pitney bowes

Spectrum™ Technology Platform
## Version 11.0
Summary of Features

Spectrum™ Technology Platform is an enterprise data quality solution that improves the completeness, validity, consistency, timeliness, and accuracy of your data through standardization, verification, and enhancement. Ensuring that your data is accurate, complete, and up to date enables your firm to better understand and connect with your customers.

Contents:

Enterprise Data Management Architecture

With Spectrum™ Technology Platform, you can build a comprehensive enterprise data management process, or you can use it as a more targeted solution. The following diagram illustrates a complete solution that takes data from its source, through data enrichment and data quality processes, feeding a master data management hub which makes a single view of the data available to multiple business applications.
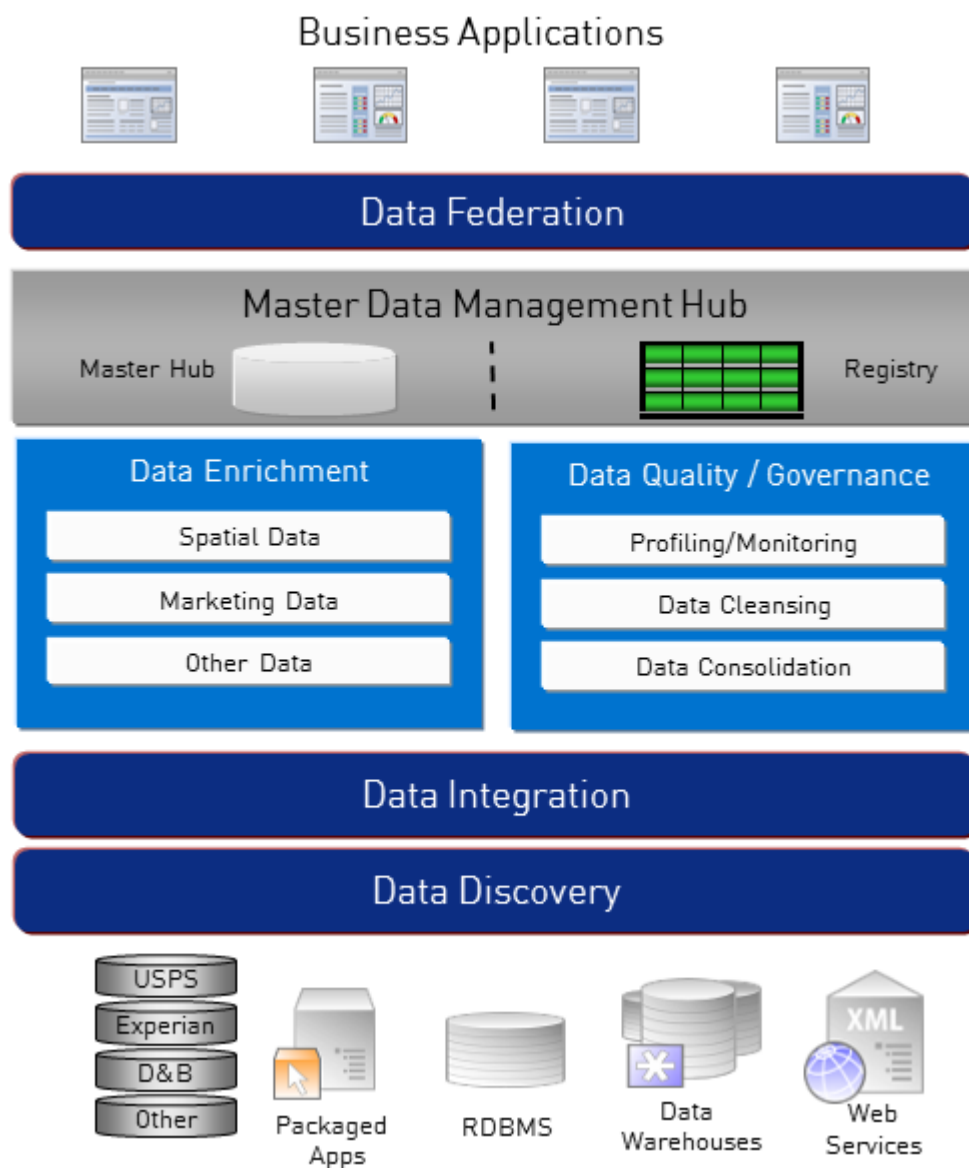
Data Discovery

Data discovery is the process of scanning your data resources to get a complete inventory of your data landscape. Spectrum™ Technology Platform can scan structured data, unstructured data, and semi-structured data using a wide array of data profiling techniques. The results of the scan are used to automatically generate a library of documentation describing your company's data assets and to create a metadata repository. This documentation and accompanying metadata repository provide the insight you need before beginning data integration, data quality, data governance, or master data management projects.

For more information on the Spectrum™ Technology Platform Data Discovery Module, contact your account executive.

Data Integration

Once you have an inventory of your data landscape, you need to consider how you will access the data you need to manage. Spectrum™ Technology Platform can connect to data in multiple sources either directly or through integration with your existing data access technologies. It supports batch and real time data integration capabilities for a variety of business needs including data warehousing, data quality, systems integration, and migration. Spectrum™ Technology Platform can access data in RDBMS databases, data warehouses, XML files, flat files, and more. Spectrum™ Technology Platform supports SQL queries with complex joins and aggregations and provides a visual query development tool. In addition, Spectrum™ Technology Platform can access data over REST and SOAP web services.

Spectrum™ Technology Platform can trigger batch processing based on the appearance of one or more source files in a specified folder. This "hot folder" trigger is useful for monitoring FTP uploads and processing them as they occur.

Some of these data integration capabilities require a license for the Enterprise Data Integration Module. For more information, contact your account executive.

Finally, Spectrum™ Technology Platform can integrate with packaged applications such as SAP and Siebel.

Data Quality/Governance

Data quality and data governance processes check your data for duplicate records, inconsistent information, and inaccurate information.

Duplicate matching identifies potential duplicate records or relationships between records, whether the data is name and address in nature or any other type of customer information. Spectrum™ Technology Platform allows you to specify a consistent set of business match rules using boolean matching methods, scoring methods, thresholds, algorithms and weights to determine if a group of records contains duplicates. Spectrum™ Technology Platform supports extensive customization so you can tailor the rules to the unique needs of your business.

Once duplicate records have been identified, you may wish to consolidate records. Spectrum™ Technology Platform allows you to specify how to link or merge duplicate records so you can create the most accurate and complete record from any collection of customer information. For example,

a single best-of-breed record can be built from all of the records in a household. The Advanced Matching Module is used to identify duplicates and eliminate them.

Data quality processes also standardize your data. Standardization is a critical process because standardized data elements are necessary to achieve the highest possible results for matching and identifying relationships between records. While several modules perform standardization of one type or another, the Spectrum™ Technology Platform Data Normalization module provides the most comprehensive set of standardization features. In addition, the Universal Name module provides specific data quality features for handling personal name and business name data.

Standardized data is not necessarily accurate data. Spectrum™ Technology Platform can compare your data to known, up-to-date reference data for correctness. The sources used for this process may include regulatory bodies such as the U.S. Postal Service, third-party data providers such as Experian or D&B, or your company's internal reference sources, such as accounting data. Spectrum™ Technology Platform is particularly strong in address data validation. It can validate or standardize addresses in 250 countries and territories around the world. There are two modules that perform address validation: the Address Now Module and the Universal Addressing Module.

To determine which one is right for you, discuss your needs with your account executive.

While Spectrum™ Technology Platform can automatically handle a wide range of data quality issues, there are some situations where a manual review by a data steward is appropriate. To support this, the Business Steward Module provides a way to specify the rules that will trigger a manual review, and it provides a web-based tool for reviewing exception records. It includes integrated access to third-party tools such as Bing maps and Experian data to aid data stewards in the review and resolution process.

Data Enrichment

Data enrichment processes augment your data with additional information. Enrichment can be based on spatial data, marketing data, or data from other sources that you wish to use to add additional detail to your data. For example, if you have a database of customer addresses, you could geocode the address to determine the latitude/longitude coordinates of the address and store those coordinates as part of the record. Your customer data could then be used to perform a variety of spatial calculations, such as finding the bank branch nearest the customer. Spectrum™ Technology Platform allows you to enrich your data with a variety of information, including geocoding (with the Enterprise Geocoding Module), tax jurisdiction assignment (with the Enterprise Tax Module), geospatial calculations (with the Location Intelligence Module), and driving and walking directions between points (with the Enterprise Routing Module).

