

State of Data '26

Что такое
«modern data stack»
в конце 2026 года?

Что там происходит в мире?

Общий знаменатель трёх вендоров: Iceberg как формат хранения семантический/контекстный слой, Postgres для состояния агентов

Google Cloud Next '26

- Knowledge Catalog — единый слой контекста над lakehouse и AlloyDB, федерация с Atlan, Collibra, DataHub;
- Cross-Cloud Lakehouse на Apache Iceberg: агенты читают данные в S3 и ADLS; федерация каталогов Glue, Unity и Horizon через Iceberg REST
- AlloyDB как serving-слой агентных приложений: lakehouse federation — чтение Iceberg и BigQuery прямо из Postgres, one-click reverse ETL, Datastream → Iceberg

Databricks Data + AI Summit

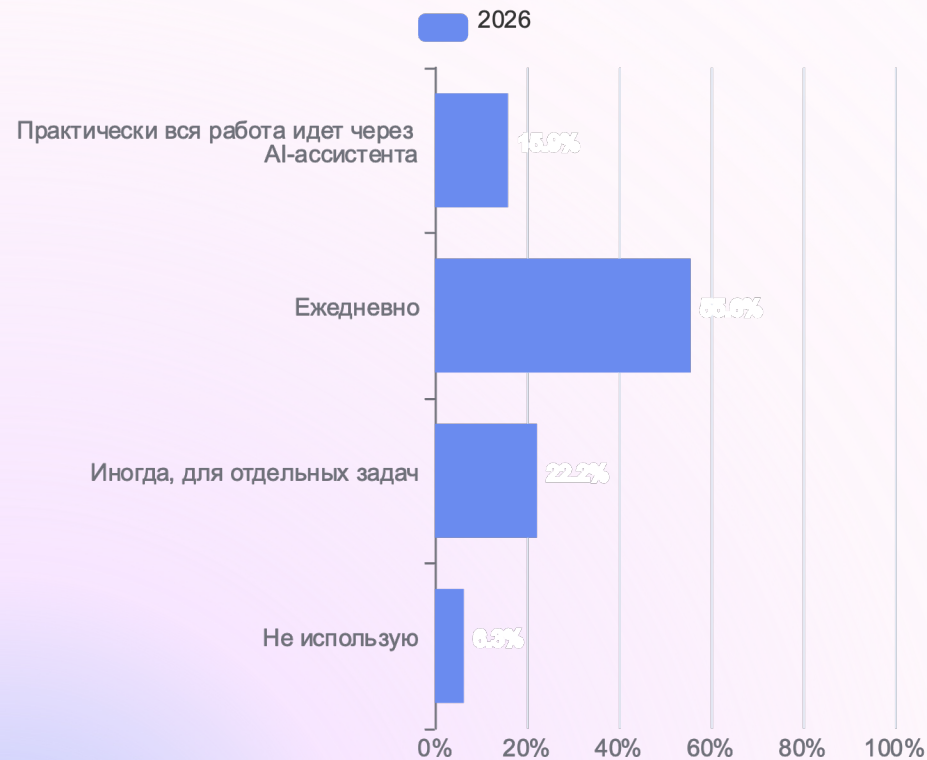
- Genie Ontology, Unity Catalog Metrics, Unity AI Gateway — семантика, метрики, агенты, модели и MCP-серверы становятся governed-объектами каталога с бюджетами и RBAC
- Lakebase (serverless Postgres) — память и состояние агентов; вместе с Lakehouse Databricks называет это «LTAP»: транзакции агентов рядом с аналитикой. Lakehouse RT — sub-100 ms над Delta/Iceberg

