White paper: Technical

pitney bowes

Customer Information Management

Effective Single Customer View implementations to answer the tough questions about your customers and their relationships

By Jeremy Peters: GISP, CIMP - Master Data Management & Data Quality Solutions Architect | Principal Consultant | Distinguished Engineer
Pitney Bowes Customer Information Management and Location Intelligence

Table of contents

Abstract

Many organizations need to create and maintain a consistent, single view of customers across their enterprise with accurate and complete customer information. Attaining a Single Customer View gives organizations the ability to have a more consistent and accurate understanding of communications with their customers. In many use cases, it's also important to understand customer information in its original contexts and how it relates to the single view of the customer. For example, businesses need to know how their master customer information relates to multiple channels of communication and transactions with their customers and prospects in order to identify and take advantage opportunities and avoid mistakes that come from knowing or not knowing accurately enough who you are interacting with.

The successful implementation of a Single Customer View can be challenging where representations of a customer are held in more than one system and customer identity and discrepancies in customer data must be resolved both within and between systems. This paper describes how the successful implementation of a Single Customer View – as well as the connected multiple views of customers that arise in different contexts – can be achieved using effective data integration, data cleaning, data enrichment, entity resolution, Master Data Management and graph database, as well as location intelligence visualization and analysis. The Single Customer View implementation is described in the context of a financial crimes and compliance use case.

Single Customer View: Overview

A Single Customer View is a holistic, consolidated and consistent representation of an organization's customer data. Customers expect interactions with an organization to reflect their current profile, history and preferences. For this reason, a Single Customer View is particularly important when organizations interact with customers through multiple channels.

However, in many use cases, it's also important to understand customer information in its original contexts and how it relates to the single view of the customer. For example, banks need to know how their master customer and external customer information relates to the customer and external customer information in the original bank transactions for financial fraud detection.

External customers are the individuals or businesses that are involved in a transaction with the bank's customers.

Implementation of a Single Customer View through Entity Resolution

Entity Resolution defined

Entity Resolution is the process of identifying and consolidating records that are determined to represent the same real-world entity. The term entity describes a person, place, object or thing. In the context of the Single Customer View, Entity Resolution cleans and links data from various data sources in order to build the most accurate profile of a customer. Geospatial Entity Resolution is the process of identifying and consolidating records that reference geospatial locations into a repository of master locations, such as a master repository of accurate and complete customer addresses and locations (latitude and longitude coordinates)

Case Study: Financial crimes and compliance

This paper outlines some of the core Single Customer View capabilities that have been implemented using the Pitney Bowes Professional Services Group SCV Accelerator Solution for fast tracking the configuration and deployment of SCV solutions based on the Spectrum Technology Platform.

The financial crimes and compliance use case for the Single Customer View will be used for the examples in this paper. In this use case, the implementation of effective banking Anti Money Laundering detection requires the identification and grouping of all payments to and from a bank by their customer and external customer. A lack of an accurate and consolidated view of a bank's customers and external customers and the breadth of their relationships to financial transactions is a root cause for inefficient and ineffective fraud detection.

An accurate and consolidated view of banks customers and external customers and their relationships can help reduce the number Transaction Monitoring events that need be looked at as well, as successfully resolve more cases when the parties involved in transactions being investigated are grouped accurately and linked to accurate master profiles of those parties. Entity Resolution, in financial crimes and compliance, is the process of identifying, consolidating and grouping transactions by the parties involved.

Spectrum Software

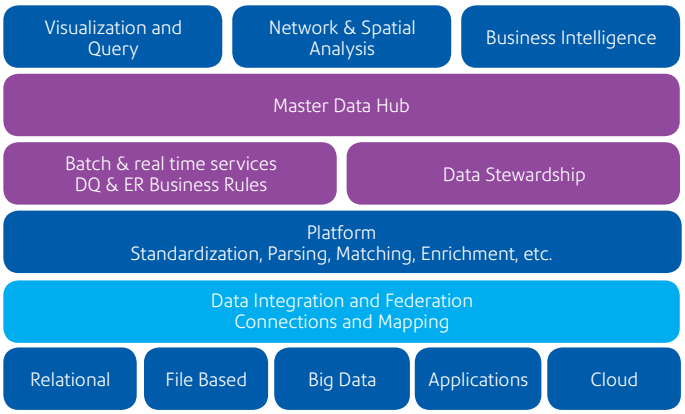
The Spectrum Technology Platform is service-oriented architecture (SOA) platform with a server that supports modules for a broad set of capabilities for data quality, master data management, analytics, location intelligence and data integration. These modules provide different functions, such as advanced parsing/normalization/ standardization, address validation, geocoding, matching/ deduplication, reporting, graph analysis, routing, spatial analysis and mapping among others. The Spectrum Enterprise Designer client tool gives developers drag and-drop capabilities to construct business-process data flows in the form of batch jobs or web services on a workflow designer. The Spectrum Platform provides data integration, processing, visualization and analysis tools for both the graph and geographic dimension of data that helps expose important insights and provide actionable information.

