



БОЛАТ ТАЙМАГАМБЕТОВ

DIS Group

**Что такое Smart Data Lake.
Как сделать так, чтобы озеро не
превратилось в болото,
Практические бизнес-кейсы.**

**Smart Data Lake для
Data Driven Компании.
Как сделать так, чтобы озеро
не превратилось в болото**

Таймагамбетов Болат, БК, DIS Group

DATA DRIVEN В КАЗАХСТАНЕ 2020-2022

Интегрированные решения

Аналитика / DS



Фабрика данных / Озера данных



Каталог & Управление данными (DG)



MDM (НСИ)



Качество данных (DQ)



Использование данных, Маркетплейс данных

Бизнес-кейсы

- Улучшение процессов
- Лучшая Рекомендация
- Cross-sell, Upsell, Отток

НСИ

- Клиенты
- Поставщики
- Продукты/Материалы
- Финансы
- Справочники

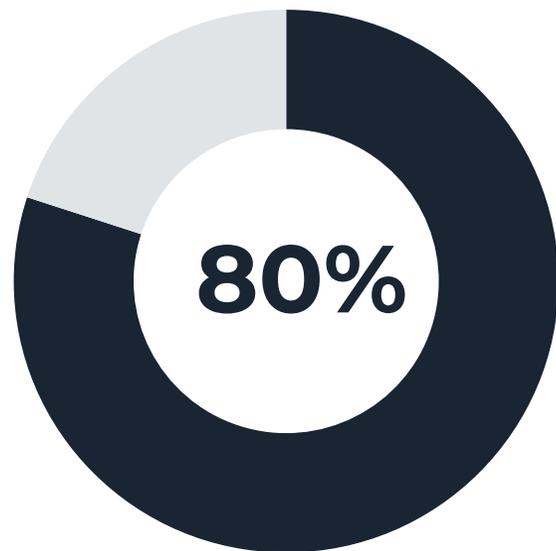
AI / ML / Data Science

- Кластеризация
- Классификация
- Рекомендации

Аналитика

- Самообслуживание
- Потоковая аналитика
- Продвинутая аналитика

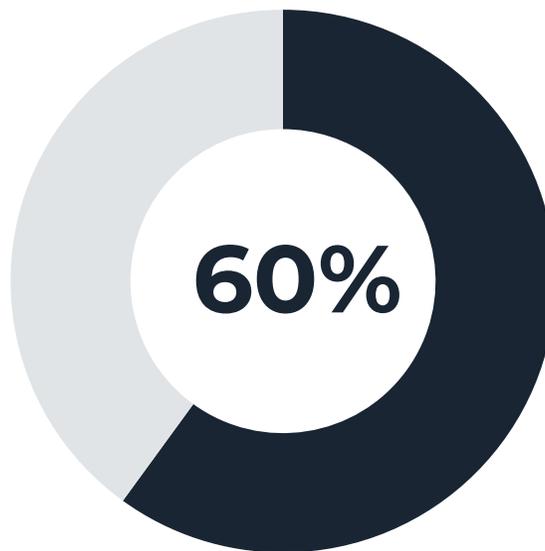
Профиль HE «Intelligent» компании



Поиск и Подготовка данных

Обычная организация затрачивает **80%** времени на поиск и подготовку данных

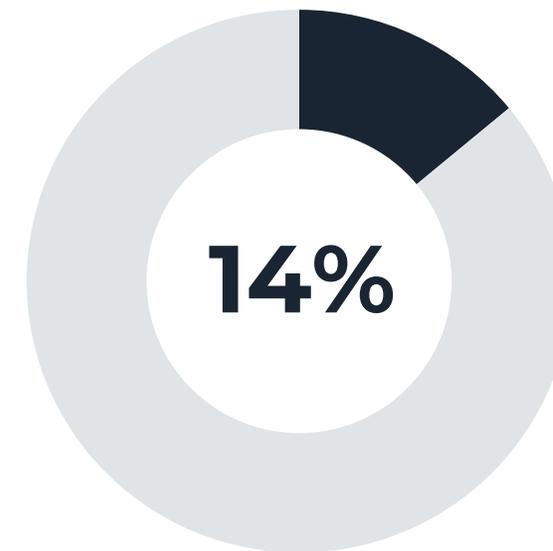
Источник: IDC, Advance Predictive Analytics Survey



Качество данных и Управление конвейерами монетизации данных

В среднем **60%** организаций сталкиваются с вызовами качества данных и управления конвейерами монетизации данных

