VERTICA

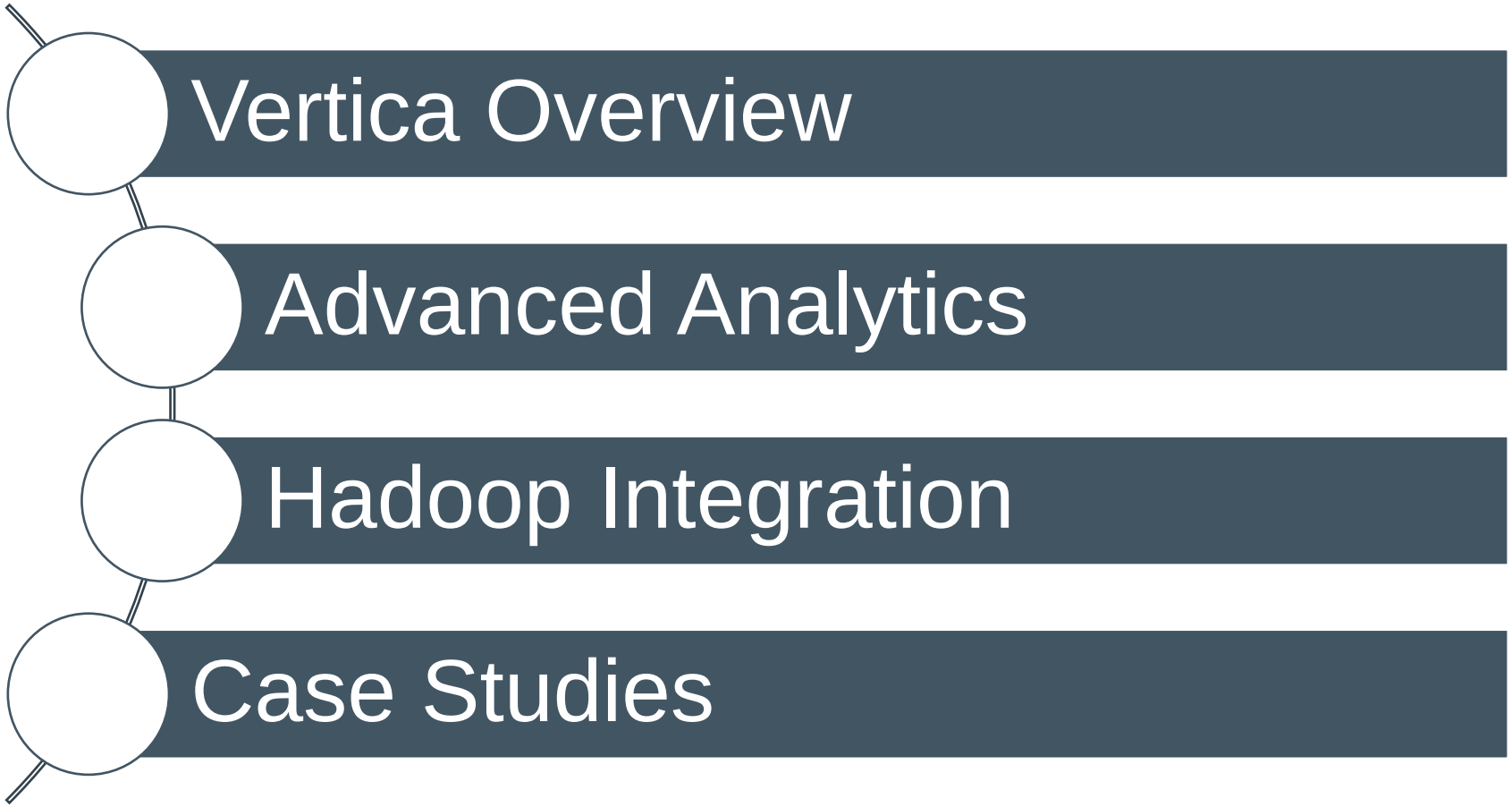
Data Warehouse Modernization

Benjamin Smith

Vertica Product Marketing
Benjamin.C.Smith@hpe.com

Hewlett Packard Enterprise

Agenda


Vertica Overview
Advanced Analytics
Hadoop Integration
Case Studies

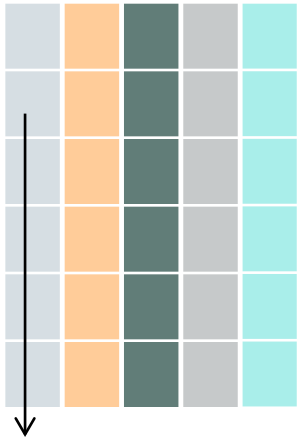
VERTICA

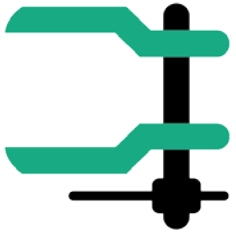
Vertica Overview

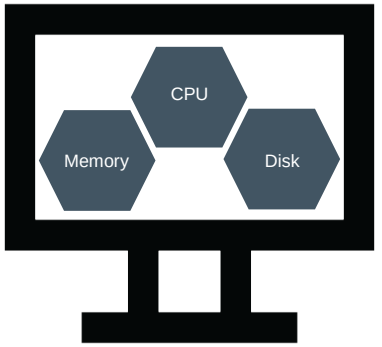
Hewlett Packard Enterprise

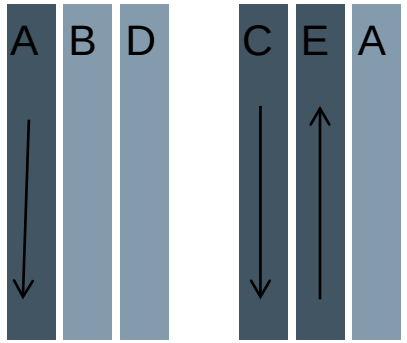
Foundations of Vertica

Columnar storage	Compression	MPP scale-out	Distributed query	Projections
Speeds query time by reading only necessary data	Lowers costly I/O to boost overall performance	Provides high scalability on clusters with no name node or other single point of failure	Any node can initiate the queries and use other nodes for work. No single point of failure	Combine high availability with special optimizations for query performance





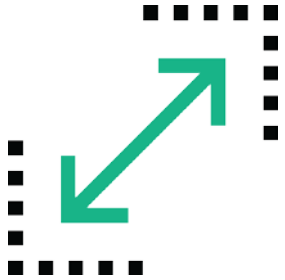




HPE Vertica analytics platform

Fast

Boost performance by 500% or more



Scalable

Handles huge workloads at high speeds

Standard

No need to learn new languages or add complexity

Costs

Significantly lower cost over legacy platforms

HPE Vertica portfolio

The broadest range of deployment and consumption models

Vertica unified platform

- Cloud
- On-Premise
- On Hadoop

HPE Vertica in the Clouds

- Deploy quickly on the choice of cloud platform
