



YOUR INDEPENDENT HPE SOFTWARE COMMUNITY



What's New and Innovative with HPE Vertica

October 13, 2016

Brought to you by Vivit's Vertica SIG Leaders



Dina R. Love



Akshar Dave



Hosted By



Dina R. Love
Senior Technical Instructor - Software Education
Americas
Hewlett Packard Enterprise
drlove@hpe.com



Welcome to the new Vivit Vertica Special Interest Group (SIG)

- SIG to host a minimum of 2 webinars per year
 - Suggestions for topics or customer implementations welcome
 - Be a guest speaker on the webinars
- Possible meet ups around conferences



Today's Speakers



Jeff Healey
Product Marketing Director - Vertica
Hewlett Packard Enterprise



Steve Sarsfield
Product Marketing Director - Vertica
Hewlett Packard Enterprise



Housekeeping

- This “LIVE” session is being recorded
Recordings are available to all Vivit members
- Session Q&A:
Please type questions in the Questions Pane



Webinar Control Panel

Toggle View Window between Full screen/window mode.

Questions





**Hewlett Packard
Enterprise**

What's New in Vertica 8 “Frontloader”

Jeff Healey, Product Marketing, HPE Vertica
Steve Sarsfield, Product Marketing, HPE Vertica

October 13, 2016



Forward-looking statements

This is a rolling (up to three year) roadmap and is subject to change without notice.

This document contains forward looking statements regarding future operations, product development, product capabilities and availability dates. This information is subject to substantial uncertainties and is subject to change at any time without prior notification. Statements contained in this document concerning these matters only reflect Hewlett Packard Enterprise's predictions and / or expectations as of the date of this document and actual results and future plans of Hewlett Packard Enterprise may differ significantly as a result of, among other things, changes in product strategy resulting from technological, internal corporate, market and other changes. This is not a commitment to deliver any material, code or functionality and should not be relied upon in making purchasing decisions.



Hewlett Packard Enterprise confidential information

This is a rolling (up to three year) roadmap and is subject to change without notice.

This Roadmap contains Hewlett Packard Enterprise Confidential Information.

If you have a valid Confidential Disclosure Agreement with Hewlett Packard Enterprise, disclosure of the Roadmap is subject to that CDA. If not, it is subject to the following terms: for a period of three years after the date of disclosure, you may use the Roadmap solely for the purpose of evaluating purchase decisions from HP and use a reasonable standard of care to prevent disclosures. You will not disclose the contents of the Roadmap to any third party unless it becomes publically known, rightfully received by you from a third party without duty of confidentiality, or disclosed with Hewlett Packard Enterprise's prior written approval.

Your Presenters

- Jeff Healey
 - Product Marketing
 - HPE Software Big Data Platform
 - Email: jeff.a.healey@hpe.com
- Steve Sarsfield
 - Product Marketing
 - HPE Software Big Data Platform
 - Email: steve.sarsfield@hpe.com

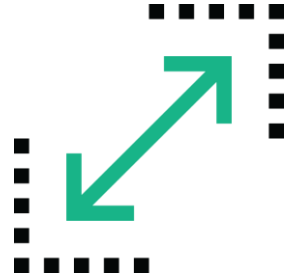


HPE Vertica Analytics Platform



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

All built on the same trusted and proven HPE Vertica Core SQL Engine

Core HPE Vertica SQL Engine

- Advanced Analytics
- Open ANSI SQL Standards ++
- R, Python, Java, Spark. Scala

HPE Vertica In the Clouds

- Get up and running quickly in the cloud
- Flexible, enterprise-class cloud deployment options
- Available on AWS and now Microsoft Azure



HPE Vertica Enterprise

- Columnar storage and advanced compression
- Maximum performance and scalability

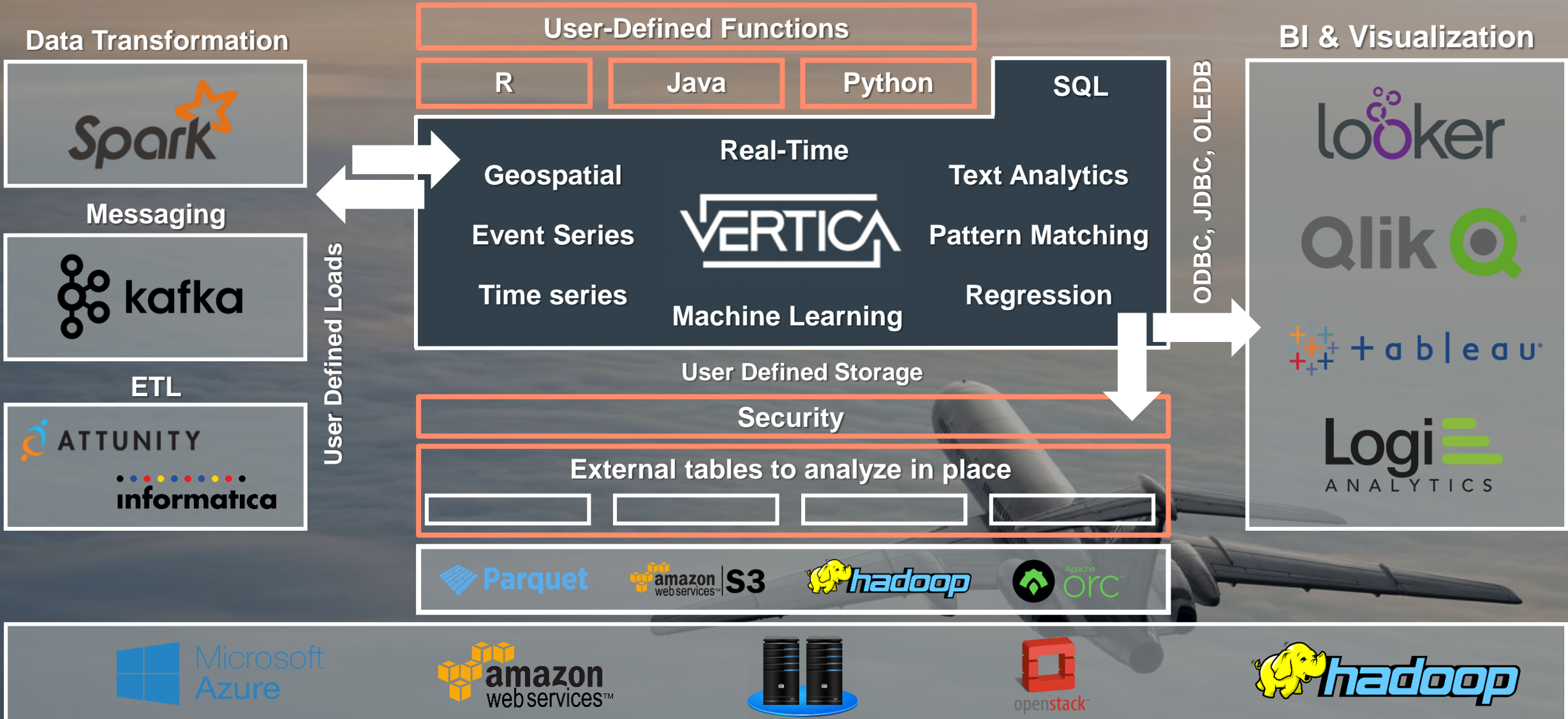


HPE Vertica for SQL on Apache® Hadoop

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



Integrated with open source innovation through an ecosystem friendly architecture



