VERTICA

Vertica Now

Quarterly Customer Webinar

January 19, 2017

Hewlett Packard Enterprise

VERTICA

Vertica Now is our brand new, customer-only quarterly webinar designed to keep you informed of new releases and updates, helpful technical resources, future roadmap capabilities, and more!

Today's Agenda

- **Vertica Updates:** Micro Focus Spin-merger **(Colin Mahony, SVP/GM Big Data Platform)**
- **What's New:** Vertica 8.0SP1 Release **(Misha Davidson, Director of Engineering)**
  - In-database analytics additions and improvements
  - Management Console for Azure and AWS
  - Open Source and format integrations
  - Scale and concurrency performance enhancements
- **Did You Know:** Geospatial Analytics **(Casey Starnes, Information Developer)**
  - Introduction
  - How Vertica approaches Geospatial Analytics (use case example)
  - Summary and resources
- **Q&A**

VERTICA

Micro Focus Spin-Merger

Colin Mahony

Hewlett Packard Enterprise

Welcome to an exciting 2017!

Colin Mahony
SVP & GM, HPE Big Data

VERTICA

Hewlett Packard Enterprise

Continuing the journey toward a pure play software business

Q: What areas are you most excited to focus on and grow within the HPE software portfolio?

The first is big data analytics. Big data analytics is in my core mission statement because it is required in every piece of software now—security, of software now—security, IT operations, application development, archiving. Customers want to be able to quickly search, find, and make search, find, and make decisions using big-data analytics. Vertica is—and will continue to be—one of the most prevalent big data analytics OEM prevalent big data analytics OEM platforms in the industry. If you look at companies that have big data as their models—Uber, Facebook, and a models—Uber, Facebook, and a number of others—Vertica is at the core of that.

Chris Hsu, EVP & GM of HPE Software
(Named CEO of newly formed company)

http://bit.ly/2jbLPJG

The New Company: Micro Focus + HPE Software

One of the world's largest pure-play software companies

- Annual Revenue **$4.5** billion*
- FTSE **100** company
- Customers **50,000**
- Employees **~20,000**

With one of the broadest IT SW portfolios - designed to address customers' challenges across:

- IT Operations
- Security
- Information Governance
- **Big Data Analytics**
- Cloud
- Linux & Open Source
- DevOps

Supported by strong leadership and operations

World-class management team comprising leaders from HPE Software and Micro Focus

Set up for success

- HPE shareholders will own 50.1% of the new company
- Senior HPE executive will serve on the board
- HPE will nominate 50% of independent directors

- Stronger go-to-market
- Deeper R&D resources
- Improved operational efficiencies and scale

… and we're just beginning… the new company will have a strong platform for M&A.

VERTICA

Hewlett Packard Enterprise

* Trailing twelve months as of April 30, 2016 for HPE's Software segment and Micro Focus including Serena Software for the full year

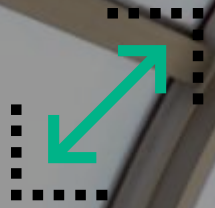
VERTICA

What's New: 8.0SP1 Release

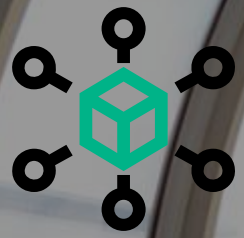
Misha Davidson

Hewlett Packard Enterprise

Vertica’s strategic focus



Performance at scale

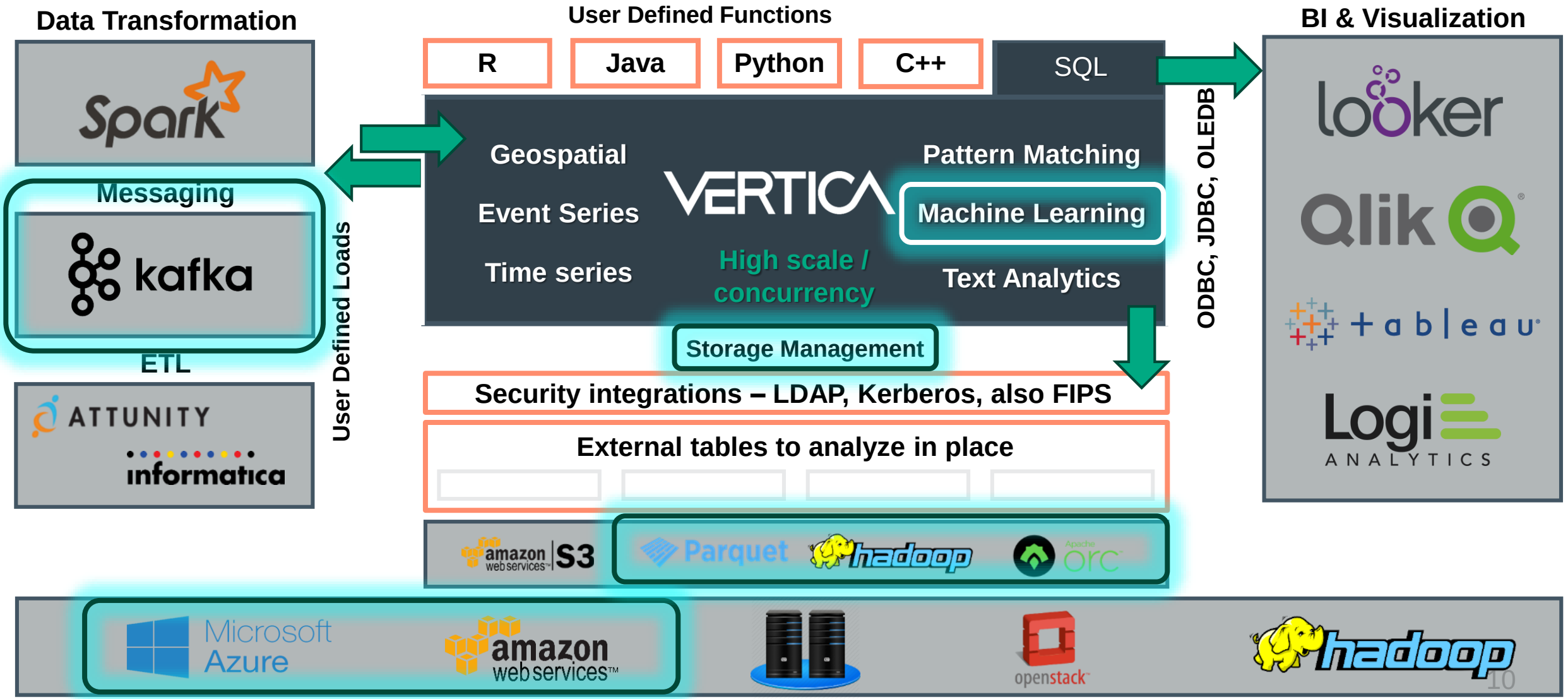


Integration: interoperation with open source / formats

Multi cloud: freedom to deploy anywhere

In-database analytics & machine learning

Vertica – Big Data Analytics integrated with OSS / CSPs through an ecosystem-friendly architecture



