

White paper



Location Intelligence

Spatial Data Lake Solution

The high-performance geospatial data processing
that powers today's big data frameworks



Power up your spatial data

Organizations will invest more than \$150 Billion in big data and business analytics in 2017, a 12 percent increase over the prior year.¹ Not surprisingly, they're looking for ways to accelerate results.

Frameworks such as Hadoop already make it possible to assemble data lakes, so you can develop applications based on these centralized data stores. Much of this structured, semi-structured and unstructured data is referenced by geographic coordinates, which means location analytics takes on an increasingly important role.

The Pitney Bowes Spatial Data Lake Solution overcomes inherent challenges, so organizations can take full advantage of spatial context. With one integrated, location-centric platform, firms can easily produce and share business insights across the enterprise.



Spectrum™ Data Lake Solution

This white paper details four technical advantages of this fully-integrated, enterprise-class platform.

- High-performance geospatial processing
- Seamless integration with business applications
- Orchestration and workflow management
- Flexible deployment options

¹International Data Corporation: Worldwide Semiannual Big Data and Analytics Spending Guide, 2017

One solution to deliver the location-based insights necessary to maximize the returns on your big data investments

The need to locate success.

Mobile apps, geotags and smart sensors generate millions of new data points every hour, all of which are referenced by location. While Hadoop delivers a framework to store and process massive amounts of data, most firms have trouble extracting value from the location data they collect. Studies show that up to 73 percent of this data goes unused in day-to-day analytics and Business Intelligence.²

A greater understanding of location-based information allows companies to manage their data more effectively. More importantly, when organizations can enrich, analyze, query and visualize data with maps, they can recognize proximity relationships in order to make better, faster decisions.

With the Pitney Bowes Spatial Data Lake Solution, organizations can take advantage of more data assets in more ways, generating insights that otherwise could not be attained.



²The Forrester Wave™: Big Data Hadoop Distributions, Q1 2016

High-performance spatial data processing at scale.

The Pitney Bowes Spatial Data Lake solution is a highly scalable, geospatial platform that allows companies to process, analyze, manage and visualize location-based data assets across organizational boundaries.

It's developed on the Pitney Bowes Spectrum™ solution for Big Data and takes advantage of key strengths and capabilities that industry leaders have relied on for over thirty years.

- Leader in data quality and address validation
- Pioneer in Location Intelligence and Geographic Information Systems
- Expertise in modern big data and cloud technologies

Spatial Data Lake integrates natively within big data environments, such as Hadoop, Spark, and NoSQL. Now organizations can process massive amount of location-based data in a highly scalable and efficient way.

Benefits at a glance

Speed and scalability

Integration with popular big data platforms such as Hortonworks, Cloudera and Amazon Web Services (AWS)

Container and micro-server oriented architecture simplifies day-to-day management

An API-oriented approach to integration

Easy integration with various enterprise computing platforms and software

Spectrum Data Lake: Technical overview

The Pitney Bowes Spatial Data Lake Solution brings big data to more knowledge workers within the organization. Through a combination of technologies and services, users can overcome challenges and take full advantage of location context with one integrated solution.

01. High-performance geospatial processing

Moves time-consuming geospatial processes into a centralized, high-performance and scalable big data framework.

02. Seamless integration with business applications

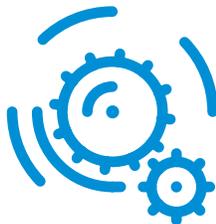
Enables users to access and consume enhanced data seamlessly in their preferred software. Deliver via desktop GIS, BI and Data Science tools, and web applications.

03. Orchestration and workflow management

Enables users to customize related settings in different processes, connecting processes to create workflows. The solution also makes it easy to manage related data store integration and add-ons.

04. Flexible deployment options

Deploys seamlessly in Hadoop or Mesos clusters, either on premise or in the cloud.

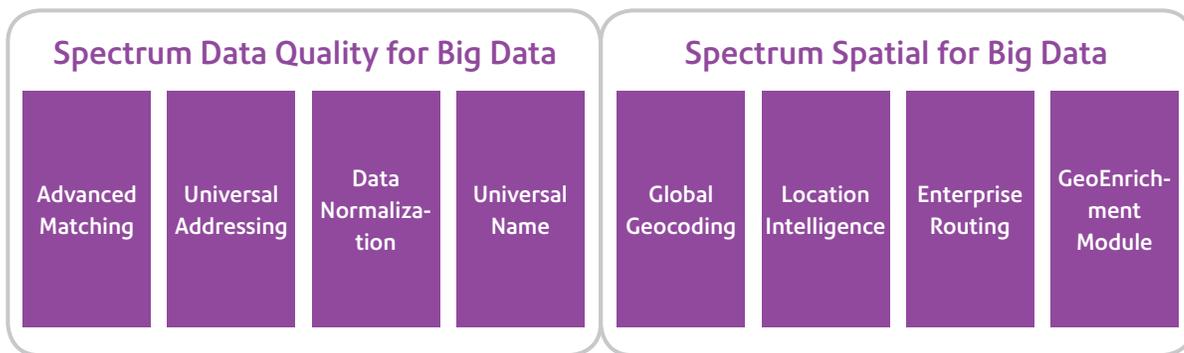


01. High-performance geospatial processing

Spatial Data Lake utilizes key components and modules within the Spectrum for Big Data product family.