Источник: IDC, Data integration and Integrity End User Survey, 2019

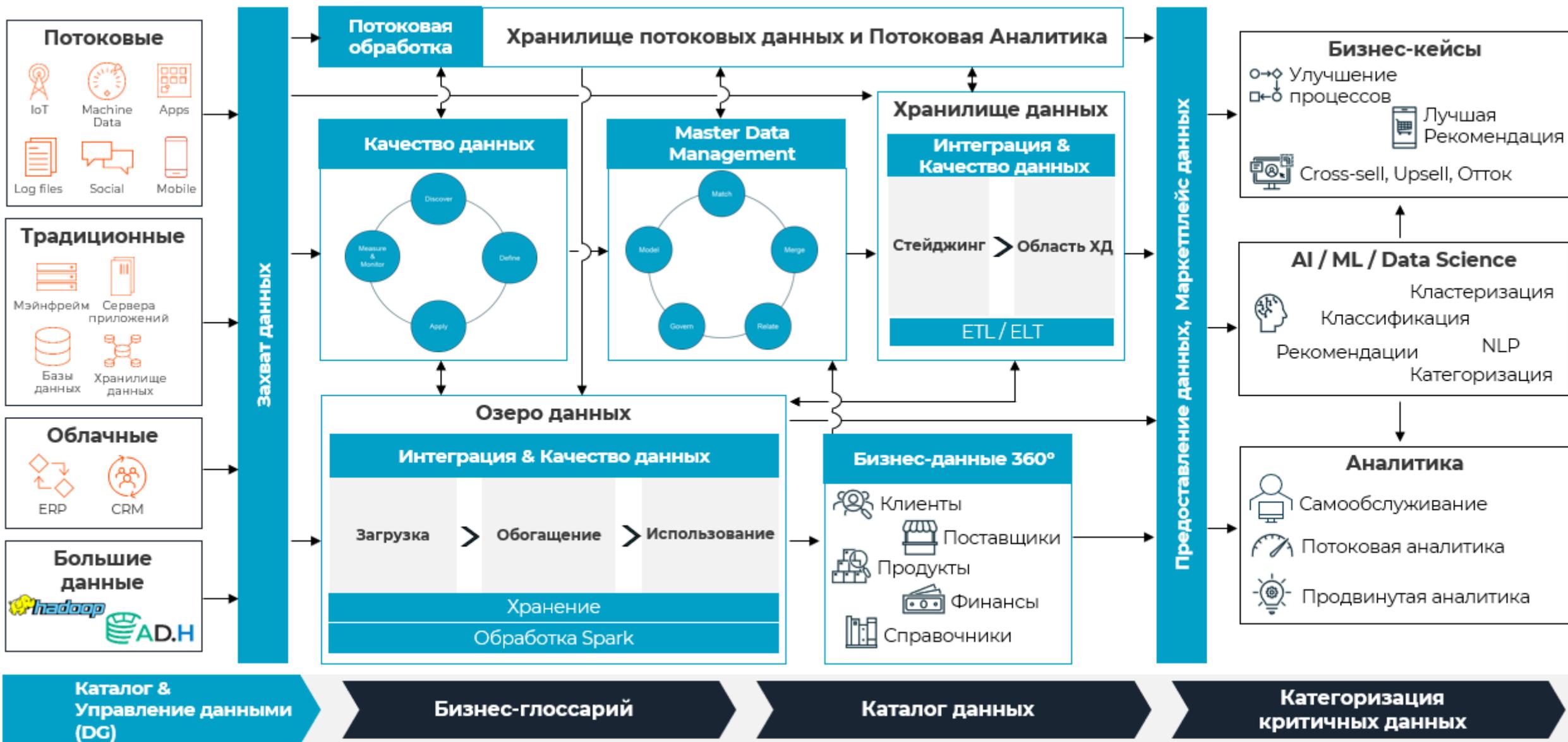


Результаты Data Science

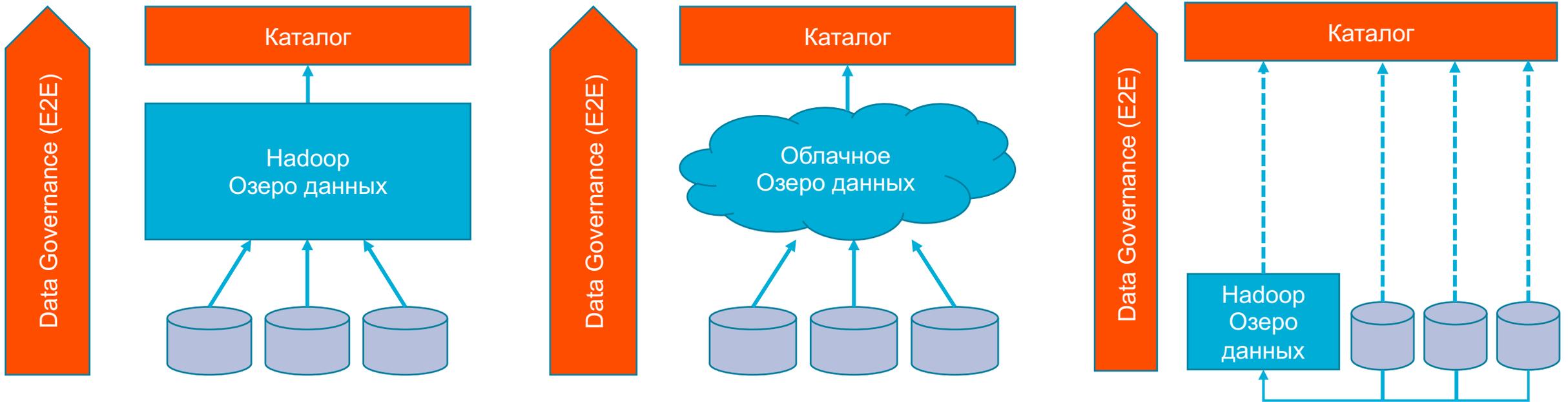
В среднем лишь **14%** организаций могут оперативно использовать результаты Data Science

Источник: IDC, Advance Predictive Analytics Survey

КОНЦЕПТ АРХИТЕКТУРЫ INTELLIGENT ENTERPRISE