In-database Analytics & Machine Learning

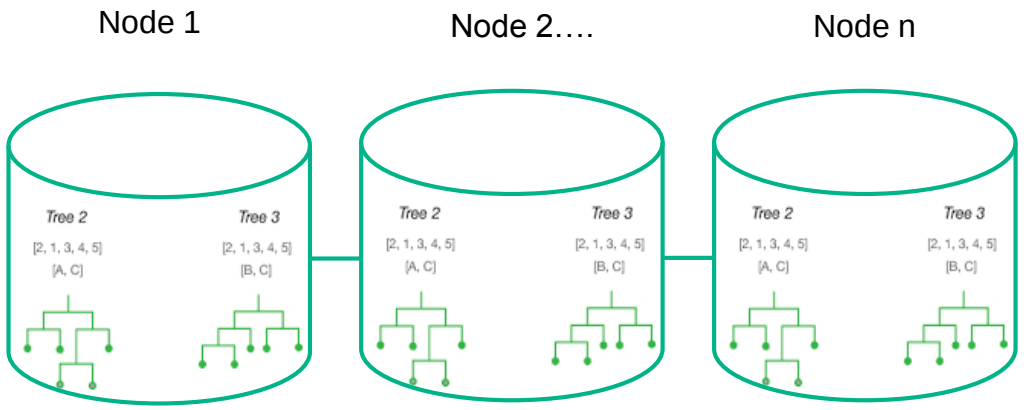
Why put Machine Learning in Vertica?

Limitations of standalone solutions for ML in Big Data

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 Vertica Server

New capabilities deliver predictive analytics at speed and scale



- Run in parallel across hundreds of nodes in a Vertica cluster
- Eliminating all data duplication typically required of alternative vendor offerings
- No need to “down-sampling” which can lead to less accurate predictions
- A single system for SQL analytics and Machine Learning

VERTICA

Hewlett Packard Enterprise

Machine Learning Pack – 8.0SP1 is much faster

Now installed with Vertica by default starting from 8.0.0

Algorithm	Model Training	Prediction	Evaluation
Linear Regression	✓	✓	✓
Logistic Regression	✓	✓	✓
K-means	✓	✓	✓
Naïve Bayes	✓	✓	✓

New in 8.1

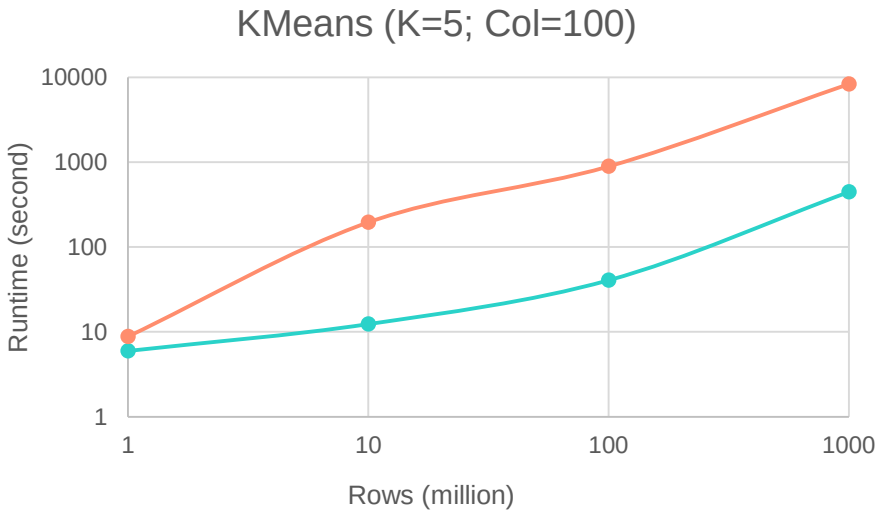
Data Preparation	Normalization	Imbalanced data processing	Sampling	Outlier Detection
	✓	✓	✓	✓

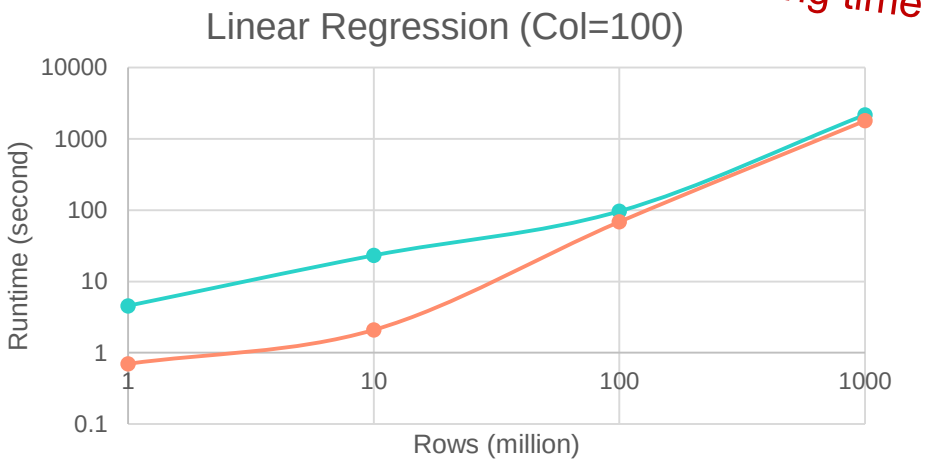
Model Management	Summarize models	Rename models	Delete models
	✓	✓	✓

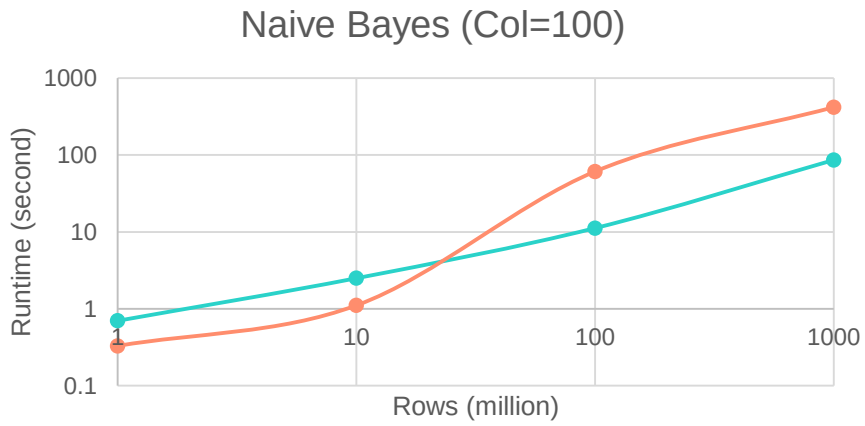
Vertica's ML libraries run faster / scale better than Spark

