



АССОЦИАЦИЯ
БОЛЬШИХ ДАННЫХ

ЭКОНОМИКА
АНО «Цифровая экономика»

FIRST RUSSIAN DATA FORUM

Целевой стек импортонезависимого ПО

Борис Рабинович, CDO Сбер

Фабрика данных как технологический центр работы с данными



Источников загружено в Фабрику данных

750+

Пользователей

17 000+

Надежность

99,99+

для транзакционных сервисов

Объем данных

160 пб

Графовые связи

17 млрд

Скорость

1 000 000+

транзакций в секунду (NRT)

Объем получаемых изменений

16-20 Тб/сутки

ЗАДАЧИ, РЕШАЕМЫЕ ФАБРИКОЙ ДАННЫХ

- ✓ аналитика и отчетность
- ✓ рекомендательные системы
- ✓ профиль клиента 360%
- ✓ оценка рисков, RWA, AML
- ✓ транзакционный скоринг AI в кредитовании
- ✓ платформа для AI, графовая платформа
- ✓ управление качеством данных на уровне Группы

Ранее типовой технологический стек работы с данными во многих компаниях выглядел так:



Машинное обучение и искусственный интеллект
Machine Learning (ML) and Artificial intelligence (AI)



Технологии и инструменты визуализации и анализа данных
Business Intelligence (BI)



Традиционные реляционные системы управления базами данных (RDBMS)



Массивно-параллельные системы управления базами данных (MPP DB)



Распределенные системы хранения и обработки данных любых форматов (HADOOP)



Системы распределенной обработки данных в оперативной памяти (InMemory)

Специализированные системы управления базами данных (GraphDB)



Технологии и инструменты интеграции и трансформации данных
Extract Transform Load (ETL)



Современный технологический стек Сбербанка по работе с большими данными

ML Space

Машинное обучение и искусственный интеллект
Machine Learning (ML) and Artificial intelligence (AI)

SDP Analytics

Технологии и инструменты визуализации и анализа данных
Business Intelligence (BI)

Традиционные реляционные системы управления базами данных (RDBMS)

Pangolin

Массивно-параллельные системы управления базами данных (MPP DB)

SDP Greenplum

Распределенные системы хранения и обработки данных любых форматов (HADOOP)

SDP Hadoop

Системы распределенной обработки данных в оперативной памяти (InMemory)

Ignite SE

Специализированные системы управления базами данных (GraphDB)

SDP Fast Graph

SDP DataQuality

Технологии и инструменты интеграции и трансформации данных
Extract Transform Load (ETL)

SDP DataFlow

- 1 Стоит задача масштабирования и **возникает потребность** удешевления стоимости владения данными
- 2 Стоит **задача обеспечения бесперебойного и непрерывного обеспечения** функционирования сервисов



НО СНАЧАЛА НАДО ПОДТВЕРДИТЬ

- **наличие** технической компетенции
- **наличие** необходимых ресурсов
- **готовность** инфраструктуры
- требуемые **сроки внедрения**

✓ В иных случаях стоит рассматривать доступные на рынке готовые решения



Технологический стек следует дополнять решениями, повышающими его производительность

1

Разделение Compute & Storage

независимо масштабируемые вычислительные ресурсы и система хранения (эффективное использование ресурсов)

2

SDP Ozone

эластичная система хранения на технологии Object Storage (снижение числа серверных стоек в 3,5 раза, снижение стоимости хранения 1 ТБ в Hadoop в 2,6 раза)

3

SDP Compute

масштабируемые вычисления на технологиях контейнеризации (высокая утилизация вычислительных ресурсов)

4

Data Fusion

виртуализация данных. Высокоскоростной доступ между различными кластерами Hadoop и Greenplum. Ноль копий, отсутствие дублирования и снижение объемов «железа» **в 10 раз**

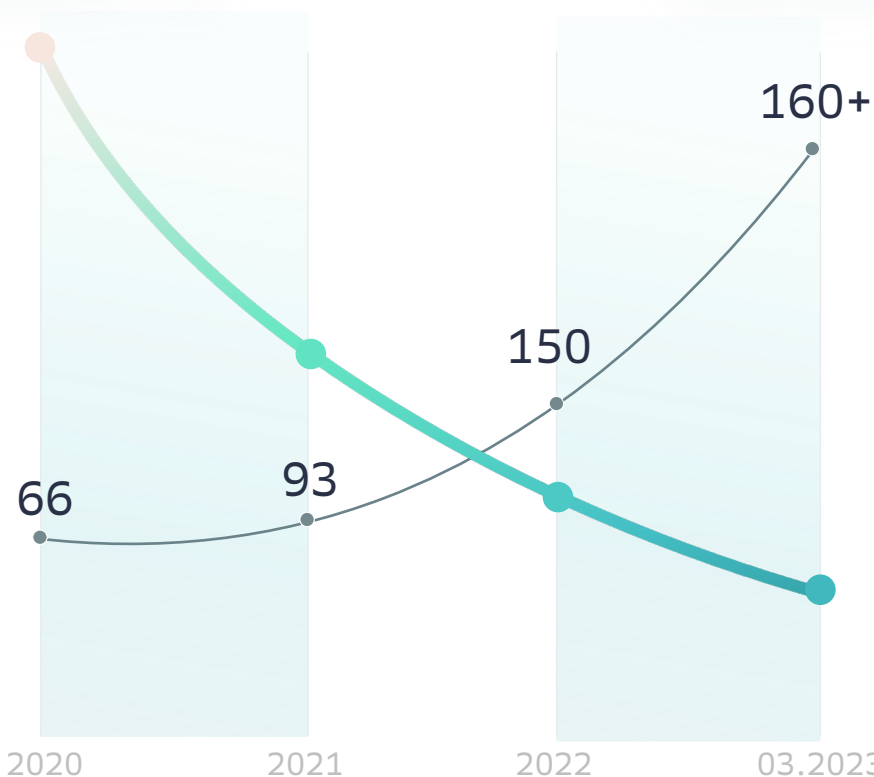
Импортозамещение: рост эффективности при снижении затрат

Снижение стоимости владения данными **в 3 раза за 2 года**

Стоимость 1 ТБ

Объем данных

ПБ



Собственная технологическая платформа и эффективное управление ресурсами позволяют **снижать затраты, несмотря на рост объема данных и растущие требования бизнеса** к качеству, скорости поставки, актуальности и полноте данных



Борис Рабинович

CDO СБЕР

 Email Rabinovich.B.II@sberbank.ru