Snowflake Summit

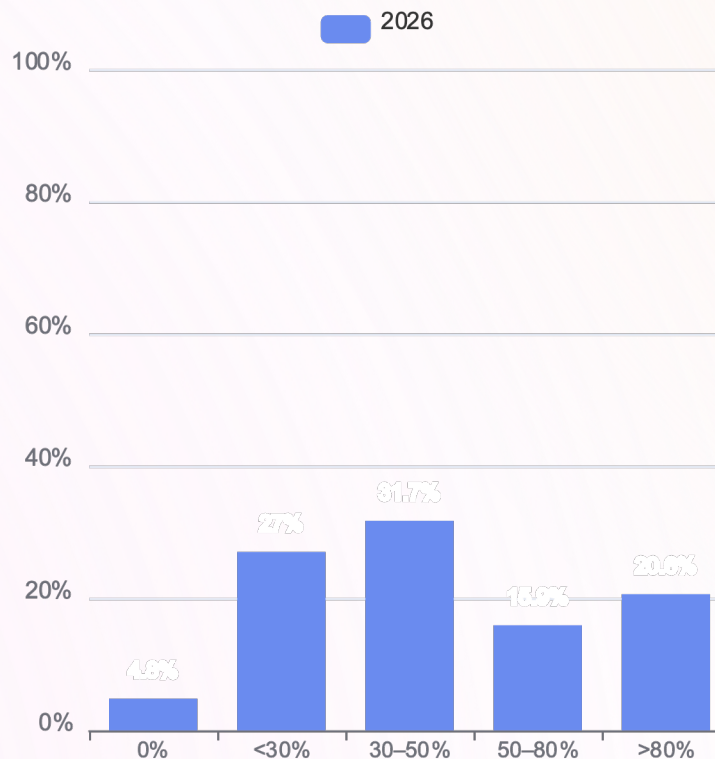
- Horizon Context + Semantic View Autopilot — сбор метаданных из Postgres, SQL Server, Tableau, Power BI, dbt и автогенерация semantic views; Open Semantic Interchange — открытый стандарт семантических моделей, 54 вендора.
- Iceberg v3 в GA: deletion vectors, row lineage, variant; двусторонний read/write через Apache Polaris;
- Snowflake Postgres + pg_lake — zero-ETL чтение и запись Iceberg-таблиц из Postgres;

И что происходит у нас: agentic ai в работе дата инженеров

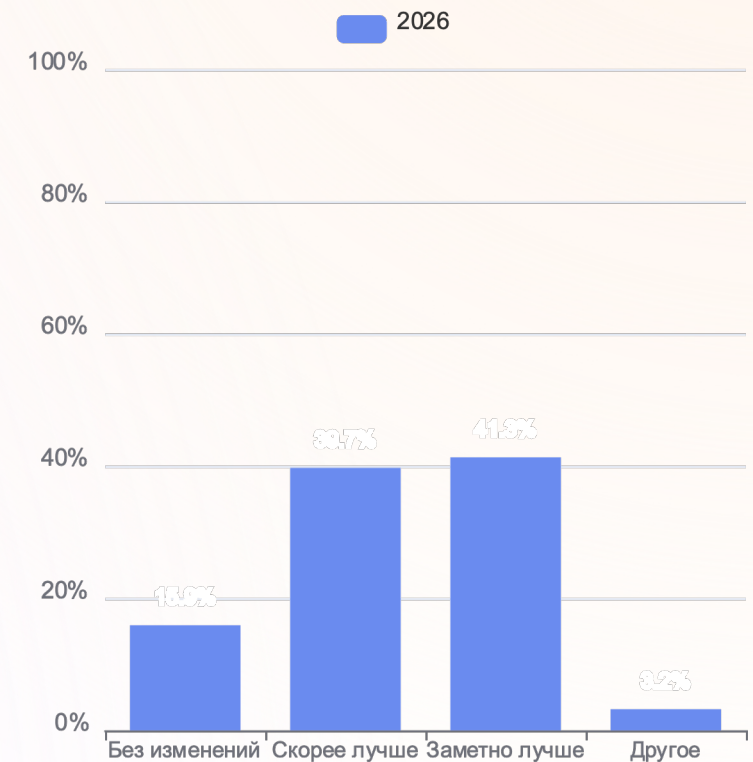
Частота использования AI-ассистентов



Доля рабочего кода, написанного с помощью AI



Влияние AI-ассистентов на продуктивность



И что происходит у нас: agentic ai в работе дата инженеров

- 71% используют AI-ассистентов ежедневно или «практически вся работа идёт через ассистента»; совсем не используют — 6%
- 37% пишут с помощью AI больше половины рабочего кода, ещё 32% — от 30 до 50%
- 81% отмечают рост продуктивности («замечто лучше» — 41%, «скорее лучше» — 40%); ухудшения не отметил никто
- Топ задач: новый код 86%, рефакторинг 77%, отладка 76%, документация 73%, разбор чужого кода 66%, SQL 63%, тесты 61%
- 90% AI-кода доходит до прода: 37% после обычного ревью, 29% без отдельных требований, 24% после усиленного ревью

Задачи, которые решают с помощью AI-ассистентов

сумма > 100%: можно выбрать несколько вариантов



ОБСУЖДЕНИЕ 23.09 / 18:00 – 18:45 (UTC+3)

Зал 4

Offline only

REC



ИИ всё сделал. А чему теперь учиться?

