



**Hewlett Packard  
Enterprise**

# Эволюция Big Data: Готовимся к изменениям

Евгений Степанов, HPE Big Data Platform

**VERTICA**

---

В современной экономике изменить мир может каждый



Крупнейшая в  
мире служба  
такси не имеет  
автомобилей

Крупнейшее в  
мире хранилище  
медиа, не  
создающая  
никакого контента

Крупнейшая в  
России торговая  
площадка без  
единой позиции на  
складе

Крупнейший в  
России  
кинотеатр без  
единого зала

---

## Как мы обычно начинаем анализировать Большие данные



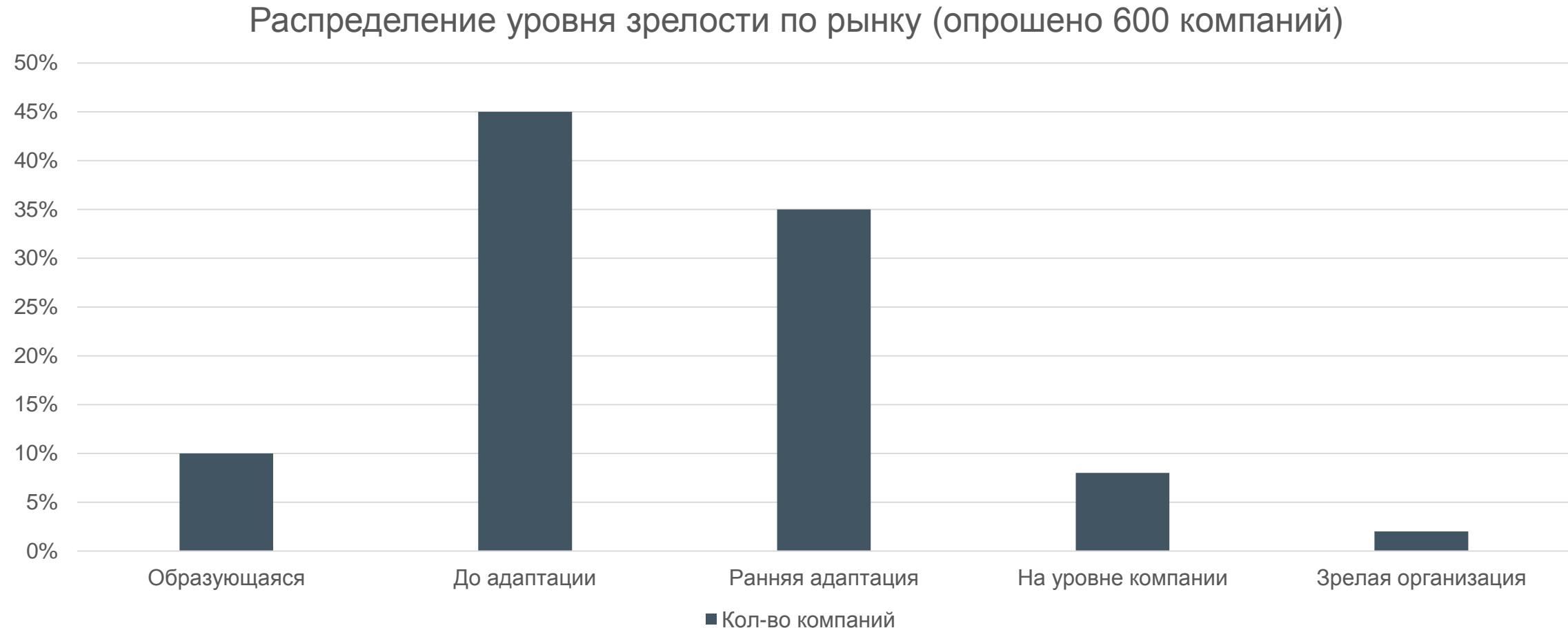
# Модель зрелости Big Data

| Уровень зрелости      | Определение   |
|-----------------------|---|
| 1. Образующаяся       | Отсутствие аналитической экосистемы                           |
| 2. До адаптации       | Начало обсуждения Больших данных, недостаток понимания задачи |
| 3. Ранняя адаптация   | Тестирование проектов по Большим данным                       |
| 4. На уровне компании | Аналитика Больших данных трансформирует бизнес                |
| 5. Зрелая             | Аналитика используется как конкурентное оружие                |

