

Современные системы хранения и анализа данных: широкие возможности платформы данных Microsoft Azure

Анна Фенюшина
Архитектор
Softline

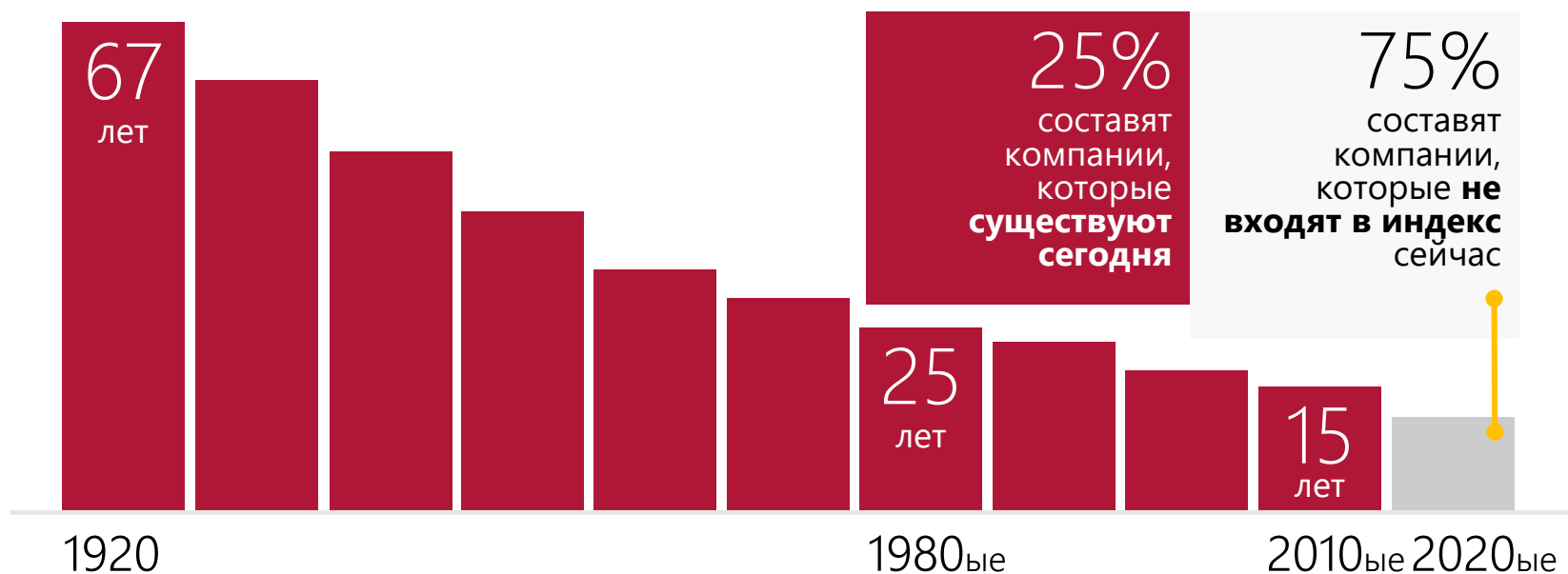


Содержание

- ➔ Данные – основа трансформации бизнеса
- ➔ Преимущества облачных решений для работы с информацией
- ➔ Обзор сервисов Microsoft Azure по работе с данными
 - Реляционные БД
 - NoSQL БД
 - SQL Data Warehouse
 - BigData: HDInsight, DataBricks и DataLake
- ➔ Организация процесса работы с данными и их интеграция из разных систем
- ➔ Работа с потоковыми данными
- ➔ Единая экосистема продвинутого анализа данных Microsoft
- ➔ Примеры использования платформы данных Microsoft Azure

Мир перемен

Средняя продолжительность жизни компании S&P 500



Source: Yale Professor Richard Foster

Трансформация бизнеса на базе анализа данных

**Дополнительные преимущества в
производительности
к 2020 году**



\$430B

будут получены организациями, способными анализировать
все соответствующие данные и предоставлять полезную
информацию

Source: IDC "FutureScape: Worldwide Big Data and Analytics
2016 Predictions," DOC #259835, November 2015

