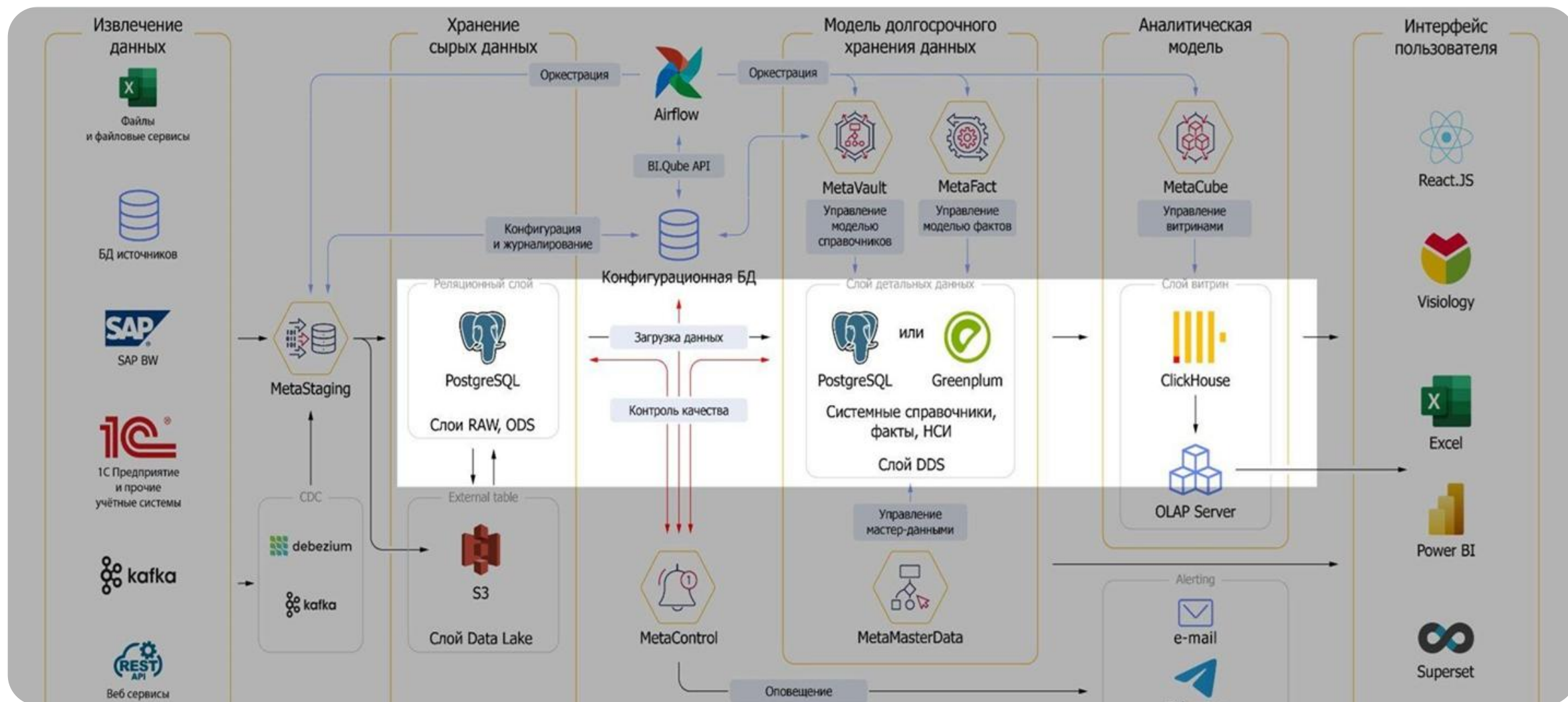


S3

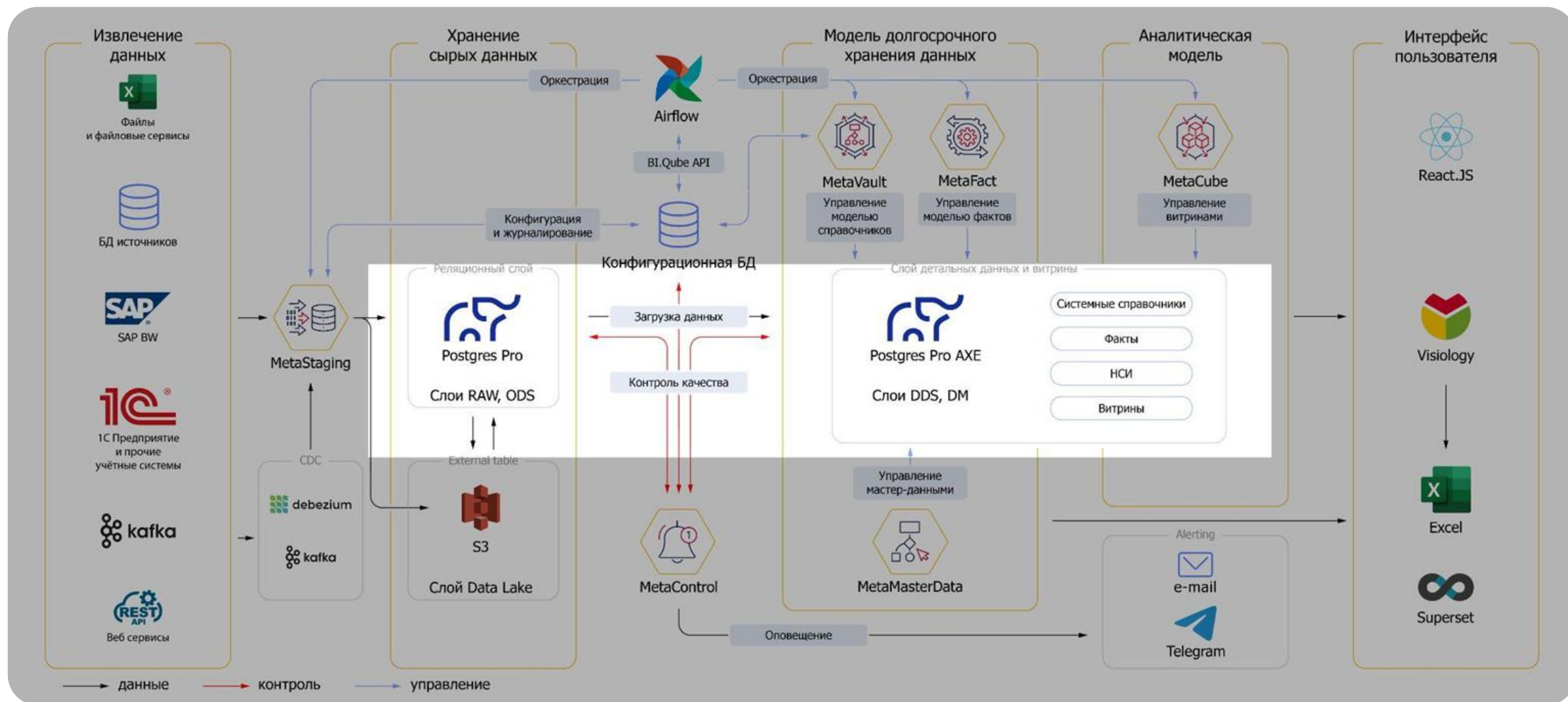


Каталог

# Типичный стек аналитики



# Стек аналитики (с Postgres Pro AXE)



## 2. Цели



# Какой продукт сделали

1. Полностью интегрированный с экосистемой Postgres Pro Enterprise инструмент для построения хранилищ данных

2. Работает из коробки на той же инфраструктуре и железе для текущих OLTP клиентов. Переиспользование экспертизы команд, кода и бизнес-логики приложений

3. Быстрый и масштабируемый (не хуже Clickhouse, Umbra, DuckDB)

4. Данные реального времени, позволяет одновременно выполнять OLTP и OLAP сценарии

5. Масштабируемый для данных любых размеров. Выбор оптимальной системы хранения под нагрузку и задачи.

6. Открытые форматы LakeHouse, данные доступны для внешних систем

# Какой продукт сделали

1. Полностью интегрированный с экосистемой Postgres Pro Enterprise инструмент для построения хранилищ данных

2. Работает из коробки на той же инфраструктуре и железе для текущих OLTP клиентов. Переиспользование экспертизы команд, кода и бизнес-логики приложений

3. Быстрый и масштабируемый (не хуже Clickhouse, Umbra, DuckDB)

4. Данные реального времени, позволяет одновременно выполнять OLTP и OLAP сценарии

