



Hewlett Packard
Enterprise

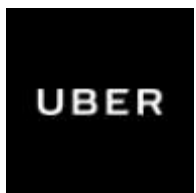
Эволюция Big Data: ГОТОВИМСЯ К ИЗМЕНЕНИЯМ

Евгений Степанов, HPE Big Data Platform

The Vertica logo, featuring the word "VERTICA" in a white, bold, sans-serif font. The logo is centered within a dark gray rectangular area, which is itself set against a larger green rectangular background.

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

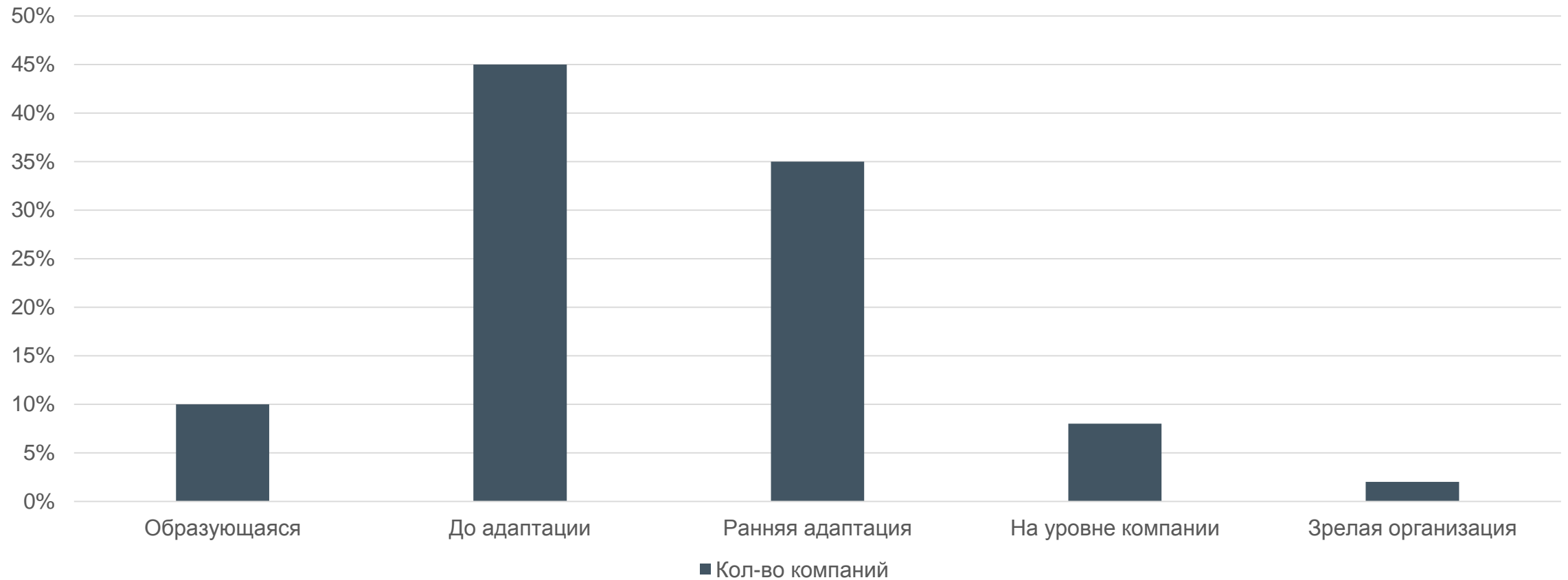
| Уровень зрелости | Определение |
|-----------------------|---------------------------------------------------------------|
| 1. Образующаяся | Отсутствие аналитической экосистемы |
| 2. До адаптации | Начало обсуждения Больших данных, недостаток понимания задачи |
| 3. Ранняя адаптация | Тестирование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