Master Data Management Hub

The Master Data Management (MDM) hub allows for rapid modeling of entities and their complex relationships across roles, processes and interactions. It provides built-in social network analysis capabilities to help you understand influencers, predict churn, detect non-obvious relationships and fraudulent patterns, and provide recommendations.
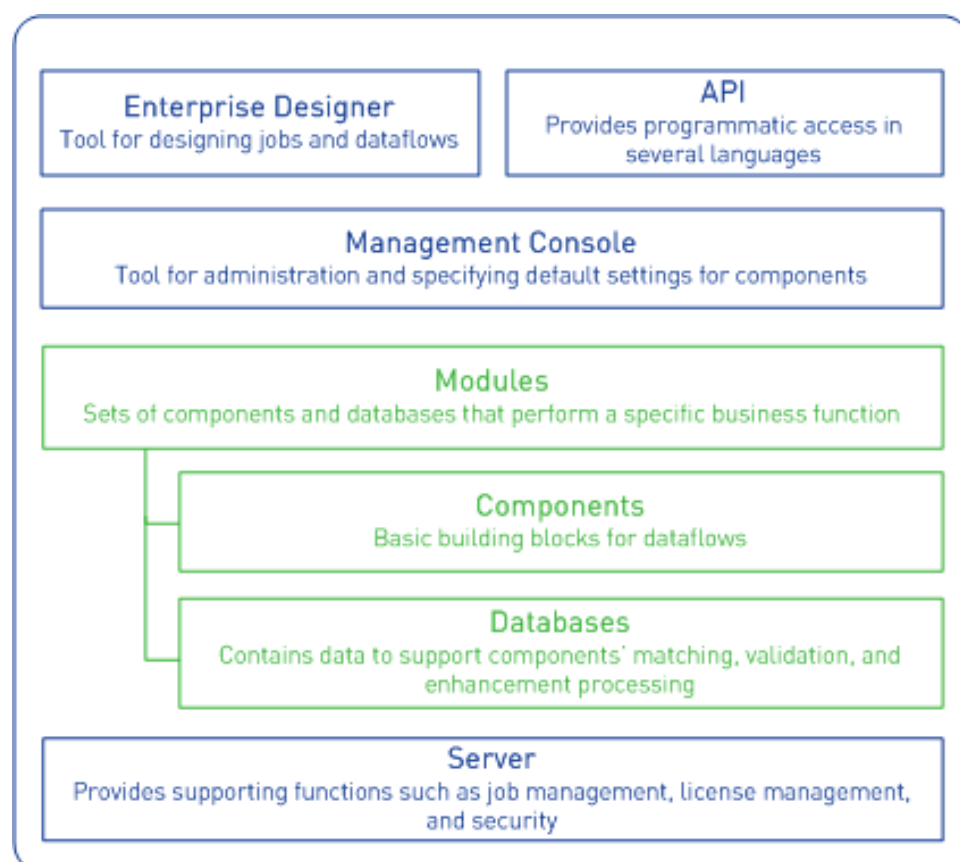
Spectrum™ Technology Platform supports two approaches to the MDM hub. In the master hub approach, the data is maintained in a single MDM database and applications access the data from

the MDM database. In the registry approach, the data is maintained in each business application and the MDM hub registry contains keys which are used to find related records. For example, a customer's record may exist in an order entry database and a customer support database. The MDM registry would contain a single key which could be used to access the customer data in both places.

The Data Hub Module provides MDM capabilities.

Spectrum™ Technology Platform Architecture

Spectrum™ Technology Platform from Pitney Bowes consists of a server that runs a number of modules. These modules provide different functions, such as address validation, geocoding, and advanced parsing, among others. The following diagram illustrates the Spectrum™ Technology Platform architecture.

Server

The foundation of the Spectrum™ Technology Platform is the server. The server handles data processing, synchronizes repository data, and manages communication between the client and the transformation modules via TCP/IP. It provides job management and security features.

Modules

Modules are sets of features that perform a specific function. For example, if you want to standardize your customers' addresses to conform to USPS standards, you would license the Universal Addressing module. If you want to determine the tax jurisdictions that apply to each of your customers, you would license the Enterprise Tax module. You can license just one module or multiple modules, depending on your specific needs. Most modules consist of "components" and databases.

Components

A component is a basic building block in a customer data quality process. Each component performs a specific function. For example, the Enterprise Geocoding module's Geocode US Address component takes an address and returns the latitude and longitude coordinates for that address; the Universal Addressing module's Get City State Province takes a postal code and returns the city and state/province where that postal code is located.

Some components must first be combined with other components into a job, service, or subflow before they can be executed. Use Enterprise Designer to create jobs, services, subflows, and process flows. For more information, see **Enterprise Designer** on page 7.

The components that you have available on your system depend on which Spectrum™ Technology Platform modules you have licensed from Pitney Bowes.

Databases

Modules often include databases that contain the data needed by the components in the module. For example, the Universal Addressing module needs to have access to USPS data in order to verify and standardize addresses. So, the Universal Addressing module comes with the U.S. Postal Database, which you must load into a location that is accessible by your Spectrum™ Technology Platform system.

Modules have both required and optional databases. Optional databases provide data needed for certain features that can greatly enhance your Spectrum™ Technology Platform process.

Management Console

The Management Console is a browser-based tool for administering Spectrum™ Technology Platform. You can use the Management Console to:

• Define the connections between Spectrum™ Technology Platform and your data
• Specify the default settings for services and flows
• Manage user accounts, including permissions and passwords
• View logs
• View licenses including license expiration information

Enterprise Designer

Enterprise Designer is a Windows-based tool for creating Spectrum™ Technology Platform jobs, services, subflows, and process flows. It utilizes an easy drag-and-drop interface to allow you to graphically create complex dataflows.
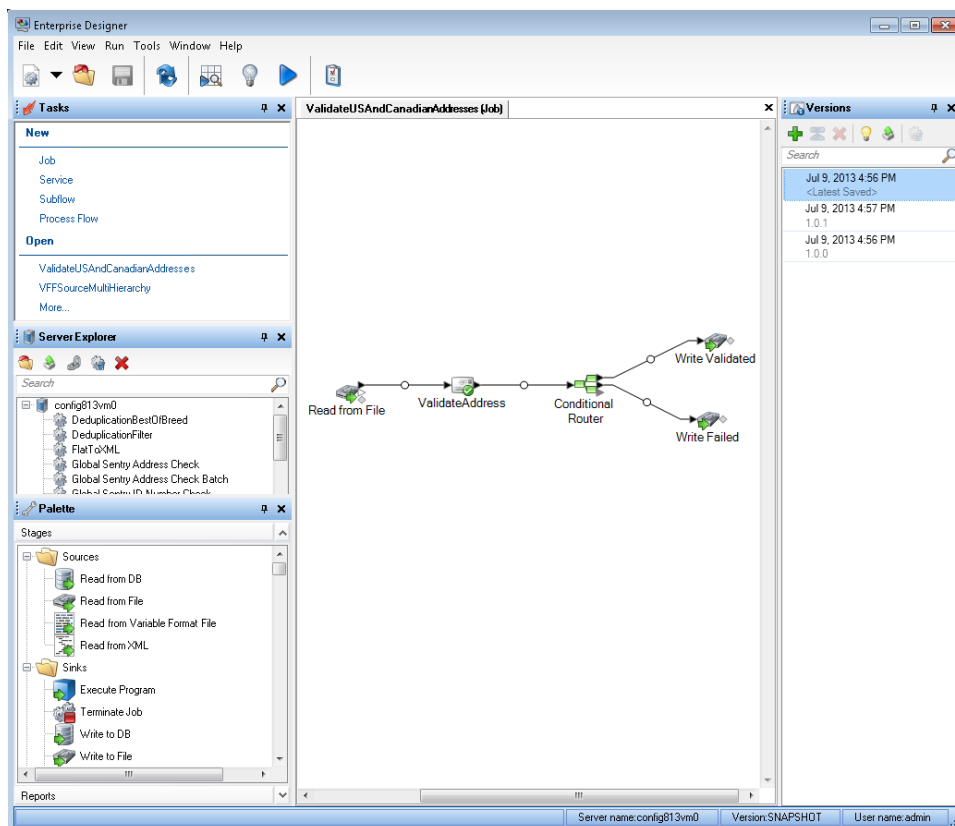
API

The Spectrum™ Technology Platform API, which is provided in the Spectrum™ Technology Platform Client API, is designed to provide simple integration, streamline record processing, and support backward compatibility of future versions. The Spectrum™ Technology Platform API can be accessed through:

- C
- C++
- COM
- Java
- .NET
- Web services

Solutions and Modules

Spectrum™ Technology Platform solutions are designed to serve many business needs. Solutions simplify the purchasing process by providing a preselected collection of modules with the capabilities necessary to improve data quality as well as meet other business needs in the most common business environments. Modules provide specific processing capabilities such as geocoding, address validation, matching, standardization, and more. While solutions provide bundled modules for common business challenges, you can also mix and match individual modules to fit your individual requirements. You can license just one module or all modules depending on your specific needs. They are listed below for convenience by solution or business need.

Data Quality/Governance

- **Address Now Module**
- **Advanced Matching Module**
- **Business Steward Module**
- **Data Normalization Module**
- **Global Sentry Module**
- **SAP Module**
- **Siebel Module**
- **Universal Addressing Module**
- **Universal Name Module**

Data Enrichment and Spatial Analysis

- **Enterprise Geocoding Module**
- **Enterprise GeoComplete Module**
- **Enterprise Routing Module**
- **Enterprise Tax Module**
- **GeoConfidence Module**
- **GeoEnrichment Module**
- **Global Geocoding Module**
- **Location Intelligence Module**

Data Integration

- **Enterprise Data Integration Module**
- **Data Federation Module**

Master Data Management

- **Data Hub Module**

Data Quality/Data Governance

Address Now Module

The Address Now Module is an address standardization and validation tool that provides comprehensive coverage for addresses outside the U.S. and Canada. Address Now is one of two address standardization and validation modules available for Spectrum™ Technology Platform. The other module is the Universal Addressing Module. The Address Now Module provides the following benefits over the Universal Addressing Module for addresses outside the U.S. and Canada:

- **Better data**—The database used by the Address Now Module is more up to date and complete in many countries than the database used by the Universal Addressing Module. This is because the Universal Addressing Module relies on data from Universal Postal Union (UPU), a body of the United Nations, for its international data and while the data coverage is extensive, the updates and the level of address details are not proactively managed by the UPU. Address Now, on the other hand, relies on data directly from the postal authorities (in most countries), plus other third-party data providers. This means that the data is more current with postal changes and is more detailed.
- **Drill-down feature**—The Address Now Module also offers drill-down capabilities to address data from any country, allowing users to rapidly enter address information without having to worry about the structure or making data entry mistakes.
- **Double-byte support**—The Address Now Module is Unicode enabled, recognizing Kanji and other double-byte characters.

Address Now Components

Address Now consists of the following components. These components can work with U.S., Canadian, and international addresses.

- **BuildGlobalAddresses**—Allows you to interactively build an address by searching for individual address elements.
- **GetGlobalCandidateAddresses**—Returns a list of addresses that are considered matches for a given address.
- **ValidateGlobalAddress**—Standardizes addresses using international postal data. ValidateGlobalAddress can also validate addresses in the U.S. and Canada but its strength is validation of addresses in other countries. If you have a significant amount of non-U.S. and non-Canadian address data, you should consider using ValidateGlobalAddress.