- Support for AWS, Azure, VMware
- Flexible, enterprise-class cloud deployment options

HPE Vertica Enterprise

- Columnar storage and advanced compression
- Maximum performance and scalability
- Flex Tables for schema-on-read

HPE Vertica for SQL on Apache® Hadoop

- Native support for ORC, Parquet, more
- Support for industry-leading distributions
- No helper node or single point of failure

VERTICA

Hewlett Packard Enterprise

Vertica Partner Ecosystem

Enabled in large part by Vertica's strong support for the SQL standard

VERTICA

Advanced Analytics

Hewlett Packard Enterprise

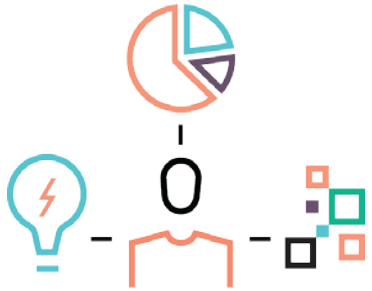
Advanced In-database Analytics

SQL ‘99

- Aggregate
- Analytical
- Window functions
- Graph
- Monte Carlo
- Statistical
- Geospatial

Allows for:

- Standard functionality that performs at scale



SQL Extensions

- Pattern matching
- Event series joins
- Time series
- Event-based windows

Allows for:

- Sessionization
- Conversion analysis
- Fraud detection
- Fast Aggregates (LAP)

SDKs

Analytics

- Java
- C++
- R

Connection

- ODBC/JDBC
- HIVE
- Hadoop
- Flex zone

Allows for:

- Machine learning
- Custom data mining
- Specialized parsers

In-database Analytics

- Regression testing
- K-means
- Statistical modeling
- Classification algorithms
- Page rank
- Text mining
- Geospatial

Allows for:

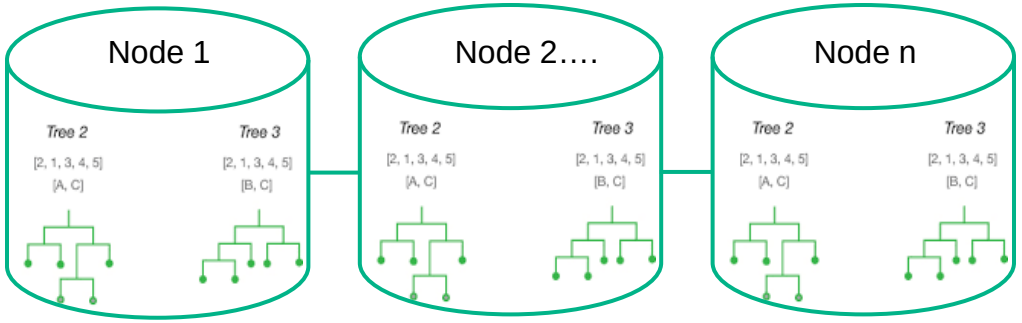
- Statistical modeling
- Cluster analysis
- Predictive analytics

