

КОРПОРАТИВНЫЙ FEATURE STORE. CASEs STUDY

BeCloud.AI

AINUR
SIDELKOVSKAYA



AINUR SIDELKOVSKAYA

Founder
BECLOUD.AI

Data Driven Management Evangelist

Создатель Smart Data Ukimet

Интегратор корпоративных экосистем данных

BECLOUD.AI

Команда, объединяющая профессионалов с опытом более 15 лет по:



Внедрению промышленных Хранилищ данных (DWH), Data Lake, Data Management & Governance



Внедрению Industrial IoT и интеллектуальных систем управления производственными процессами



Углубленной аналитике (Data science, Data mining, Machine Learning) и визуализации данных (Business Intelligence)



Разработке стратегий по внедрению AI-решений, Data-driven менеджменту и трансформации компании

Case Study.

Создание Feature store для БВУ СНГ



ЗАДАЧА

Создать единое пространство для работы с клиентскими данными и совместного использования DSML признаков.

ВЫЗОВЫ

- ✓ Каждая команда DS имела собственный sandbox. Алгоритм расчетных атрибутов был индивидуальным, и чаще всего использовался под конкретную модель
- ✓ Описание велось каждой командой отдельно в рамках своего sandbox-а
- ✓ Наличие дублей в данных и открытых вопросов к качеству данных, используемых в DSML
- ✓ Действующее хранилище построено на классической SQL СУБД и загружено операционными задачами по отчетности. Параллельная работа DSML team существенно снижает производительность СУБД КХД

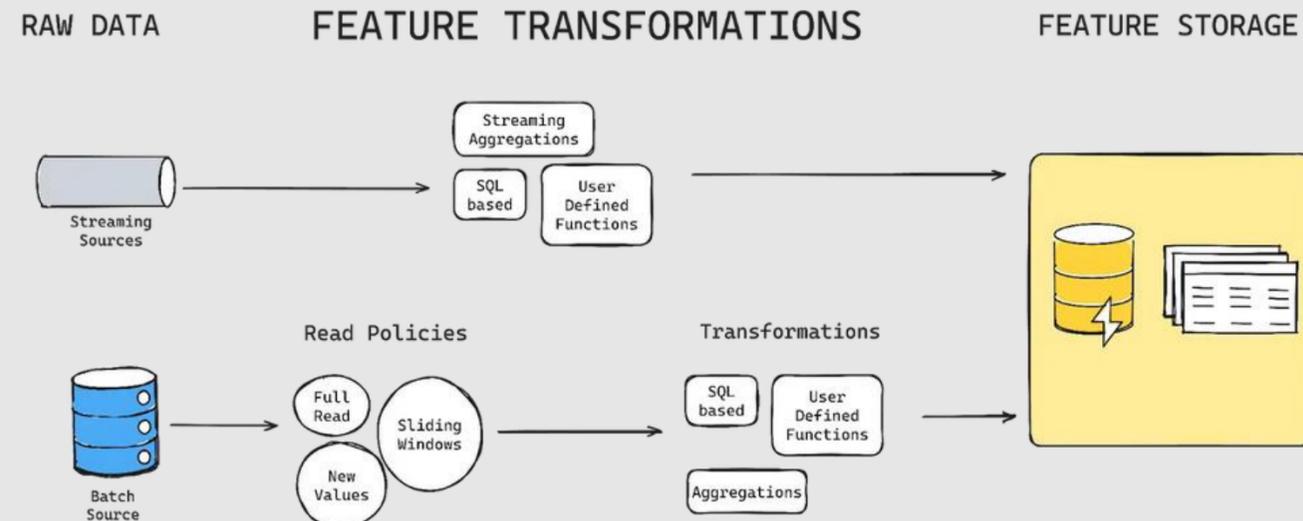
Case Study

Создание Feature Store

Выбор решения: определена СУБД **Vertica Analytics Platform** по результатам промышленного тестирования и сравнения с другими аналитическими СУБД

Решение: Создание **Feature store** на базе **Vertica**

Результаты 1 этапа:



Case Study

Создание Feature Store

Эффекты:

- ✓ Снижение нагрузки на DS team по сбору и очистке данных
- ✓ Выше производительность за счет разделения задач КХД и Feature Store
- ✓ Бизнес пользователи и DS team видят фичи в едином окне без обращения к DE на создание датасетов
- ✓ Обработка клиентских данных на Vertica осуществляется в 100 раз быстрее чем на классической СУБД КХД

10 000+ атрибутов и фичей

Подключение внешних данных

Подключение стриминговых источников

Стриминговая аналитика

FEATURE STORE



FEATURE STORE.

ПОЧЕМУ
СЕЙЧАС?

DS команды собирают собственные датасеты,
создают алгоритмы расчета атрибутов и фичей
под конкретную модель



Количество команд
Data Scientist-ов в компании
X количество маркетинговых
кампаний и DSML моделей

Одноразовые датасеты

Невозможно пере-
использовать фичи другими
командами Data Scientist-ов



Регулярное
автоматизированное
обновление данных

Возможность
повторного
использования
качественных данных

Сложность ведения документации

Нет описания расчетных
показателей, отсутствие
версионности



Актуальный бизнес-
гlossарий и каталог
данных в едином окне

Нерациональное использование ресурсов

Повторение расчетных процессов
Низкое доверие к качеству данных



Оптимизация ресурсов за
счет использования
"единой точки правды"

5 КОМПОНЕНТОВ FEATURE STORE.

Трансформация

Создание таблиц
преагрегаторов для
получения нескольких
вариаций фичей

Управление хранением

Создание версий на уровне
индивидуальных фичей.
Возможность накапливания
истории по фичам

Каталог

Централизованный реестр,
который содержит
информацию откуда, в
каком формате берутся
данные, как рассчитываются

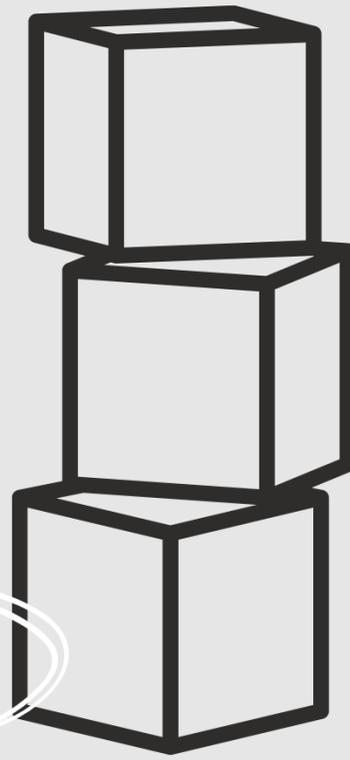
Мониторинг

Мониторинг качества
данных для модели,
выявление аномалий

СУБД

**Полностью колоночная
MPP СУБД со встроенными
DSML алгоритмами**

DSML-те в самой СУБД!



VERTICA
by **opentext**[™]

BeCloud.AI
Let your data work