Дополнительный операционный доход

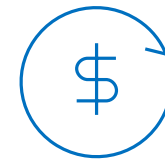


8%

сгенерирован наиболее цифровизованными
предприятиями

Source: Keystone Strategy 2016

**Дальнейшее снижение
затрат**



14%

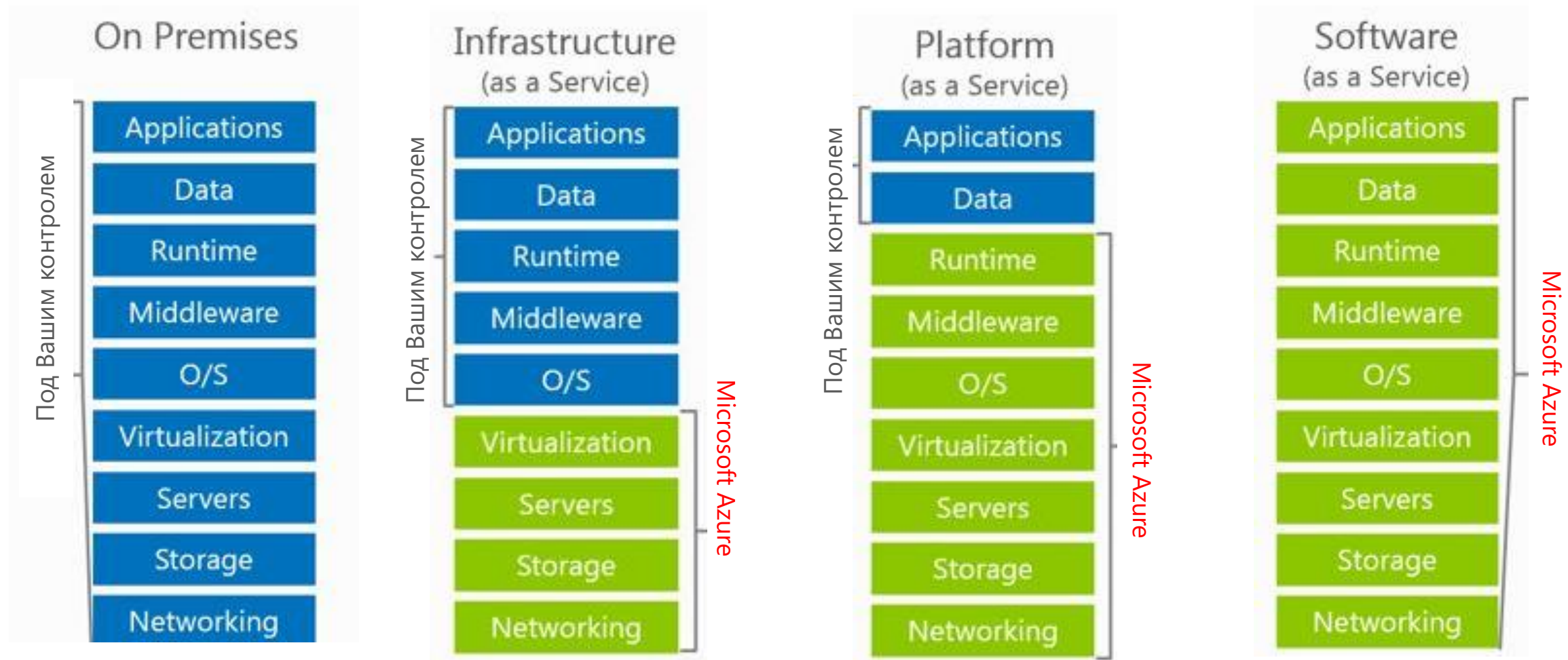
будет получено за счет снижения
CAPEX, стоимости ПО, ИТ
и оптимизации трат

Source: Gartner "Cost Optimization Secrets: Plan & Manage Cloud Services
for 14% Savings, on Average," ID: G00299793, March 2016