5. Масштабируемый для данных любых размеров. Выбор оптимальной системы хранения под нагрузку и задачи.

6. Открытые форматы LakeHouse, данные доступны для внешних систем

# Какой продукт сделали

1. Полностью интегрированный с экосистемой Postgres Pro Enterprise инструмент для построения хранилищ данных

2. Работает из коробки на той же инфраструктуре и железе для текущих OLTP клиентов. Переиспользование экспертизы команд, кода и бизнес-логики приложений

3. Быстрый и масштабируемый (не хуже Clickhouse, Umbra, DuckDB)

4. Данные реального времени, позволяет одновременно выполнять OLTP и OLAP сценарии

5. Масштабируемый для данных любых размеров. Выбор оптимальной системы хранения под нагрузку и задачи.

6. Открытые форматы LakeHouse, данные доступны для внешних систем

# Какой продукт сделали

1. Полностью интегрированный с экосистемой Postgres Pro Enterprise инструмент для построения хранилищ данных

2. Работает из коробки на той же инфраструктуре и железе для текущих OLTP клиентов. Переиспользование экспертизы команд, кода и бизнес-логики приложений

3. Быстрый и масштабируемый (не хуже Clickhouse, Umbra, DuckDB)

4. Данные реального времени, позволяет одновременно выполнять OLTP и OLAP сценарии

5. Масштабируемый для данных любых размеров. Выбор оптимальной системы хранения под нагрузку и задачи.