Applied Machine Learning at Scale

Current Barriers to Machine Learning and Predictive Analytics

Added Cost	Requires Down Sampling	Slower Time to Development	Slower Time to Deployment
– Additional hardware required for building predictive models	– Cannot process large data sets due to memory and computational limitations, resulting in inaccurate predictions	– Higher turnaround times for model building/scoring and need for moving large volumes of data between systems	– Inability to quickly deploy predictive models into production

Building Machine Learning into the Core of Vertica



Machine learning functions run in parallel across hundreds of nodes in a Vertica cluster

- Machine Learning algorithms, such as k-means and regression, built into the core of Vertica

- Advanced predictive modeling runs within the database eliminating all data duplication typically required of alternative vendor offerings

- Traditional approaches can't handle many data points forcing data scientists to "down-sample" leading to less accurate predictions

- A single system for SQL analytics and Machine Learning

In-Database ML Algorithms

Algorithm	Model Training	Prediction	Evaluation
Linear Regression	✓	✓	✓
Logistic Regression	✓	✓	✓
K-means	✓	✓	✓
Naïve Bayes	✓	✓	✓
Support Vector Machine (SVM)	✓	✓	✓

Model Management	Summarize models	Rename models	Delete models
	✓	✓	✓

Data Preparation	Data Normalization
	✓

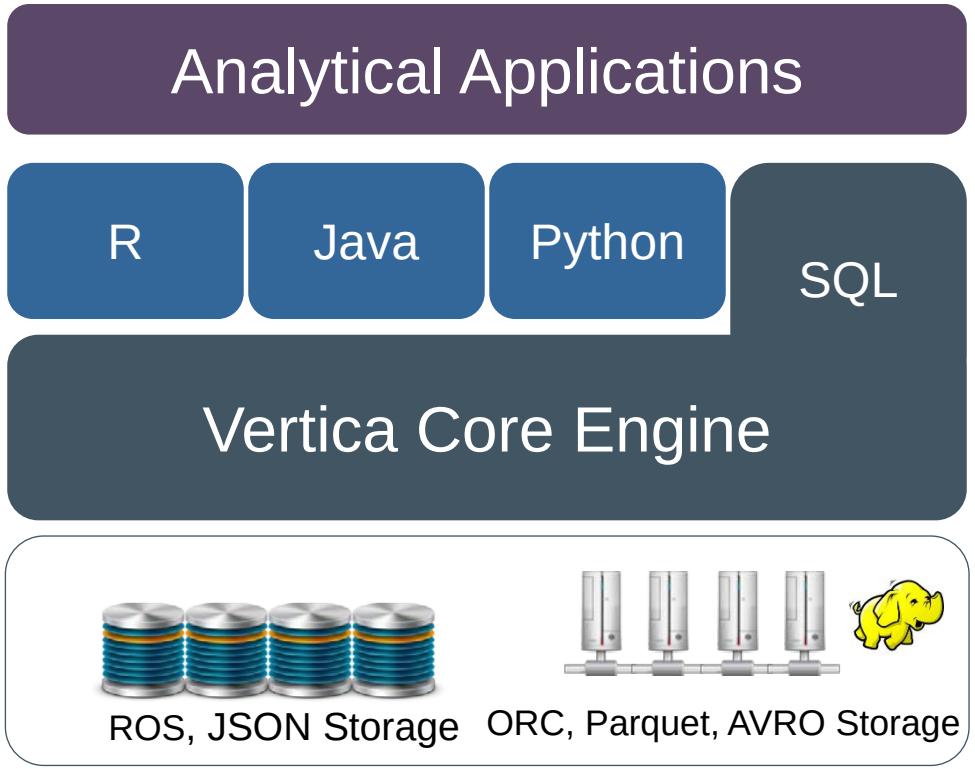
VERTICA

Hadoop Integration

Hewlett Packard Enterprise

HPE Vertica for SQL on Apache Hadoop®

- Same Vertica MPP Columnar Architecture
- Base ANSI SQL
- Co-Located with Hadoop
- Data Query Across Parquet, ORC, JSON, and many other format
- Optimized stack for query speed
  - Parallelized query planning and execution
  - Vectorized data loading, asynchronous data access
- Hadoop Agnostic