ИИ уже помогает решать математические задачи, писать код и проводить исследования. Что в этой реальности остается задачей человека? Нужны ли нам прежние знания и навыки, чему должен учить университет и каким будет ученый будущего?

Спикеры



Евгений Ильюшин
Okko

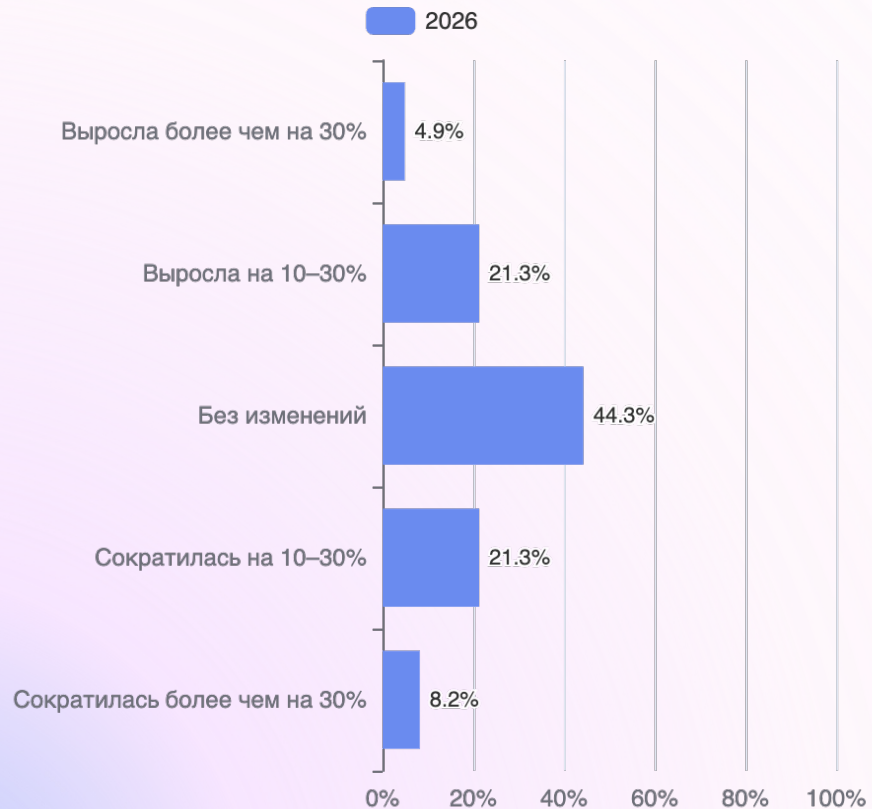