Leading data-driven organizations rely on Vertica



- Promotional testing
- Claims analyses
- Patient records analyses
- Clinical data analyses
- Fraud monitoring
- Financial tracking
- Tick data back-testing
- Behavior analytics
- Clickstream analyses
- Network analyses
- Customer analytics
- Compliance testing
- Loyalty analysis
- Campaign management

Solid Track Record of Delivering Innovation

- Columnar Store
- Aggressive Data Compression
- MPP Architecture
- HA Architecture

11/2013

Flex Tables (Schema on Read) Live Aggregate Projections

(In-memory optimization for frequent queries on discrete data)

7/2014

Geospatial & Social Analytics

SQL on Hadoop

10/2015

Fast ORC Reader

Kafka Support

Deploy and Analyze Anywhere Unified Architecture, In-Database Analytics, Spark, More

2016

- ANSI SQL Compliant
- Java, Python, R APIs
- ACID Compliance
- No Single Point of Failure
- Management Console
- Database Designer
- Projections and Optimizations

Crane



Dragline



Excavator



Frontloader



Introducing HPE Vertica 8 “Frontloader”

Deploy and Analyze Anywhere



Performance at Exabyte Scale

Unified architecture delivers frictionless data movement and orchestration for thousands of concurrent users



Advanced, In-Database Analytics

Applied Machine Learning, geospatial analytics, and more



Any Cloud Design and Open Source Innovation

MSFT Azure cloud support, native Hadoop ORC and Parquet readers enable data analysis-in-place, optimized Spark and Kafka integration



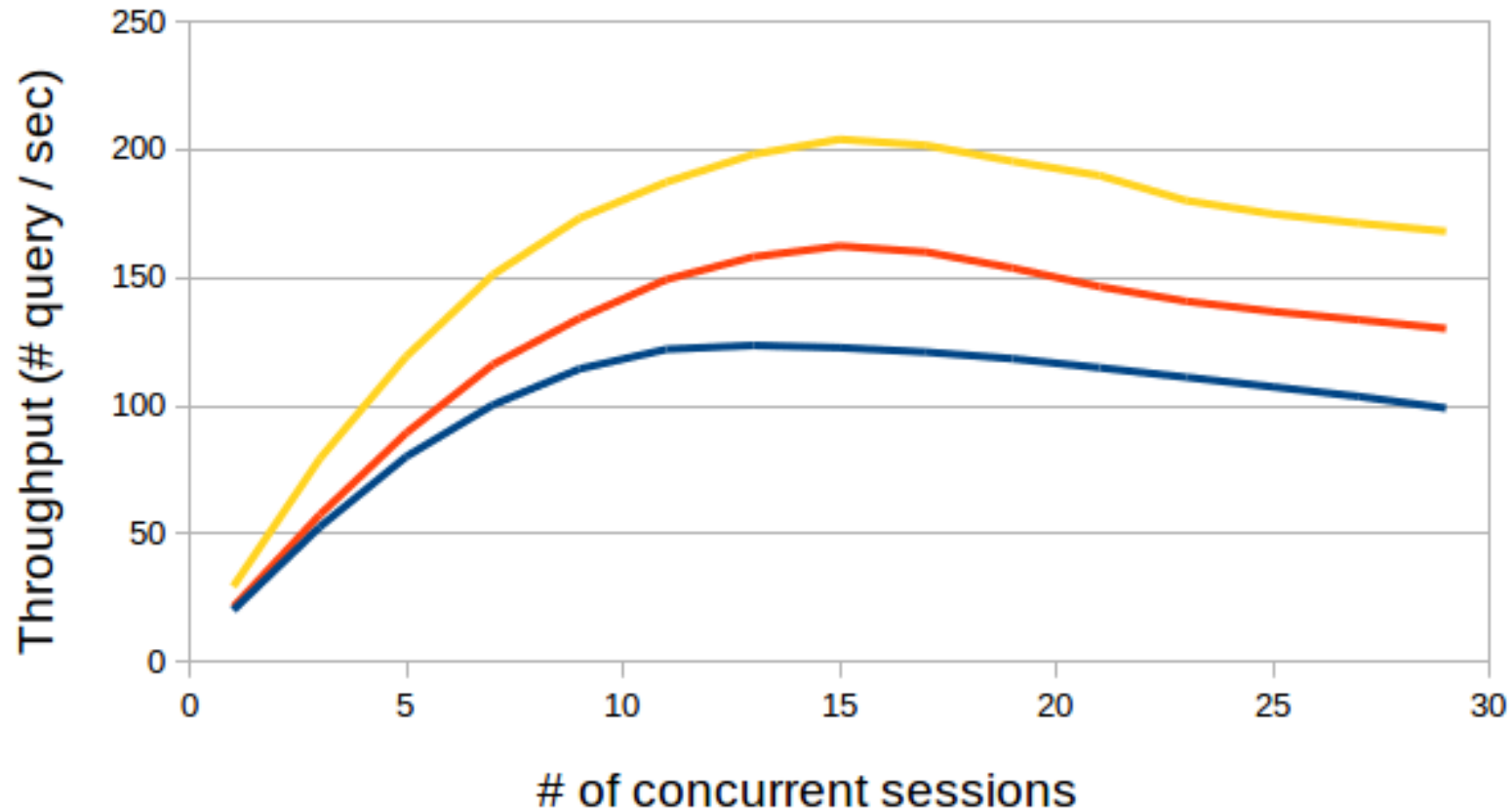
Best-Fit
Engineering



Don't Compromise Big Data Analytics. Empower the Data-Driven Organization.

Boosting Core Analytical Performance

Concurrent Query Execution



Boosting Core Analytical Performance

Latest Innovations

Directed Queries

Users can specify with greater precision how a query should run and which projection should be used. Users can disagree with the automatic query plan and use their own.

Resilient Queries

Vertica will execute queries with a modified query plan when a node fails, optimized for the impaired environment. Ensures that queries run fast no matter what challenges occur.



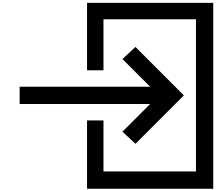
Boosting Loading Performance

Latest Innovations

Apportioned Loading

Boosts performance of data loading when data is wide – hundreds or thousands of columns. Users can load data up to 700% faster by using cluster more efficiently.