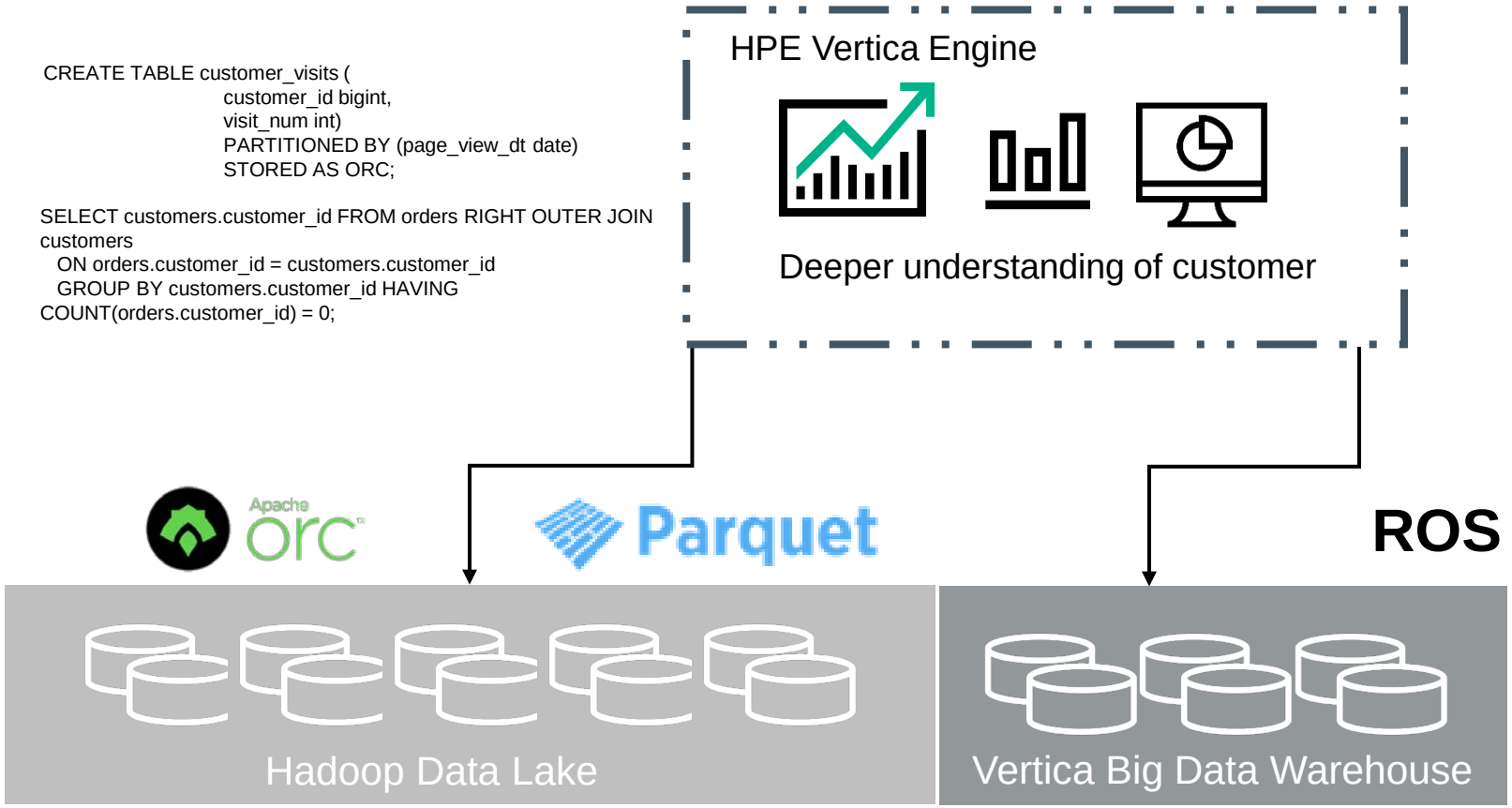
VERTICA

Hewlett Packard Enterprise

Vertica's Unique Value to Expand the Data Warehouse

```
CREATE TABLE customer_visits (
                    customer_id bigint,
                    visit_num int)
                    PARTITIONED BY (page_view_dt date)
                    STORED AS ORC;

SELECT customers.customer_id FROM orders RIGHT OUTER JOIN
customers
  ON orders.customer_id = customers.customer_id
  GROUP BY customers.customer_id HAVING
COUNT(orders.customer_id) = 0;
```



Customer information in Hadoop

Customer information in Data Warehouse

Querying data that sits BOTH in the data warehouse and Hadoop is our unique value.

Most solutions require that you move the data.

- Leveraging Web Logs to gain customer insight
- Sensor and IOT data for pre-emptive service
- Marketing Programs Tracking