Ведущие



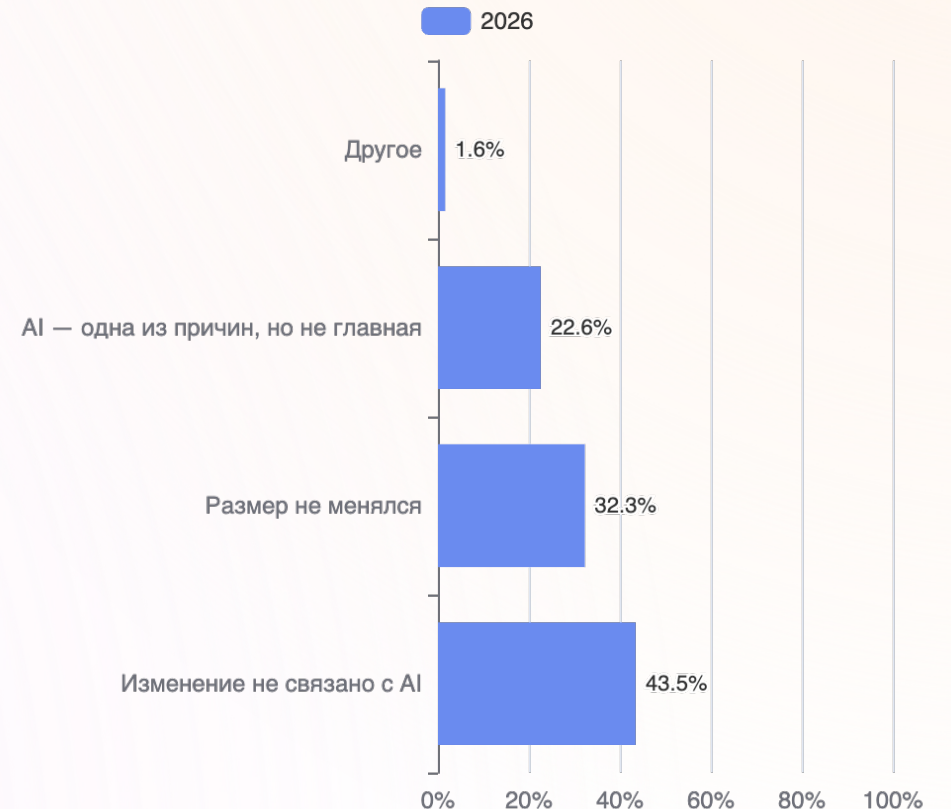
Александра Мурзина
Яндекс Go

И что происходит у нас: влияние на headcount

Как изменился размер дата-команды за последний год?

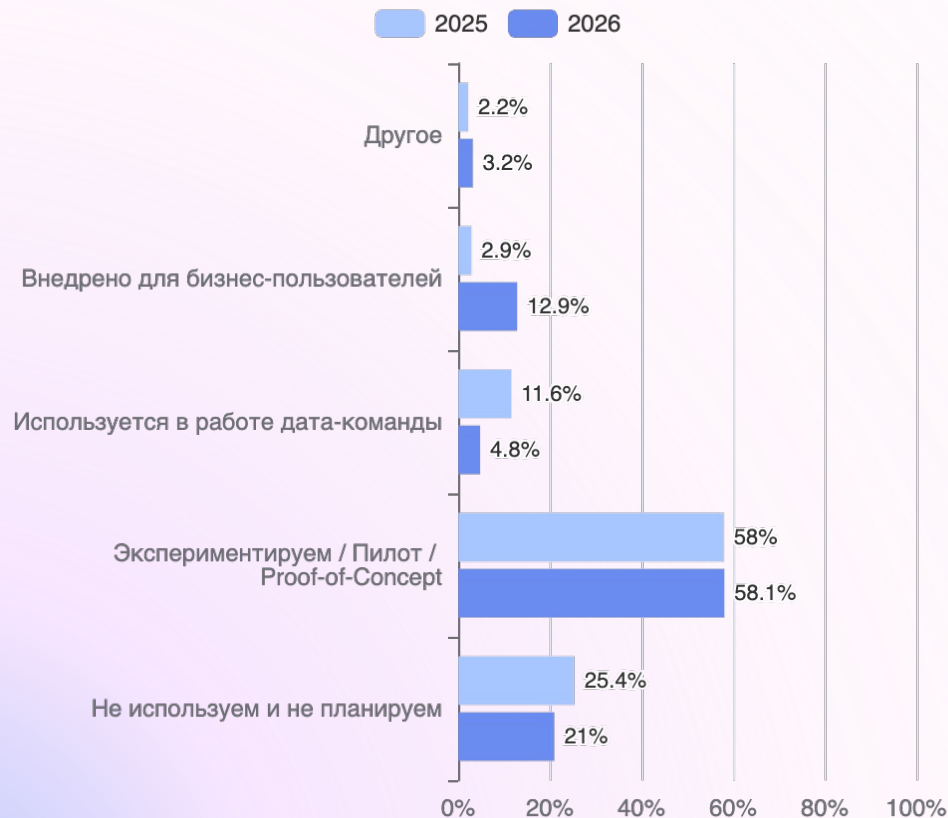


Связано ли изменение размера команды с внедрением AI?