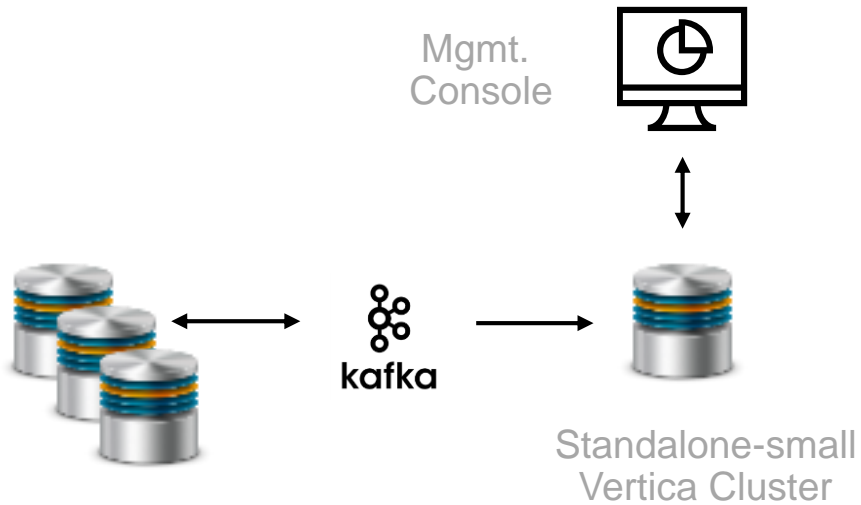
- Reduced contention and increased concurrency of the planning process further speed up data load
- Especially applicable/effective for wider tables with > 1000 columns



Platform Features - Monitoring

Advanced System Monitoring

- Advanced System Monitoring Enables organization to easily set up a stand-alone monitoring system that does not rely on production cluster



Production Vertica Cluster

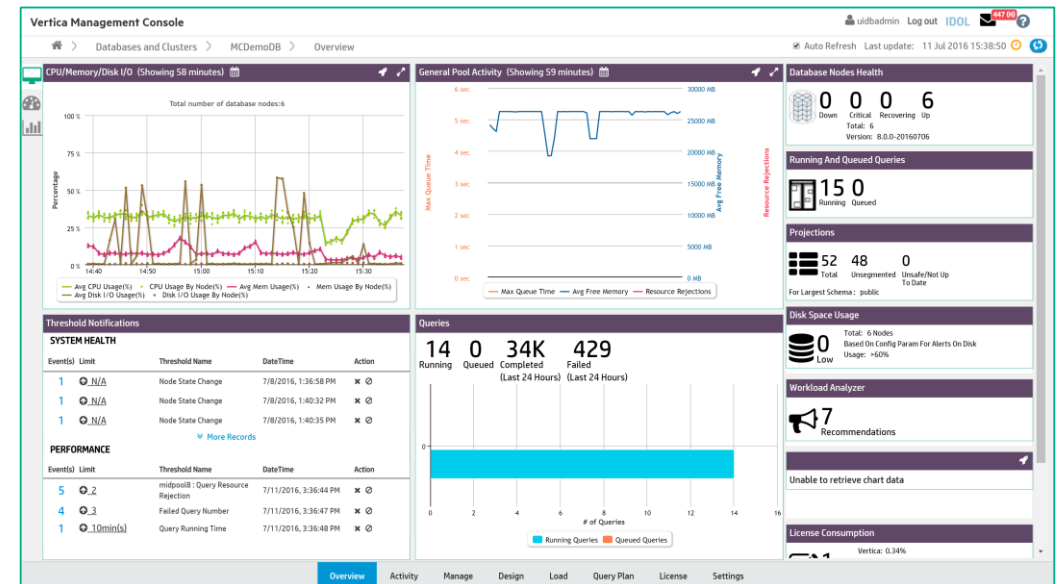
Hewlett Packard Enterprise

Kafka Monitoring in management Console

- Kafka Monitoring in Management Console Added screens to Vertica Management Console (MC) showing load health and rate.

Cloud Monitoring in Management Console

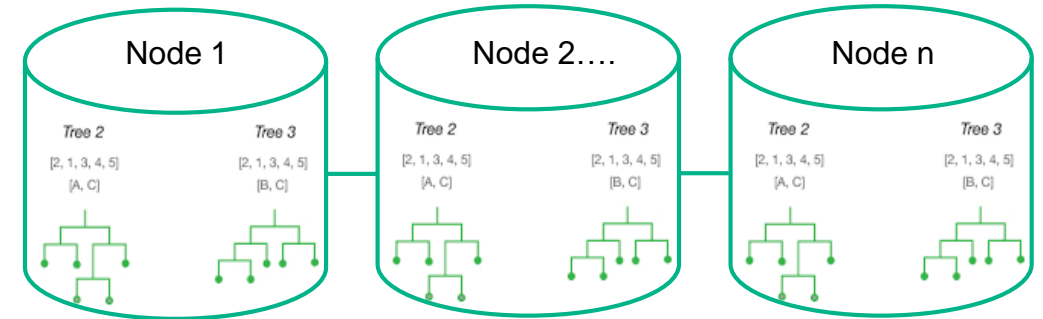
- Vertica Management Console (MC) part of the AMI



Building Machine Learning into the Core of Vertica

New capabilities deliver predictive analytics at speed and scale

- 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



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



Sample Use Cases:

- Customer churn prediction
- Identify fraudulent insurance claims



Apache Spark Connector

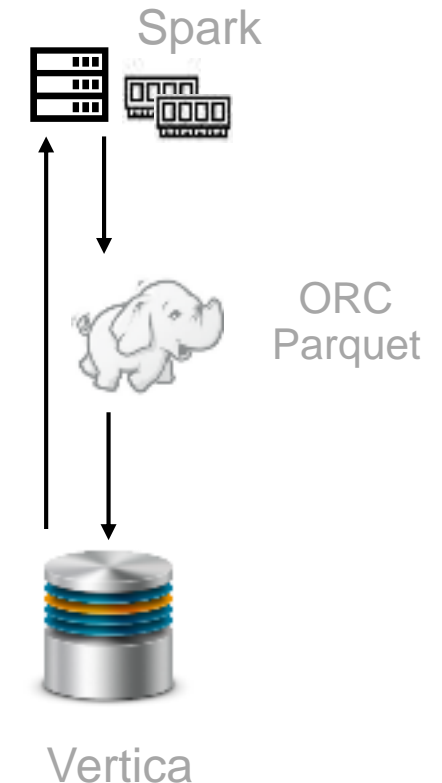
Bi-Directional, Parallel, High-Speed Data Movement

Challenge: Ingest Spark data into HPE Vertica for high-speed SQL Analytics and push Vertica data into Spark for machine learning analytics

