HPE Vertica & Hadoop
Tapping Innovation to Turbocharge Your Big Data
#SeizeTheData
The HPE Vertica portfolio
One Vertica Engine – running on Cloud, Bare Metal, or Hadoop Data Nodes

Core HPE Vertica SQL Engine
- Advanced Analytics
- Cloud, Bare Metal, & Data Lake
- Open ANSI SQL Standards ++
- R, Python, Java, Spark, Scala

HPE Vertica OnDemand & AMI
- Get up and running quickly in the cloud
- Amazon AWS
- 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 Hadoop
- Native support for ORC, more
- Support for industry-leading distributions
- No helper node or single point of failure
Vertica on the Data Lake
The same Vertica Engine – Now using Open File Formats on Shared Storage

Not a Data Mart or EDW 2.0 – The Data Lake is *Open* and *Elastic* at Enabling Economics
– ORC, Parquet, AVRO, JSON: Stage, Retain, and use data in original formats, at high velocity
– ETL Process & BI Tools: “Lift & Shift” to access Enabling Economics without implementation barriers
– Leverage the Tools, Training, and Expertise within lines of business

**ETL / Transformation Throughput**
– Leverage Vertica efficiency and rack aware networking to supercharge legacy ETL

**Discovery & Ad-Hoc**
– Combine data sets which previously were too big to house together cost effectively
– Understand the value of data without a Procurement Process

**HPE Vertica for SQL on Hadoop**
– Native support for ORC, more
– Support for industry-leading distributions
– No helper node or single point of failure
Vertica Enterprise on “Bare Metal”
The same Vertica Engine – Now using Open File Formats on Shared Storage

Transition Highly Leveraged Data Sets

– Traditional Vertica, managing fault tolerant storage on rack servers with SAS disks
– Enterprise features: backup, replication, data lifecycle management
– Perfect for BI tools and Application APIs which require consistent, predictable response time

HPE Vertica Enterprise
– Columnar storage and advanced compression
– Maximum performance and scalability
– Flex Tables for schema-on-read

Agility and Elasticity

– Discover Value in the Lake / Curate to Production
– “Right Size” server count with elastic provisioning

HPE Vertica for SQL on Hadoop
– Native support for ORC, more
– Support for industry-leading distributions
– No helper node or single point of failure
Vertica OnDemand & Cloud
The same Vertica Engine – With the Infrastructure When & Where you need it

Spin up Low Volume Production
– Discover on Lake / Deploy on cost effective Virtual Private Cloud or Hybrid Cloud
– Sunset legacy workloads
– Variety of Supported AMI Images / Machine types appropriate for any workload

Where you want it
– Your data is in S3? No problem. Run Vertica where your data lives
– Data Governance requirements? No problem. OnDemand available in the geographies you need

HPE Vertica OnDemand & AMI
– Get up and running quickly in the cloud
– Amazon AWS
– 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 Hadoop
– Native support for ORC, more
– Support for industry-leading distributions
– No helper node or single point of failure
HPE Vertica for SQL on Hadoop features and benefits

Query data, no matter where it is located

– Install HPE Vertica directly on your Hadoop infrastructure
– ORC, Parquet, Avro, Vertica ROS and JSON supported
– Full-functionality ANSI SQL
– 100% of TPC-DS queries
  – No helper node or single point of failure
– Competitive price point
HPE Vertica Enterprise – Proven

- Advanced ANSI SQL Analytics
- Massively Parallel Processing
- Column Orientation
- Application Integration
- Highly Available
- Automatic DB Designer
- Advanced Compression
- Management Console

<table>
<thead>
<tr>
<th>ANSI SQL ENGINE</th>
<th>ANSI SQL ENGINE</th>
<th>ANSI SQL ENGINE</th>
<th>ANSI SQL ENGINE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Columnar Formats (ROS, Flex)</td>
<td>Columnar Formats (ROS, Flex)</td>
<td>Columnar Formats (ROS, Flex)</td>
<td>Columnar Formats (ROS, Flex)</td>
</tr>
<tr>
<td>File System (EXT 4)</td>
<td>File System (EXT 4)</td>
<td>File System (EXT 4)</td>
<td>File System (EXT 4)</td>
</tr>
<tr>
<td>Commodity Hardware (x86)</td>
<td>Commodity Hardware (x86)</td>
<td>Commodity Hardware (x86)</td>
<td>Commodity Hardware (x86)</td>
</tr>
</tbody>
</table>