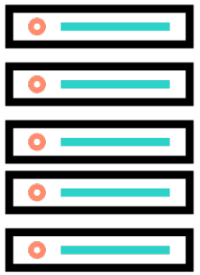
VERTICA

Hewlett Packard Enterprise

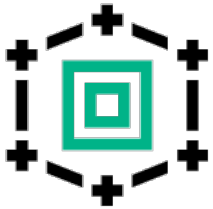
Proof is in the pudding

99 TPC-DS Queries

HPE Proliant DL 380 gen 9
Five nodes



3 TBs of test data

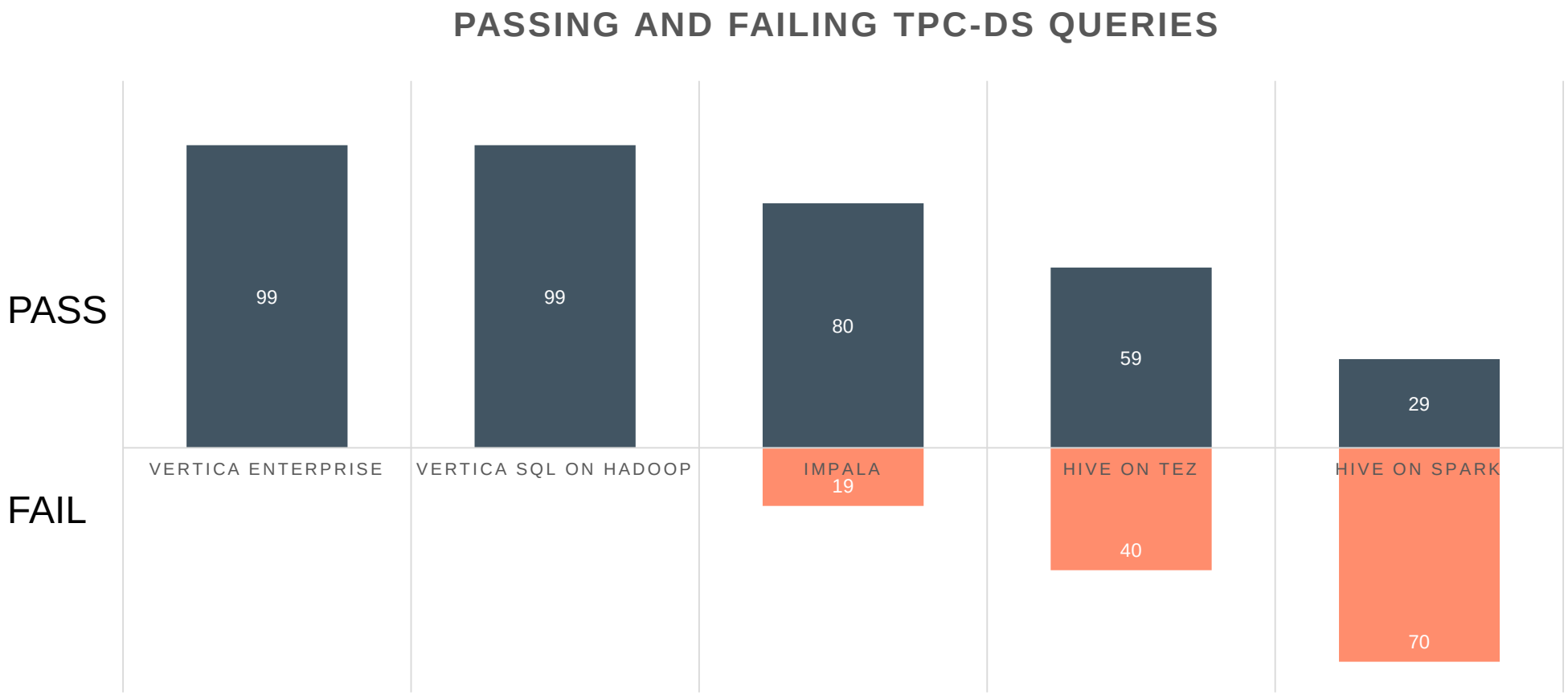


ROS, Parquet and ORC format.