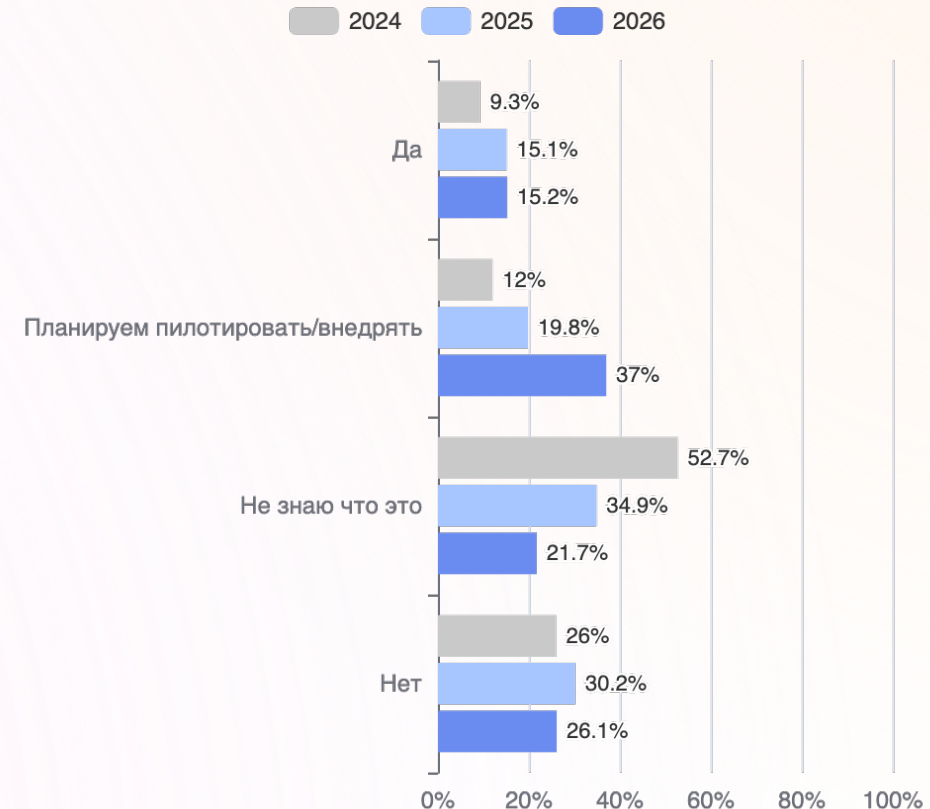


И что происходит у нас: text-to-sql для пользователей

Инструменты text-to-SQL для пользователей



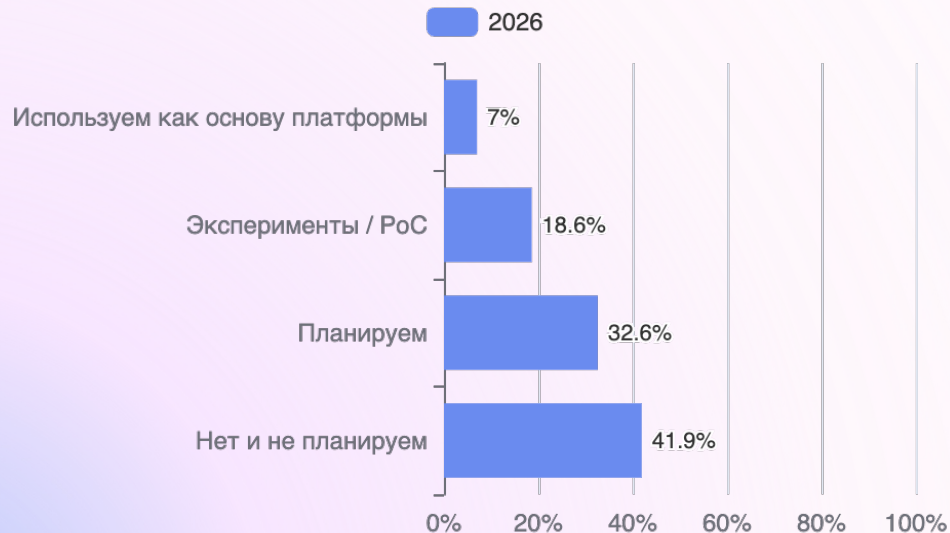
Semantic Layer



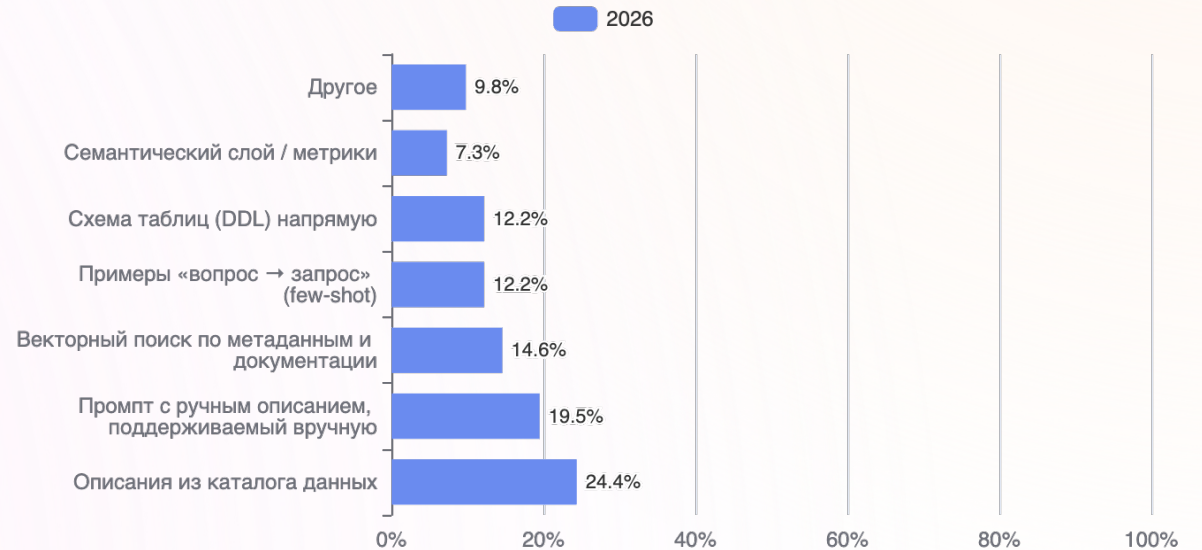
И что происходит у нас: слой контекста

- Онтология или граф знаний поверх данных: 42 % — нет и не планируем, 33 % планируют, 19 % — PoC, только 7 % — основа платформы
- Контекст для LLM-агентов: описания из каталога 24 %, ручной промпт 20 %, векторный поиск по метаданным 15 %, few-shot 12 %, DDL 12 %, семантический слой — 7 %