In cases where ValidateGlobalAddress returns multiple address matches for a given input address, you can use GetGlobalCandidateAddresses to return the address stack. GetGlobalCandidateAddresses is designed to return additional information from the postal databases to help you determine which of the returned addresses is the best match.

Address Now Database

The Address Now database contains postal data from all supported countries. You can install the entire database or the data for specific countries only. The database is installed on the server. This database is available by subscription from Pitney Bowes and is updated monthly.

Advanced Matching Module

The Advanced Matching Module matches records between and/or within any number of input files. You can also use the Advanced Matching Module to match on a variety of fields including name, address, name and address, or non-name/address fields, such as social security number or date of birth.

Advanced Matching Module Components

The Advanced Matching Module consists of the following stages:

- **Best-of-Breed**—This stage selects a best-of-breed record from the duplicates cluster by selecting a primary/template record and then using this record to build a best composite record. This record is then returned as the survivor record.
- **Candidate Finder**—This stage obtains the candidate records that will form the set of potential matches that the Transactional Match stage will evaluate. Candidate Finder is used in conjunction with Transactional Match.
- **Duplicate Synchronization**—This stage allows you to specify which fields from a collection of records will be copied (posted) to the corresponding fields of all records in the collection.
- **Filter**—This stage allows you to specify the criteria that records must satisfy in order to be retained or removed from a collection of duplicate records, either for further processing downstream or to be written to your output file.
- **Interflow Match**—This stage identifies matches between similar data records across multiple input streams.
- **Intraflow Match**—This stage identifies matches between similar data records within a single input stream.
- **Match Analysis**—Use Match Analysis to analyze and compare results from matching stages to give you a better understanding of the outcome of the matching process and assist you in improving your matching results.
- **Match Key Generator**—This stage creates a non-unique key shared by all like records. Records that share the same key can then be grouped together for comparison.
- **Private Match**—This stage enables two entities to compare datasets and identify common records, or customers, without compromising sensitive information. By using an encrypt function security is retained while a match function is performed, and then a decrypt function shows the output of the matched data.
- **Transactional Match**—This stage matches suspect transactions against a database using the Candidate Finder Stage to query and return potential candidate records.

- **Write to Search Index**—This stage enables you to create a full-text index based on the data coming in to the stage. Having this data in a dedicated search index results in quicker response time when you conduct searches against the index from Candidate Finder.

Big Data Quality SDK

The Big Data Quality SDK helps you create, configure and run MapReduce jobs for Data Quality operations on a Hadoop platform. Using the SDK, you can create and execute the jobs directly on a Hadoop platform, thus eliminating network delays and running distributed Data Quality processes in cluster, resulting in a drastic improvement in the performance.

The following modules are supported in the Big Data Quality SDK:

1. Advanced Matching Module
2. Data Normalization Module
3. Universal Name Module

SDK Usage

This SDK can currently be used through:

1. Java APIs
2. Hive User-Defined Functions

Introduction to the Business Steward Module

The Business Steward Module is a set of features that allow you to identify and resolve exception records. Exception records are records that Spectrum™ Technology Platform could not confidently process and that require manual review by a data steward. Some examples of exceptions are:

- Address verification failures
- Geocoding failures
- Low-confidence matches
- Merge/consolidation decisions

The Business Steward Module provides a browser-based tool for manually reviewing exception records. Once exception records are manually corrected and approved, they can be reincorporated into your Spectrum™ Technology Platform data quality process.

Many of the actions that take place within the Business Steward Module are reflected in the audit log, a Management Console tool that records user activity. The following actions are included in the log:

- Adding exceptions in the Write Exceptions stage
- Deleting exceptions in the Read Exceptions stage and the Business Steward Portal Manage Exceptions page
- Assigning exceptions in the Business Steward Portal Exception Manager
- Retrieving exceptions in the Business Steward Portal Exception Editor
- Updating exceptions in the Business Steward Portal Exception Editor

• Revalidating exceptions either when clicking Save and Revalidate in the Business Steward Portal Exception Editor

Read more about the audit log in the Spectrum™ Technology Platform Administration Guide for the WebUI.

Business Steward Module Components

The Business Steward Module consists of:

• **Exception Monitor**—A stage that evaluates records against a set of conditions to determine if the record requires manual review by a data steward. When records meet those conditions, this stage can send an email notifying recipients of the exceptions.
• **Write Exceptions**—A stage that writes the exception records to the exception repository. Once exception records are in the exception repository they are available for review by a data steward.
• **Exception Dashboard and Editor**—A browser-based tool that displays a dashboard of summary statistics and charts to help you understand the kinds of exceptions that occur in your data. It is also an editor that allows you to modify exception records and approve them for reprocessing.
• **Manage Exceptions**—A browser-based tool that enables you to review and manage exception record activity for all users.
• **Data Quality Performance**—A browser-based tool that provides information on trends within your exception records and enables you to identify key performance indicators and send notifications when certain conditions have been met.
• **Read Exceptions**—A stage that reads records from the exception repository into a dataflow. This stage allows you to reprocess exception records that have been corrected by a data steward.

Data Normalization Module

The Data Normalization Module examines terms in a record and determines if the term is in the preferred form.

• **Advanced Transformer**—This stage scans and splits strings of data into multiple fields, placing the extracted and non extracted data into an existing filed or a new field.
• **Open Parser**—This stage parses your input data from many cultures of the world using a simple, but powerful parsing grammar. Using this grammar, you can define a sequence of expressions that represent domain patterns for parsing your input data. Open Parser also collects statistical data and scores the parsing matches to help you determine the effectiveness of your parsing grammars.
• **Table Lookup**—This stage evaluates a term and compares it to a previously validated form of that term. If the term is not in the proper form, then the standard version replaces the term. Table Lookup includes changing full words to abbreviations, changing abbreviations to full words, changing nick names to full names or misspellings to corrected spellings.
• **Transliterator**—Transliterator converts a string between Latin and other scripts.

Data Normalization Module Components

The Data Normalization Module consists of:

- **Advanced Transformer**—This stage scans and splits strings of data into multiple fields, placing the extracted and non extracted data into an existing filed or a new field.
- **Open Parser**—This stage parses your input data from many cultures of the world using a simple, but powerful parsing grammar. Using this grammar, you can define a sequence of expressions that represent domain patterns for parsing your input data. Open Parser also collects statistical data and scores the parsing matches to help you determine the effectiveness of your parsing grammars.
- **Table Lookup**—This stage evaluates a term and compares it to a previously validated form of that term. If the term is not in the proper form, then the standard version replaces the term. Table Lookup includes changing full words to abbreviations, changing abbreviations to full words, changing nick names to full names or misspellings to corrected spellings.
- **Transliterator**—Transliterator converts a string between Latin and other scripts.

Global Sentry Module

The Global Sentry Module matches your records against government watch lists so that you can identify countries, individuals, and organizations in your data that have financial sanctions imposed on them by governments around the globe. The Global Sentry Module addresses the specific requirements of the USA PATRIOT Act and similar laws.

The Global Sentry Module identifies matches between your records and entities on the following government watch lists:

- Australian Government Department of Foreign Affairs and Trade
- Bank of England
- European Union - Common and Foreign Security Policy
- Office of Superintendent of Financial Institutions of Canada
- U.S. Bureau of Industry and Security - Denied Persons List
- U.S. Bureau of Industry and Security - Unverified List
- U.S. Treasury – Office of Foreign Assets Control
- U.S Department of State Directorate of Defense Trade Controls - List of Statutorily Debarred Parties

Global Sentry provides optimized rules for matching and scoring, along with a knowledge base of names from 120 cultures and associated gender and name variant data to identify matches between your data and the watch lists. Matches are performed against sanctioned country names, names, addresses, ID numbers, and other information such as date of birth to provide an overall risk score. This information allows your organization to make a decision on whether or not to block a particular transaction.

The Global Sentry Module uses stages from the Universal Name Module, Data Normalization Module, and Advanced Matching Module.

Global Sentry Components

Global Sentry deploys five dataflow templates that you can modify in Enterprise Designer. Each dataflow consists of various components that were installed from the Spectrum™ Technology Platform, Universal Name, Data Normalization and Advanced Matching modules.

The names of the dataflows are:

- Global Sentry
- Global Sentry Name Check
- Global Sentry Address Check
- Global Sentry ID Number Check
- Global Sentry Other Data Check

Microsoft Dynamics CRM Module

The Spectrum™ Technology Platform connector for Microsoft Dynamics CRM provides address validation features that help ensure that accurate addresses are entered into your Microsoft Dynamics CRM records. Optionally, the Spectrum™ Technology Platform connector can also determine the latitude and longitude coordinates of an address, which is a process called *geocoding*.

This guide contains instructions for administrators on how to integrate Spectrum™ Technology Platform with Microsoft Dynamics CRM. It also contains instructions for users on how to validate and geocode addresses.

SAP Module

The Spectrum™ Technology Platform SAP Module optimizes the quality of your customer, vendor and partner records so you can improve operations across the enterprise.

Because SAP is such a far-reaching application, introducing accurate address data provides equally far-reaching benefits. With data quality for over 220 countries, the SAP Module identifies and manages duplicate records, standardizes and validates addresses, auto-populates missing fields, and gives you a single, integrated view of your customers.