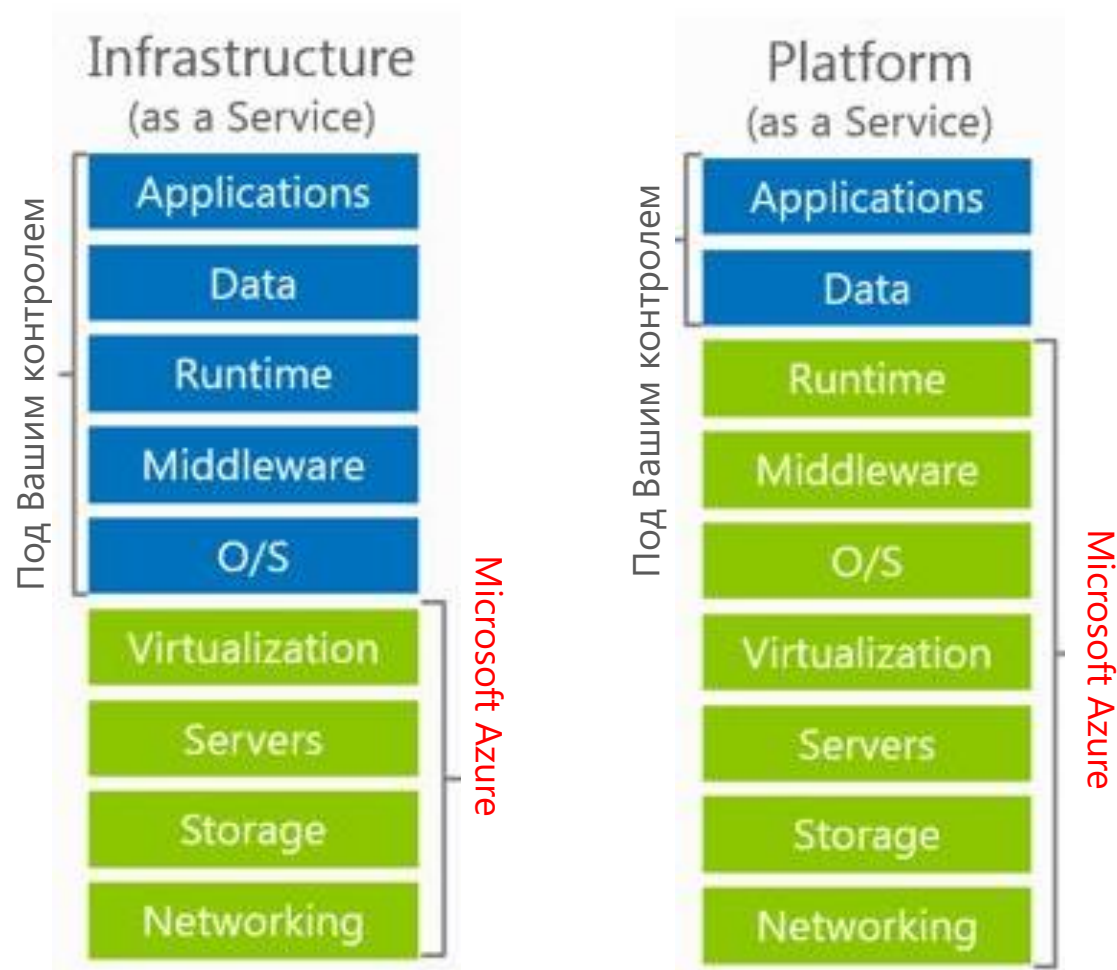
Культура данных



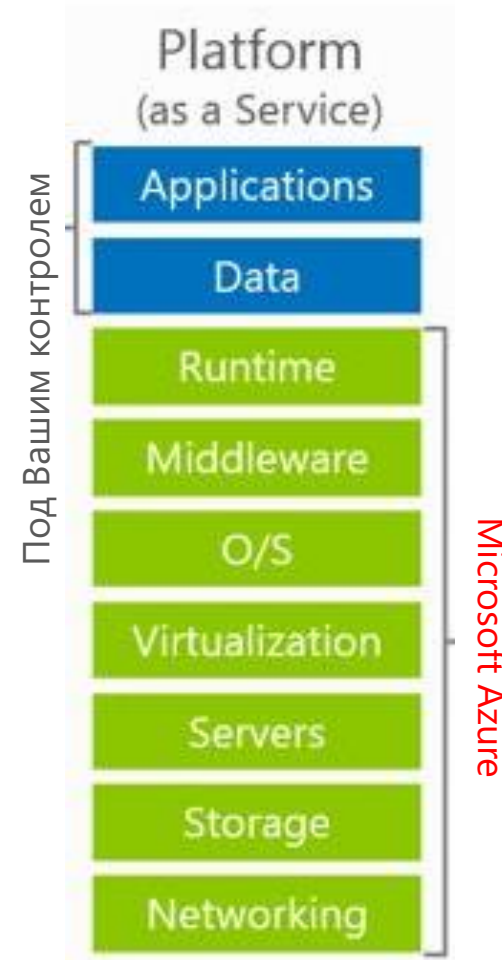
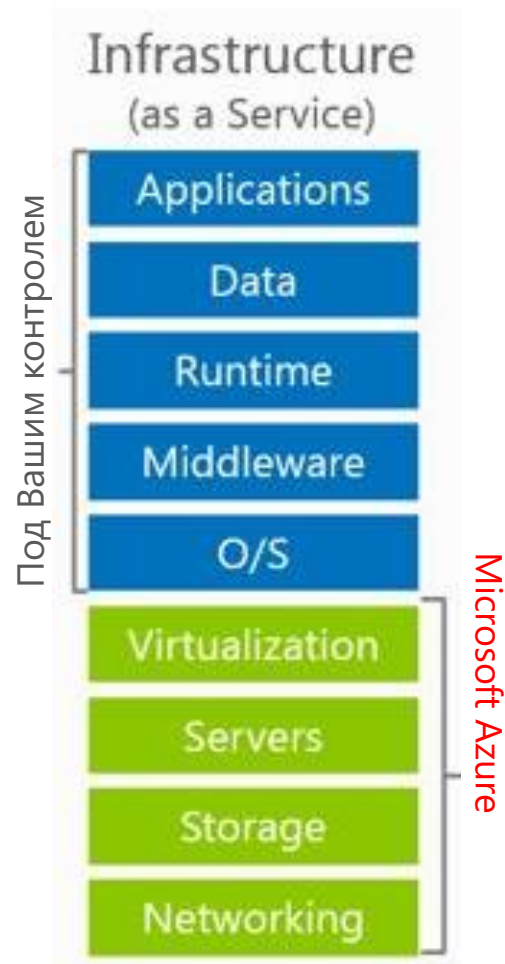
IaaS и PaaS



IaaS и PaaS



IaaS и PaaS



SQL в Облаке: Azure SQL и Managed Instance



Azure SQL Database

Отдельная

Вариант развертывания с предсказуемой производительностью рабочей нагрузки

↓
Подходит для приложений, которым требуется гарантия ресурсов на уровне базы данных

Эластичный пул

Модель общих ресурсов оптимизирована для большей эффективности мультитенантных приложений

↓
Подходит для SaaS-приложения с несколькими базами данных, которые могут совместно использовать ресурсы на уровне базы данных, обеспечивая более высокую экономическую эффективность