Commodity Hardware (x86)
HPE Vertica for SQL on Hadoop

- Same Vertica MPP Columnar Architecture
- Base ANSI SQL
- Co-Located with Hadoop
- Data Query Across Parquet, ORC, JSON, and many other formats
- Hadoop Agnostic
**VSQLH is reliable and performant**

In this industry-standard benchmark:
- 43/99 queries will not run in other solutions that work unchanged in VSQLH
- Some queries run up to 7X faster on HPE SQL on Hadoop
- Concurrency is better supported
- Security with Kerberos
New ORCFile Reader

- Open Source Project jointly developed by Vertica and Hortonworks Engineers
- Query close to where data resides
- Column Pruning, Predicate Pushdown
Vertica Management Console

- Manage multiple clusters from a single web-based console
- Real-time view of database activity and cluster status
- Correlate system and database activity
Third-party integrations

- pentaho
- BusinessObjects
- snaplogic
- GOLDEN GATE
- informatica
- QlikView
- looker
- BusinessObjects
- Tableau
- Logi
- Talend
- Tamr
- HPE Vertica

HPE Vertica
ANSI SQL ENGINE

HPE Vertica
ANSI SQL ENGINE

HPE Vertica
ANSI SQL ENGINE

HPE Vertica
ANSI SQL ENGINE
Key differentiators
No limits data exploration

- **Richest and proven SQL query engine**
  Dramatically reduce complexity of your data architecture with a time-tested, unified query engine

- **Enterprise-ready, stable, trusted**
  Improve productivity with a robust, highly optimized, enterprise-ready SQL engine and first-class support

- **Simplicity and fast insights in data exploration**
  Leverage our Database Designer to simplify the management of your queries

- **Easy path to Vertica Enterprise**
  Perform data discovery in Hadoop and boost analytics performance with Vertica EE

- **Rich ecosystem**
Trial: Take a free test drive
http://goo.gl/rp4fzi

Vertica for SQL on Hadoop Test Drive
An enterprise-ready SQL engine that delivers the most comprehensive analytics on your Hadoop distribution.

Click on the test drive link below to spin up a temporary Hadoop cluster in the Amazon cloud with HP Vertica for SQL on Hadoop and preloaded data. You can access the cluster for four hours. We encourage you to first run some basic queries with the preloaded data, then challenge the query engine with queries that you’ve been unable to execute with other SQL engines for Hadoop.

Key Benefits

- Install Vertica directly on your Hadoop infrastructure to query and gain value from Hadoop data.
- Supports data in many formats such as ORCFile, Parquet, Avro, as well as Vertica RDS and JSON.
- Supports many Hadoop distributions, including Hortonworks, MapR and Cloudera.
- Delivers full-functionality ANSI SQL for query completeness on all your queries.
- No helper node or single point of failure.

Take a Test Drive Now!
HPE Vertica Connector for Apache Spark
Beta download at https://saas.hpe.com/marketplace/haven/hpe-vertica-connector-apache-spark

– Ingest terabytes of Spark in-memory data into HPE Vertica for interactive SQL analytics
– Operationalize Spark machine learning models using HPE Vertica extensible Big Data platform
– Ingest Spark streaming data to HPE Vertica for low-latency interactive SQL analytics
HPE Vertica Connector for Apache Spark
Operationalize Spark Machine learning models in HPE Vertica

- Business Problem
- Prepare Data
- Explore Data
- Model and Evaluate
- Operationalize and Monitor

Vertica standalone cluster
Vertica SQL on Hadoop
Vertica Cloud Deployments
HPE Vertica Connector for Apache Spark
Apache Kafka + Spark + HPE Vertica for both Batch and Streaming Analytics
Save Apache Spark RDD and DataFrame in HPE Vertica
HPE Vertica native efficient parallel bulk loading
HPE Vertica Native Data Source for Apache Spark
Simple and fast HPE Vertica data access in Spark in-memory analytics libraries
Operationalize Predictive Models
Deploy Spark ML Models in PMML for in-database scoring

- Classification models
- Regression models
- Clustering models
HPE Vertica Connector for Spark

Store data in RDDs in HPE Vertica native columnar format for high-speed interactive SQL analytics

Fast access to HPE Vertica data in Spark RDDs and data frames with filter and predicate push-down
On-premise data access