Spectrum™ for Big Data

Enabling big data frameworks with data quality and geospatial technology



Unique geospatial processing

Data quality plays an important role, but Pitney Bowes offers much more. With our Spatial Data Lake Solution, organizations can apply the most advanced, comprehensive range of geospatial operations natively within big data frameworks like Spark and Hadoop. Capabilities include:

Geometry operations

- Predicate functions
 - Disjoint, intersects, overlaps, within, etc.
- Validity and measurement functions
 - Area, length, perimeter, distance, closest points, etc.
- Processing functions
 - Buffer, intersection, convex-hull, union

XY-point data processing

- Flexible and powerful points to grid aggregation
 - Aggregated data into multiple level hexagon, rectangle or square grids

- Polygon aggregation
 - Aggregate in geometry boundaries (e.g. zip code)
 - Aggregate in calculated boundaries (e.g. road buffer)
- Point to line processing
 - Snapping a point records to a road segmentation

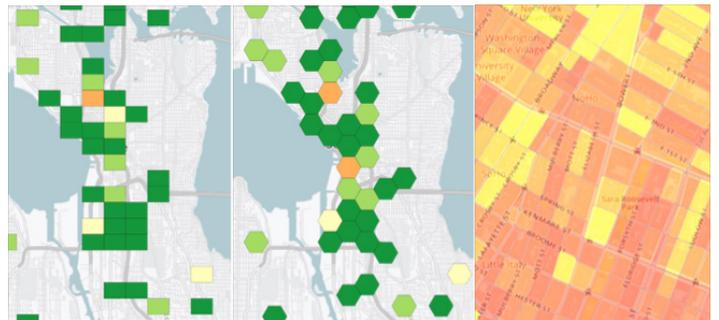


Figure 1: Visualization showing different XY-point data aggregation methods (rectangle, hexagon, polygon)

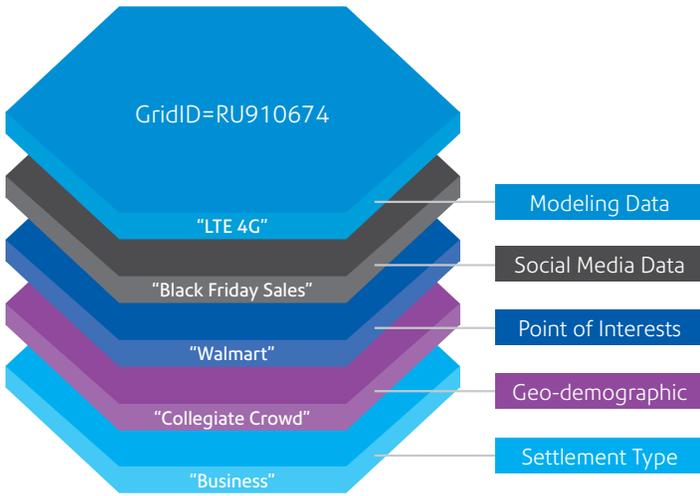


Figure 2: grid based geo-enriching method

Powerful geo-enrichment capabilities

Geo-enrichment processing

- Grid-based enrichment
 - Aggregate points to a grid (rectangle, hexagon, polygon) and attached related context information such as geo-demographic information within the grid
- Geospatial calculations
 - Find nearest POIs (e.g. shop, firestation)
 - Point in Polygon of an address to a flood zone

Map tiling

- Map visualization (See Figure 1)
 - Create vector tiles (MVT) in batch
 - Create raster tiles (e.g. PNG, MRR) in batch (future release)

Geocoding and routing

- Spectrum™ Geocoding for Big Data
 - Global forward geocoding
 - Global reverse geocoding
- Spectrum™ Routing for Big Data
 - Point-to-point distance and time
 - Isochrones
 - Isodistance

02. Seamless integration with business applications

The Spatial Data Lake solution enables users to access and consume geospatial-processed data easily. You can serve processed data and map tiles across a wide variety of user interfaces, including desktop GIS, BI tools, Data science tools, as well as web or mobile map APIs.

Desktop GIS integration

Enables Desktop GIS solutions such as MapInfo Pro™ software to access processed data and map tiles using JDBC connection or map services.