6. Открытые форматы LakeHouse, данные доступны для внешних систем

# Какой продукт сделали

1. Полностью интегрированный с экосистемой Postgres Pro Enterprise инструмент для построения хранилищ данных

2. Работает из коробки на той же инфраструктуре и железе для текущих OLTP клиентов. Переиспользование экспертизы команд, кода и бизнес-логики приложений

3. Быстрый и масштабируемый (не хуже Clickhouse, Umbra, DuckDB)

4. Данные реального времени, позволяет одновременно выполнять OLTP и OLAP сценарии

5. Масштабируемый для данных любых размеров. Выбор оптимальной системы хранения под нагрузку и задачи.

6. Открытые форматы LakeHouse, данные доступны для внешних систем

# Какой продукт сделали

1. Полностью интегрированный с экосистемой Postgres Pro Enterprise инструмент для построения хранилищ данных

2. Работает из коробки на той же инфраструктуре и железе для текущих OLTP клиентов. Переиспользование экспертизы команд, кода и бизнес-логики приложений

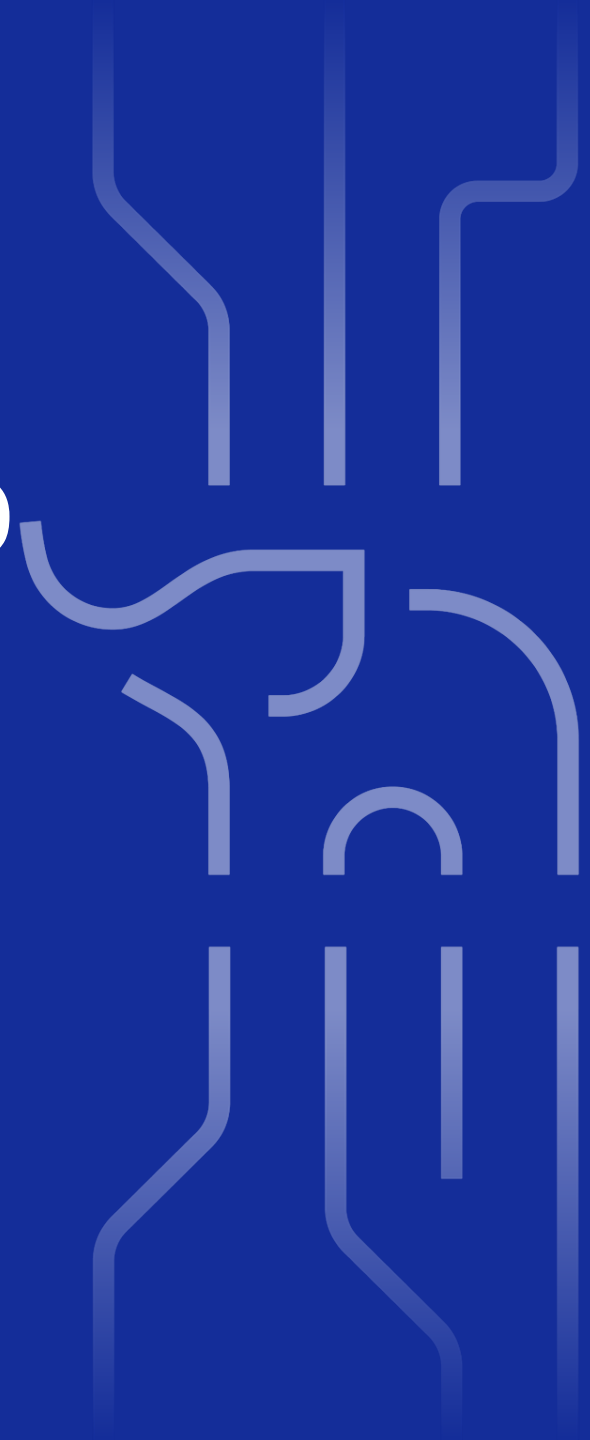
3. Быстрый и масштабируемый (не хуже Clickhouse, Umbra, DuckDB)

4. Данные реального времени, позволяет одновременно выполнять OLTP и OLAP сценарии