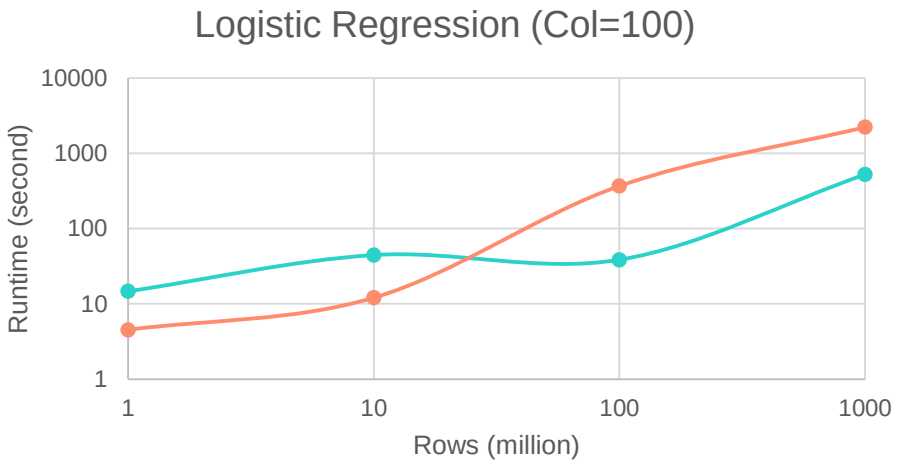
Benchmark between Vertica 8.01 and Spark 2.01

For Spark, the runtime doesn't include the data loading time from HDFS









VERTICA

Hewlett Packard Enterprise

Multi cloud: freedom to deploy anywhere

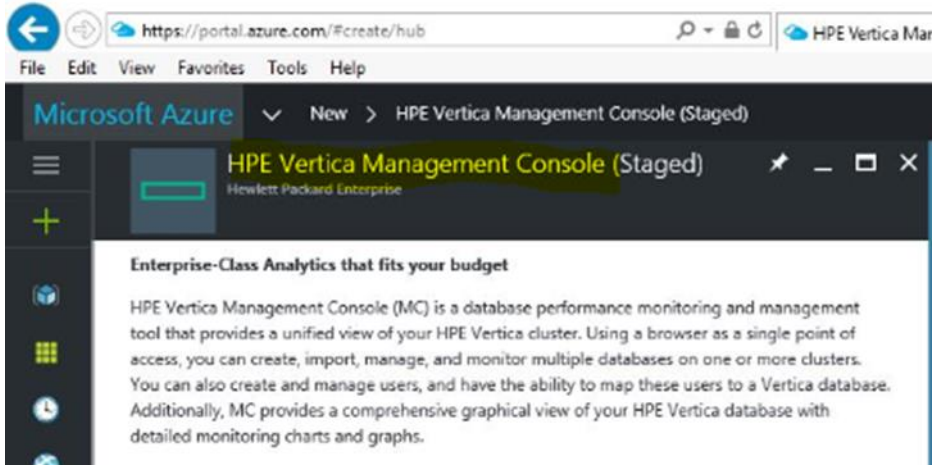
AWS / Azure

Vertica multi cloud

New in 8.0SP1 – Management Console (MC) support

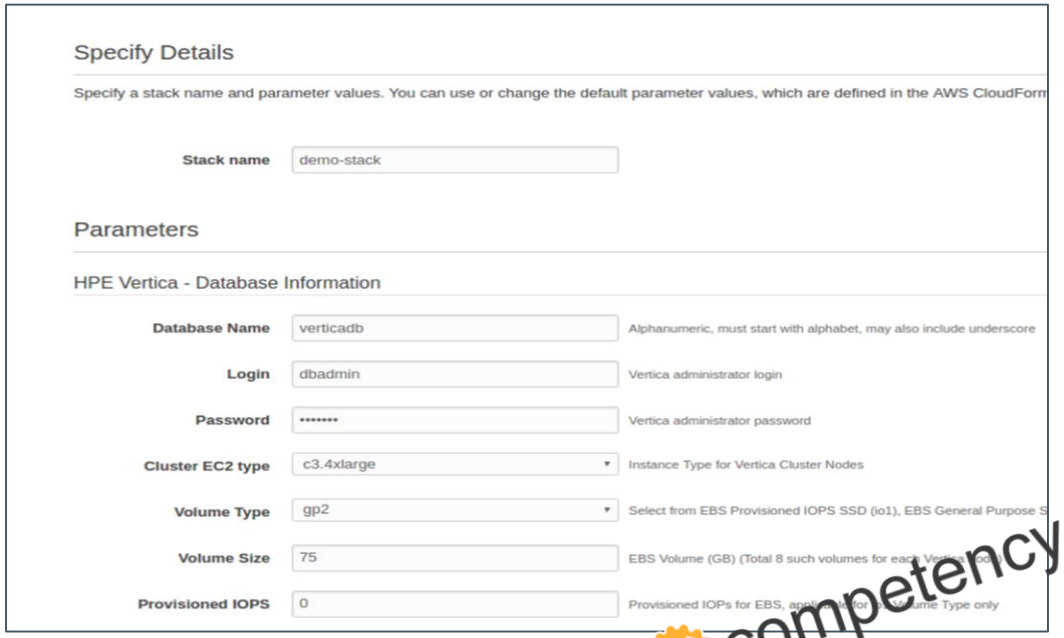
Management Console is now available in Azure

Can be instantiated alongside a Vertica cluster running in Azure as well



3 nodes + MC Cloud Formation Template now works in AWS

In about 15 minutes, the new 3 node Vertica cluster with MC is ready. All access information can be viewed in the create stack's outputs tab.