The following diagram illustrates how Spectrum™ Technology Platform integrates with SAP.
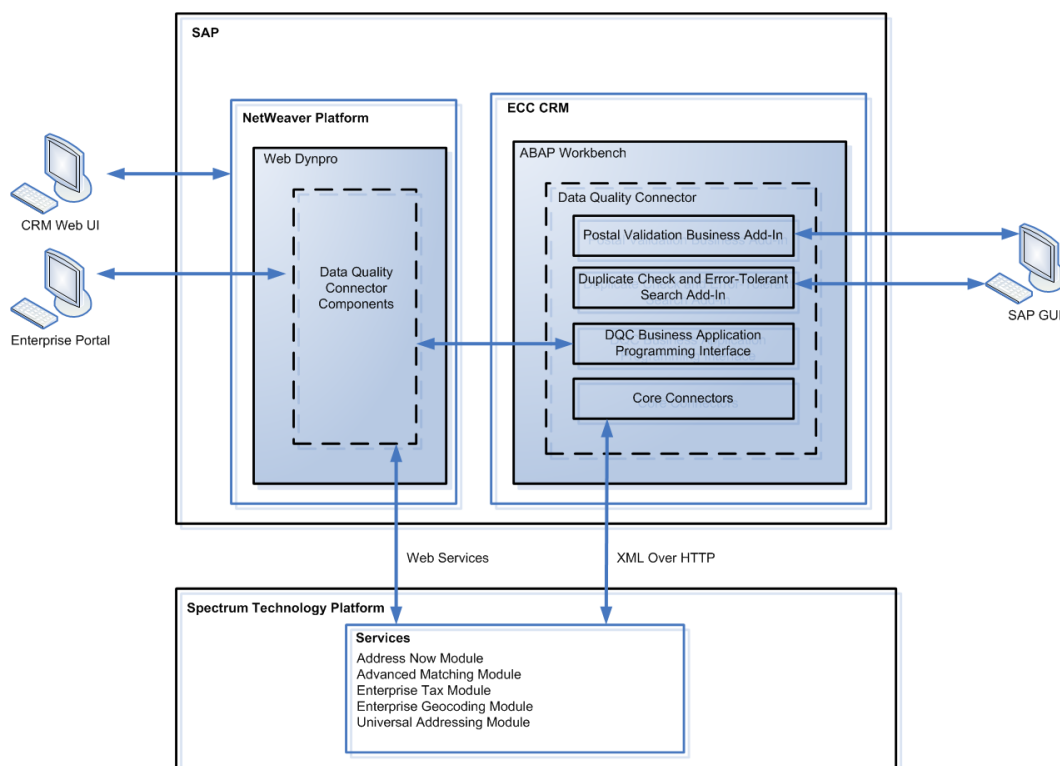
Siebel Module

The Spectrum™ Technology Platform Siebel Module identifies and manages duplicate records, standardizes and validates addresses, standardizes and validates names, and auto-populates missing fields. Records in your Siebel system are checked against known, up-to-date reference data from sources such as regulatory bodies (for example, the United States Postal Service), third-party data providers (for example, Dun & Bradstreet) or your company's internal reference sources (for example, accounting data).

The Siebel Module is available in two implementations: Siebel Data Quality (SDQ) and Non-SDQ. The SDQ implementation uses the Siebel Data Quality Universal Connector to connect Siebel to Spectrum™ Technology Platform. The Non-SDQ implementation utilizes a custom connector to connect Siebel to Spectrum™ Technology Platform.

SDQ Architecture
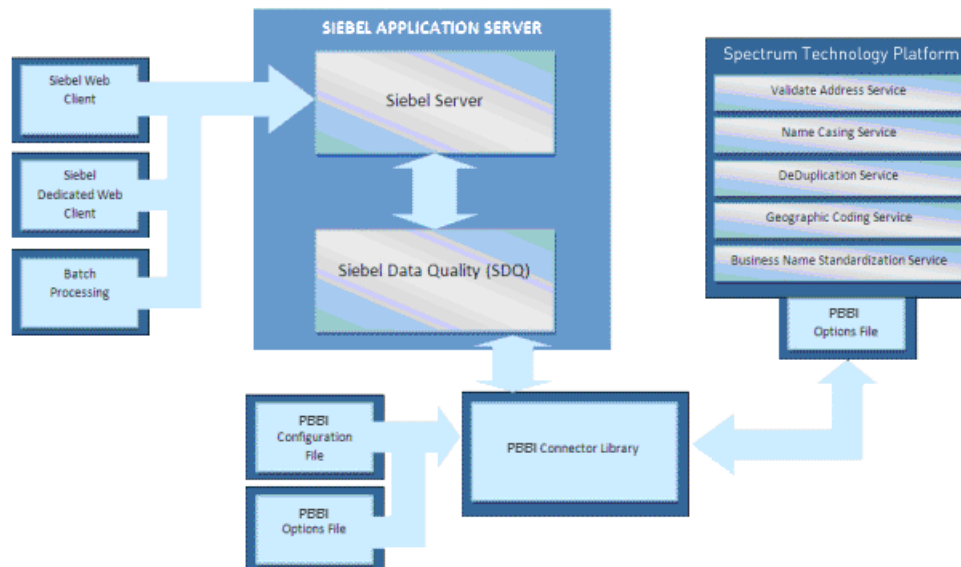
Non-SDQ Architecture
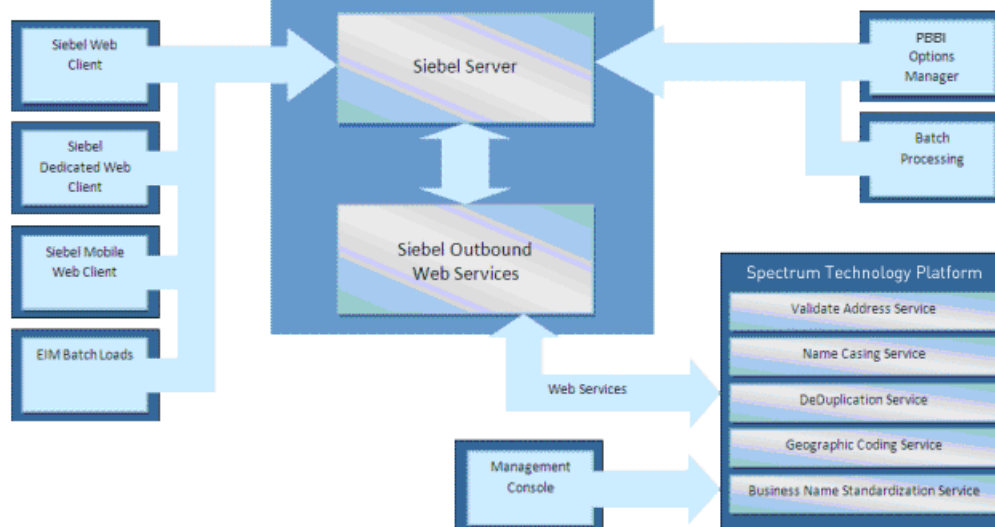
The following table illustrates the differences between SDQ and Non-SDQ implementation.

Table 1: Supported Web Clients

	SDQ	Non-SDQ
Web Client	Yes	Yes

	SDQ	Non-SDQ
Dedicated Web Client	Yes	Yes
Mobile Web Client	No	Yes
Open UI	Yes	Yes

Data cleansing may be enabled for business addresses, personal addresses, and prospect addresses. De-duplication may be enabled for accounts, contacts, business addresses, personal addresses[1], and prospects. Settings are global for interactive mode and can be unique for each batch job. Settings for the Siebel Module are managed using options and configuration files for SDQ. For Non-SDQ, Siebel's Options Manager and the Spectrum™ Technology Platform Management Console are used to set options and configuration.

Universal Addressing Module

The Universal Addressing Module is an address quality module that can standardize and validate addresses, improving the deliverability of mail. The Universal Addressing Module can ensure that your address data adheres to quality standards established by the postal authority. An address that adheres to these standards is more likely to be delivered in a timely manner. In addition, mailers who follow these standards can qualify for significant postage discounts. For information on discounts for U.S. mail, refer to the USPS *Domestic Mail Manual* (*DMM*) available at **www.usps.com**. For information on discounts for Canadian mail, refer to the Canada Post website at **www.canadapost.ca**. For information on discounts for Australian mail, refer to the Australia Post website at **www.auspost.com.au**.

The Universal Addressing Module can be used in batch mode, realtime mode, or as a hosted service, depending on which option you have licensed. The batch version of the Universal Addressing Module is CASS Certified™ by the USPS®. It is also AMAS certified by Australia Post.

The Universal Addressing Module is one of two address quality modules available for Spectrum™ Technology Platform. The other address quality module, the Address Now Module, provides enhanced support for addresses outside the U.S. and Canada, including validation for more countries and double-byte support. If you have a large amount of international address data, you may want to consider using the Address Now Module for address standardization and validation.

Universal Addressing Components

The Universal Addressing Module consists of the following components. These components can work with U.S., Canadian, Australian, and international addresses as long as you are licensed for the appropriate database (if you are running Universal Addressing in your own environment) or hosted service (if you are utilizing Universal Addressing through the Pitney Bowes hosted services).

---

[1] Personal Address is supported only for Siebel Industry Applications

- **Auto Complete Loqate**—Offers real-time entry of address data and returns instant results based on each character entered into the form, ensuring only accurate data is entered into the database.
- **Get Candidate Addresses**—Returns a list of possible matches for a given address.
- **Get Candidate Addresses Loqate**—Returns a list of possible matches for a given address using a Loqate engine and database.
- **Get City State Province**—Returns the city and state/province for a given postal code.
- **Get City State Province Loqate**—Returns the city and state/province for a given postal code using a Loqate engine and database.
- **Get Postal Codes**—Returns the postal codes for a given city.
- **Get Postal Codes Loqate**—Returns the postal codes for a given city using a Loqate engine and database.
- **Validate Address**—Standardizes and validates addresses using U.S., Canadian, and international postal data.
- **Validate Address AUS**—Standardizes and validates addresses using Australian postal data.
- **Validate Address Global**—Validate Address Global provides enhanced address standardization and validation for addresses outside the U.S. and Canada. Validate Address Global can also validate addresses in the U.S. and Canada but its strength is validation of addresses in other countries. If you process a significant number of addresses outside the U.S. and Canada, you should consider using Validate Address Global.
- **Validate Address Loqate**—Validate Address Loqate standardizes and validates addresses using postal authority address data. Validate Address Loqate can correct information and format the address using the format preferred by the applicable postal authority. It also adds missing postal information, such as postal codes, city names, state/province names, and so on.