VERTICA

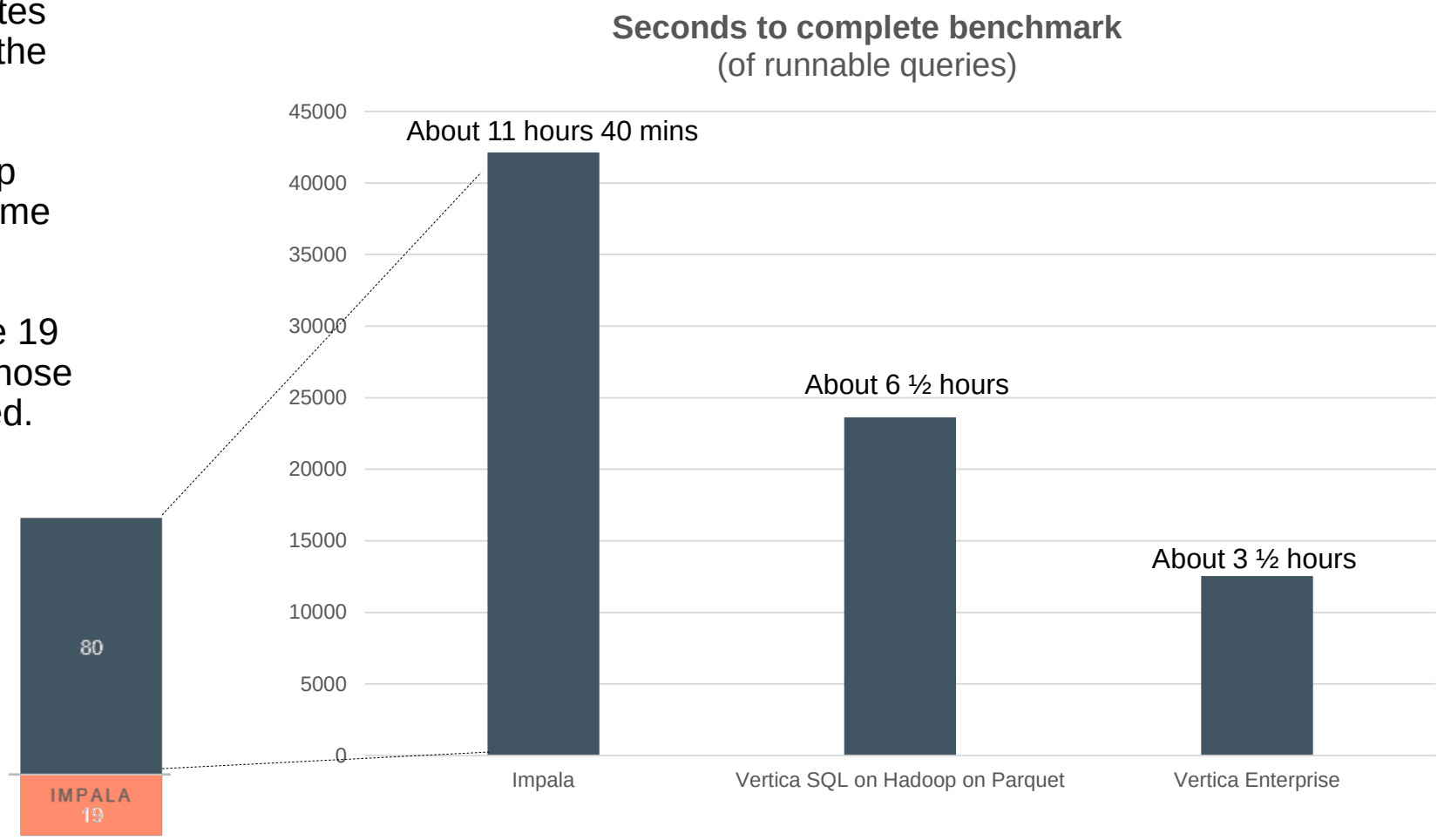
Hewlett Packard Enterprise

Other SQL on Hadoop can’t run all the queries



Boosting Cloudera with Vertica

- Vertica Enterprise completes the benchmark in 30% of the time of Impala
- Vertica for SQL on Hadoop completes in 56% of the time as Impala
- Impala could not complete 19 of the 99 queries at all. Those queries were not compared.