VERTICA

Hewlett Packard Enterprise

aws competency

Integration: interoperation with open source / formats

Hadoop / Kafka

Vertica SQL on Hadoop integration

Vertica is fast and Hadoop is big; our integration brings together fast and big

Vertica offers only full-featured query engine for Hadoop

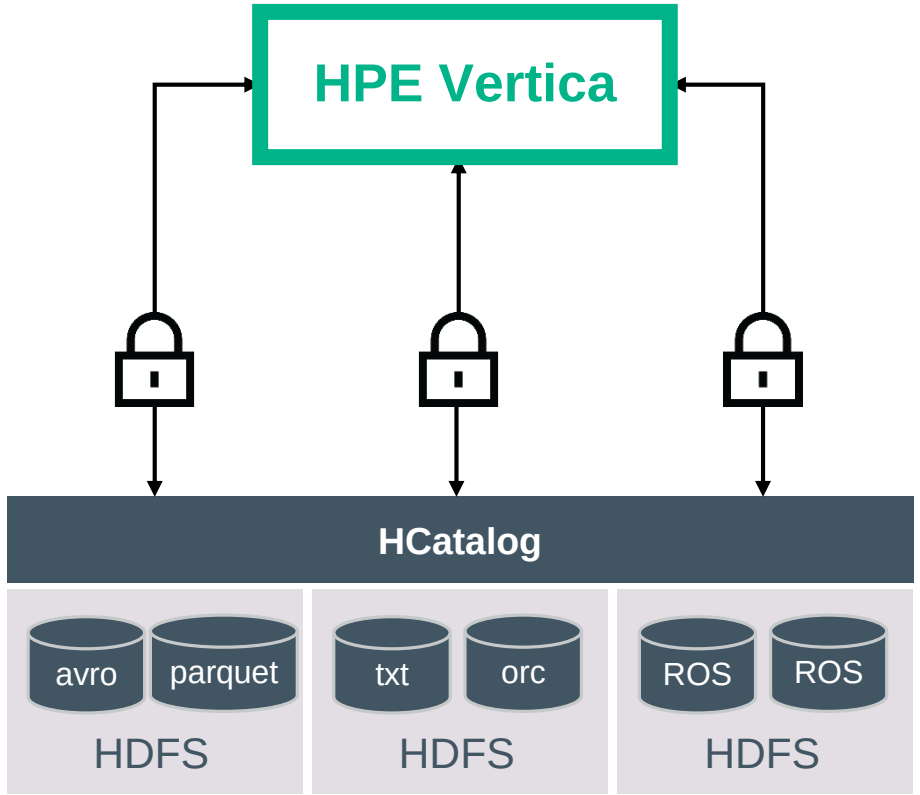
– Analyze data in place from HDFS, both in Parquet and ORC
– Integrated with HCatalog, is Hadoop distribution agnostic
– Rich SQL support – same core engine for both deployments
– High performance C++ libhdfs connectivity, parallel readers
– Integrated with Kerberos security
– Support for running on Hadoop machines, or dedicated cluster

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 in 8.0SP1

– Support concurrent access to multiple HDFS clusters
– Added support for HA Name Nodes
– Expose syntax for Hive-style partition pruning for ORC and Parquet readers



MAPR Hortonworks cloudera CERTIFIED

VERTICA Hewlett Packard Enterprise

*Results are reflective of internal testing by HPE in a controlled environment. Actual performance may vary depending on application, installation, environment, datasets, and other factors

Vertica - Kafka integration

Scalable, Fault Tolerant load with exactly-once semantics

Vertica schedules loads to continuously consume from Kafka

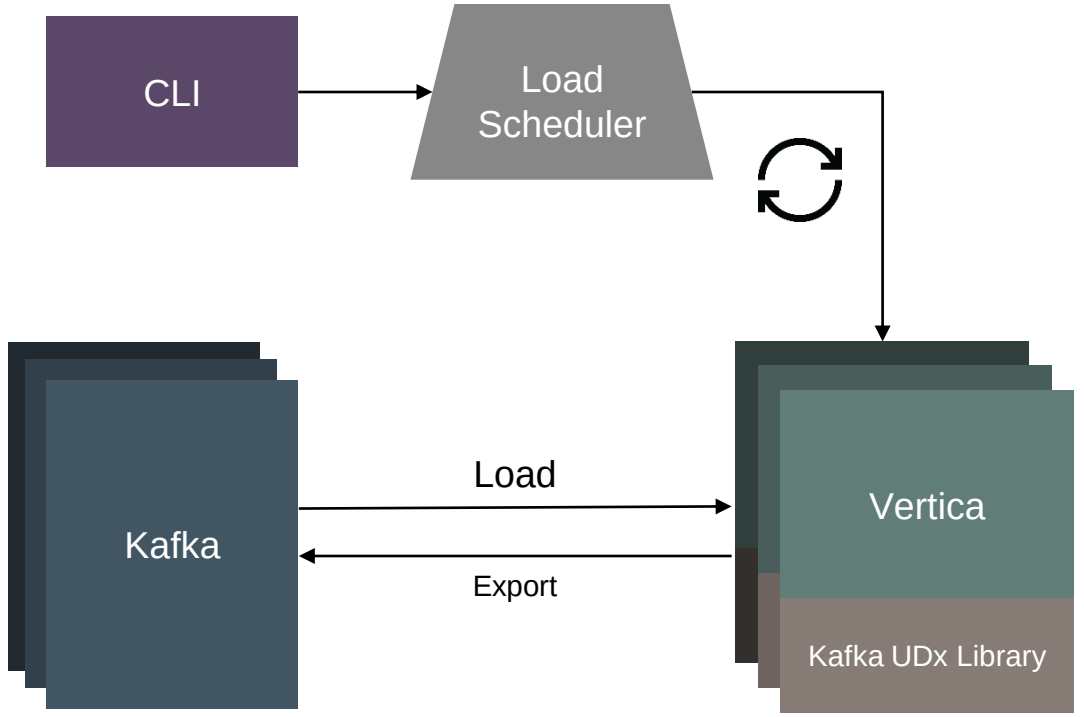
- JSON, Avro, and custom parsers
- Handles bursty data, dynamically prioritizes topics according to their throughput
- In-database monitoring with admin GUI

Terabytes per hour loaded, no tuning

- 3 Node Vertica Cluster (24 cores, 252 GB RAM each), 1 Kafka Broker

New in 8.0SP1

- Easier to use, automated configuration / validation of Kafka topics and brokers