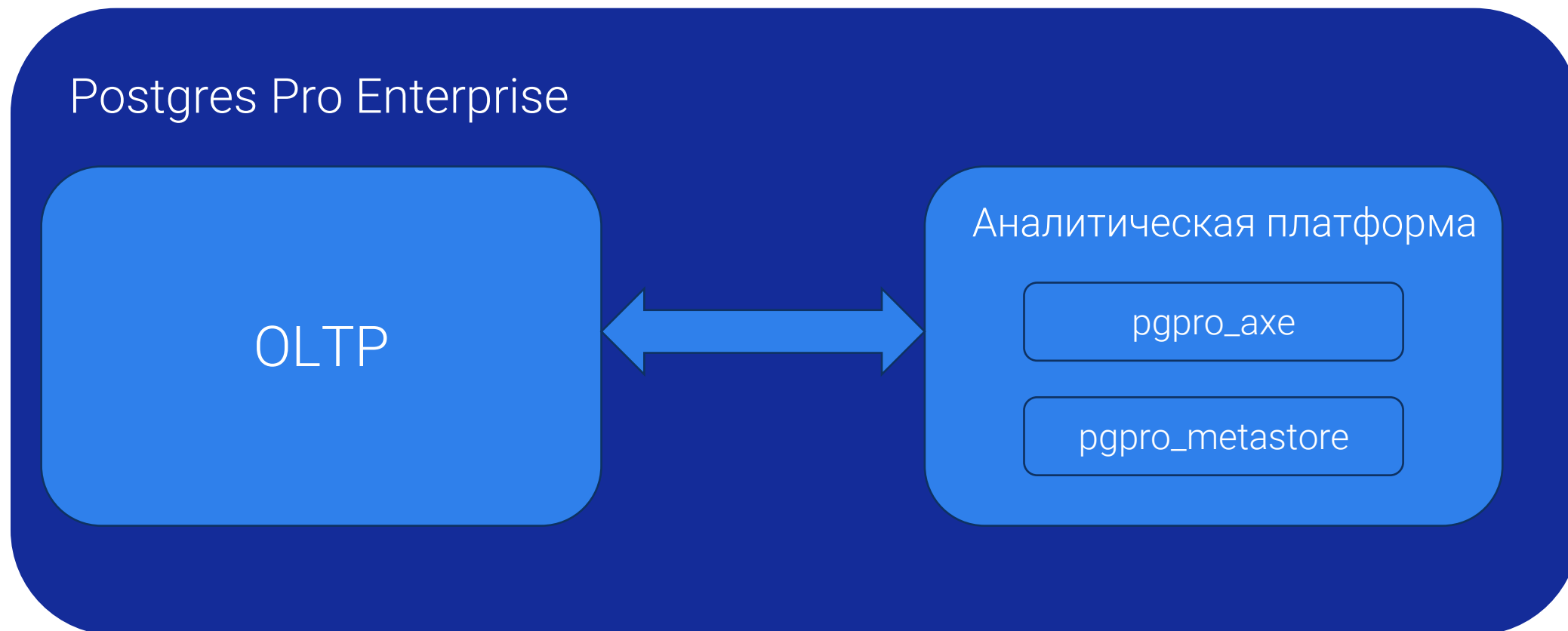
5. Масштабируемый для данных любых размеров. Выбор оптимальной системы хранения под нагрузку и задачи.

6. Открытые форматы LakeHouse, данные доступны для внешних систем

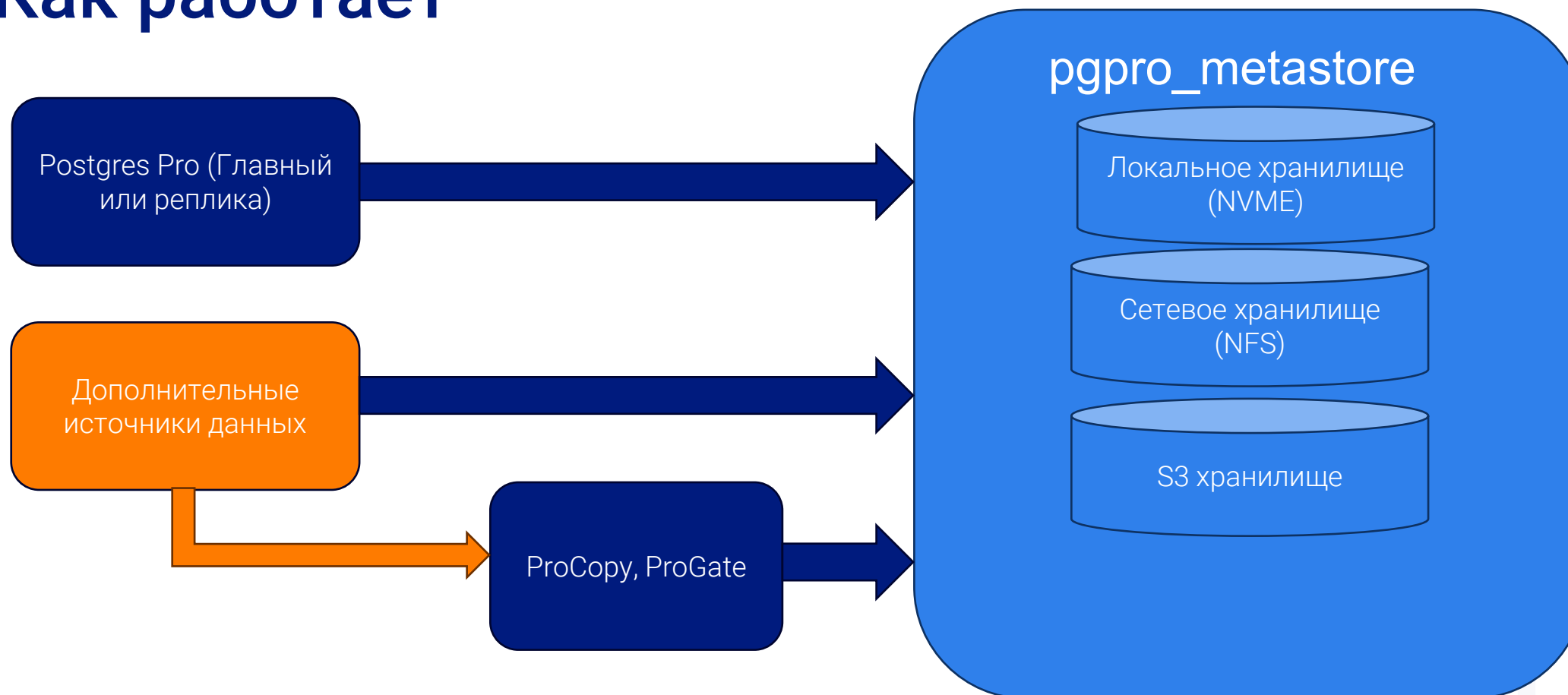
# 3. Аналитическая платформа Postgres Pro AXE