# Модель зрелости Big Data

| Уровень зрелости      | Характеристики  |  |
|-----------------------|---|--|
| 1. Образующаяся       | <ul style="list-style-type: none"><li>- Низкая информированность, отсутствие финансирования, отсутствие мотивации на аналитику</li><li>- Отсутствие архитектуры на уровне компании</li></ul>      | <ul style="list-style-type: none"><li>- Наличие хранилища данных, но не как части инициативы Big Data</li><li>- Аналитика сосредоточена в бизнес-функциях</li></ul>          |
| 2. До адаптации       | <ul style="list-style-type: none"><li>- Посредственное вовлечение руководства, начало диалога</li><li>- Рассмотрение вариантов инфраструктуры</li></ul>   | <ul style="list-style-type: none"><li>- Разрозненные группы хранилищ</li><li>- Разрозненные аналитические функции</li><li>- Зарождение управления качеством данных</li></ul> |
| 3. Ранняя адаптация   | <ul style="list-style-type: none"><li>- Большее вовлечение на уровне организации</li><li>- Использование новых технологий</li><li>- Большие объемы структурированных данных</li></ul>             | <ul style="list-style-type: none"><li>- Усложнение аналитики, например, предиктивная аналитика</li><li>- Появление команды контроля качества данных</li></ul>                |
| 4. На уровне компании | <ul style="list-style-type: none"><li>- Большие данные – часть стратегии компании</li><li>- Создана унифицированная архитектура</li><li>- Эффективное использование разрозненных данных</li></ul> | <ul style="list-style-type: none"><li>- Аналитика встроена в бизнес-процессы организации</li><li>- Документированы политики управления данными</li></ul>                     |
| Зрелая                | <ul style="list-style-type: none"><li>- Превалирует аналитическая культура</li><li>- Внедрена экосистема на уровне компании</li></ul>   | <ul style="list-style-type: none"><li>- Аналитика автоматизирована и интегрирована во все процессы компании</li></ul>  |

# Текущее состояние рынка



---

Coming soon..



---

# Основные проблемы при росте объемов данных

- Ограничения производительности и масштабирования
- Ограничения функциональности
- Высокая стоимость решения
- Проблемы отказоустойчивости

---

# Ограничение масштабирования и производительности

## Типичные решения

- Вертикальное масштабирование
- Добавление проприетарных вычислительных и коммуникационных компонент (апплаенсы)
- Использование технологий *in-memory*

## Результат

- Потолок масштабирования
- Высокая стоимость решения
- Высокая стоимость расширения
- Высокая стоимость решения
- Низкая производительность
- Ограничения масштабирования

# Ограничение функциональности

## Типичные решения

- Использование проприетарного процедурного языка для решения несвойственных для хранилища задач
- Нестандартная функциональность (коннекторы, преобразователи)

## Результат

- Вендор лок
- Внутренний вендор лок, высокая стоимость сопровождения

---

# Высокая стоимость решения

## Типичные решения

–Использование open-source продуктов при отсутствии экспертизы

## Результат

–Высокая скрытая стоимость (необходимо развитие новой экспертизы, проблема болота данных, высокая стоимость сопровождения)