**Streaming**
Kafka, Trickle (Insert/Update)

**Schema on read**
Flex Zone: JSON, CSV, Text, Social Media

**Batch**
ODBC/JDBC, Bulk COPY, LCOP, ETL: Pentaho, Attunity, Informatica, Talend, ET AL

**Unstructured**
IDOL: Video, Audio, Voice Recognition, Facial Recognition

**Hadoop**
ORC Reader, MapR NFS, HIVE Serializer: HDFS, Parquet, AVRO
Start with the Event Bus

Operational Analytics
On-the-fly observations, complex events, and heads-up analytics

Messaging Bus
Publish-Subscribe services for sharing data

Batch Analytics
Longer term analysis, discovery, complex predictions and deep analytics

Live Data Streams
Web Logs
Application Data
IOT data

Databases
Hadoop
Cloud
Take the Insights to Production

- **Operational Analytics**
  - On-the-fly observations, complex events and heads-up analytics
  - Spark, MemSQL, etc.

- **Batch Analytics**
  - Publish-Subscribe services for sharing data
  - HPE Vertica
  - Longer term analysis, discovery, complex predictions and deep analytics

- **Messaging Bus**
  - Kafka. Also Talend ESB, Mulesoft, Informatica, other proprietary solutions

- **Live Data Streams**
  - Web Logs
  - Application Data
  - IOT data

- **Databases**
  - Hadoop
  - Cloud

- **Hewlett Packard Enterprise**
Trends & HPE Big Data Platform
Obvious: structured data analytics trends continue

– Data continues to grow
– Hardware gets more capable
– Those who demand the highest levels of performance can generally forecast their capacity demands, and are willing to dedicate hardware will continue to choose high-performance MPP databases
– “Cloud” offerings of MPP databases will continue to mature, but will not fully displace dedicated offerings
– In-memory processing of “small” big data will continue
– The market will continue to consolidate
HPE’s “Big Data” portfolio
You should consider HPE for all your “Big Data” needs because:

– “Big Data” is “everything”… HPE has something for ALL of your data
– The HPE portfolio has unmatched breadth vs. any other vendor
  – Hardware
  – Software
    – IDOL
    – Partners & FOSS
  – Services
– It’s all shipping, and it all has customer proof points
– HPE is a solid company and is committed to “Big Data”
– HPE has the technology, people, and partners to get you on track
– Who has a stronger portfolio?
The “data lake” paradigm will build

- Data will first be kept in cheap file storage (HDFS, at worst)
  - It isn’t necessarily “unstructured,” just not yet structured

- A variety of tools will be available to structure & analyze data
  - Dumb storage isn’t enough
  - Unstructured, self-structured, structured, graph, aggregated, all on one platform

- The variety of software and tools to treat data will continue to expand
  - A pyramid of choices will align costs of keeping and curating the data with the value of the data

- Enlightenment on “democratization” will be reached… but we are not there yet
  - Collaboration, sharing, etc.

- Tools will be available to prioritize critical processes & data
  - Resource provisioning
  - Break out dedicated “marts”, etc.

- The new “one size fits all”?
Archiving, ECM, Records Management, ControlPoint

Data Lake
Smarter, Safer

Vertica SQL
UDX/SDK/API
IDOL
R
Hadoop Tools

Vertica
Hadoop
Tools

IDOL
R
Hadoop
Connectors
(Flume, Storm, Sqoop, Scribe)
Hadoop connectors
(Vertica)
Vertica connectors
CEP/streaming
EL (Structured, PolyStructured)
ArcSight
IDOL Connectors (KeyView)

Ingest

On Premise/Cloud

Store & explore

Govern & protect

Serve

Virtualization tools
Web portal
eDiscovery
APIs/ODBC/JDBC/R
Search
Reports
Dashboard

Infrastructure - Hardware, Cloudburst, BackupData Protector

Smarter, Safer
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 OnDemand & AMI
Get up and running with proven Data Warehouse as a Service in minutes.
Install your Vertica license on select cloud hardware to create cluster on AWS.

Vertica AMI: http://amzn.to/1OLVe4C
VOD: https://saas.hpe.com/buy/vertica-on-demand

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

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
Thank you
bryan.whitmore@hpe.com

#SeizeTheData
Advanced Analytics World Tour

#SeizeTheData