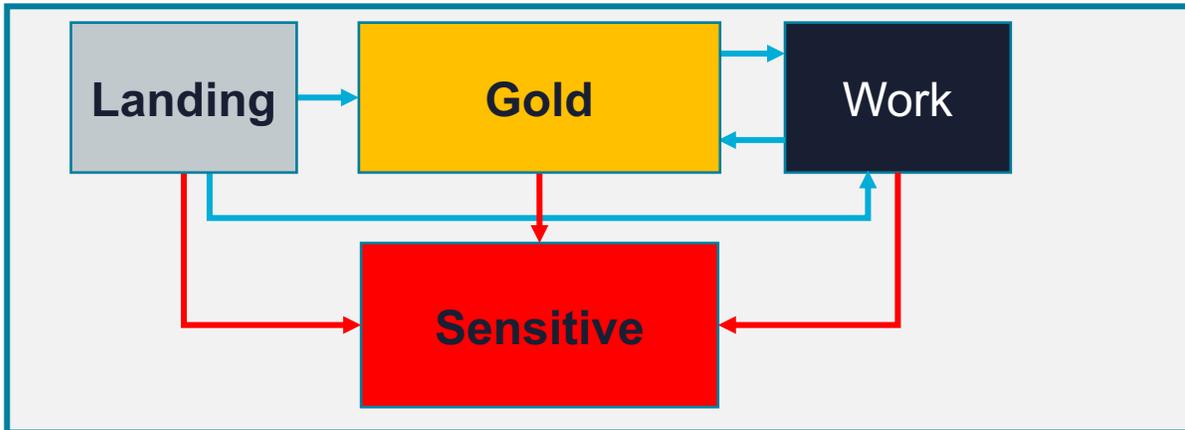
Умное Озеро данных vs Болото Данных



Озеро данных без бизнес-гlossария, каталога данных, средств управления качеством и критичными данными, маркетплейс данных = Болото Данных

Зонирование Data Lake и ролевое распределение DG

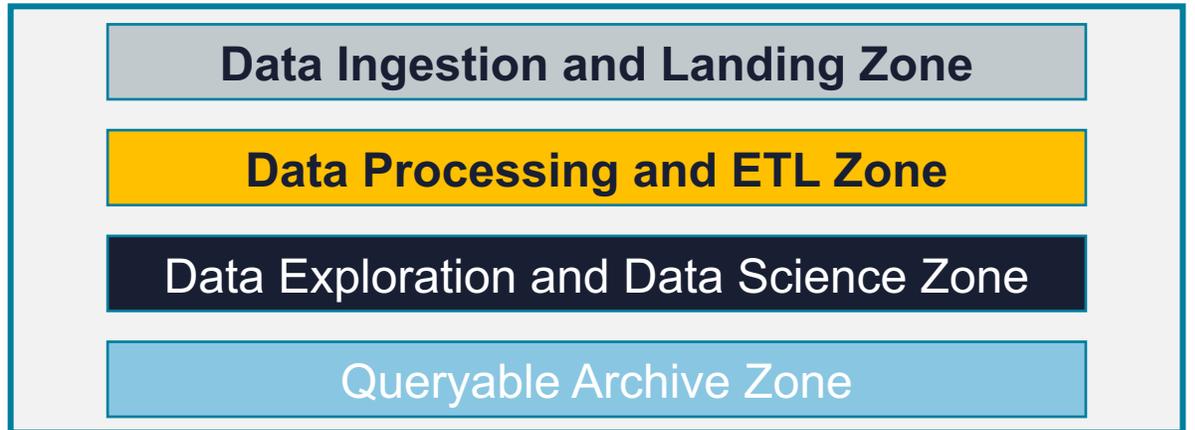
Enterprise Big Data Lake (есть ограничения)