Performance at scale

Operations / Data Management

More efficient and scalable rebalance in 8.0SP1

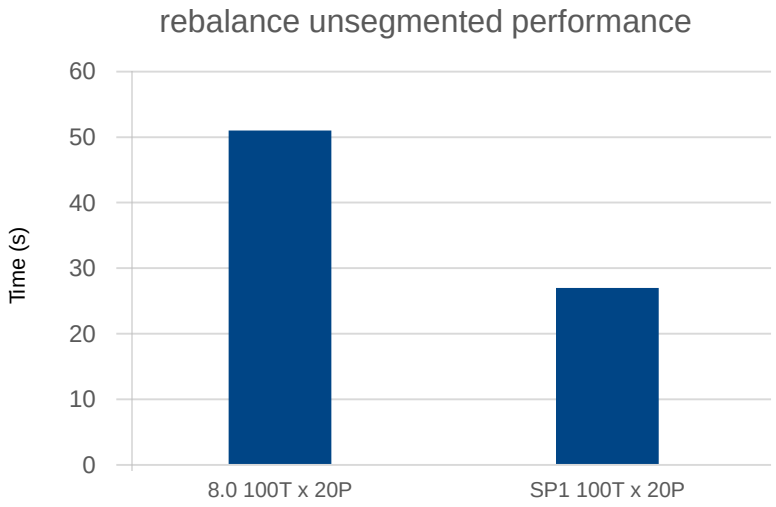
Improved performance and scale

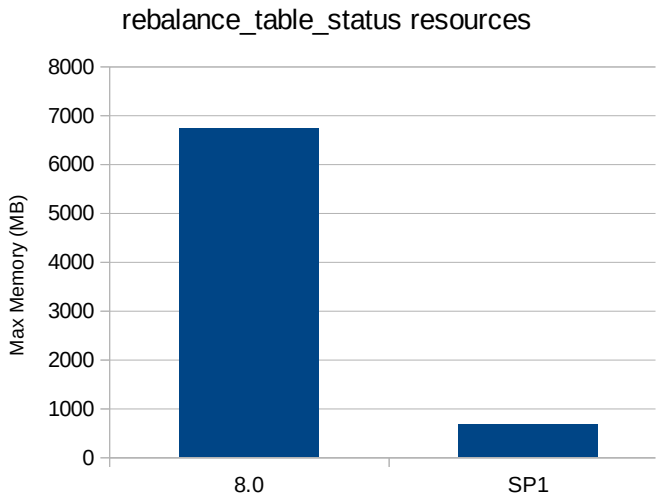
E.g., rebalance from 4 to 14 nodes improvements in 8.0SP1:

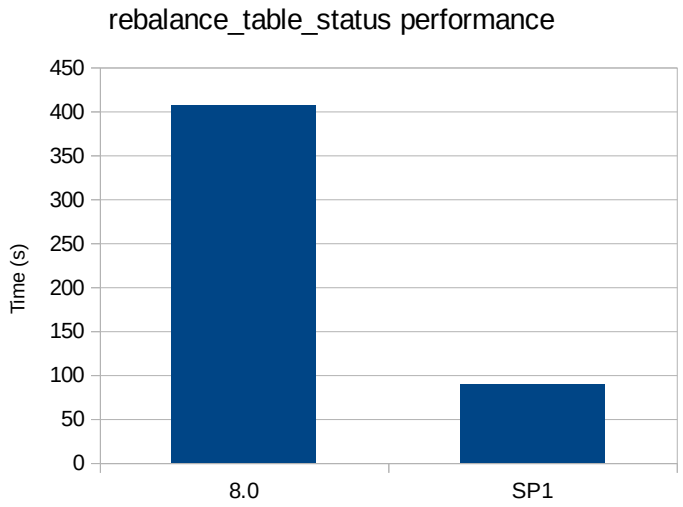
– 10-100 Tables, 20-200 Projections

Improved monitoring tables / queries

Both execution time and memory footprint of monitoring operations are dramatically reduced







VERTICA

Hewlett Packard Enterprise

VERTICA

Did You Know: Geospatial Analytics

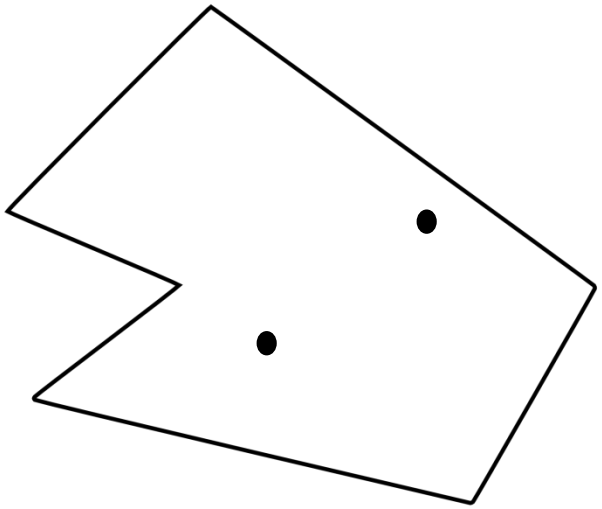
Casey Starnes

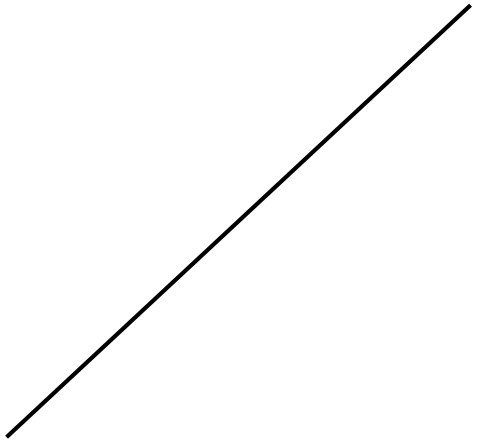
HPE Vertica Geospatial Analytics

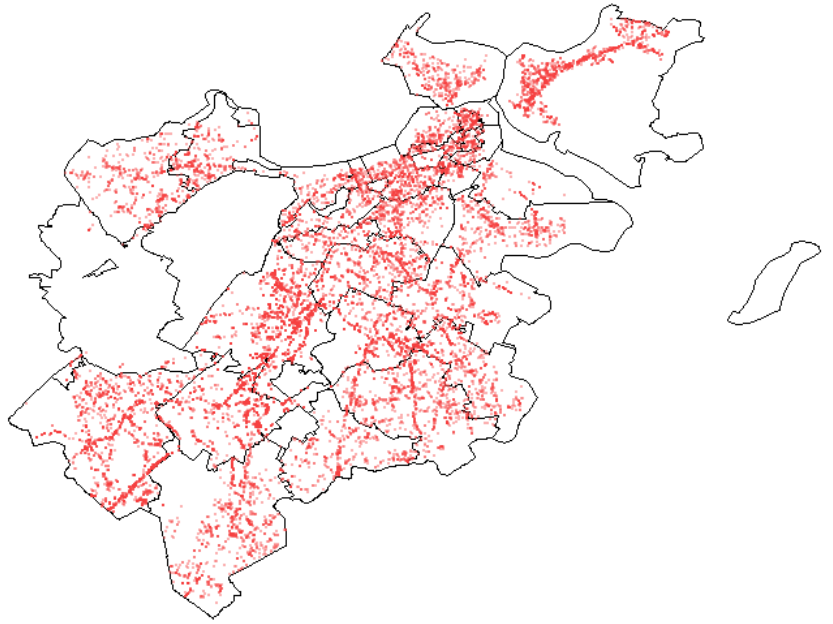
Quick Overview

What do I use Geospatial Analytics for?