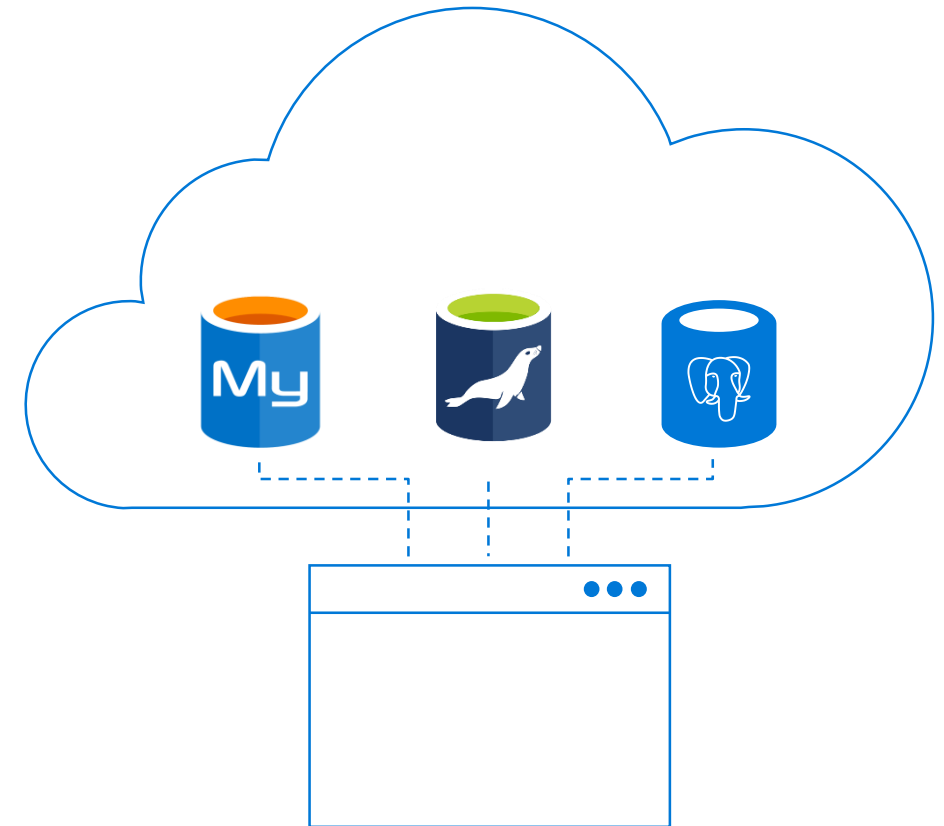
Управляемый экземпляр

Возможность развертывания в экземпляре с высокой совместимостью с SQL Server и всеми преимуществами PaaS

↓
Подходит для масштабной модернизации с минимумом усилий

Реляционные СУБД: PostgreSQL, MySQL, MariaDB

- Работа в БД по вашему выбору
- Community editions
- Создавайте со встроенной высокой доступностью
- Установить в минутах, масштаб на лету
- Поддерживать производительность благодаря адаптивным улучшениям
- Отдыхайте спокойно с беспрецедентной безопасностью



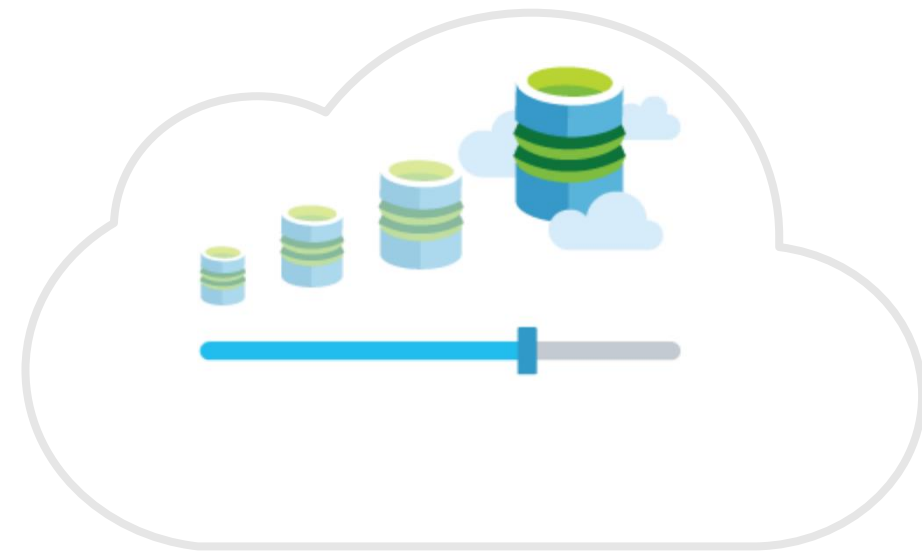
Lift and shift приложений
на MySQL и PostgreSQL

Миграция WordPress
и Drupal

Создавайте приложения с
OSS DB по выбору

Azure SQL Data Warehouse

Производительность и безопасность в масштабах в облаке



Unrivaed power



Производительность запросов в 10 раз выше, чем в традиционных хранилищах

Unlimited scale



Независимое и безграничное масштабирование хранилища и вычислений

Fast time to value



Полная интеграция с Microsoft и сторонними сервисами

Trusted & reliable



Повышенная безопасность благодаря шифрованию, аудиту, VNET и соответствию требованиям