Universal Name Module

To perform the most accurate standardization you may need to break up strings of data into multiple fields. Spectrum™ Technology Platform provides advanced parsing features that enable you to parse personal names, company names, and many other terms and abbreviations. In addition, you can create your own list of custom terms to use as the basis of scan/extract operations.

Universal Name Module Components

The Universal Name Module consists of:

- **Name Parser (Deprecated)**—This component breaks down personal names, business names, and other terms in the name data field into their component parts. These parsed name elements are then subsequently available to other automated operations such as name matching, name standardization or multi-record name consolidation.

  Attention: The Name Parser stage is deprecated and may not be supported in future releases. Use Open Name Parser for parsing names.

- **Name Variant Finder**—This component works in either a first name or last name mode in order to query the names database and return name variations. Name Variant Finder can limit the number of results by gender and/or culture. Spectrum™ Technology Platform includes a base

names file. Add-on names files of other cultures are available and can be deployed by copying the appropriate JAR file into the `modules/tables/ext` folder. Arabic names are available as an add-on. The data in this file is by Nomino, Inc.

- **Open Name Parser**—The open name parser breaks down personal and business names and other terms in the name data field into their component parts. These parsed name elements are then subsequently available to other automated operations such as name matching, name standardization, or multi-record name consolidation.

Data Enrichment and Spatial Analysis

Enterprise Geocoding Module

The Enterprise Geocoding Module performs address standardization, address geocoding, and postal code centroid geocoding. You can enter an address and get outputs such as geographic coordinates, which can be used for detailed spatial analysis and demographics assignment. You can also enter a geocode, a point represented by a latitude and longitude coordinate, and receive address information about the provided geocode.

Enterprise Geocoding Stages

Enterprise Geocoding Module consists of the following stages. The specific stages you have depend on your license.

- **Geocode Address AUS**—Takes an address in Australia and returns latitude/longitude coordinates and other information.

  Note: Geocode Address AUS has been deprecated. GNAF PID Location Search is the only stage that is used from Geocode Address AUS. For all other Australia geocoding functions, use the Geocode Address Global component.

- **Geocode Address GBR**—Takes an address in Great Britain and returns latitude/longitude coordinates and other information.

  Note: Geocode Address GBR supports the GBR AddressBase Plus data source. Use Geocode Address Global for the GBR Streets (TomTom) data source.

- **Geocode Address Global**—Takes an address in any supported country and returns latitude/longitude coordinates and other information. Geocode Address Global geocodes addresses only from countries you have licensed. It does not support Australia and Great Britain.
- **Geocode Address World**—Takes an address located in any of the supported countries and returns the city centroid or, for some countries, postal centroid. Geocode Address World cannot geocode to the street address level.
- **Geocode Address Africa** — Provides street-level geocoding for many African countries. It can also determine city or locality centroids, as well as postal code centroids for selected countries.

- **Geocode Address Middle East** — Provides street-level geocoding for many Middle East countries. It can also determine city or locality centroids. Middle East supports both English and Arabic character sets.
- **Geocode Address Latin America** — Provides street-level geocoding for many Latin American countries. It can also determine city or locality centroids. There is postal code coverage for selected countries.
- **Geocode US Address**—Takes an input address and returns latitude/longitude coordinates and other address information.
- **GNAF PID Location Search**—Identifies the address and latitude/longitude coordinates for a Geocoded National Address File Persistent Identifier (G-NAF PID).
- **Reverse APN Lookup**—Takes an Assessor's Parcel Number (APN), Federal Information Processing Standards (FIPS) county code, and FIPS state code and returns the address of the parcel.
- **Reverse Geocode US Location**—Takes as input a geocode (latitude and longitude coordinate) and returns the address of the location.

Enterprise GeoComplete Module

The Enterprise GeoComplete Module automatically suggests addresses or Points of Interest (POIs) based on your input. You can search on address, POI name or POI Categories. This "as you type" recognition allows for searching based on your location. Enterprise GeoComplete supports batch and real time data integration capabilities for a variety of business needs, including type ahead of addresses and points of interest.

Note: POI data is provided in separate data bundles and must be separately purchased.

Enterprise GeoComplete is delivered as a Spectrum™ Technology Platform module and can be built into Spectrum stages just like any other component.

GeoComplete Stage

The Enterprise GeoComplete Module consists of the following stage.

- **GeoComplete** —- Automatically suggests addresses as you type and immediately returns candidates based on your input. The GeoComplete Module can also return Points of Interest (POIs), but POI data must be separately purchased.

Global Geocoding Module

The Global Geocoding Module provides geocoding capabilities for over 120 countries in a single, integrated module. The Global Geocoding Module performs address standardization, address geocoding, and postal and geographic centroid geocoding. You can enter an address and get outputs such as geographic coordinates, which can be used for detailed spatial analysis and demographics assignment. You can also enter a geocode, a point represented by a latitude and longitude coordinate, and receive address information about the provided geocode. The Global Geocoding Module removes the country code as an input requirement for reverse geocoding.

Global Geocoding Stages

The Global Geocoding Module consists of the following stages.

- **Global Geocode** - This stage performs forward geocoding, using input addresses and returning location data and other information.
- **Global Reverse Geocode** - This stage performs reverse geocoding, using input coordinates and returning address information that is the best match for that point.

Enterprise Routing Module

The Enterprise Routing Module provides the ability to obtain driving or walking directions, calculate drive time and drive distance, and identify locations within a certain time or distance from a starting point.

Enterprise Routing Stages

- **Get Travel Boundary**—Allows you to obtain polygons corresponding to an isochrone or isodistance calculation. An isochrone is a polygon or set of points representing an area that can be traversed in a network from a starting point in a given amount of time. An isodistance is a polygon or set of points representing the area that is a certain distance from the starting point.
- **Get Travel Cost Matrix**—Calculates the travel time and distances between an array of start and end locations.
- **Get Travel Directions**—Allows you to route from one point to another point or from one point to multiple other points.

Enterprise Tax Module

The Enterprise Tax Module determines the tax jurisdiction for an address. The Enterprise Tax Module takes an address, standardizes it, and then matches the address to an exact physical location, returning latitude/longitude coordinates with the correct place code for the address. This solution greatly reduces the inaccuracies associated with 9-digit and 5-digit ZIP Code-based matching. The Enterprise Tax Module can also calculate latitude/longitude coordinates for individual address locations, including the use of interpolation and offset.

For companies that deliver goods or services to locations that don't have a recognizable address, the Enterprise Tax Module can accept latitude/longitude coordinates as input to return the tax jurisdictions and corresponding GeoTAX Key values. Examples of this type of location include bill boards, street lights, communications towers or new housing developments.

The Enterprise Tax Module uses a database of tax jurisdictions provided by TomTom. This data, which is collected through an ongoing research program and updated regularly, provides current jurisdictional boundary information down to the municipal and special tax district levels.

In addition to determining the tax jurisdiction for an address or location, with the optional Pitney Bowes Sales and Use Tax Rate file, the Enterprise Tax Module can return sales and use tax rates for each of the assigned tax jurisdictions as well as the total tax rate for the assigned locations.

There is also the option to use other software, such as Vertex or Taxware, to cross-reference and supply tax rate data.

Note: The Enterprise Tax Module processes only U.S. addresses.

Enterprise Tax Stages

The Enterprise Tax Module consists of the following stages:

• **Assign GeoTAX Info** - Takes an input address and returns census, latitude/longitude, and tax information about the address. Assign GeoTAX Info utilizes Pitney Bowes' GeoTAX technology.
• **Calculate Distance** - Takes two latitude/longitude coordinates as input and computes and returns the distance between the coordinates.
• **Reverse GeoTAX Info Lookup** - Takes an input latitude/longitude coordinate and returns census and tax information about the location. This service will not return an address for the location.

GeoConfidence Module

The GeoConfidence Module is used to determine the probability that an address or street intersection is within a given area. The module takes an address or intersection's location (determined by Geocode US Address), converts that location to a point, line, or polygon (depending on the precision of the match), then compares that shape with a database of known shapes to see if the two overlap, and the percentage overlap. For example, you could use the GeoConfidence Module to make decisions on a flood zone rating based on how much overlap there is between an address's location and the flood zone data. Anything greater than a 95% overlap with a 100-year flood zone may indicate that the address is in the flood zone. Conversely, anything less than 95% could cause your business process to send the address to exception processing that might include a manual review.

An address or intersection can be geocoded to a point, an address along a street segment (an array of street segment points), ZIP + 4 centroid, ZIP + 2 centroid, or ZIP Code centroid (polygons). You can use these shapes (points, lines, or polygons) to compare with other shapes to determine overlap, which can be used to determine a risk or probability.

Different geoconfidence polygons are generated depending on the GeoConfidence result returned by the Enterprise Geocoding Module. Refer to the Enterprise Geocoding Module documentation for more information about the GeoConfidence information returned by the Enterprise Geocoding Module.

The GeoConfidence Module supports U.S. locations only.

Note: GeoConfidence uses services provided by the Enterprise Geocoding and Location Intelligence modules.

GeoConfidence Stages

GeoConfidence deploys three dataflows that you can modify in Enterprise Designer. Each dataflow consists of various components that were installed with the Enterprise Geocoding and Location Intelligence modules.

For information about each component in the installed dataflows, see the relevant component chapter in the *Spectrum™ Technology Platform User's Guide*.

The names of the dataflows are:

- **GeoConfidenceSurface** This is the dataflow that creates the geoconfidence surface that can be used for further analysis. The input is the GeoConfidence information that is returned from the Enterprise Geocoding Module. Currently, only the Geocode US Address stage can return this information.
- **CreatePointsConvexHull** This is a subflow that is used by the GeoConfidenceSurface template. You should not need to make any changes to this subflow.
- **FloodRiskAnalysis** This is an example dataflow.