Solution: Apache Spark Connector

Benefits:

- Ingest terabytes of Spark in-memory data to Vertica for batch ETL and rich SQL analytics
- Ingest Spark streaming data to Vertica for low-latency interactive SQL analytics in Vertica, e.g. sensor data analysis
- Push Vertica data in Spark for analysis or combining with other data
- Fast, scalable data transfer, exploiting Vertica's parallelism, HDFS connectivity, and fluency with open source data formats



Multi-Cloud Deployments

More Functionality in Amazon cloud

- Integrated AWS credentials with Vertica roles – Seamless access to S3
- When node health check fails for at least <threshold> minutes, send notification or reboot node
- When disk usage > 60% for at least <threshold> minutes, send notification
- Configure load balancer when AWS CloudFormation instantiates Vertica cluster



HPE Vertica Now Runs on Microsoft Azure

A screenshot of the Microsoft Azure portal. The top navigation bar includes "Microsoft Azure", "SALES 1-800-807-1189", "MY ACCOUNT", "PORTAL", and "SEARCH". Below the navigation bar, there are links for "Why Azure", "Solutions", "Products", "Documentation", "Pricing", "Partners", "Blog", "Resources", and "Support", along with a "FREE ACCOUNT" button. The main content area features a dark blue header with the HPE Vertica logo and the text "HPE Vertica Analytics Platform by Hewlett Packard Enterprise". A prominent green "Deploy" button is visible. Below the header, the text reads "Bring Your Own License enabled." and "Enterprise-Class Analytics that fits your budget". A paragraph describes the platform's capabilities, including fine-tuning, BI/reporting, and data ingestion. A section titled "This solution will automatically deploy an HPE Vertica cluster, allowing you to login and start using it." is followed by a list of four steps: 1. Connect to the Public IP address via SSH using the DBAdmin account. 2. Type the command `admintools` to launch the administration tools. 3. Enter the license key or select the Community Edition license. 4. Accept the End User License Agreement. The page also includes a "VERSION: 1.0.1" and a "Useful Links" section with a link to "Get started with Vertica".

HPE Vertica for SQL on Hadoop

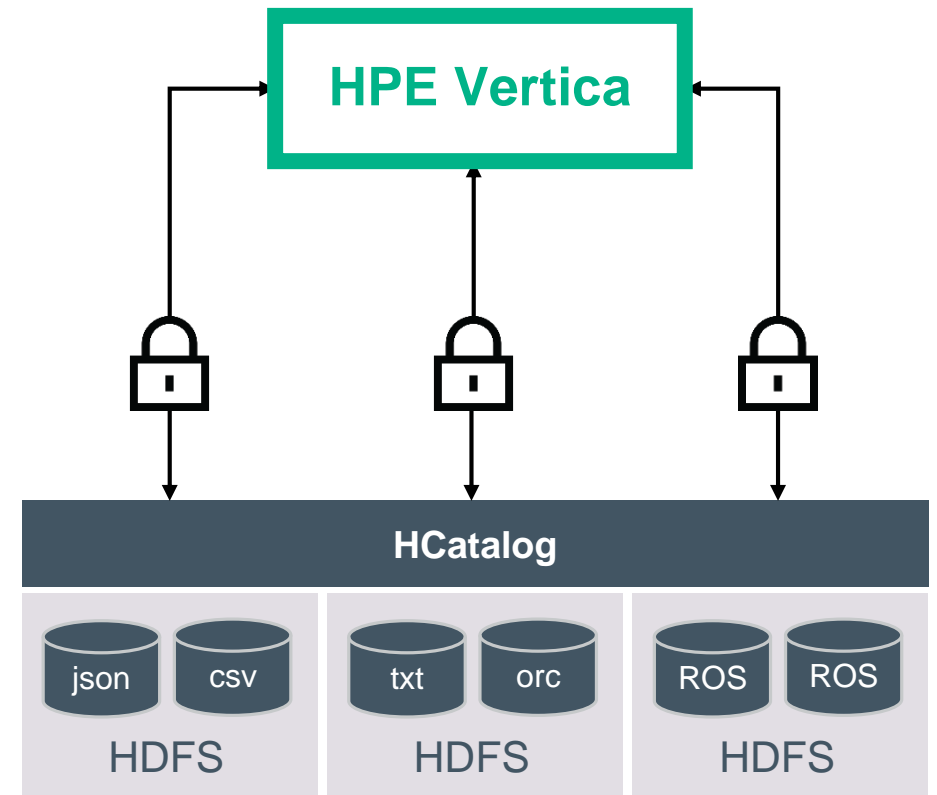
Vertica is fast and Hadoop is big; VSQLOH brings together fast and big.

HPE Vertica for SQL on Hadoop offers the only full-featured query engine on Hadoop

- Data Exploration in-place in HDFS
- Integrated with HCatalog
- Data Query Across ORC, Vertica ROS, and Cloudera Parquet **New!**
- Same Core Engine – full SQL support
- Hadoop distribution agnostic
- Integrated with Kerberos Security