NoSQL - CosmosDB

Глобально распределенная, масштабируемая, многомодельная служба баз данных

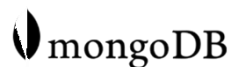
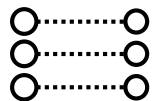


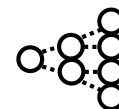
Table API



Key-value



Column-family



Document



Graph

Эластичность хранилища
и пропускной способности

Гарантированная
низкая задержка

Пять четко
определенных моделей

Глобально
распределенная

SLAs

Cloud | Security | Big Data | Mobility

Digital Transformation and Cybersecurity Solution Service Provider

softline® 25
Years in IT

HDInsight



Apache Storm
Streaming



Apache Spark
In-Memory



Apache Kafka
Real-Time



Apache Hadoop
Map Reduce



Apache HBase
NoSQL



Apache Hive LLAP
SQL

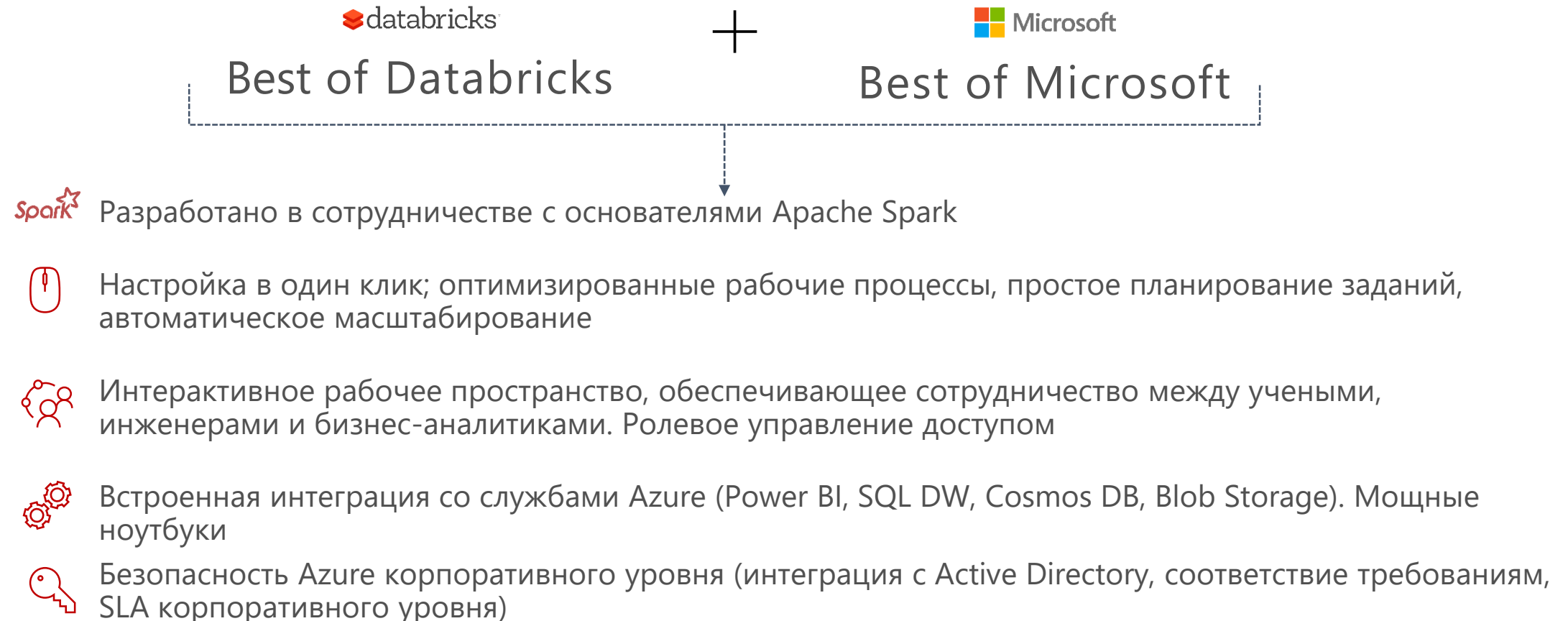
HDInsight

- Масштабируете до петабайт по запросу
- Обработка неструктурированных и псевдо-структурированных данных
- Разработка на Java, .NET и другие
- Развертывание Apache Hadoop кластера за минуты
- Разделение данных и вычислительных ресурсов
- Не нужно покупать и обслуживать железо