# Компоненты для построения хранилищ данных



# Как работает



# Набор компонент платформы

## Расширения Postgres

```
graph TD; A[Расширения Postgres] --> B[pgpro_axe]; A --> C[pgpro_metastore];
```

### pgpro\_axe

Векторный движок Postgres Pro с поддержкой разных систем хранения (NVME, NFS, S3) и Parquet

(01)

### pgpro\_metastore

Каталог файлов аналитических таблиц и средства репликации таблиц Postgres

(02)

Pgpro Copy — репликация из дополнительных источников

# Расширение pgpro\_ахе

Векторный движок Postgres Pro:

1. Быстрые запросы над колоночными источниками данных (Parquet файлы и временные колоночные таблицы)

3. Объединение данных HEAP таблиц Postgres и колоночных источников

2. Практически полная поддержка синтаксиса Postgres

4. Поддержка работы с объектными хранилищами (S3)

# Колоночный формат Parquet

1. Использует сжатие и кодирования однотипных данных

2. Содержит статистику по данным для projection и column pruning

3. Стандарт де факто для аналитики. Поддерживается всеми популярными инструментами.

4. Подходит любое дешевое хранилище (локальные диски, сетевое, S3).

5. Нет проблем с обработкой холодных данных неограниченного размера.

6. Линейное масштабирование по мощности в случае общего каталога.

# Расширение pgpro\_metastore

Хранилище аналитических таблиц и средства репликации таблиц Postgres

1. Хранит метаинформацию аналитических таблиц в БД по спецификации DuckLake

2. Может использовать внешний сервер Postgres

3. Формат хранения — Parquet файлы

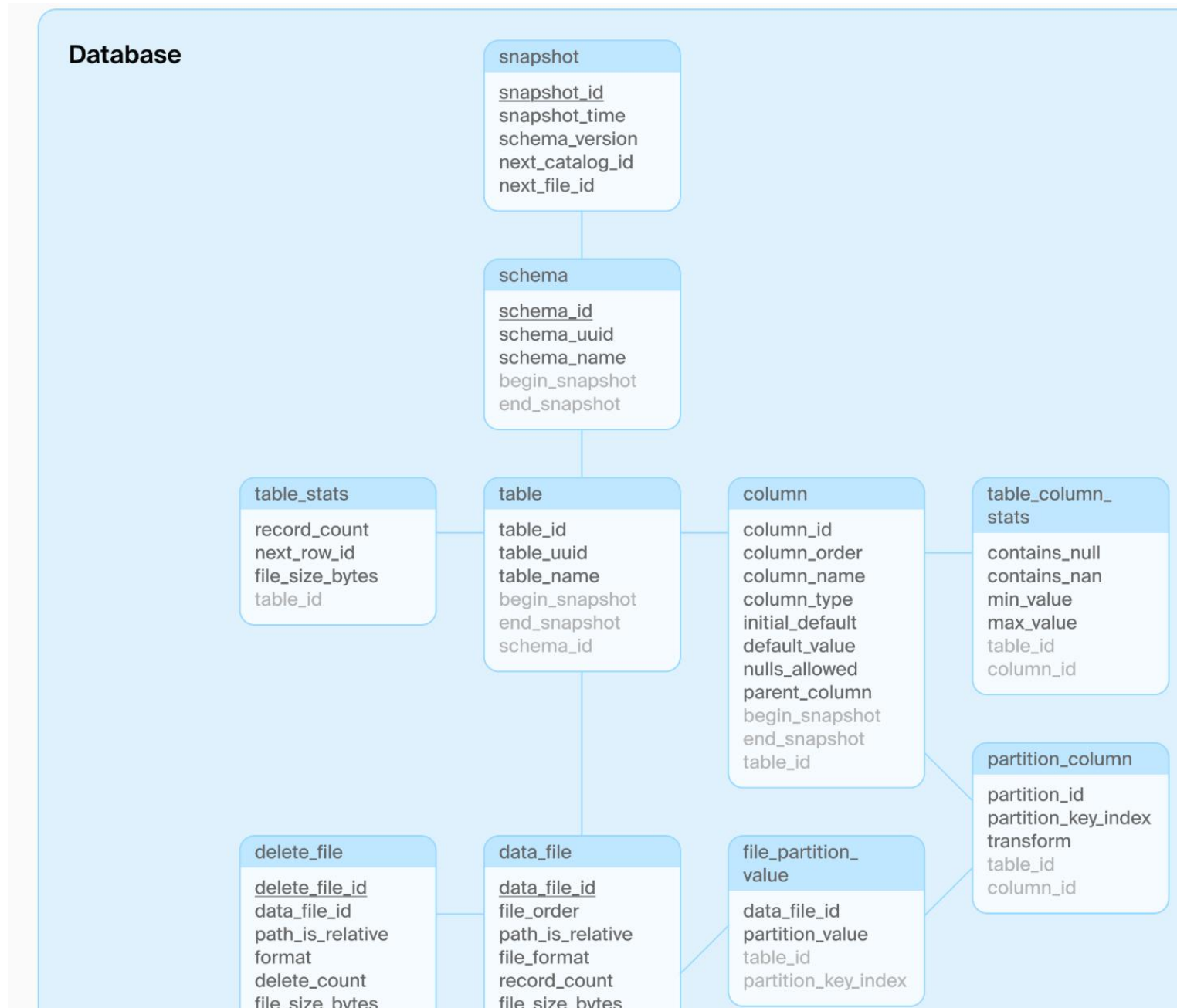
4. Типы хранилища — файловая система и S3