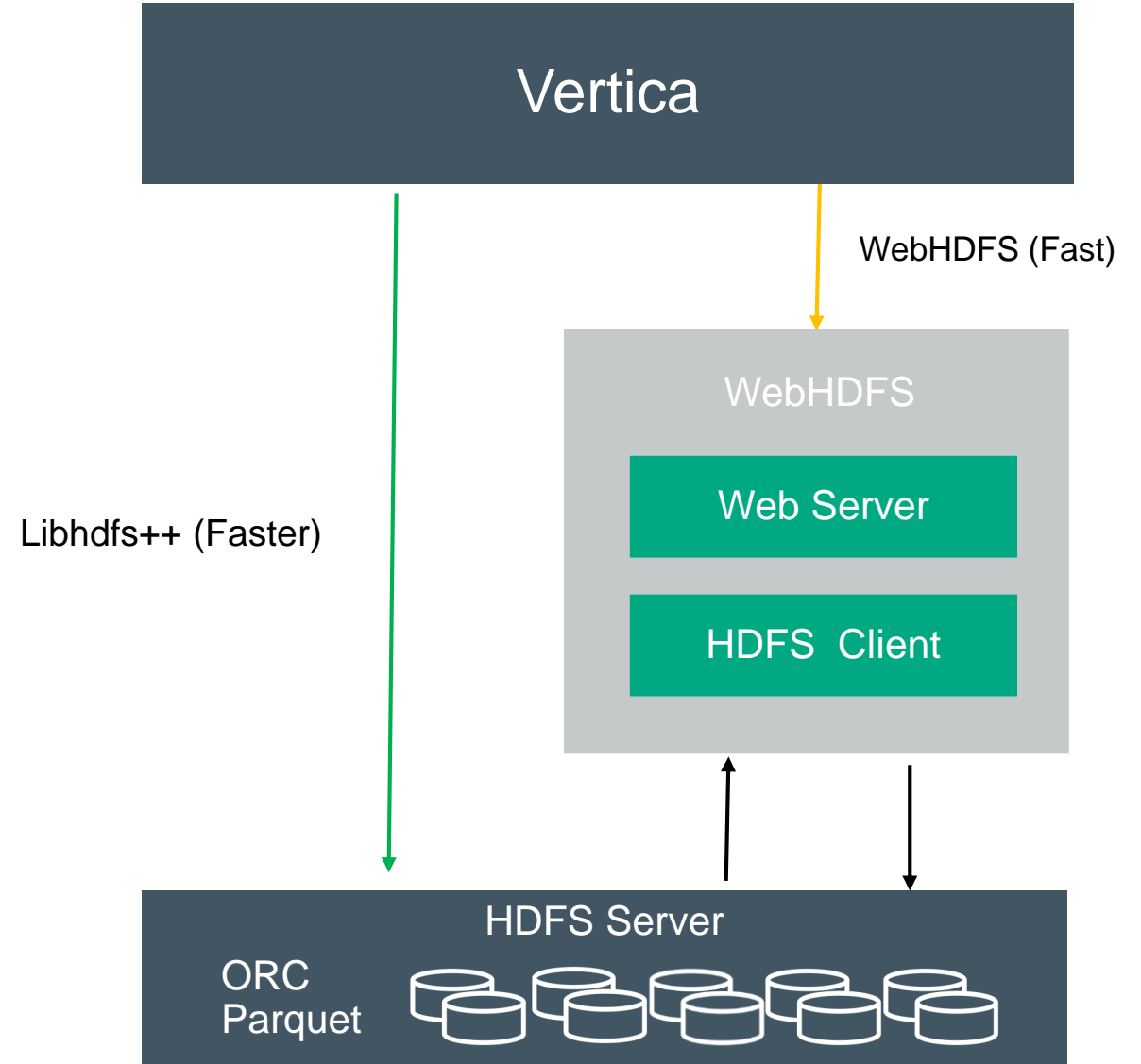
TPC-DS Benchmark results

- 1.6X faster than Impala 5.3.3 on average
- Ran all 99 queries with no modification vs. Impala's 67



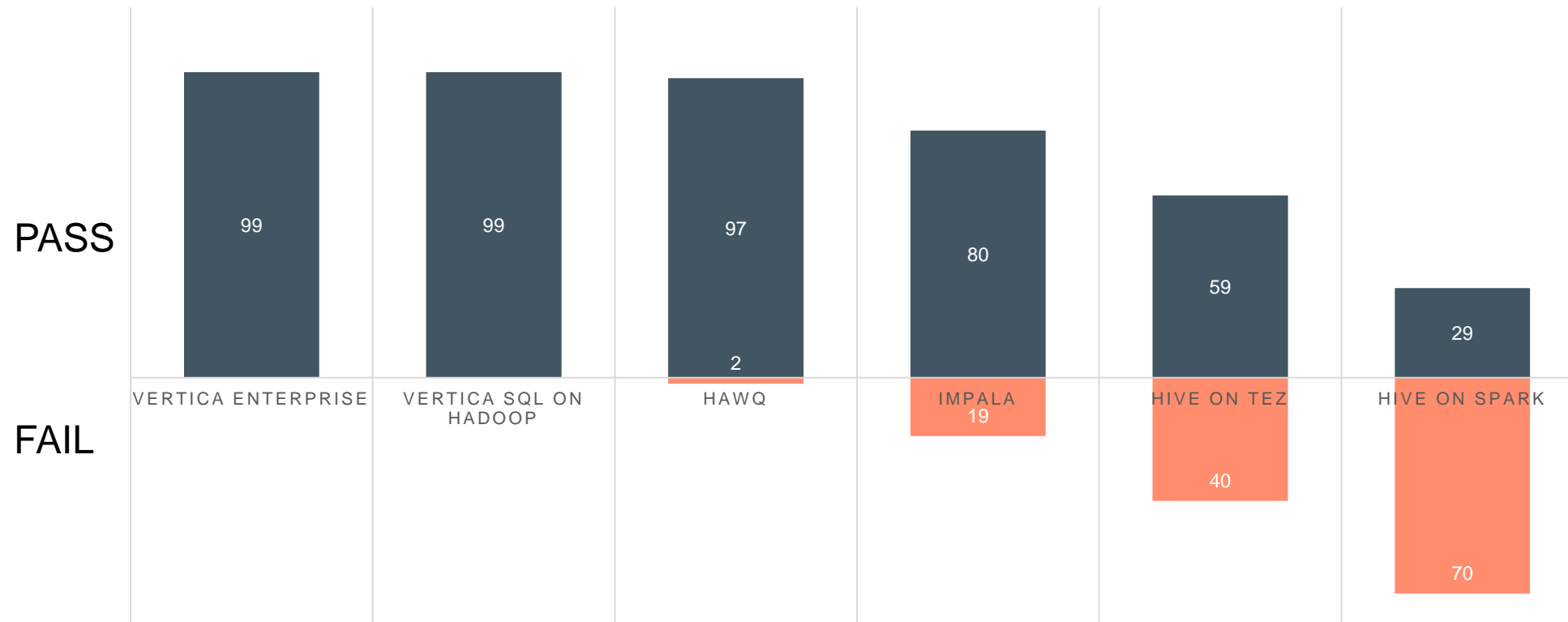
New ORC Reader, Parquet Reader and Libhdfs++

- Hadoop has Parquet and ORC as two popular formats for columnar data
- Developed libParquet and libOrc in collaboration with Cloudera and HortonWorks respectively
 - Query close to where data resides
 - Column Pruning, Predicate Pushdown
 - Hive partition pruning and SQL syntax to read Hive partitioned data
- Reliable and high speed libhdfs++ for HDFS data access
 - Alternate to WebHDFS interface
 - Implemented in collaboration with Apache community