Data Bricks



Быстрая и простая аналитическая платформа для совместной работы на основе Apache® Spark™, оптимизированная для Azure

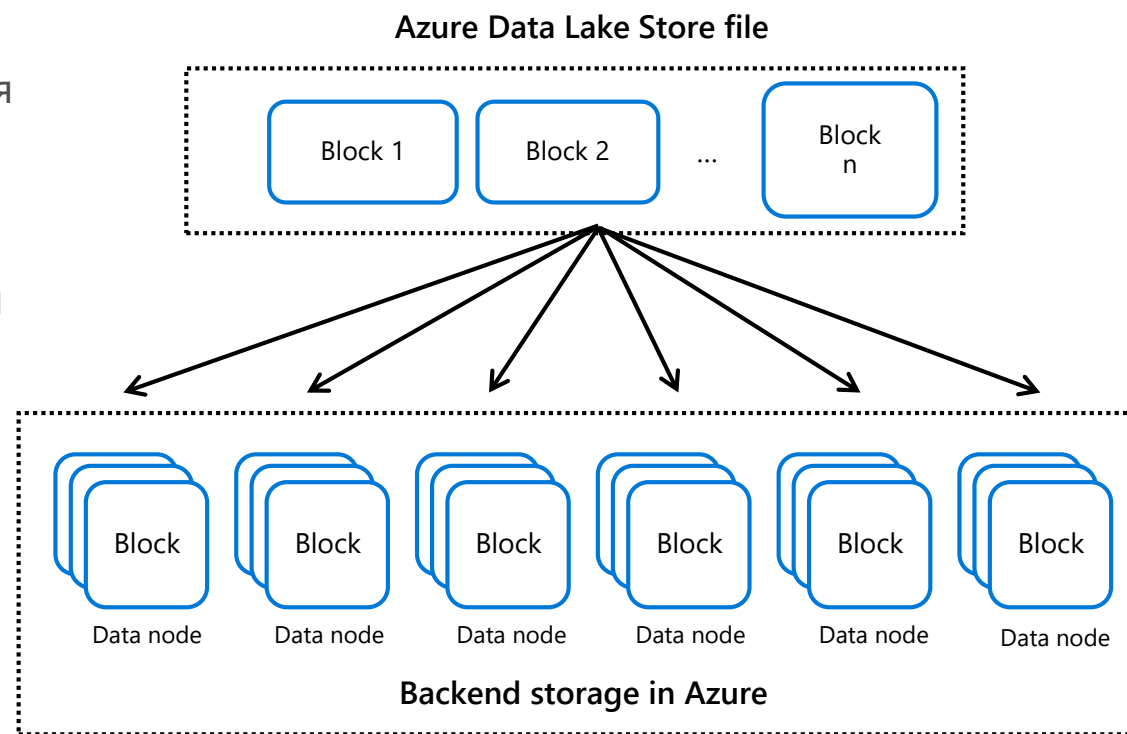


Data Lake Store и Data Lake Analytics



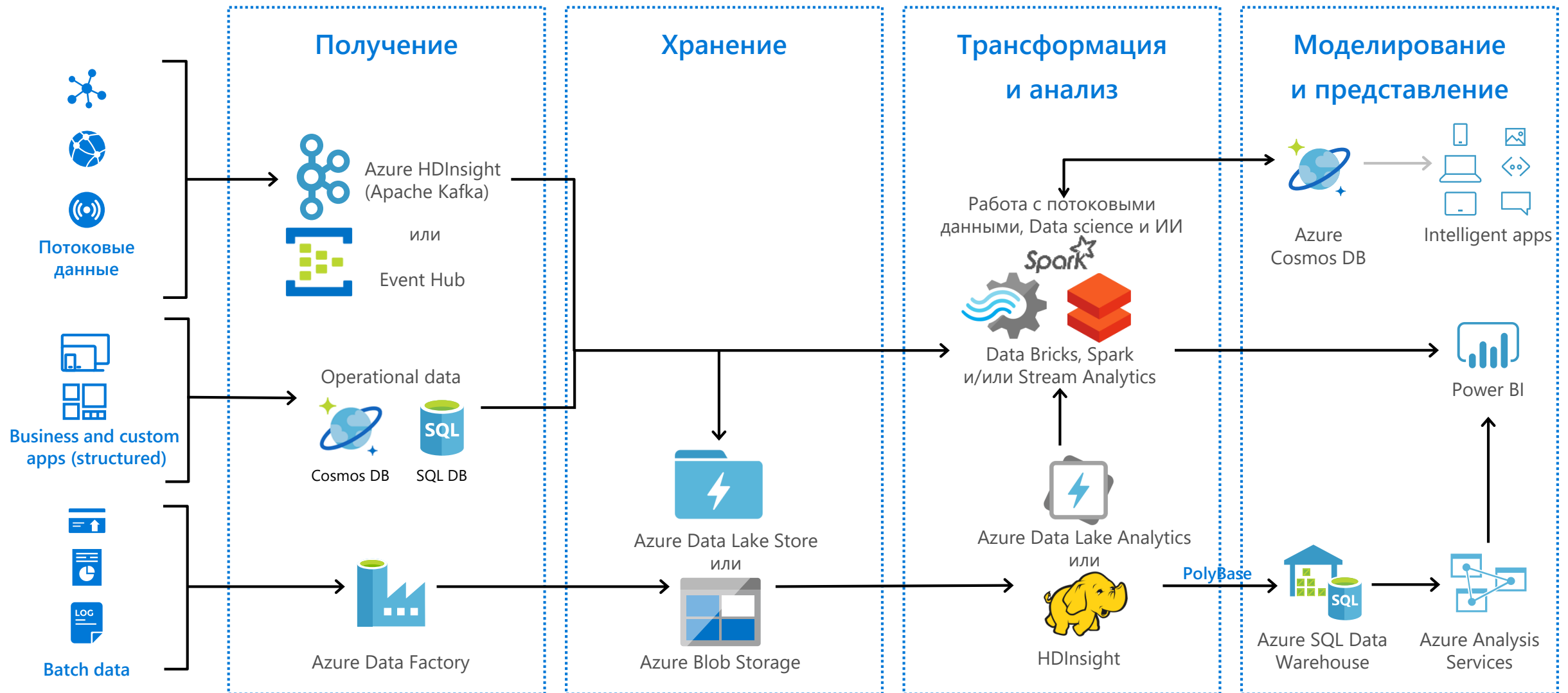
Высоко масштабируемая параллельная файловая система в облаке, оптимизированная для анализа больших данных

- Нет ограничений по количеству файлов, размеру отдельных файлов, общему объему хранимых данных, продолжительности хранения данных или скорости загрузки
- Для загрузки потоковых данных могут использоваться рабочие нагрузки с низкой задержкой и высокой пропускной способностью.
- Хранит все типы данных
- Совместим с Hadoop через REST API WebHDFS. Поддерживается ведущими дистрибутивами Hadoop и HDInsight



- Анализ данных любого типа и размера
- Быстро разрабатывать, отлаживать и оптимизировать
- Использование U-SQL
- Управляется и сопровождается с SLA
- Динамически масштабируется
- Безопасность Azure Active Directory
- Построен на YARN, разработан для облака

Интеграция и работа с потоковыми данными



Единая экосистема продвинутого анализа данных Microsoft



Azure Data Factory



Azure Import/Export service



Azure CLI



Azure SDK



Azure IoT Hub



Azure event hubs



Kafka on Azure HDInsight



Azure SQL DB



Azure Cosmos DB



Azure SQL data warehouse



Azure Analysis Services



Power BI



Azure Blob Storage



Azure Data Lake Store



Azure Data Lake Analytics



Azure HDInsight



Azure Databricks



Azure ML



ML Server



Azure Databricks



Azure Search



Azure Data Catalog



Azure Stream Analytics



Azure HDInsight



Azure Databricks



Bot service



Cognitive services



Azure ExpressRoute



Azure Active Directory



Azure network security groups



Azure key management service



Operations Management Suite



Azure Functions



Visual Studio



GO GLOBAL



GO CLOUD



GO INNOVATIVE