Spectrum™ Spatial Analyst

Supports web based visualization within a standard, off-the-shelf web mapping platform.

BI tools integration

Enables tools such as Tableau to access processed data and map tiles using JDBC connections or APIs.

Data science tools

Integrates with modern data science tools including Apache® Zeppelin™ notebook.

APIs

Users can employ a series of Data and Map Tile APIs to create a customized web or mobile interface.

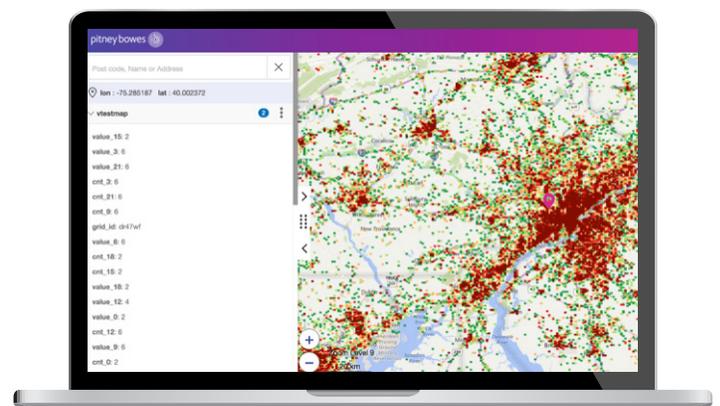


Figure 3: Spectrum Spatial Analyst integration

03. Orchestration and workflow management

Spatial Data Lake enables users to customize related settings in different processes, connecting processes to create workflows. The solution also makes it easy to manage related data store integration and add-ons.

Process customization

The solution makes it easy for users to control related settings in geospatial processes. In this way, users can quickly create customized applications.

Examples of these settings include:

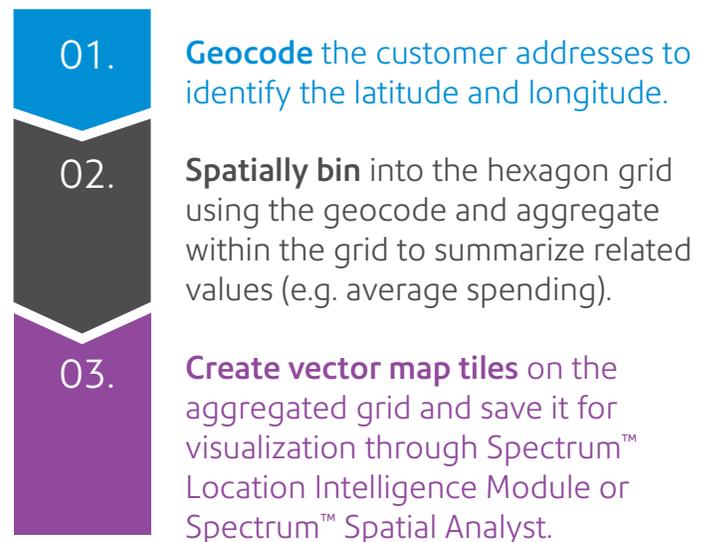
- Geocoding: precision filtering such as street interpolation, etc.
- Routing: major road only, off-road options, etc.
- Grid aggregation: type, level, etc.
- Map tiling: tile size, default style, etc.

Users can also control and set their own business logic in related processes. For example, when a user aggregates XY-point records with different attributes from mobile devices into grids, they often want to calculate and select different statistics on selected attributes (e.g. mean or median of wireless signal strength).

Workflow management

This capability enables users to chain different processes in the Spatial Data Lake to create a workflow for a specific task.

For example: A user wants to create a hexagon grid-based map visualization of 100 million customer addresses. Within Spatial Data Lake, they can chain the necessary processes to create a specific workflow:



Data store integration and controls

Supported integration includes but is not limited to:

- Read and write to HDFS
- Read and write to S3
- Read and write to Cassandra
- Read and write to relationship databases like SQL server, PostgreSQL

You can also easily control the data storage format and schema. For example, BI tools like Tableau can only read geometry objects in a customized WKT format. Spatial Data Lake can use a customized format add-on to write the aggregated data in the precise format required by Tableau.

04. Flexible deployment options

Organizations can deploy Spatial Data Lake in Hadoop or Mesos clusters, either on premise or in the cloud.

For example, the figure below illustrates the architecture in which Spatial Data Lake in a Mesos cluster on AWS VPC.