Онтология / граф знаний поверх данных



Как LLM-агенты получают контекст о данных

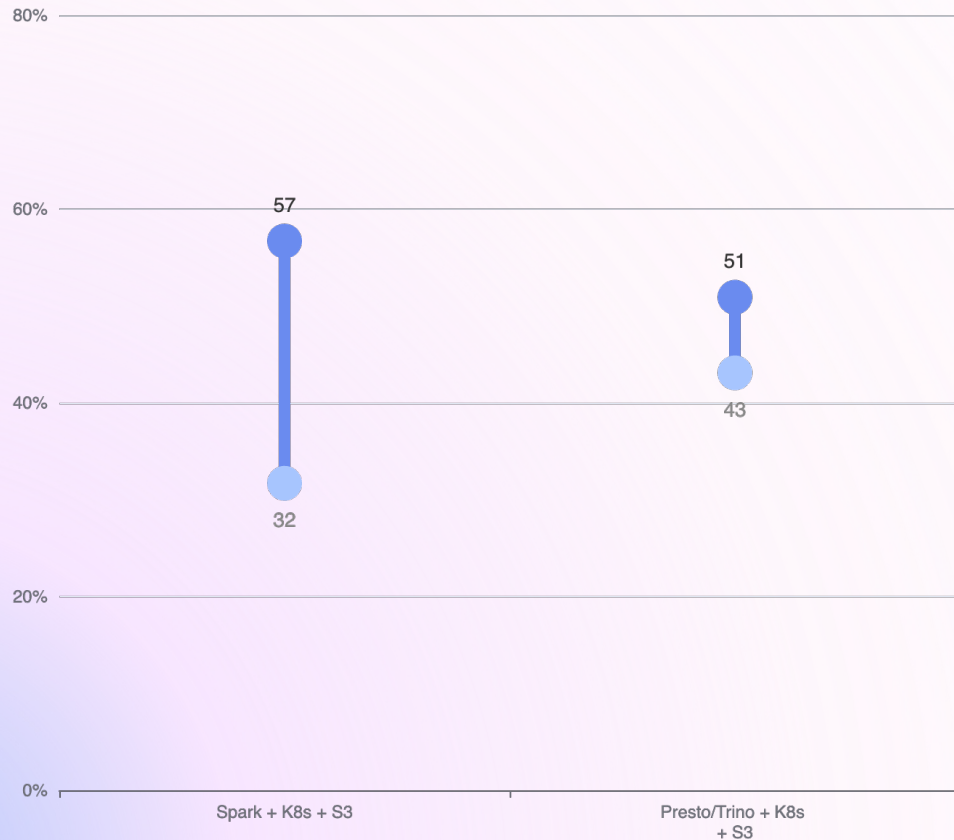


И что происходит у нас: lakehouse и iceberg

Основной стек для хранения и обработки данных: динамика net score

Net score = Доля (Используем+Планируем – Использовали в прошлом – Не интересуем), без «Даже не слышал» в знаменателе

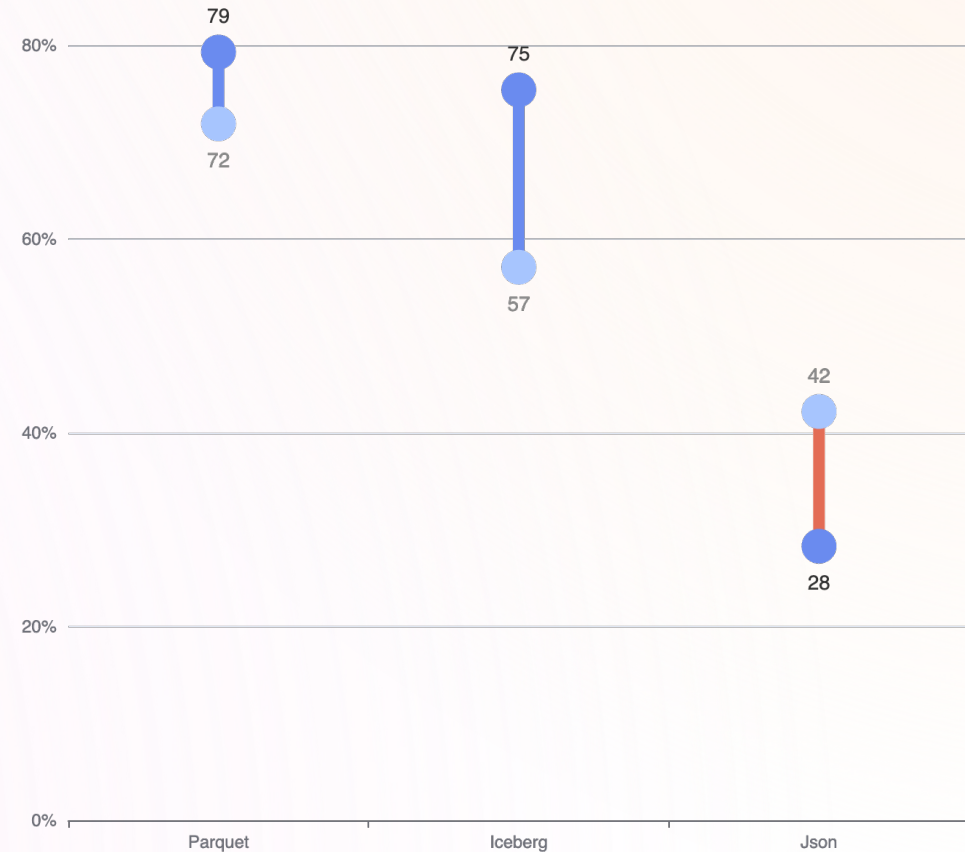
● 2025 ● 2026



Основной формат хранения данных: динамика net score

Net score = Доля (Используем+Планируем – Использовали в прошлом – Не интересуем), без «Даже не слышал» в знаменателе

● 2025 ● 2026



И что происходит у нас: Внезапное

- Hadoop неожиданно решил воскреснуть (но это не точно)
- Greenplum всех утомил (тоже не точно)