Use Geospatial Analytics to store and query vector data (points, lines, polygons) in Vertica.







HPE Vertica Geospatial Analytics

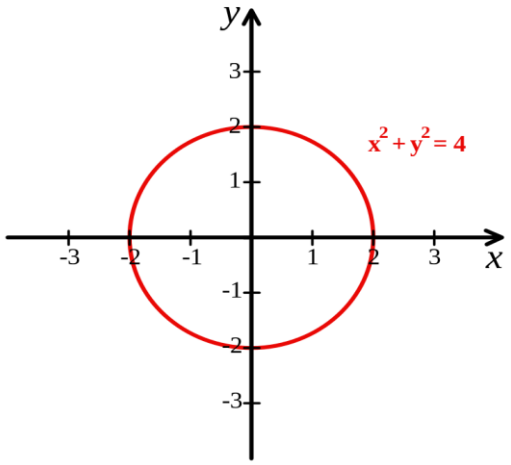
Quick Overview

What data types are supported?

Vertica supports two spatial data types that can be used to store geographical objects such as points, lines, and polygons.

– GEOMETRY: Used to store planar data. It is generally used to store X,Y coordinates in a two-dimensional space.

– GEOGRAPHY: Used to store spherical (round-earth) data, or an object in the WGS84 coordinate system. It is used to store longitude and latitude coordinates that represent points, lines, and polygons on the Earth's surface.



HPE Vertica Geospatial Analytics

Quick Overview

What functions are supported?

Vertica provides 65+ built-in functions for spatial analysis. The functions allow for the creation, comparison, analysis, and retrieval of spatial data.

ST_<function_name>
Compliant with OGC standards.
(40+ functions)

- `ST_Area`
- `ST_Distance`
- `ST_Intersects`
- `ST_IsValid`

STV_<function_name>
Vertica specific; not OGC compliant.
(25+ functions)

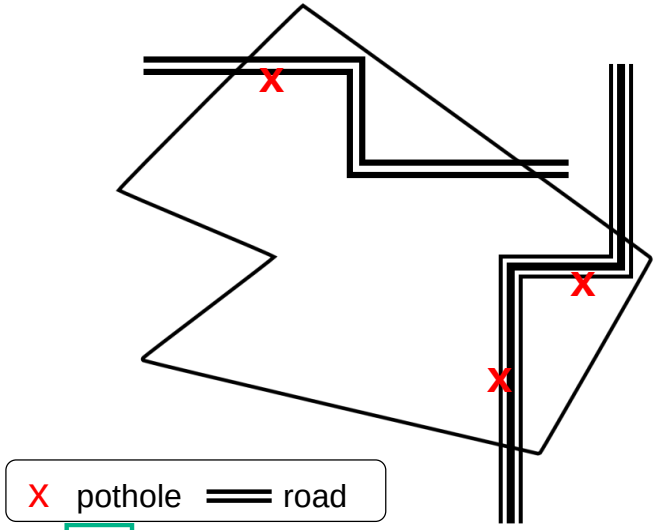
- `STV_Create_Index`
- `STV_NN`
- `STV_Intersect`
- `STV_Export2Shapefile`

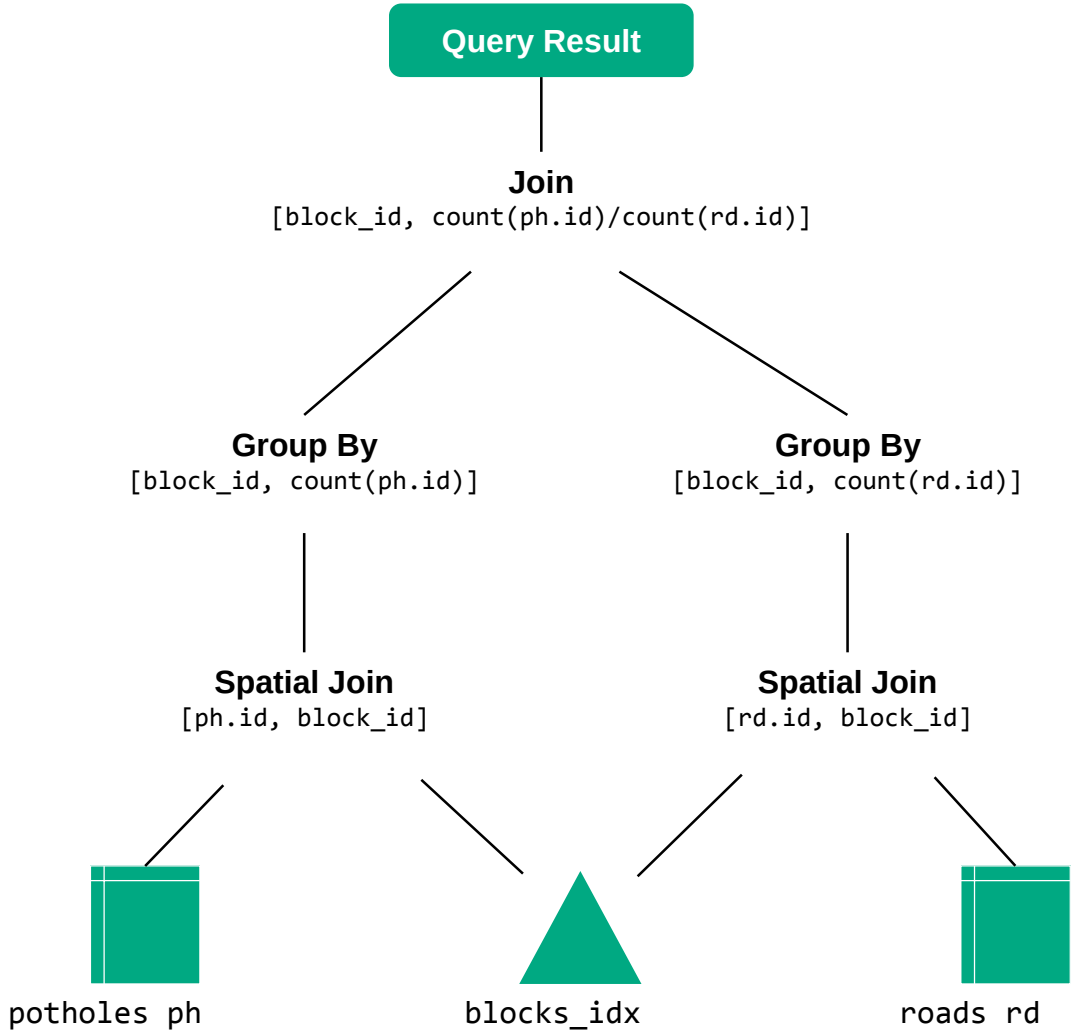
HPE Vertica Geospatial Analytics

Quick Overview

How do spatial indexes in Vertica work?