HPE Vertica Tops in TPC-DS Benchmarks

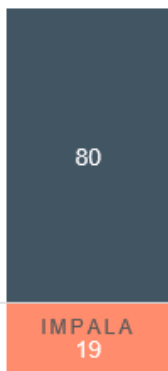
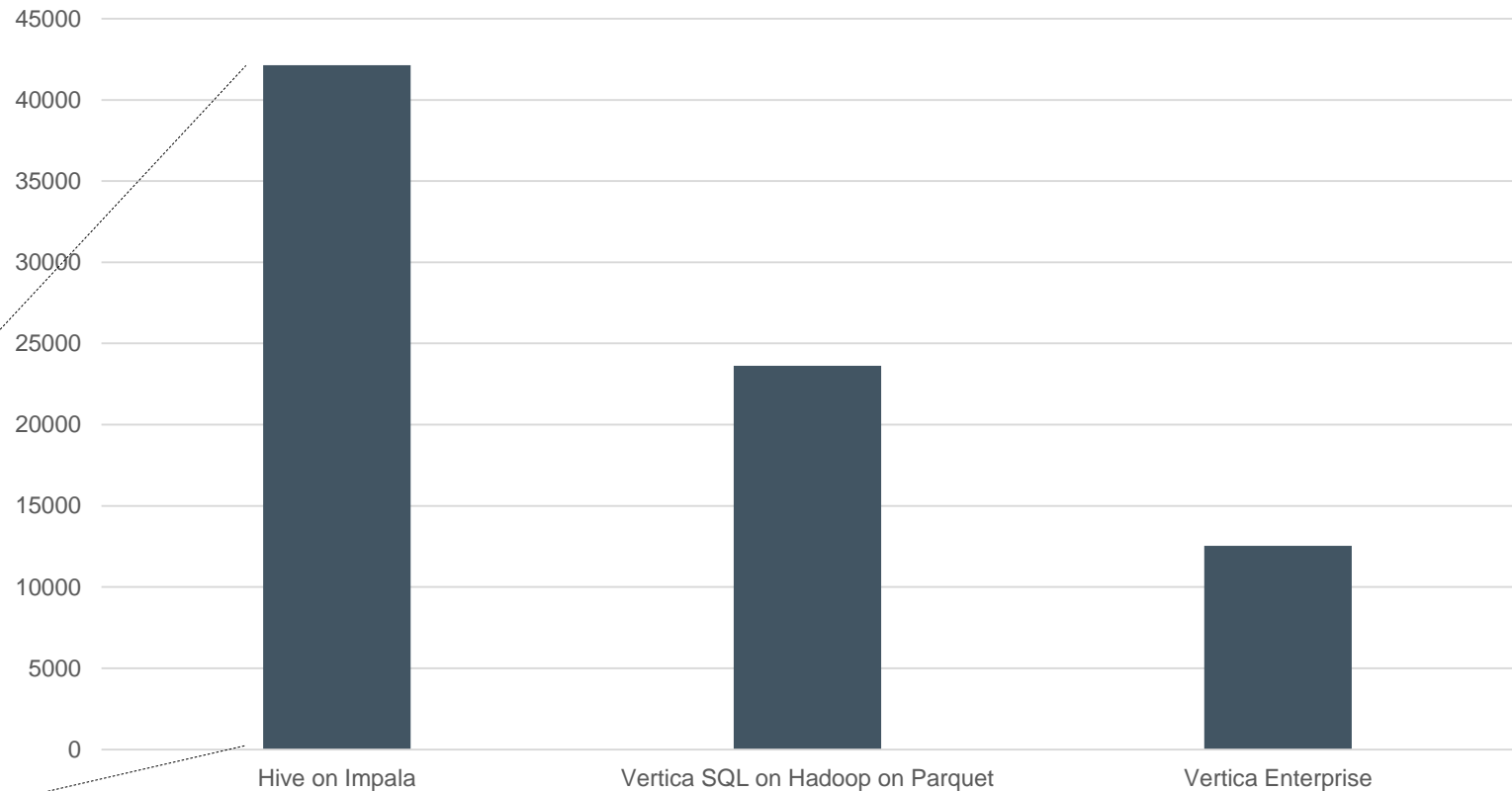
PASSING AND FAILING TPC-DS QUERIES



Boosting Cloudera with Vertica

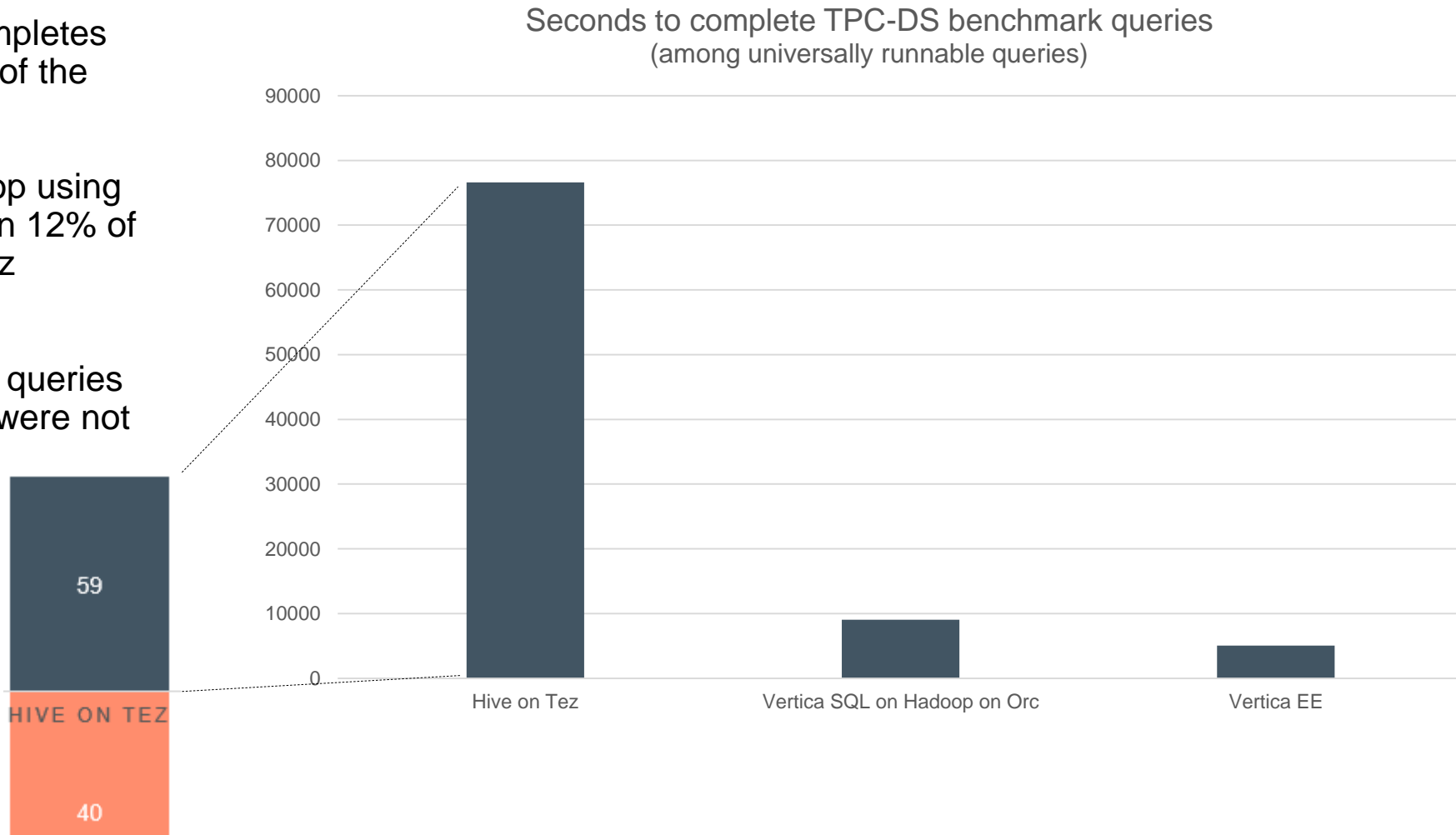
- Vertica Enterprise completes the benchmark in 30% of the time of Hive on Impala
- Vertica for SQL on Hadoop completes in 56% of the time as Hive on Impala
- Impala could not complete 19 of the 99 queries at all. Those queries were not compared.