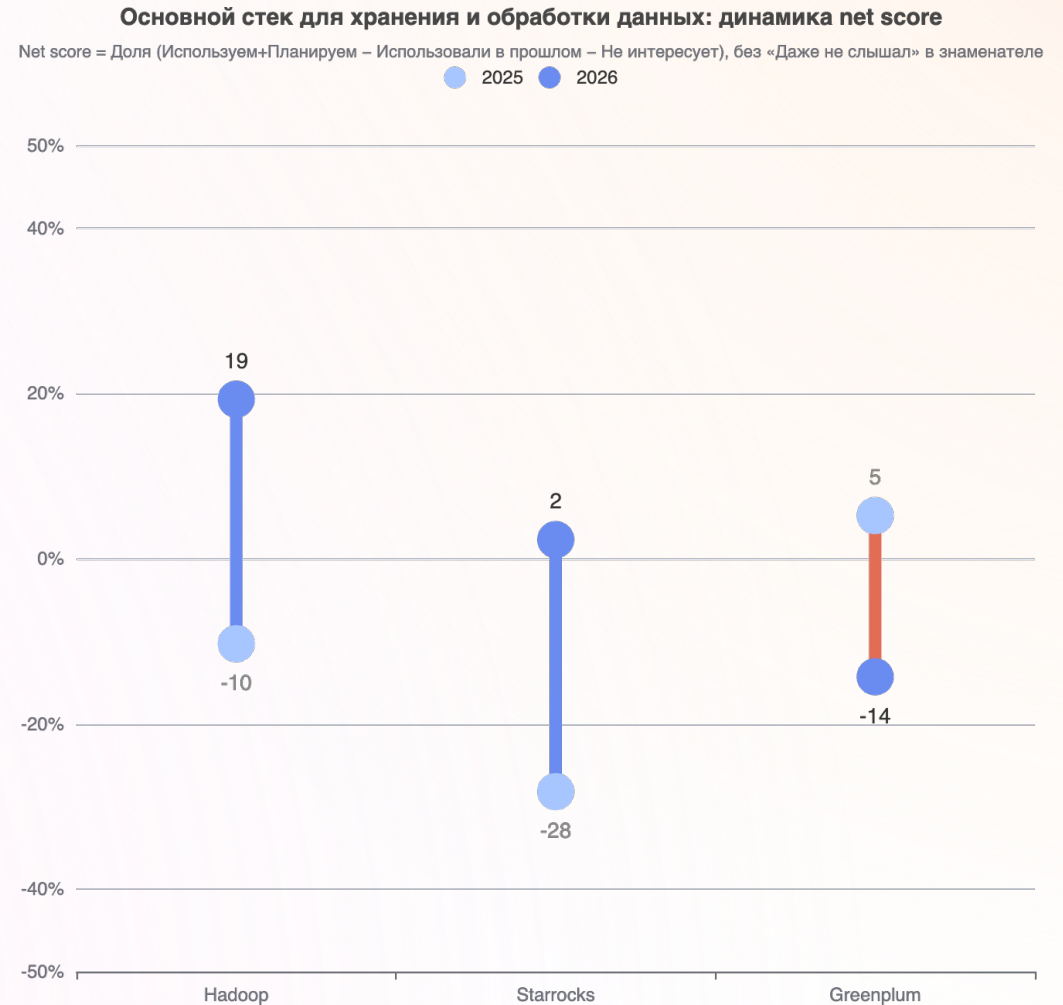
Обсуждение

Эволюция аналитических СУБД — на какой стадии развития находится Greenplum и что идет ему на смену?

Без занудства и слайдов обсудим, стоит ли в 2026 году брать Greenplum (или какой-нибудь его форк) для нового проекта.

Леонид Борчук Yandex Cloud
Дмитрий Инокентьев GlowByte
Дмитрий Немчин Т-Банк
Александр Мошурa РТК ИТ
Дмитрий Зуев Positive Technologies

Offline only REC



И что происходит у нас: стабильное



И что происходит у нас: непонятное

Обсуждение



ETL без боли: Low-Code, No-Code или просто попросить AI Agent?

На этой панельной дискуссии эксперты столкнут лбами три лагеря: сторонников чистого кода, апологетов Low/No-Code и евангелистов AI.

Дмитрий Руднев Т-Банк
Александра Попова Positive Technologies
Игорь Мосягин shrimpsizemoose AB
Бронислав Житников Positive Technologies

Offline only

REC

Инструменты загрузки данных: динамика net score

Net score = Доля (Используем+Планируем – Использовали в прошлом – Не интересует), без «Даже не слышал» в знаменателе

● 2025 ● 2026



Спасибо!