Spatial indexes are only used when performing spatial joins using STV_Intersect





VERTICA

Hewlett Packard Enterprise

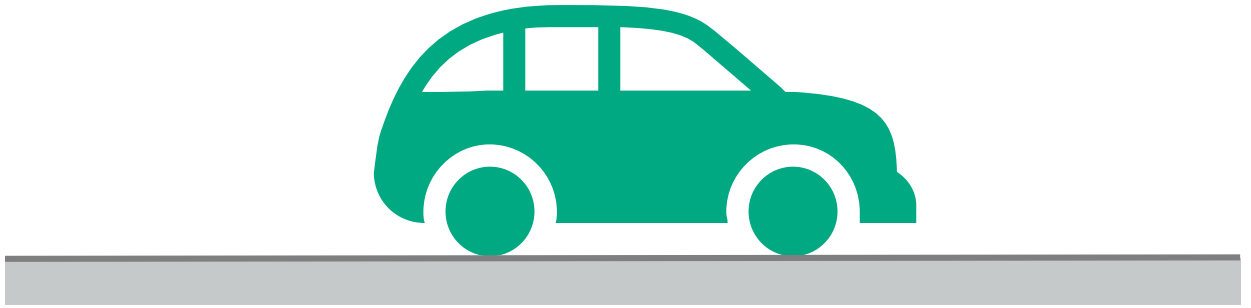
Boston Potholes

Use Case

Boston Potholes - Use Case

Potholes are a fact of life in New England

Boston Potholes - Use Case



Potholes need to be fixed

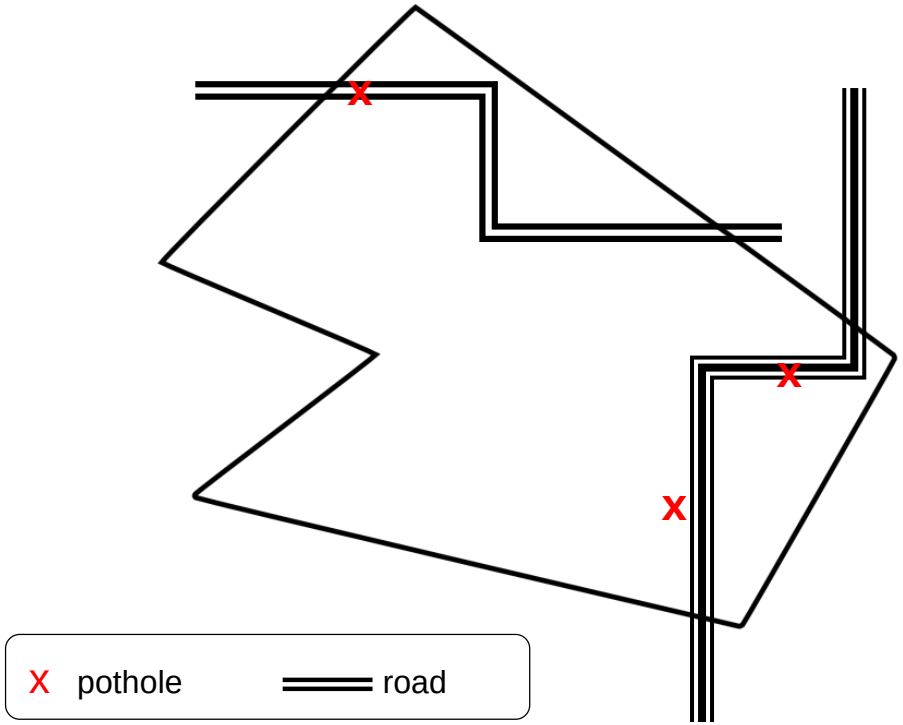
Boston Potholes - Use Case

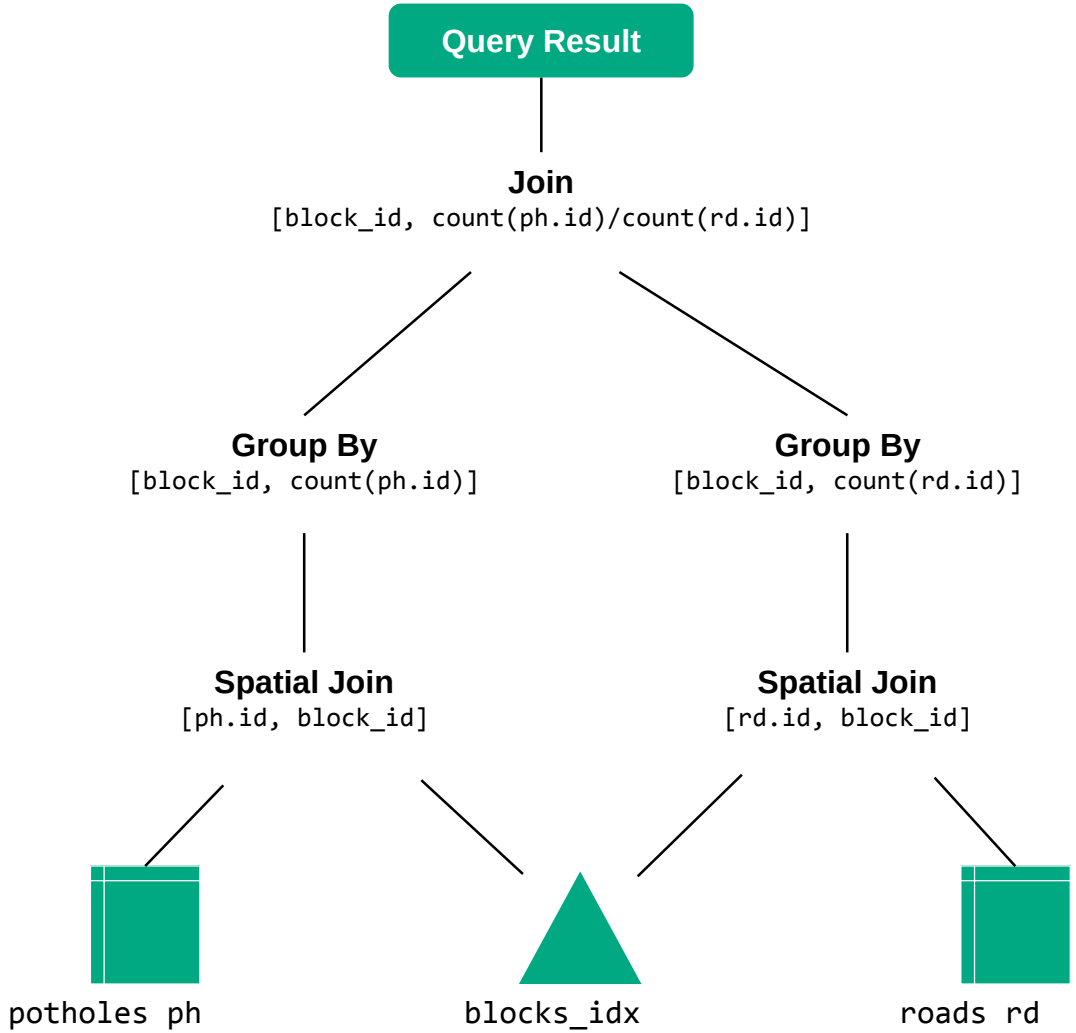
Where in Boston are the most potholes fixed?