Seconds to complete benchmark
(of runnable queries)



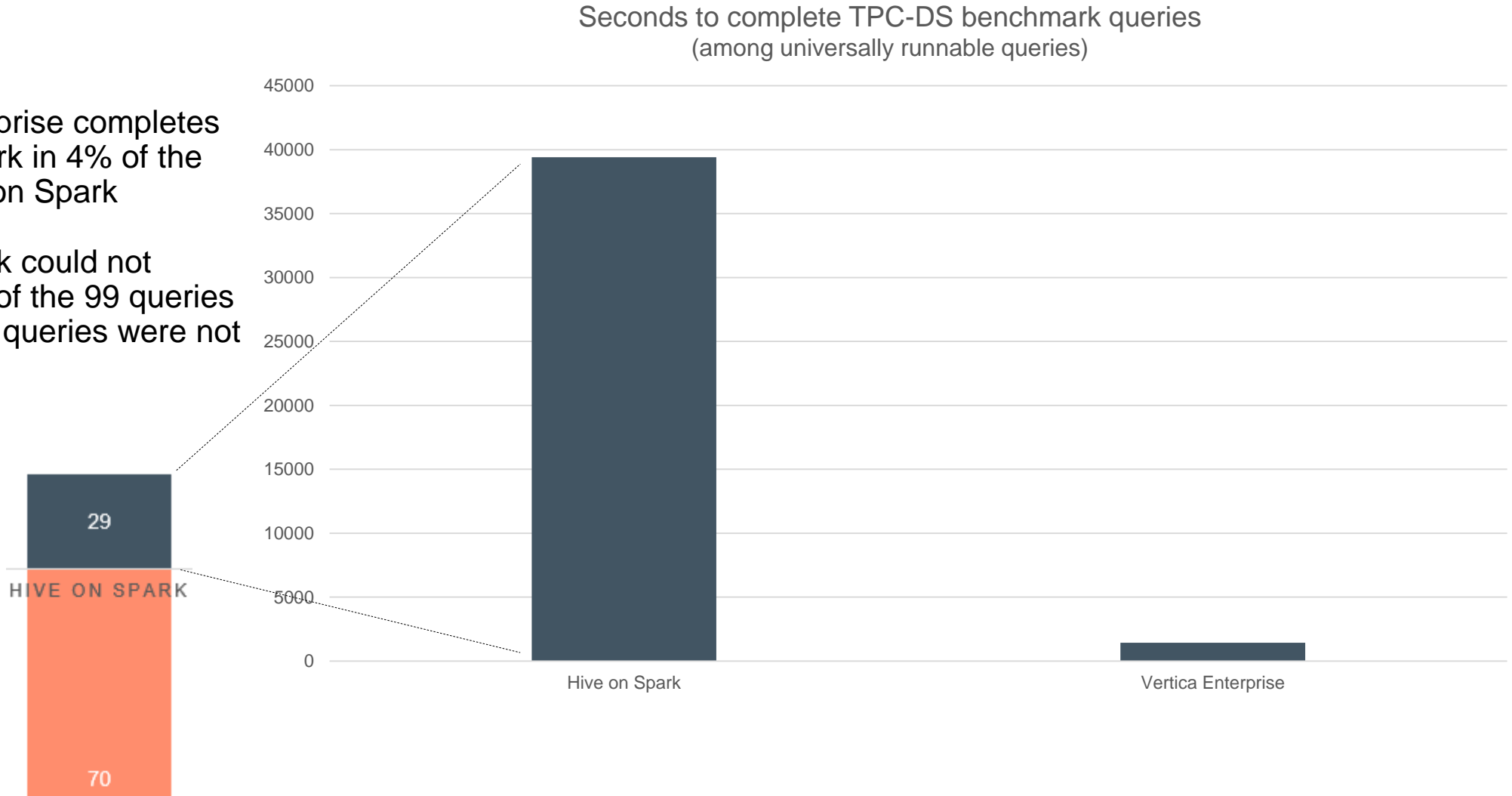
Boosting Apache/Hortonworks with Vertica

- 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.



Boosting Apache Spark with Vertica

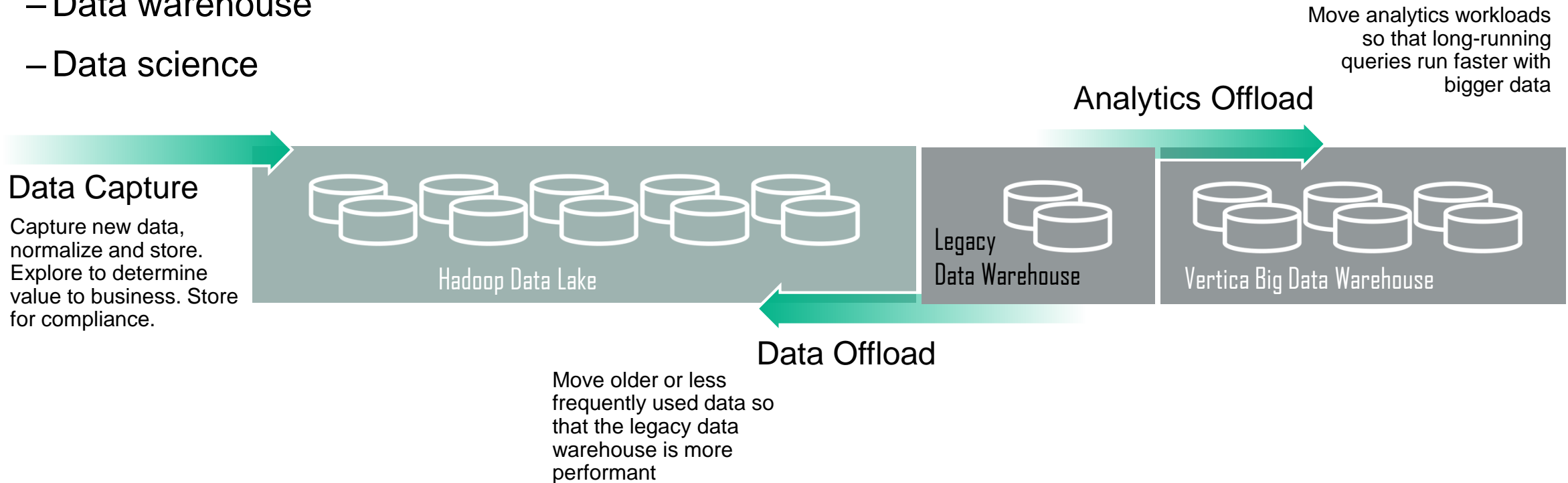
- Vertica Enterprise completes the benchmark in 4% of the time of Hive on Spark
- Hive on Spark could not complete 70 of the 99 queries at all. Those queries were not compared.