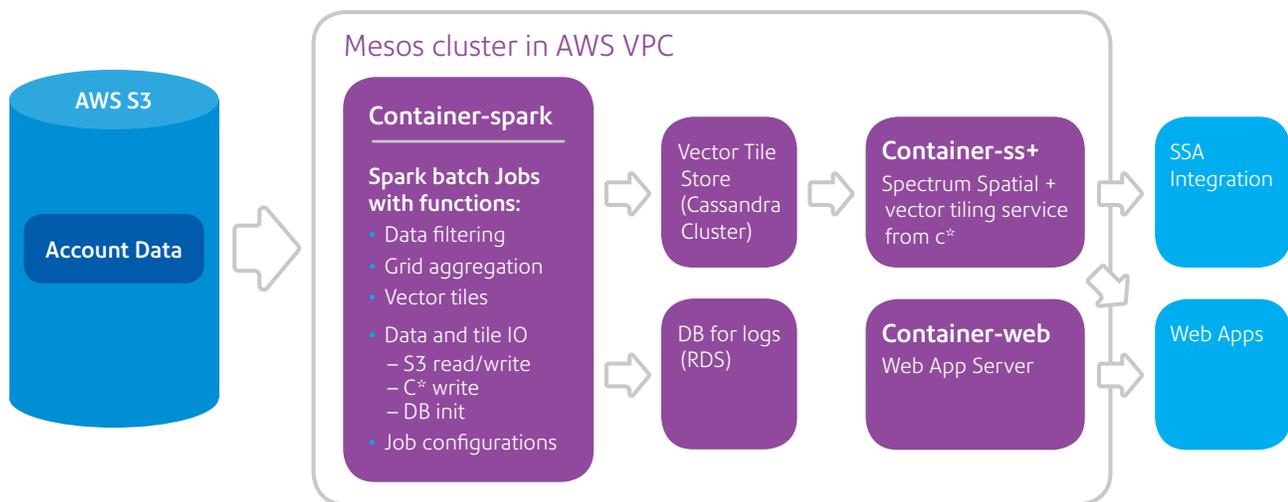


Figure 4: Mesos container based deployment example

In this example, the Spatial Data Lake processing engine is packed into a Docker container and could be launched by the Marathon container orchestration platform. This processing engine is developed on top of Spectrum Location Intelligence SDK with customer-specific business logic. The deployment also leverages the client's Cassandra database for pre-generated vector tile storage and caching, and a PostgreSQL database for meta-data and log management.

Pitney Bowes experts can work with you to illustrate and deploy an appropriate architecture based on your unique business needs.

Power up your spatial data with Pitney Bowes.

Spatial Data Lake is built upon a legacy of address validation, data quality and Location Intelligence. Today, Pitney Bowes powers billions of transactions across the world of commerce. Ranked among the PwC Global 100 Software Leaders, our solutions fuel the digital transformation of the world's top banks, insurers, telecoms, retailers and government agencies.

Our technologies and services are consistently recognized for their innovative, world-class capabilities. Recent recognition includes:

<p>A leader in The Forrester Wave™: Master Data Management, Q1 2016</p>	<p>A leader in The Forrester Wave™: Geospatial Analytics Tools And Platforms, Q3 2016</p>	<p>A leader in The Forrester Wave™: Customer Analytics Solutions, Q1 2016</p>	<p>CRN Partner Program 5-Star Rating</p>

Now, these capabilities integrate natively within big data technologies, such as Hadoop, Spark, and NoSQL. With the Spatial Data Lake solution, organizations can process massive amounts of location-based data in a highly scalable and efficient way. To learn more, please contact us (see a listing of local offices on the next page).



Product details

This technical white paper provides an overview of the technical advantages of this fully-integrated, enterprise-class platform. For details on the specific Pitney Bowes products and services incorporated in the Spatial Data Lake solution, please speak with your account manager or visit pitneybowes.com.

Spatial Data Lake incorporates products, related integration services and a web interface, which include:

- Spectrum™ for Big Data plus Software Development Kit (SDK)
 - Spectrum™ Location Intelligence for Big Data
 - Spectrum™ Geocoding for Big Data (optional)
 - Spectrum™ Routing for Big Data (optional)
- Spectrum™ Platform modules
 - Spectrum™ Location Intelligence Module
 - Spectrum™ Spatial Analyst
- MapInfo Pro™
- Proprietary global data sets
 - Streets
 - Boundaries
 - Points of interest
 - Demographics
- Location Intelligence APIs

United States

3001 Summer Street
Stamford, CT 06926-0700
800 327 8627
pbsoftware.sales@pb.com

Europe/United Kingdom

The Smith Centre
The Fairmile
Henley-on-Thames
Oxfordshire RG9 6AB
0800 840 0001
pbsoftware.emea@pb.com

Canada

5500 Explorer Drive
Mississauga, ON L4W5C7
800 268 3282
pbsoftware.canada.sales@pb.com

Australia/Asia Pacific

Level 1, 68 Waterloo Road
Macquarie Park NSW 2113
+61 2 9475 3500
pb.apac@pb.com

For more information,
visit us online: pitneybowes.com