Pothole spatial analysis

Block Pothole Density:

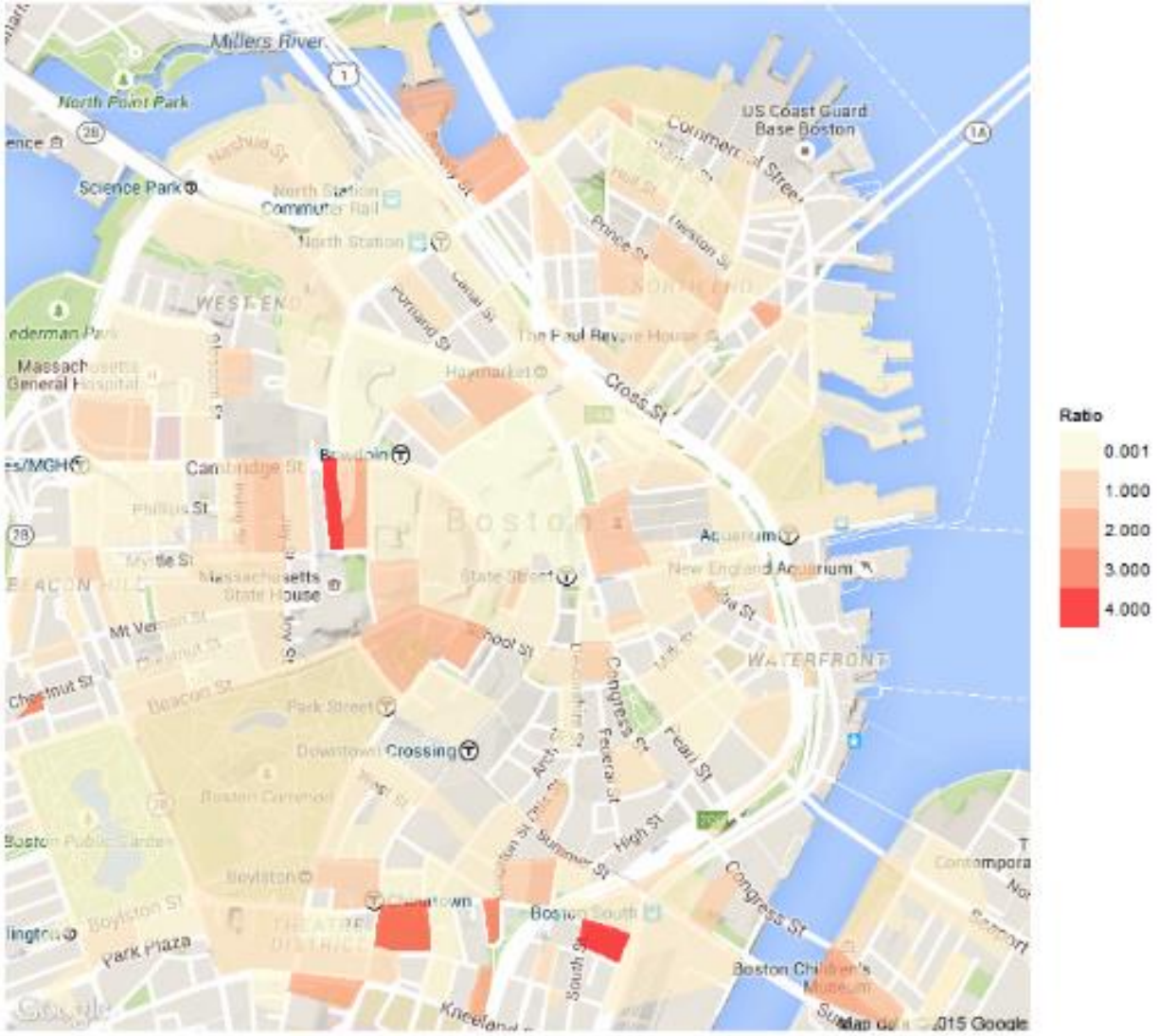
$$\frac{\text{\# potholes}}{\text{\# roads}}$$





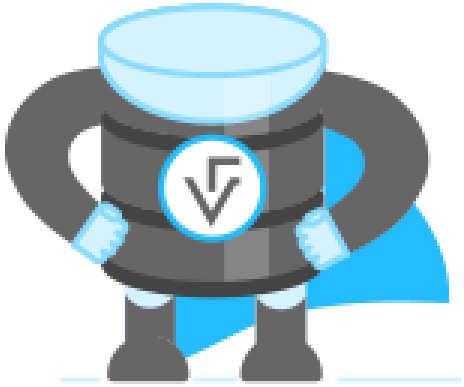
Block Pothole Density = 3/2

Boston Potholes - Use Case



D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

Geospatial Analytics - Summary



- Spatial data types follow the OGC standard
- 65+ geospatial functions
- Perform fast spatial joins using STV_Intersect
  - Spatial indexes can only be used with STV_Intersect
- Leverages the Vertica MPP architecture for scalability

- In-database geospatial analytics
  - Spatial data types and SQL functions
  - Inclusion of geographical dimension in data analysis

- Additional Resources and Examples: https://github.com/vertica/Vertica-Geospatial

VERTICA

Q&A

Do you have additional feedback for our team?

BigDataPlatformCustomerCareTeam@HPE.com

Hewlett Packard Enterprise