Data Warehouse and the HPE Ecosystem

Complex infrastructure that drives:

- Data warehouse
- Data science



Opportunity Map



HPE Vertica SQL on Hadoop

HPE Vertica Enterprise

Embracing an Open Source Architecture

Apache Spark, Hadoop and Kafka Integration

Optimal Use Case

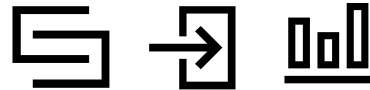
- Deep Analysis
- Massive scale
- Many concurrent users

HPE Vertica



Kafka

Spark



Hadoop



Features:

- Vertica performs optimized data load from Spark
- Spark runs queries on Vertica data

Optimal Use Case

- Small, fast running queries
- ETL and complex event processing
- Operational analytics

Features:

- Share data between applications that support Kafka
- Data streaming into Vertica

Optimal Use Case

- Data lake
- Warm, cold storage
- Data discovery
- ETL

Features:

- Analyze-in-place without data movement via native ORC and Parquet readers
- Any Hadoop
- Run ON the Hadoop cluster or ON Vertica cluster

What's New in Vertica 8

Top technology enhancements

Spark Integration

- Spark is now integrated with Vertica
 - Feed data in Spark with Vertica for ML models
 - Connector to access data in Spark

Hadoop Integration

- Reading files from Hadoop is now faster and more secure than ever
 - Fast parquet file reader
 - Kerberos security
 - Faster HDFS reader (LibHDFS++)

Machine Learning

- Machine learning algorithms are now built into the database
 - k-means
 - logistic regression
 - Linear regression
- More to come

Kafka Enhancements

- Kafka has been further enhanced with a new user interface to monitor streaming data
- Share data between Spark, Hadoop or any application that supports it.

Performance at Exabyte Scale

- Reduced footprint on disk
- Higher throughput and concurrency
- More efficient memory usage
- Improved workload management through increased parallelism

Benefits

Users can set up an infrastructure where they use Spark for queries in memory and Vertica for longer running analytics on larger data volumes.

Users can reduce costs by leveraging low-cost Hadoop servers for data storage, ETL and more.

Users can leverage the power of the Vertica cluster for predictive analytics.

Users can securely share data in their infrastructure using Kafka's publish and subscribe capabilities.

Users can also more easily stream data into Vertica for operational analytics.

Try HPE Vertica Today!

HPE Start-Up Essentials Program

Grow your start-up business with Big Data analytics – free to companies with <10 employees for the initial years.



<http://go.saas.hpe.com/bigdata/accelerator>

HPE Vertica Community Edition

Download and install community edition. Manage and analyze up to 1 TB of data across three nodes for an unlimited time.



<https://my.vertica.com/community>

HPE Vertica AMI

Install your Vertica license on select cloud hardware to create cluster on AWS.



Vertica AMI: <http://amzn.to/1OLVe4C>

HPE Vertica for SQL on Hadoop

Apply to join the Early Adopter Program for a 90-day trial with premium level technical presales and implementation support directly from Vertica R&D.



<https://hp.orbitera.com/c2m/customer/testDrives/index>



Q&A

jeff.a.healey@hpe.com

steve.sarsfield@hpe.com



Discover
2016
London
Nov 29 -
Dec 1

Discover 2016 London

- Excel London

- In today's idea economy, success favors enterprises that can reinvent to outpace the competition. Discover 2016 can put your enterprise in the lead. Come to the intersection of digital transformation, technology and ideas. Accelerate what's next for your business.
- Join more than 10,000 IT executives, architects, engineers, partners and innovators from around the globe. Accelerate your unique digital transformation journey. Explore how to deliver seamless experiences, harness data, mitigate risk, boost agility and enable workforce productivity.
- Dive into Discover 2016.

Register Now and receive your member discount
hpe.com/discover



Top ten reasons to attend

1



Get ready to transform your IT

Transforming to a hybrid infrastructure, protecting your digital enterprise, empowering the data-driven organization and enabling workplace productivity can position your enterprise for success. Learn how Hewlett Packard Enterprise is the best strategic digital partner equipping you for transformation.

2



Hear from the Hewlett Packard Enterprise leadership team and other industry visionaries

Learn about our four Transformation Areas and hear directly from President and Chief Executive Officer Meg Whitman and the executive team on how Hewlett Packard Enterprise is helping you go further, faster.

3



Accelerate next with Hewlett Packard Enterprise Transformation Advisors

Bring your digital transformation to life through one-on-one consultations with our solution architects on hand to answer your questions and address your challenges to help you accelerate what's next for your business.

4



Seize valuable opportunity

Time-to-value is the new name of the game, and the ability to capitalize immediately is key to survival. Through the Transformation Zone and Innovation and Spotlight Sessions, you'll learn how HPE can help you transform your organization to significantly reduce the time-to-value of your IT investments.

5



Accelerate transformation with IT consumption

The velocity of change in today's IT world is daunting. Driving innovation while still running the business is essential. Through a series of Spotlight, Theater and Breakout Sessions, learn how you can fund transformation, map your investment strategy to your IT roadmap and achieve faster time-to-market.

6



Take a sneak peek at tomorrow's technologies

Customers and partners under CDA can get early access to preview upcoming Hewlett Packard Enterprise products and solutions.

7



Find answers to your biggest business and IT challenges

Discover 2016 is the perfect place to find solutions, seek new innovations and learn how Hewlett Packard Enterprise can help your enterprise succeed.

8



Explore the Transformation Zone

Interact with experts and peers and learn more about Hewlett Packard Enterprise's most innovative technologies. See the latest products and solutions while speaking with the engineers and architects who created them.

9



Expand your network and grow relationships

Engage with other IT professionals, schedule formal onsite meetings at the Hewlett Packard Enterprise Meeting Center or visit the Guru Bar to chat with Hewlett Packard Enterprise experts.

10



Tap into the power of our partners

Meet with Discover 2016 sponsors, including Intel and Microsoft, in the Transformation Zone. See how collaboration and our growing network is bringing more opportunity and advantages for you.

Thank you

www.vivit-worldwide.org