---

# Проблемы отказоустойчивости

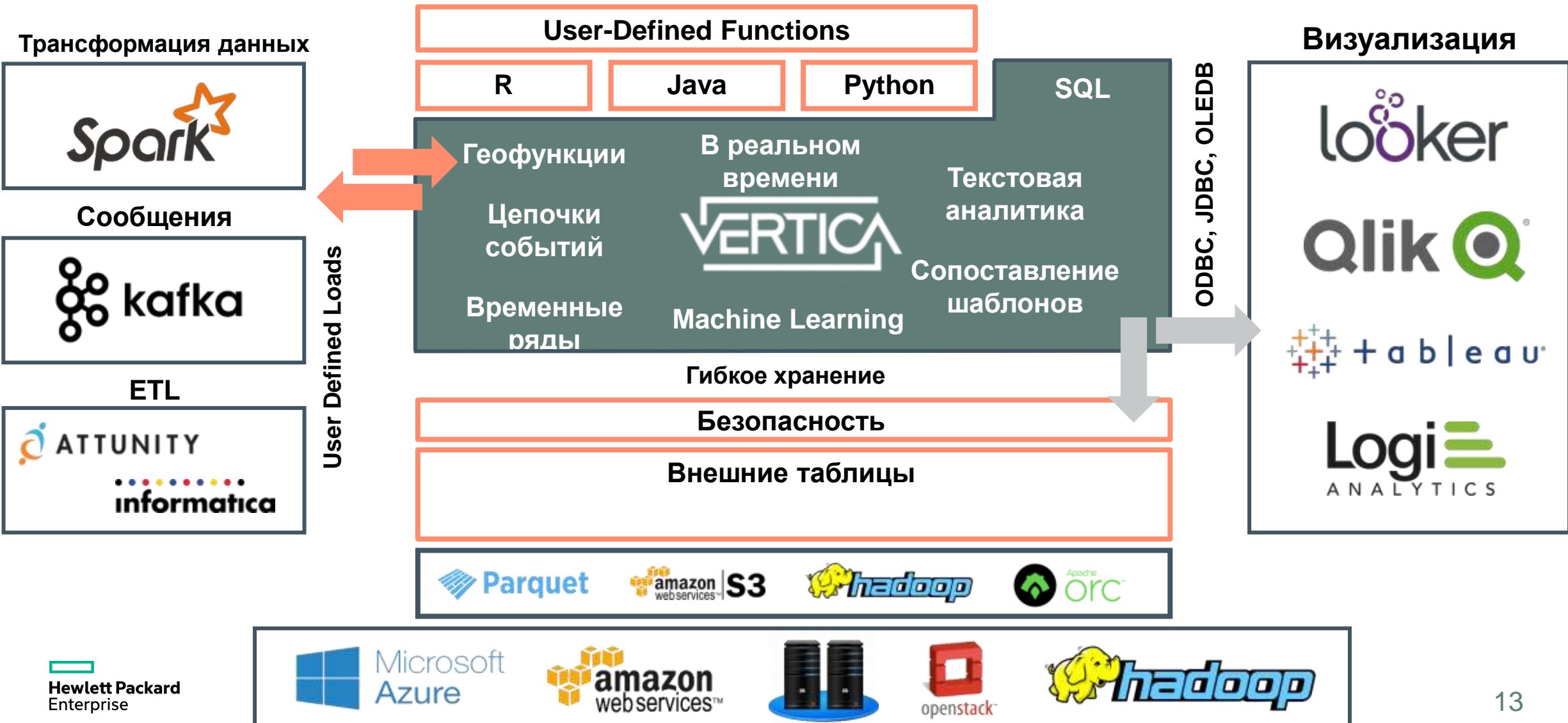
## Типичные решения

–Разработка отказоустойчивого решения при внедрении

## Результат

–Высокая сложность и стоимость итогового решения

# Платформа, интегрированная с инновациями через открытую архитектуру



---

# 3000 клиентов, начиная с 2006



# Facebook

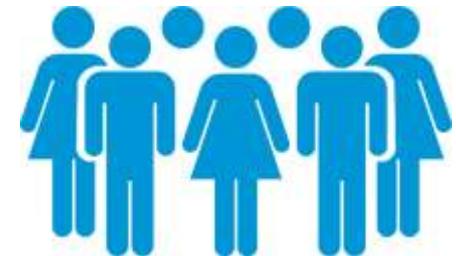
- Крупнейшая в мире инсталляция HP Vertica > 20ПБ (570 узлов)
- Общий объем данных в Hadoop – 1,7 Экзабайта (55000 узлов)
- Расширение аналитического окна индивидуальной рекламы от 30 дней до года
- Данные загружаются каждый час, а не раз в день, как раньше
- 158 млрд строк в день загружаются ежедневно в HP Vertica
- Запросы выполняются за секунды



# Avito



- Общий объем хранилища 76ТБ
- Автоматическая модерация размещаемых на площадке Avito.ru объявлений.
- 5 млн объявлений в день
- 80% из многомиллионного числа публикуемых объявлений блокируется из-за недопустимого контента.
- Задержка попадания транзакционных данных в хранилище составляет не более 20 минут
- Используется инструментарий бизнес-аналитики Tableau



---

## Полезные ссылки

- Каталог «Платформа и решения HPE для Больших данных»  
<http://bit.ly/1Nxsmx5>
- Статья «Обзор HPE Vertica 8 (Frontloader)»: <https://habrahabr.ru/post/309550/>
- Бесплатная версия HPE Vertica Community Edition: <http://my.vertica.com>
- Вебинар Авито, 4 года с Vertica:  
[https://www.brighttalk.com/webcast/13185/242829?utm\\_campaign=webcasts-search-results-feed&utm\\_content=Avito&utm\\_source=brighttalk-portal&utm\\_medium=web](https://www.brighttalk.com/webcast/13185/242829?utm_campaign=webcasts-search-results-feed&utm_content=Avito&utm_source=brighttalk-portal&utm_medium=web)



**Hewlett Packard  
Enterprise**

Спасибо

[evgeny.stepanov@hpe.com](mailto:evgeny.stepanov@hpe.com)