Data Engineer
Data Scientist, Business Analyst
Data Scientist
Data Steward, Security Officer

Landing
Gold
Work
Sensitive

Gartner Best Practice**



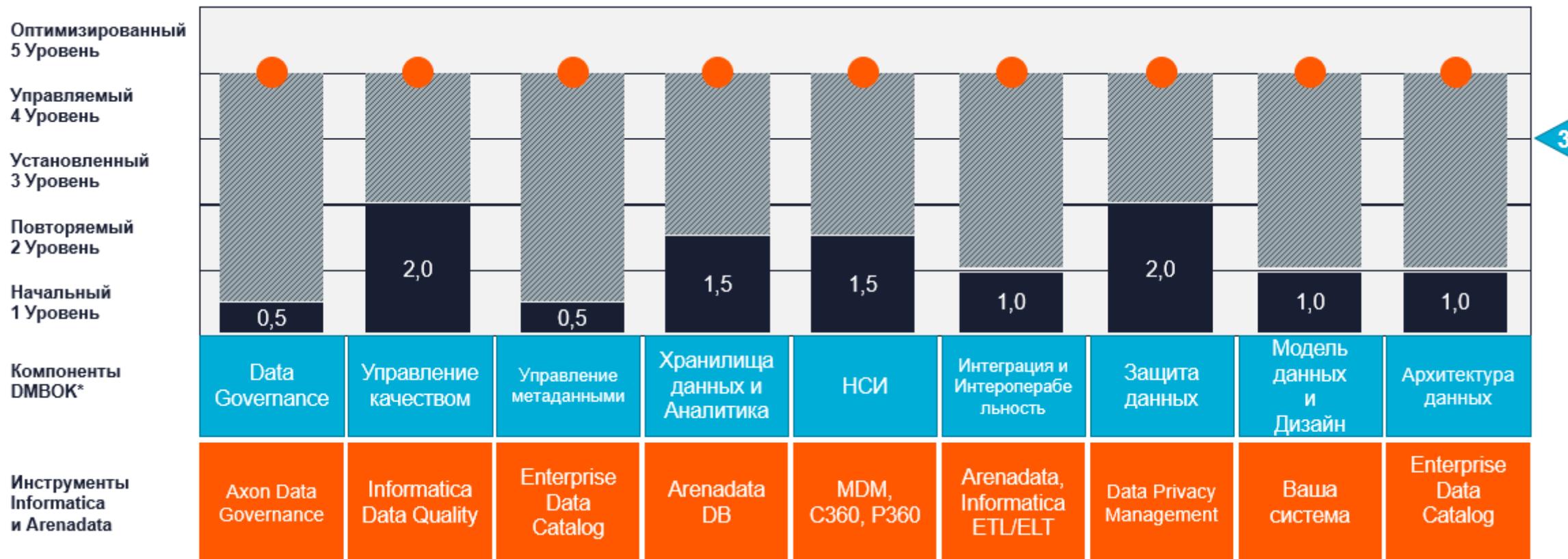
<ul style="list-style-type: none"> - Минимальный контроль и DG - Обеспечение отсутствия «чувствительных данных»
<ul style="list-style-type: none"> - Строгий контроль и DG - Доверенные, курируемые данные, управляемые данные - Data Lineage, Качество данных
<ul style="list-style-type: none"> - Минимальный контроль и DG - Обеспечение отсутствия «чувствительных данных»
<ul style="list-style-type: none"> - Строгий контроль и DG - Ограниченный доступ

* Источник: Gartner - How to Avoid Data Lake Failures (2018)

Ключевые аспекты для оценки зрелости Озера данных

АСПЕКТ	ОЦЕНКА
Люди и компетенции	
Процессы Data Governance и Data Management	
Организационная структура	
Защита и безопасность данных	
Интерфейсы	
Платформа	
Данные	

Пример оценки зрелости аспекта Процессы «Управления данными»



Легенда

● - Пример «Целевое состояние»
 - Пример «Текущее состояние»
 - Пример «Разрыв» - Инвестиции/Усилия до цели
 ◀ 3 - Минимальный уровень

* Пример, перечислены не все компоненты

Фабрика данных. Народный Банк

Высокий приоритет ■ ■ ■ Низкий приоритет





**СПАСИБО
ЗА ВНИМАНИЕ!**