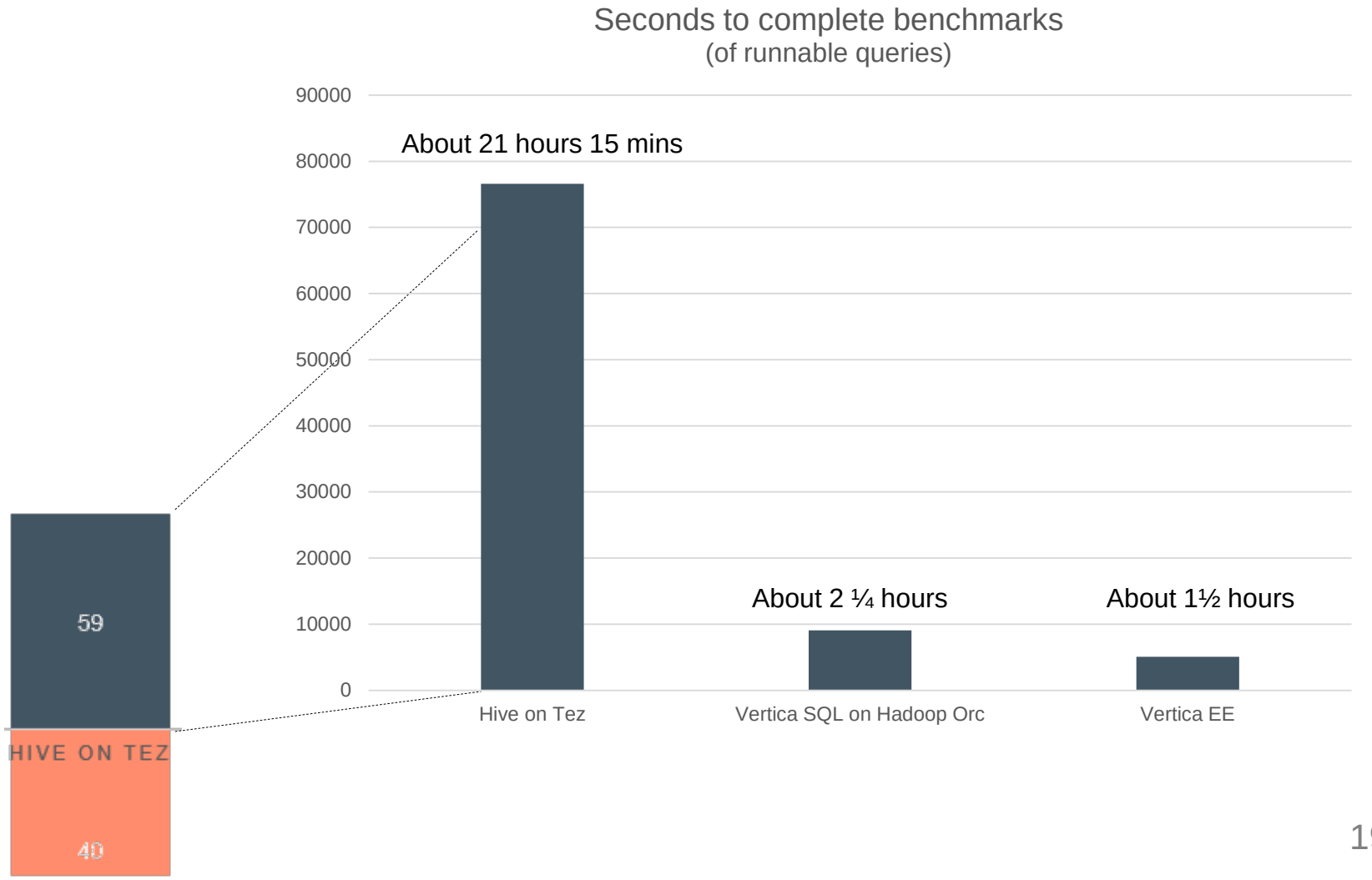


VERTICA

Hewlett Packard Enterprise

Boosting Apache/Hortonworks with Vertica

- Vertica Enterprise completes the benchmark in 7% of the time of Hive on Tez
- Vertica SQL on Hadoop using ORC files completes in 12% of the time of Hive on Tez
- Hive on Tez could not complete 40 of the 99 queries at all. Those queries were not compared.