GeoEnrichment Module

The GeoEnrichment Module functions as a delivery mechanism for accessing a data source and returning data. Given a lookup key as input, the GeoEnrichment Module accesses a registered database and returns additional attribute data for the referenced record. The type of attribute data is determined by the installed dataset(s).

The GeoEnrichment Module is delivered as a Spectrum module and can be built into Spectrum stages. The Spectrum™ Technology Platform must already be installed.

Location Intelligence Module

The Location Intelligence Module provides spatial analysis that allows you to determine the relationship between a location and areas or points of interest. You can perform a variety of spatial calculations with the Location Intelligence Module, such as area, union, intersection, point in polygon, and more.

Location Intelligence Stages

- **Find Nearest** – Locates the points of interest (POI) that are nearest to a given location.
- **Query Spatial Data** – Allows you to query a spatial database using MI SQL functions that can, for example, search for a point within a given polygon.
- **Read Spatial Data** – Allows you to access spatial data in a variety of commonly-used spatial data formats using MI SQL functions.
- **Spatial Calculator** – Performs string and geometry conversions as well a variety of calculations on spatial data, such as determining the area of a polygon or distance between two points.
- **Spatial Union** – Returns a geometry object which represents the union of two input geometry objects.
- **Write Spatial Data** – Allows you to insert data into a named table or update data in a named table as output of a job. You can also delete data.
- **Closest Site** – Determines which sites are closest to a given location.
- **Point In Polygon** – Determines whether or not a location resides in a given area.

Location Intelligence Module Databases

Location Intelligence Module databases define regions or points of interest. You choose the databases that contain the information you are interested in using in your spatial analysis. Pitney Bowes has an extensive library of databases available for use with the Location Intelligence Module. For a listing of the available databases, see **www.pbinsight.com/products/data**. The Location Intelligence Module works with databases in Centrus GSB format and MapInfo TAB format.

Routing Databases

Routing databases define regions or points of interest. You choose the databases that contain the information you are interested in using in your routing dataflow. Each country has its own database, named in this format: Location Intelligence Module - <Country> - Database. Each database also has its own country code. For example, the name of the Austrian batch database is "Location Intelligence Module - Austria - Database," and the Austrian batch country code is "A1T." Databases for the following countries are available. Each database requires a separate license. For more information on licensing, contact your account executive.

Table 2: Routing Databases and Country Codes

Product	Country Code
Location Intelligence Module - US - Batch	ERB
Location Intelligence Module - US - Realtime	ERR
Location Intelligence Module - US - Web	ERW
Location Intelligence Module - US - Database	U1S
Location Intelligence Module - Canada - Database	C1A
Location Intelligence Module - Austria - Database	A1T
Location Intelligence Module - Belgium+Luxembourg - Database	B1E
Location Intelligence Module - Denmark - Database	D1K

Product	Country Code
Location Intelligence Module - Finland - Database	F1I
Location Intelligence Module - France - Database	F1R
Location Intelligence Module - Germany - Database	D1E
Location Intelligence Module - Ireland - Database	I1E
Location Intelligence Module - Italy - Database	I1T
Location Intelligence Module - Netherlands - Database	N1L
Location Intelligence Module - Norway - Database	N1O
Location Intelligence Module - Poland - Database	P1L
Location Intelligence Module - Portugal - Database	P1T
Location Intelligence Module - Spain - Database	E0S
Location Intelligence Module - Sweden - Database	S1E
Location Intelligence Module - Switzerland+Lichtenstein - Database	C1H
Location Intelligence Module - United Kingdom - Database	G1B
Location Intelligence Module - Australia - Database	A1U

Product	Country Code
Location Intelligence Module - Malaysia - Database	M1Y
Location Intelligence Module - New Zealand - Database	N1Z
Location Intelligence Module - Singapore - Database	S1G
Location Intelligence Module - Thailand - Database	T1H

Insurance

Table 3: Location Intelligence Module Databases for the Insurance Industry

Database Name & Description	Required or Optional	Supplier
Coal Mines The Coal Mines database identifies the approximate location of Problem Areas containing public health, safety, and public welfare problems created by past coal mining. This database provides information needed to implement Title IV Abandoned Mine Reclamation, of the Surface Mining Control and Reclamation Act (SMCRA) of 1977. One of the major uses of this data set is for the reporting of annual Abandoned Mine Land Program accomplishments to Congress. The data can be used to conduct geographic analysis on national and large regional scales. The Coal Mines database uses a subset of data contained in the Office of Surface Mining (OSM) Abandoned Mine Land Inventory.	Optional	Pitney Bowes; updated as needed
Coastal Boundaries This database identifies the coastal areas at high risk for coastal flooding for the Atlantic, Gulf, Pacific, Alaska, Hawaii, and Puerto Rico shorelines. It contains the feature name (e.g. Bodega Bay), as well as its major water name (e.g. Gulf of Mexico, Tampa Bay, and Atlantic Ocean). The Coastal Boundaries database uses data from the TANA Dynamap/2000 database.	Optional	Pitney Bowes; updated as needed

Database Name & Description	Required or Optional	Supplier
Earthquake Faultlines This database contains data on the generalized major earthquake faults in the contiguous United States. This data is intended primarily for regional and national analysis, rather than for more detailed analysis in specific areas. The data should be displayed and analyzed at scales appropriate for 1:7,500,000-scale data. The Earthquake Faultlines database uses data from the U.S. Geological Survey.	Optional	Pitney Bowes; updated as needed
FEMA Flood Plain Boundary This database identifies the areas at risk for flood damage in the United States. It contains certain features of the Federal Emergency Management Agency (FEMA) Flood Insurance Rate Map (FIRM) product. The data has been developed by scanning the existing FIRM hardcopy and assigning vectors in a schematic overlay of flood risks. The FEMA Flood Plain Boundary database uses the Q3 Flood boundary data from the Federal Emergency Management Agency.	Optional	Pitney Bowes; updated annually
Florida Sinkholes This database identifies the location of known sinkholes in Florida. The Florida Sinkholes database uses data from the State of Florida Department of Environmental Protection.	Optional	Pitney Bowes; updated as needed
GDT Windpool Boundary This database identifies the areas at high risk for wind and hail damage, primarily those at risk from hurricanes. The data covers Florida, Louisiana, Mississippi, North Carolina, South Carolina, Texas, Alabama, and Georgia.	Optional	Pitney Bowes; updated annually
Hail Event Data This database identifies the position of 86,675 historical hail events in the United States. The Hail Event database is point data and is based on the hail event point data from the National Weather Service.	Optional	Pitney Bowes; updated as needed
Hurricane Tracks This database identifies the tracks of North Atlantic hurricanes with landfall in the United States for all major storms from 1851 through 2005. Major storms are those that made landfall in the United States and that were	Optional	Pitney Bowes; updated as needed

Database Name & Description	Required or Optional	Supplier

classified on the Saffir-Simpson Hurricane Scale as category 3, 4, or 5 at the time of landfall. Land falling storms are defined as those storms whose center is reported to have either crossed or passed directly adjacent to the U.S. coastline, and which came ashore with tropical storm intensity or greater (sustained surface winds of 34 knots or 39 miles per hour or greater). Each storm track contains the 6-hourly (0000, 0600, 1200, 1800 UTC) center locations, name, date, category, wind speed, winds in MPH, pressure, location and BTID. Coverage includes the United States (eastern seaboard, Florida, Gulf States, Gulf of Mexico, Northern Mexico, Caribbean, and the Atlantic Ocean.

The Hurricane Tracks database is intended for geographic display and analysis at the national level, and for large regional areas. The data should be used only at scales appropriate for 1:2,000,000-scale data. No responsibility is assumed by the National Oceanic and Atmospheric Administration, the U.S. Geological Survey, or Pitney Bowes in the use of these data.

The Hurricane Tracks database uses data from the National Oceanic and Atmospheric Administration, National Hurricane Center.

Landslides

This database identifies point locations of costly individual landslide events in the 50 United States and Puerto Rico. Costly events are defined as those where there was public or private property damage or loss of human life. Landslides are defined in a broad manner, to include most types of gravitational mass movement such as rock falls, debris flows, and the failure of engineered soil materials. Causes of landslides include earthquakes, reservoir draw-downs, and heavy precipitation.

In several cases, the points in this database locate the nearest town or other geographic feature. In all cases, the locations should be considered approximate.

The Landslides database uses the landslide locations determined from published landslide maps and coordinates.

(Required or Optional: Optional — Supplier: Pitney Bowes; updated as needed)

Seismic

This database identifies the probability that ground motion will reach a certain level. This database shows peak horizontal ground acceleration (the fastest measured change in speed, for a particle at ground level that is moving horizontally due to an earthquake) with a 10% probability of exceedance in 50 years. Values are given in %g, where g is acceleration due to gravity, or 9.8 meters/second^2. The lines of equal hazard, which are the lines between the polygons, were determined by interpolating from a grid of equally spaced points in latitude and longitude. Each point was weighted based on the seismic hazard at that location. The grid spacing is 0.1 degrees for Alaska and the conterminous United States, and 0.02 degrees for Hawaii.