5. Поддерживает гранулярные права доступа к объектам хранилища

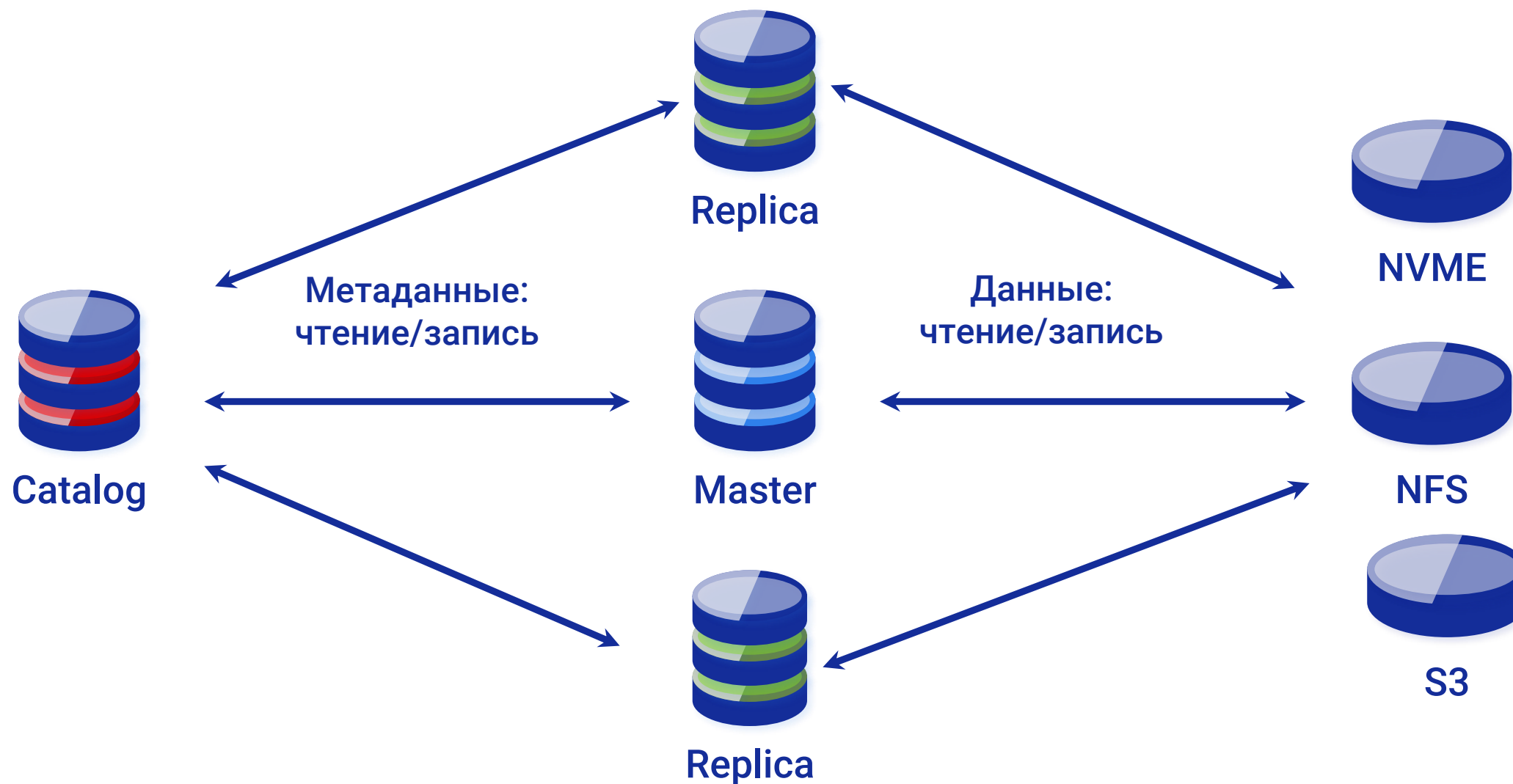
6. Поддерживает транзакционные изменения аналитических таблиц