Модель зрелости Big Data

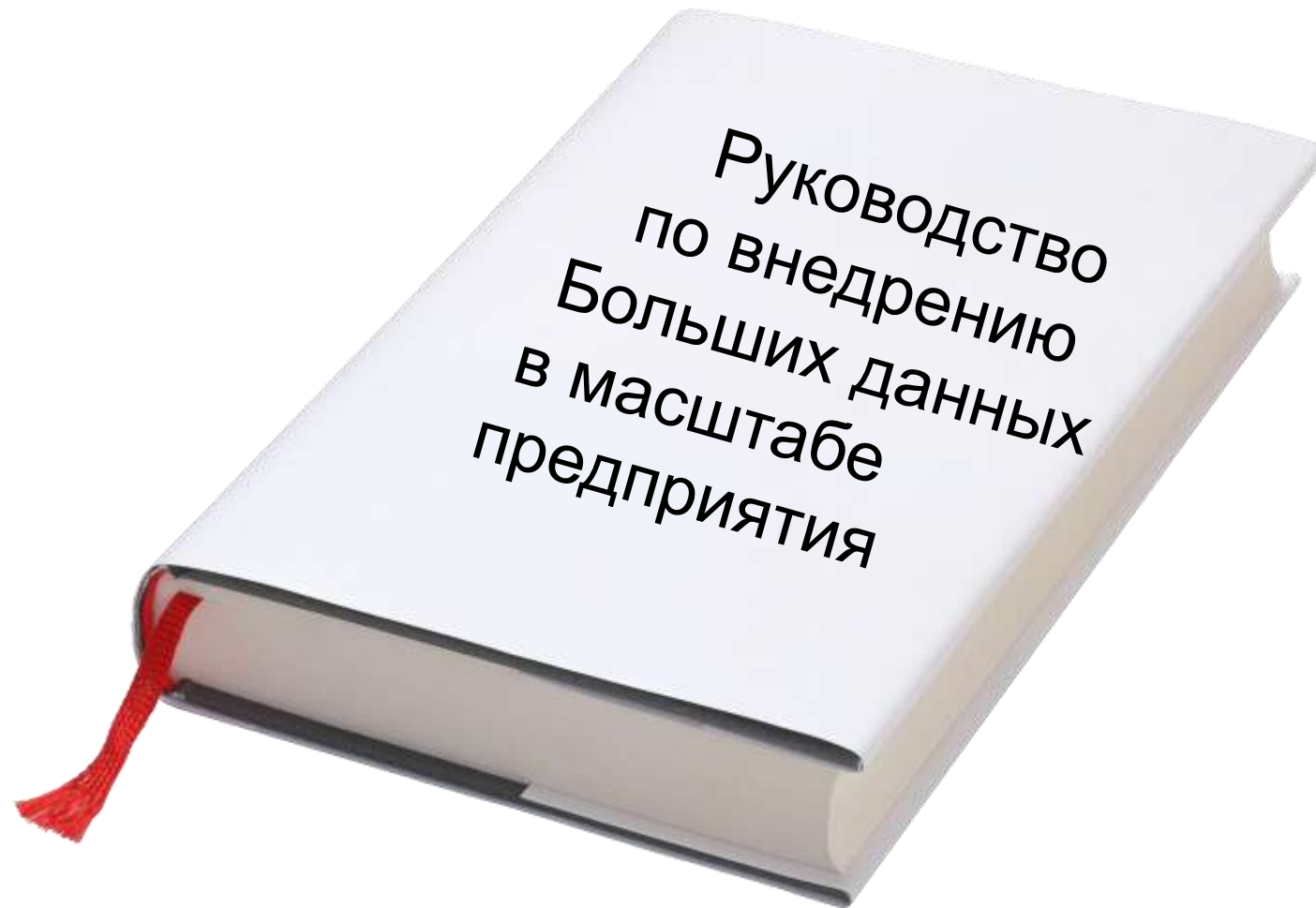
| Уровень зрелости | Характеристики | |
|-----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Образующаяся | <ul style="list-style-type: none">- Низкая информированность, отсутствие финансирования, отсутствие мотивации на аналитику- Отсутствие архитектуры на уровне компании | <ul style="list-style-type: none">- Наличие хранилища данных, но не как части инициативы Big Data- Аналитика сосредоточена в бизнес-функциях |
| 2. До адаптации | <ul style="list-style-type: none">- Посредственное вовлечение руководства, начало диалога- Рассмотрение вариантов инфраструктуры | <ul style="list-style-type: none">- Разрозненные группы хранилищ- Разрозненные аналитические функции- Зарождение управления качеством данных |
| 3. Ранняя адаптация | <ul style="list-style-type: none">- Большее вовлечение на уровне организации- Использование новых технологий- Большие объемы структурированных данных | <ul style="list-style-type: none">- Усложнение аналитики, например, предиктивная аналитика- Появление команды контроля качества данных |
| 4. На уровне компании | <ul style="list-style-type: none">- Большие данные – часть стратегии компании- Создана унифицированная архитектура- Эффективное использование разрозненных данных | <ul style="list-style-type: none">- Аналитика встроена в бизнес-процессы организации- Документированы политики управления данными |
| 5. Зрелая | <ul style="list-style-type: none">- Превалирует аналитическая культура- Внедрена экосистема на уровне компании | <ul style="list-style-type: none">- Аналитика автоматизирована и интегрирована во все процессы компании |