(Required or Optional: Optional — Supplier: Pitney Bowes; updated as needed)

Database Name & Description	Required or Optional	Supplier
The data is intended for analysis at the national level, and for large regional areas. The data should be analyzed at scales appropriate for 1:2,000,000-scale data.		
Vegetation Zones This database identifies vegetation growth and change for much of California's forest and range lands. This data can be used to assess current fire conditions, monitor changes over time, and determine management and underwriting options. Each vegetation zone contains cover type, update date, and update cause information. Sutter County, California does not contain any vegetation data because they do not participate in the ongoing vegetation mapping project. The Vegetation Zones database uses data from the California Department of Forestry and Fire Protection and LANDSAT TM imagery at a scale of 1:24,000.	Optional	Pitney Bowes; updated as needed
Volcanoes This database identifies the location of over 250 Holocene volcanoes in or near the United States. Holocene volcanoes are those that are thought to have been active within the last 10,000 years. Coverage includes the United States, Canada, Mexico, Caribbean, West Indies, Iceland, and the Atlantic. This data is intended primarily for analysis at the national level, and for large regional areas. The data should be displayed and analyzed at scales appropriate for 1:2,000,000-scale data. The Volcanoes database uses data from the Global Volcanism Program of the Smithsonian Institute.	Optional	Pitney Bowes; updated as needed

Telecommunications

Table 4: Location Intelligence Module Databases for the Telecommunications Industry

Database Name & Description	Required or Optional	Supplier
Cellular Market Areas This database identifies the Cellular Market Areas (CMAs). CMAs were created from the 1980 Metropolitan Statistical Areas (MSAs) defined by the Office of Management and Budget, the Gulf of Mexico, and Rural Service Areas (RSAs) established by the FCC which do not cross state borders.	Optional	Pitney Bowes; updated as needed

Database Name & Description	Required or Optional	Supplier

The regions cover the United States, Puerto Rico, U.S. Virgin Islands, Guam, American Samoa, and Northern Marianas Islands.

The Cellular Market Area database uses the area definitions from the Federal Communications Commission.

ILEC Boundaries

The Incumbent Local Exchange Carrier (ILEC) Boundaries database identifies the geographic area within which a telecommunications organization has been granted a Certificate of Convenience, or operating authority, to provide local exchange telephone service. The ILEC is also the "carrier of last resort" in that they are obligated to serve that specific geography as a result of regulatory agreements.

Optional — Pitney Bowes; updated quarterly

LATA Boundaries

The Local Access Transport Area (LATA) Boundaries database identifies geographical areas within which a local telephone company may offer local or long distance service. LATA boundaries were established in the 1984 Consent Decree, as part of the divestiture of AT&T local exchange operations.

Optional — Pitney Bowes; updated quarterly

NPA Boundaries

The Number Plan Areas (NPA) Boundaries database identifies the geographical areas of every NPA in use for end office destination codes on the public switched telecommunications network. The database reflects boundary changes both where new NPAs split from existing NPAs and where new NPAs overlay existing NPAs within which a local telephone company may offer local or long distance service.

Optional — Pitney Bowes; updated quarterly

POP Points

The Point of Presence (POP) Points database identifies the locations of all POPs in the US to enable closest site determinations. The term Point of Presence traditionally refers to the switch locations of long distance (inter-LATA or inter-exchange) telecommunications carriers that reside within a specific geographic area.

Optional — Pitney Bowes; updated quarterly

Rate Center Boundaries

Rate Centers are areas within which rates are uniform for various types of local exchange services. The Rate Center Boundaries database identifies rate centers served by at least one incumbent end office destination code (NPA/NXX range) as published in an industry-standard data feed. "Point"

Optional — Pitney Bowes; updated quarterly

Database Name & Description	Required or Optional	Supplier
rate centers, such as educational or governmental institutions, are not included.		
Regional PCS Areas This database identifies the Regional Narrowband PCS service areas. These areas are an aggregation of Major Trading Areas (MTA) into 5 regions. The regions cover the United States, Puerto Rico, Guam, Northern Marianna Islands, U.S. Virgin Islands, and American Samoa. The Regional PCS Area database uses the area definitions from the Federal Communications Commission.	Optional	Pitney Bowes; updated as needed
Wire Centers/Switches Data This database contains wire center boundary information including Wire Center Boundary polygons and Switch Center Points data.	Optional	Pitney Bowes; updated quarterly

Government/Regulatory

Table 5: Location Intelligence Module Databases for Government and Regulatory Applications

Database Name & Description	Required or Optional	Supplier
Washington DC Wards This database identifies Washington DC Wards. The Washington DC Wards database uses the Political Wards definitions from the Washington DC city government and is aligned with the TIGER street network data.	Optional	Pitney Bowes; updated as needed
Washington DC Quadrants This database identifies Washington DC quadrants (northeast, northwest, southeast, and southwest). The Washington DC Quadrants database uses the quadrant definitions from the Washington DC city government and is aligned with the TIGER street network data.	Optional	Pitney Bowes; updated as needed
Congressional Districts	Optional	Pitney Bowes; updated annually

Database Name & Description	Required or Optional	Supplier
This database identifies U.S. congressional districts. It concatenates the State/County FIPS ID with the Congressional District number, hyphen delimited (-), to provide a unique ID string that provides positive identification to the state, county, and district level. The Congressional Districts database uses data from the Congressional boundary definitions from the U.S.Geological Survey.		
Voting Districts This database identifies voting districts in the U.S. The Voting Precincts database uses boundary definitions from the U.S. Census Bureau's TIGER/Line data.	Optional	Pitney Bowes; updated semiannually
Airport Boundaries This database identifies airport property and runways. Coverage includes the U.S. and Puerto Rico.	Optional	Pitney Bowes; updated annually
Airport Points This database identifies facilities in the United States, Puerto Rico, and the U.S. Virgin Islands that handle at least 250 passengers per year. The Airport Points database uses the point definitions from the Federal Aviation Administration.	Optional	Pitney Bowes; updated as needed
School Districts This database identifies Elementary, Secondary, and Unified school districts in the United States. The School Districts database uses the TIGER boundary definitions from the U.S. Census Bureau's TIGER/Line data.	Optional	Pitney Bowes; updated semiannually
GDT ZIP Code Boundary This database identifies the boundaries of United States Postal Service ZIP Codes.	Optional	Pitney Bowes; updated quarterly
GDT County Boundary This database identifies the boundaries of counties in the United States and Puerto Rico.	Optional	Pitney Bowes; updated annually

Database Name & Description	Required or Optional	Supplier
MCD Boundary Data This database identifies the boundaries of Minor Civil Divisions in the United States and Puerto Rico. Minor Civil Divisions (MCDs) are the primary governmental or administrative divisions of a county in many states. MCDs represent many different kinds of legal entities with a wide variety of governmental and/or administrative functions. MCDs are variously designated as American Indian reservations, assessment districts, boroughs, election districts, grants, locations, magisterial districts, parish governing authority districts, plantations, precincts, purchases, road districts, supervisors' districts, towns, and townships. The U.S. Census Bureau recognizes MCDs in 28 states and Puerto Rico. The District of Columbia has no primary divisions, and the District of Columbia is considered equivalent to an MCD for statistical purposes. In some states, all or some incorporated places are not part of any MCD. These places also serve as primary legal subdivisions and have a unique FIPS MCD code that is the same as the FIPS place code. In other states, incorporated places are part of the MCD in which they are located, or the pattern is mixed—some incorporated places are independent of MCDs and others are included within one or more MCDs.	Optional	Pitney Bowes; updated annually
Municipal Boundaries This database identifies the areas of incorporated municipalities in the United States. For additional reference, unincorporated areas recognized by the U.S. Census Bureau (Census Designated Places), are also included with the database. **Note:** This database is not appropriate for taxation and insurance applications.	Optional	Pitney Bowes; updated quarterly
U.S. Quadrangle Boundary Data This database identifies the boundaries of 56,617 Quadrangles in the continental United States and Hawaii. The U.S. Quadrangle Boundary database uses the 1:24,000 scale Quadrangle boundary data from the U.S.Geological Survey.	Optional	Pitney Bowes; updated as needed
Time Zones This database identifies the six standard time zones for the United States, and a seventh zone, which encompasses Puerto Rico and the U.S. Virgin Islands. It also shows the areas in each zone that do not observe Daylight Savings Time (DST). Additional descriptive information includes the zone name and the offset from Coordinated Universal Time or Greenwich Mean Time (GMT).	Optional	Pitney Bowes; updated as needed

Database Name & Description	Required or Optional	Supplier
The Time Zones database uses data from the U.S. Geological Survey.		
State Legislative Boundaries This database identifies legislative districts for each individual state. The data set provides accurate state Senate and Assembly boundaries and their attributes. These datasets are consistent with boundaries defined in legislation to create new legislative boundaries. Currently, the State Legislative Boundaries only contains data for New York. The data is provided by the State of New York.	Optional	Pitney Bowes; updated as needed

Non-U.S. Data

Table 6: Location Intelligence Module Databases Non-U.S. Locations

Database Name & Description	Required or Optional	Supplier
Canada Features This database identifies national parks, rivers and lakes, railways, and major airports within Canada. The Major Airport point layer contains information such as the names of the major airports of Canada, class of the airport, and its regional code which is based on territory or providence. The Railways polygon layer contains information such as railways or ferry information and the company that owns each segment. The Rivers and Lakes polygon layer contains information such as English and French name of Canadian rivers and lakes. The Canada Features database was created using ArcINFO at a 7.5 million base scale.	Optional	Pitney Bowes; updated as needed
World Provinces This database identifies provincial boundaries, Province Name, Country Name, and Continent Name for 8,290 provinces around the world.	Optional	Pitney Bowes; updated as needed

Data Integration

Enterprise Data Integration Module

The Enterprise Data Integration Module provides the features you need to connect to data in multiple sources either directly or through integration with your existing data access technologies. With the Enterprise Data Integration Module you can use Spectrum™ Technology Platform for data warehousing, data quality, systems integration, and migration.