# Схема каталога

21 таблица + расширения от Postgres Pro Axe



# Масштабируемость



# Технологические преимущества

- В ~ 10-30 раз быстрее
- В ~ 5-10 раз меньше объем данных быстрее
- Выбор оптимальной системы хранения для горячих и холодных данных
- Эффективный масштабируемый каталог в виде набора SQL таблиц с хорошо документированной схемой
- Неограниченная масштабируемость – как по вычислениям, так и по системе хранения

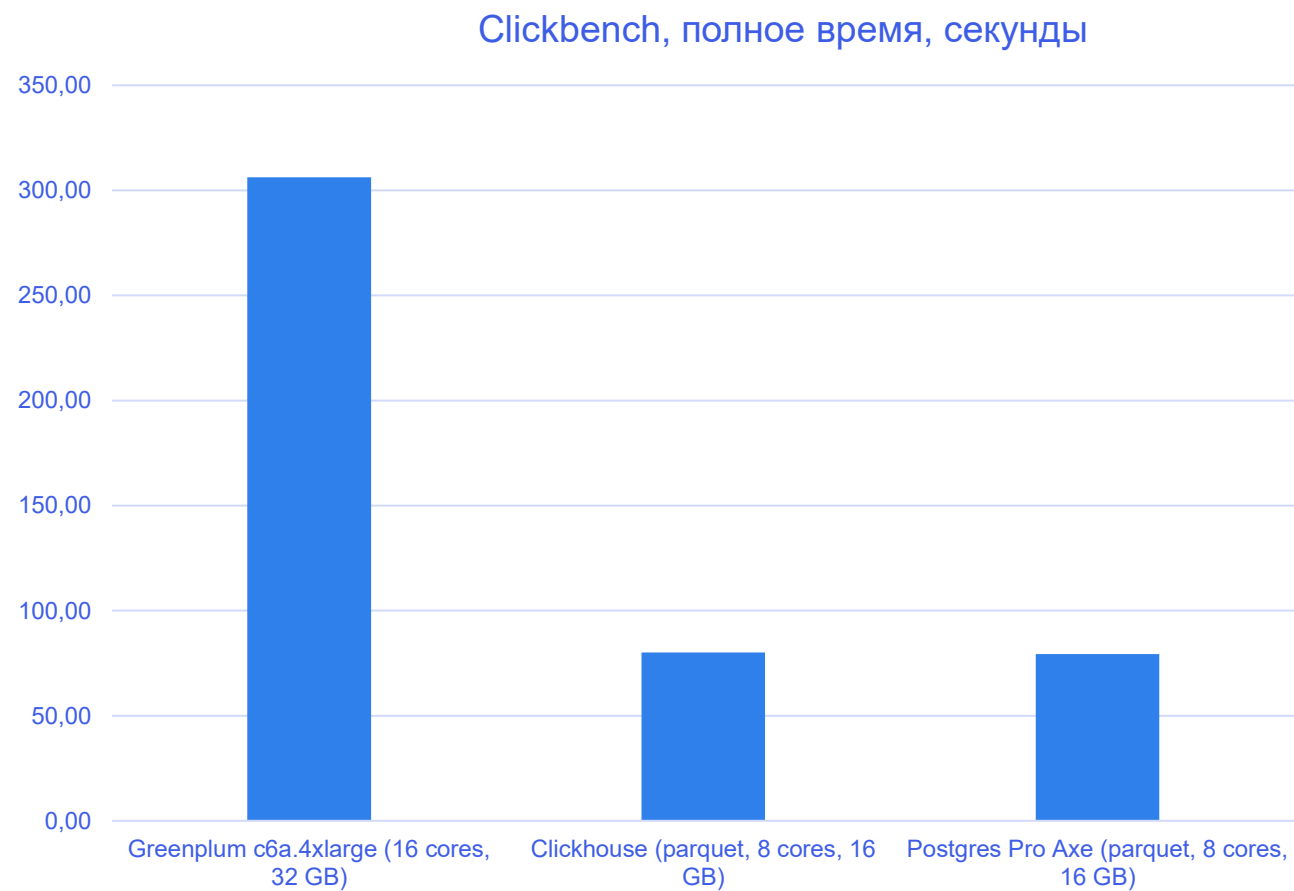
\*Среднее преимущество в тестах Clickbench, TPC-H/TPC-DS по сравнению с Postgres на такой же мощности железа