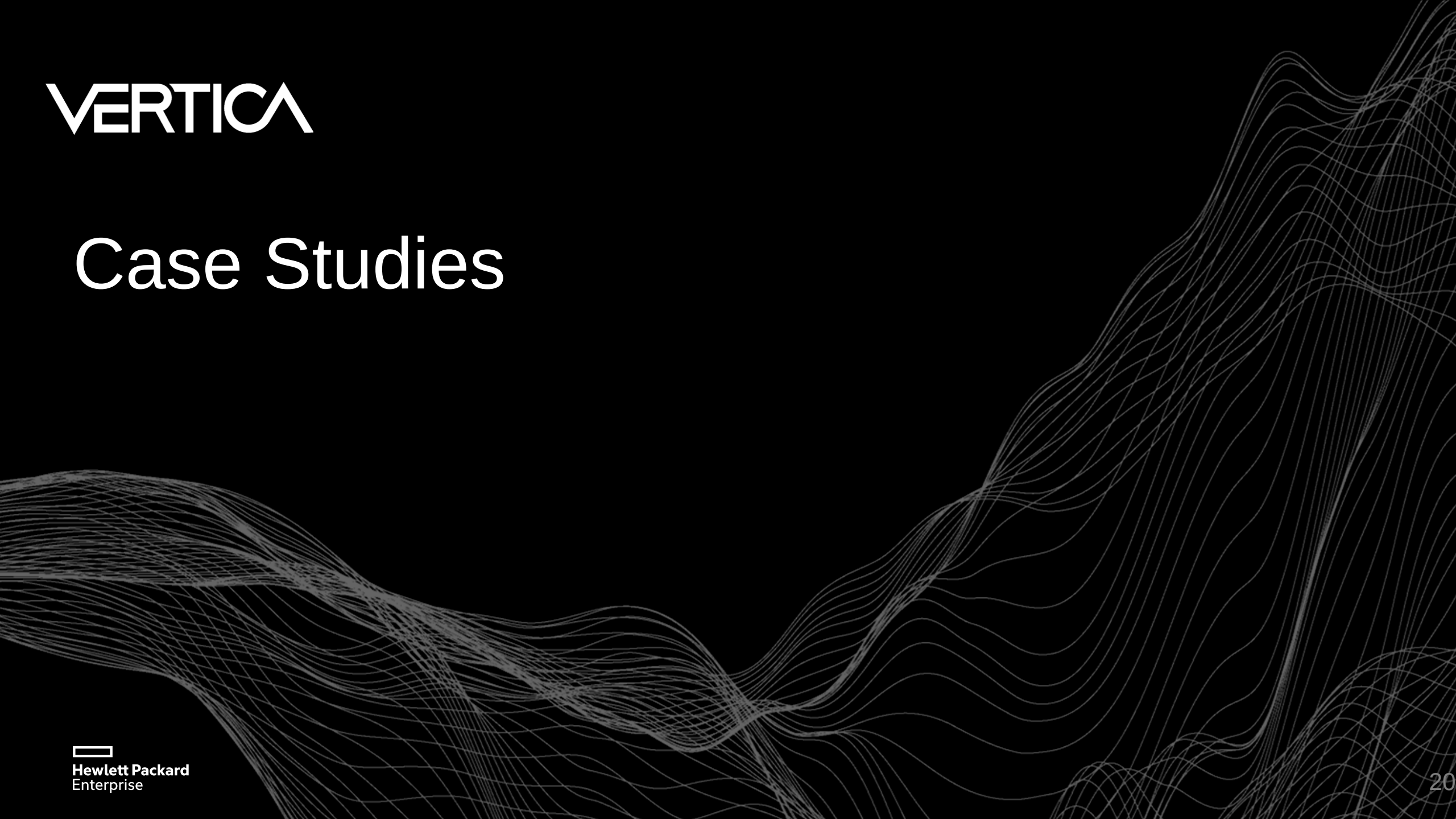


VERTICA

Hewlett Packard Enterprise

VERTICA

Case Studies



Hewlett Packard Enterprise

Anritsu

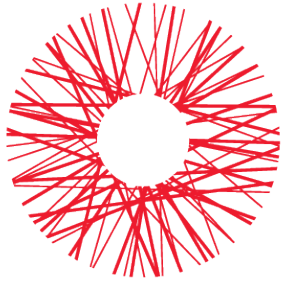
Delivering predictive network analytics for Telecommunications companies

The Challenge

- Data storage requirements increasing exponentially, but customers expecting analytic response times in seconds, not minutes or hours
- With their previous Oracle system, enlarging storage was complicated and time consuming.

The Solution

- Vertica has provided Anritsu with the technology necessary to implement predictive analytics solutions that have only been theoretical until now
- Realized rapid ROI after implementing Vertica in place of a legacy Oracle solution: 351% ROI with a payback of just 4 months



NUCLEUS RESEARCH

ROI: **351%**

Payback: **4 months**

Annual Benefit: **$3+ million**

VERTICA

Hewlett Packard Enterprise

Using customer analytics to eliminate level 1 and level 2 support

The Challenge

- Perform analytics on over 1.5 petabytes of customer product and performance metadata to fine-tune and continue leading-edge product development and evolution

The Solution

- Leverage Vertica for operational analytics to engage in ongoing communications and with customers about storage environments and optimizations
- Use analytics to understand distribution of customers' workloads and how customers access storage, which helps it design storage solutions that match its customers' use patterns

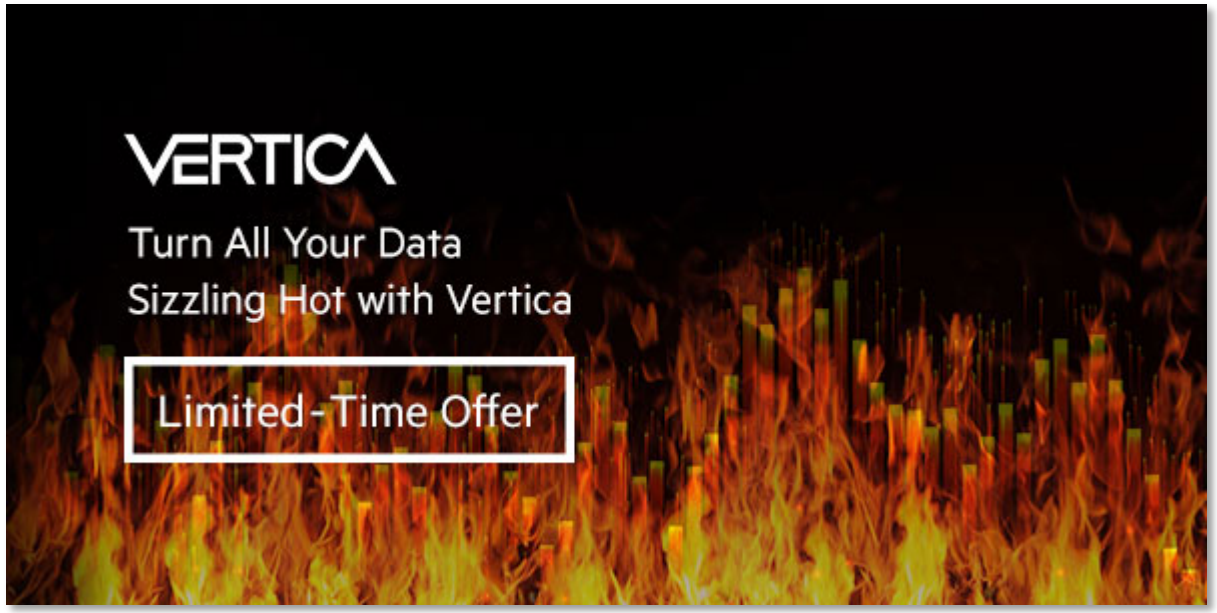
NUCLEUS RESEARCH

ROI: **287%**

Payback: **6 months**

Annual Benefit: **$13.6 million**

Vertica “Hot Data” Promotion



For a limited time, get free nodes of Vertica for SQL on Hadoop when you buy Vertica Premium Edition.

This is a great way to perform data discovery and ad-hoc analysis on all of your data!

Learn more at **www.vertica.com/hotdata**

Thank You

Benjamin.c.smith@hpe.com

Hewlett Packard Enterprise