Enterprise Data Integration Module Components

The Enterprise Data Integration Module includes the following stages:

- **Call Stored Procedure**—Call Stored Procedure is a source stage that executes a stored procedure in a database, and returns the results of the stored procedure call as input for the dataflow. Use Call Stored Procedure when you want to get data from a database using a database's stored procedure rather than a query to a table or view.
- **Field Selector**—Field Selector allows you to choose which fields to pass to the next stage in the dataflow. You can use Field Selector to remove unwanted fields from a dataflow. For example, if you have created a new field by combining the data from two fields, and you no longer need the two source fields, you can use Field Selector to retain only the new field and remove the two source fields from the dataflow.
- **Generate Time Dimension**—Generate Time Dimension creates date records, one for each day of the date range you specify. You can then write these records to a time dimension table in a database using the Write to DB stage. The time dimension table can then be used to perform accurate calculations based on a time period. For example, sales by quarter, budget spend by quarter, and revenue by day are all analyses that require a time dimension. Time dimension tables also enable you to account for fiscal years or non-standard quarters in the analysis.
- **Query Cache**—Query Cache looks up data in a cache based on values in one or more dataflow fields and returns data from matching records in the cache, adding the cache record's data to the record in the dataflow. Looking up data in a cache can improve performance compared to looking up data in a database.
- **Read From DB**—The Read From DB stage reads data from a database table or view as input to a dataflow. Read From DB is available for jobs and subflows but not services.
- **Read From HL7**—Read from HL7 File reads Health Level Seven (HL7) data from a text file as input to a dataflow. HL7 is a messaging standard used in the healthcare industry to exchange data between systems. For more information about HL7, go to **www.hl7.org**.
- **Read from Spreadsheet**—Read from Spreadsheet reads data from an Excel spreadsheet as input to a dataflow. It supports both .xls and .xlsx file formats.
- **Read From Variable Format File**—Read from Variable Format File reads data from a file containing records of varying layout. Each record is read in as a list field. You can specify the tag that indicates the parent record type, and all other record types will become list fields under the parent.

- **Read From XML**—The Read from XML stage reads an XML file into a job or subflow. It defines the file's path and data format, including XML schema and data element details.
- **Transposer**—Transposer converts columns to rows. Transposing data is the opposite of pivoting data using the Group Statistics stage, which transforms row data into columns.
- **Write to Cache**—Write to Cache loads output from a dataflow into a global cache, making the data available for lookup from the Query Cache stage. Using a global cache for data lookups improves performance compared to lookups to databases.
- **Write to DB**—The Write to DB stage writes the output of a dataflow to a database.
- **Write to Spreadsheet**—Write to Spreadsheet writes data to an Excel spreadsheet as output from a dataflow.
- **Write to Variable Format File**—Write to Variable Format File writes records of varying layout to a file.
- **Write to XML**—The Write to XML stage writes the output of a job or subflow to an XML file.

Data Federation Module

The Spectrum™ Data Federation Module allows you to create an abstracted virtual data source providing a unified access to diverse heterogeneous data sources through a uniform user interface. Using the module, you can collate the data from multiple data sources and organize it as a specific representation of your business model.

Core Components - Spectrum™ Data Federation

The key components of Spectrum™ Data Federation are **Data Source Connections**, **Baseviews**, **Metaviews**, and **Virtual Data Sources**.

Data Source Connection	Allows you to connect to and access any number of autonomous data sources simultaneously from within the Spectrum™ Data Federation Module.
Baseview	Provides a uniform representation of the underlying data derived from all the configured Data Source Connections.
Metaview	A logical representation of specific business entities created by collating chunks of data derived from one or more Baseviews.
Virtual Data Source	Provides a logical data source to group the user-defined business entities or Metaviews generated. Multiple Virtual Data Sources can be created, and deployed as needed for your business requirements.

Visual Insights Module

Spectrum™ Visual Insights provides the data visualization capability to represent data from heterogeneous data sources using dashboard, reporting and charting tools. It allows you to take informed business decisions based upon data analysis available in the form of reports and charts.

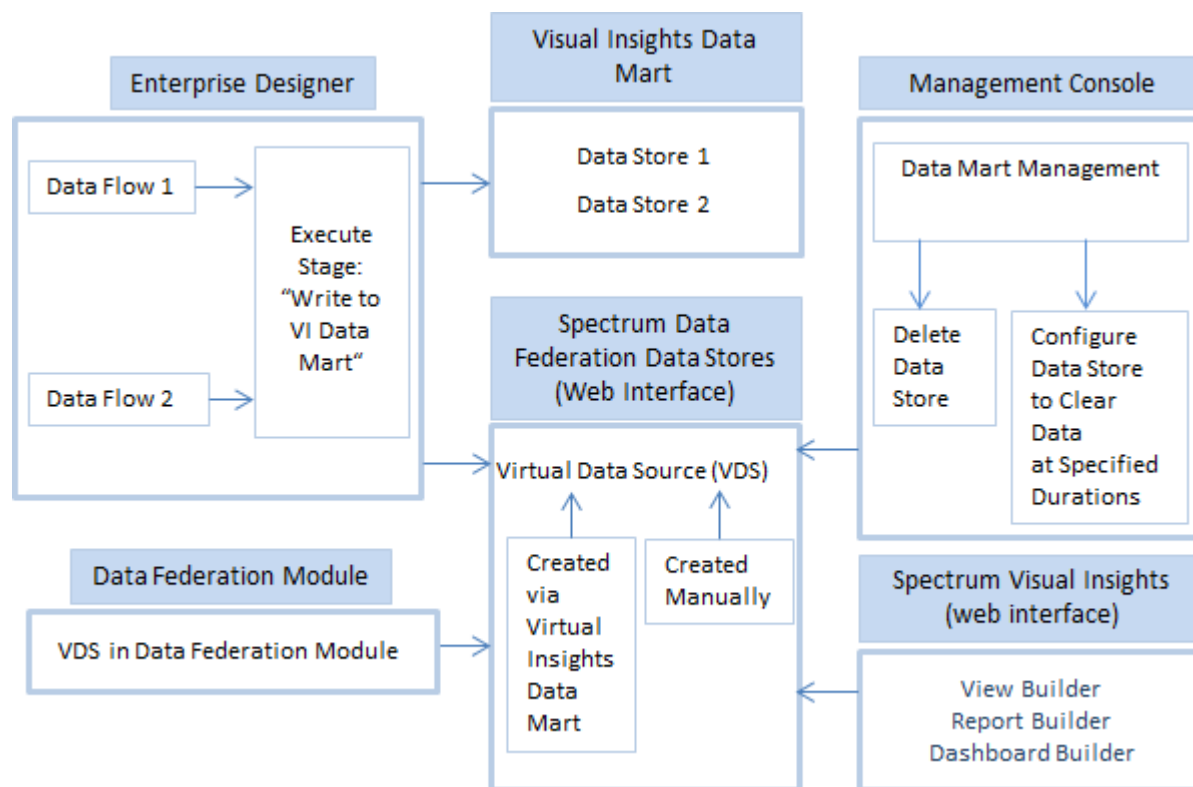***Spectrum™ Visual Insights - Architecture***

Figure 1: Spectrum™ Visual Insights Architecture

The table below explains in detail the Spectrum™ Visual Insights architecture components:

Components	Process	Task	Description
Enterprise Designer	Execute Write to VI Data Mart stage by running a dataflow	To create data stores in Visual Insights Data Mart	When executed, the Write to VI Data Mart Stage copies data from various data sources and stores it in the form of the data stores into Visual Insights Data Mart. Parallelly, a Virtual Data Source (VDS) is created which acts as the input to connect to the data stores.
			Note: Refer Enterprise Data Integration Guide for more details on Write to VI Data Mart stage.

Components	Process	Task	Description
Visual Insights Data Mart	On executing Write to VI Data Mart stage, data stores are copied to Visual Insights Data Mart	To create data stores	Visual Insights Data Mart acts as a container that stores the data stores. **Note:** Through Virtual Data Source (VDS), the Spectrum™ Visual Insights web interface connects with the Visual Insights Data Mart, to utilize its data stores for report generation.
Spectrum™ Visual Insights (web interface)	Sign-in into Spectrum™ Visual Insights (web interface). Refer **Spectrum™ Data Federation Datastores** for more details.	To visualize the data in report/chart format	Visualization tool that utilizes data in Visual Insights Data Mart as input and provides reports/charts/graphs as output.
Management Console	Sign-in into Management Console	To manage data stores	Provides data mart management capabilities such as deleting data stores, configuring data stores to be cleared at specific intervals.

Master Data Management

Data Hub Module

The Data Hub Module provides a persistent repository to help you manage and understand your most critical data assets. It supports Master Data Management and Business Intelligence initiatives. The Data Hub Module is built on top of a graph database that allows companies to rapidly capture and evolve data models based on real-world complex relationships that may span processes, interactions, hierarchies, roles and domains, and extract actionable insight to drive business outcomes.

Data Hub Components

The Data Hub Module consists of the following five components:

• **Write to Hub**—A sink stage that allows you to intuitively create a model using input data to define entities, relationships, and properties. Upon execution Write to Hub loads the data into the hub.
• **Import to Hub**—A stage that uses two incoming channels of data, one for entities and one for relationships, to define a new model or populate an existing model. Includes an optional outgoing error port that collects records not successfully processed by the dataflow.
• **Read from Hub**—A source stage that uses a saved or new query to read the data inside an existing model. It then returns that data as fields in your dataflow's output stage and makes it available for use with other stages or processes.

- **Query Hub**—An intermediate stage that uses incoming data rows to define queries that extract specific entities and relationships from a model. For example, Query Hub can be used as part of a service to understand a customer's influence score within the network or determine if a customer record already exists in the hub.
- **Merge Entities**—A stage that accesses data from an existing model and enables you to merge two or more entities into one.
- **Split Entity**—A stage that accesses data from an existing model and enables you to split one entity into two or more new entities.
- **Delete from Hub**—A stage that enables you to delete entities and relationships from an existing model.
- **Relationship Analysis Client**—A web browser tool that provides a visual interface for viewing relationships and hierarchies within the hub, discovering hidden or non-obvious relationships, creating what-if scenarios, performing temporal or geospatial analysis, creating rules-driven event triggers, running centrality algorithms to determine influence score either against the entire network or against the data being visualized within the client.
- **Data Hub Browser**—A discovery tool where you can search the contents of a model by browsing the results of a natural-language inspired query based on the model's metadata.