Текущее состояние рынка

Распределение уровня зрелости по рынку (опрошено 600 компаний)



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

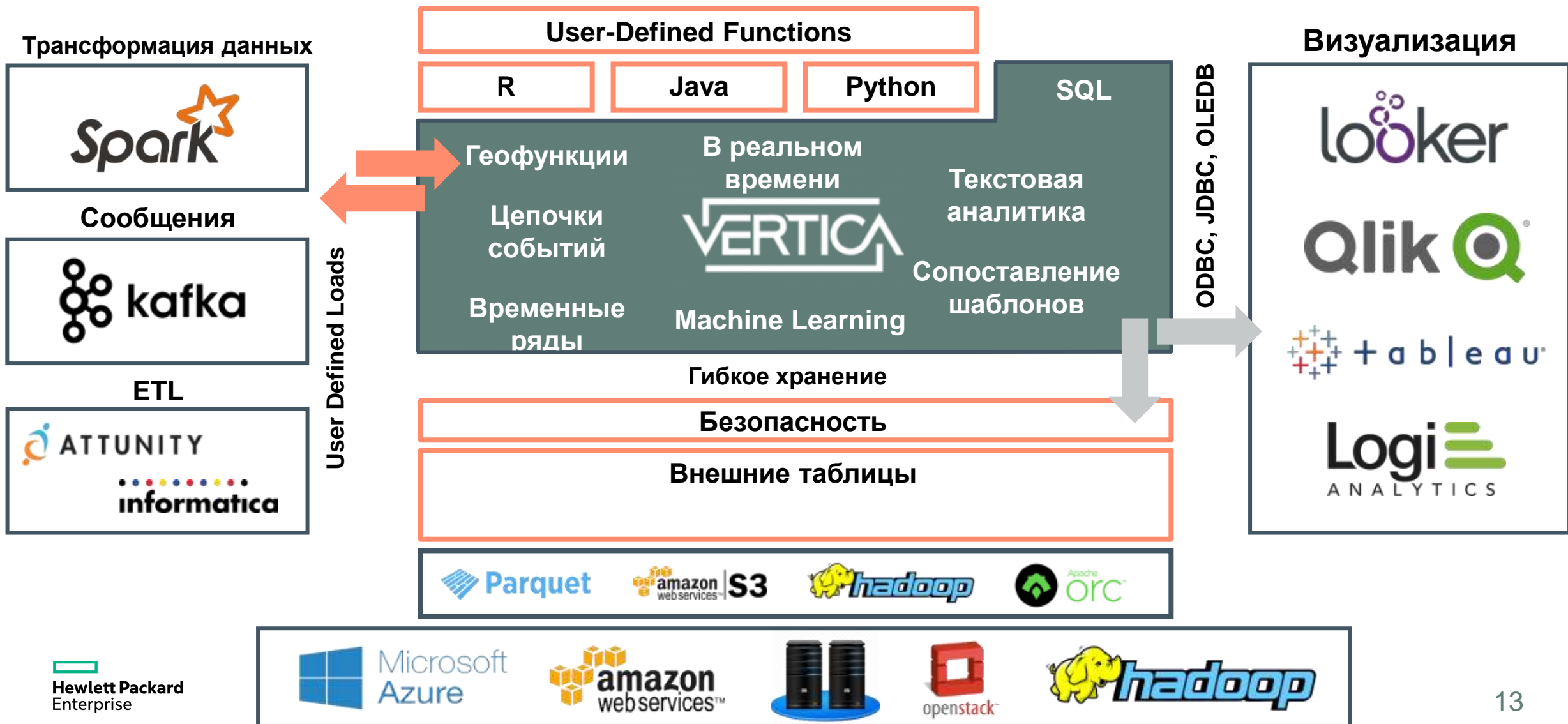
Типичные решения

– 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



Hewlett Packard
Enterprise

Спасибо

evgeny.stepanov@hpe.com