\*\*По сравнению с размером HEAP таблиц

# 4. Тесты производительности и примеры использования



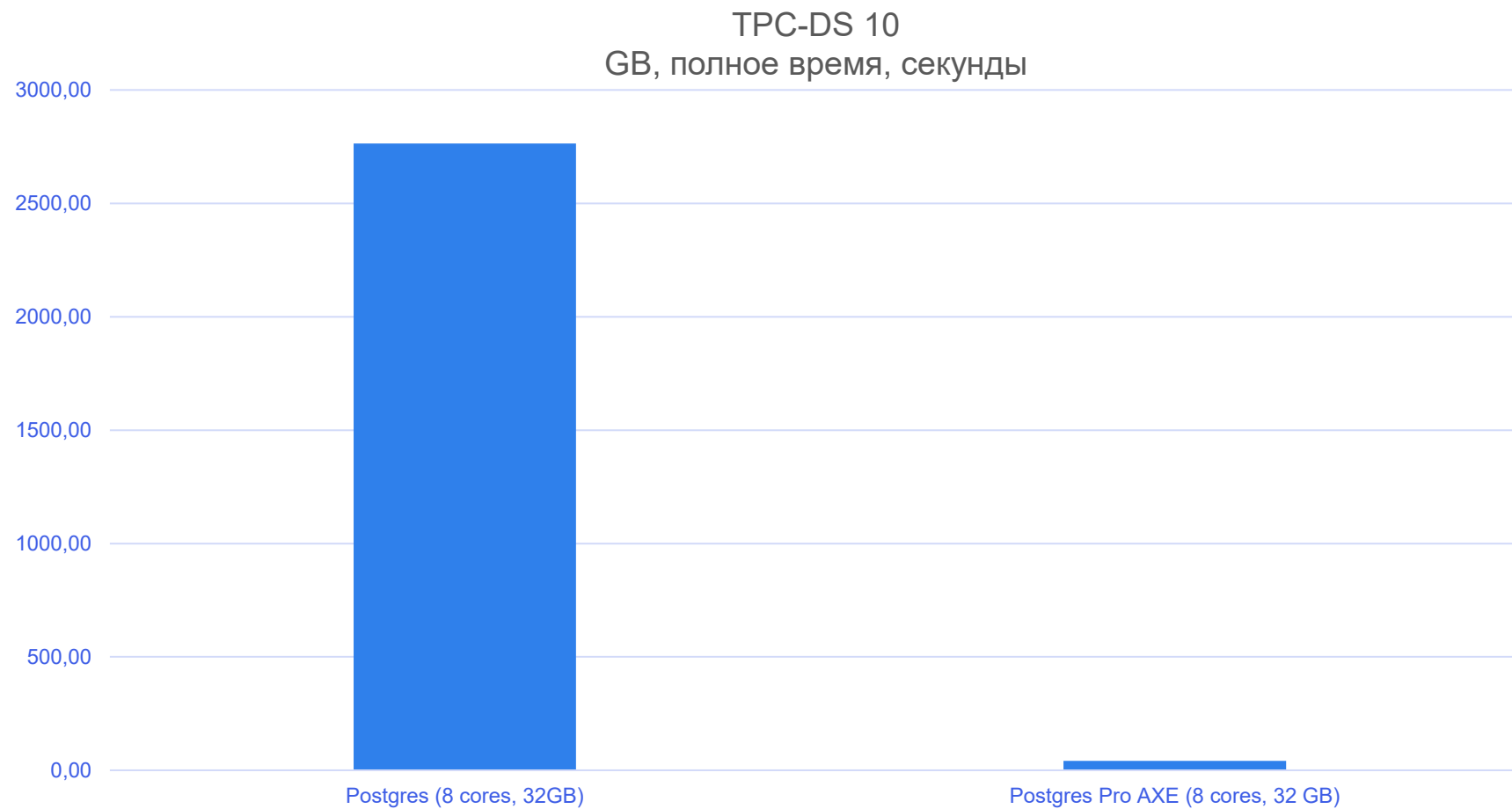
# Тесты Clickbench



# Тесты TPC-H



# Тесты TPC-DS



# Примеры сценариев использования

1. Снять аналитическую нагрузку с главного OLTP сервера

3. Использование ресурсов реплик Postgres Pro для обработки аналитической нагрузки

2. Эффективное хранение и быстрый поиск исторических данных

4. Построение полноценного аналитического хранилища на базе методологии Data Vault.

# 5. Дорожная карта



# Дорожная карта

Релиз Postgres Pro Ахе запланирован на январь 2026. Ведется бета тестирование.

1. Быстрая SIMD конвертация  
PG Heap таблиц на лету в колонки

2. Запросы к аналитическим таблицам  
pgpro\_metastore без удержания горизонта  
транзакции PG

3. Table Access Method поверх таблиц  
pgpro\_metastore

4. Быстрый параллельный CDC на основе  
WAL файлов (без использования слотов  
репликации)

5. Интеграция с продуктами Postgres Pro:  
Tengri, Backup, Shardman

# Преимущества

1. Полностью интегрированный с экосистемой Postgres Pro Enterprise инструмент для построения хранилищ данных

2. Работает из коробки на той же инфраструктуре и железе для текущих инсталляций Postgres Pro. Использование экспертизы команд, кода и бизнес-логики приложений

3. Быстрый и масштабируемый (не хуже Clickhouse, Umbra, DuckDB)

4. Данные реального времени, позволяет одновременно выполнять OLTP и OLAP сценарии

5. Масштабируемый для данных любых размеров. Выбор оптимальной системы хранения под нагрузку и задачи.

6. Открытые форматы LakeHouse, данные доступны для внешних систем

# Главное преимущество

Снижение TCO – меньше зоопарк решений, железа, проблем в разработке и поддержке.

Возможность легко строить хранилища данных на базе работающей экосистемы Postgres Pro.

# Как попробовать

Участие в бета тестировании – [sales@postgrespro.ru](mailto:sales@postgrespro.ru)

**Отвечу на ваши вопросы!**