In some of these implementations, Spectrum's Metadata Modeling Data Federation/ Virtualization capabilities are used to execute queries against multiple data sources to create virtual integrated views of the data in memory. Spectrum's DNM, UNM, UAM, EGM and LIM modules and related data sets are used to parse, normalize, standardize, validate, geocode and enrich customer profile information including personal name, addresses, account numbers, bank codes, national IDs, DOB, email addresses and phone numbers. Spectrum's AMM and Data Hub modules are used to match, consolidate and update the master customer graph database repository with any existing or new customer profile information. The web based Business Steward Portal client enables a business steward to review, correct and re-process records that failed automated processing or were not processed with a sufficient level of confidence.

The Spectrum Data Hub Module Visualization and Relationship Analysis (RAC) clients enable users to search, view, analyze and edit customer related entities and relationships in a graph database in a graphical manner to update information and identify relationships and trends. The Spectrum Visual Insights Business Intelligence Dashboard provide data visualization and data analysis

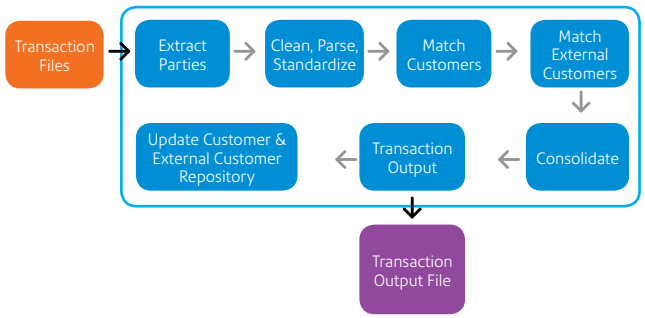
capabilities to represent Key Performance Indicators (KPIs) for customer data using dashboard, reporting and charting tools. Spectrum Spatial Analyst Connect provides integrated map and tabular based visualization, search, analysis and editing capabilities for customer information to help build a more complete and accurate customer profile, identify location anomalies, verify customer place of residence or business visually and flag proximity risk, as well as cross-border situations.

Figure 01



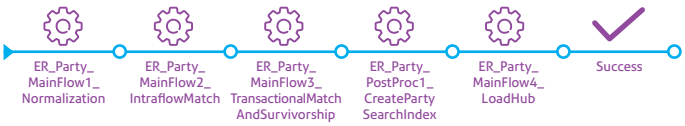
Pitney Bowes Professional Services Group SCV Accelerator Solution components

Figure 02



Example: Financial Transaction customer and external customer consolidation process

Figure 03



Example: SCV Accelerator Solution based customer consolidation Spectrum Process flow for batch processing

Customer data integration

In many customer implementations, multiple different data sources containing customer information are mapped to the same logical customer model so the same entity resolution process can master customer data from multiple existing sources based on a common input data schema. These data sources technologies can include relational and analytical DBMS (e.g Oracle and data warehouse appliances, respectively), big data non-relational data management platforms (e.g. Hadoop platform), NoSQL data store (e.g., graph database, such as Neo4j), applications (e.g. SAP), cloud (e.g. Azure) and text based (e.g. XML).

These data sources can be integrated into a single logical model via batch ETL, real time web service requests and/or Data Federation/ Virtualization among other integration methodologies. In this example, Spectrum's Metadata Insights Modeling Data Federation/ Virtualization capabilities are used to execute queries against multiple data sources to create a virtual integrated logical view of the data in memory. Specifically:

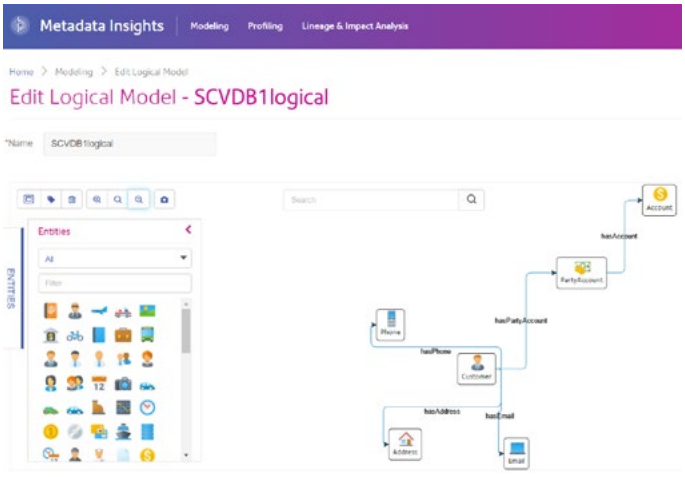
- Connections to the various different data sources are defined and tested in Spectrum Management Console.
- Physical customer data models are defined using these data source connections and Spectrum's Metadata Insights Modeling graphical user interface to represent the tables, views and columns needed from the customer data assets.
- The relevant customer data from the various data sources are profiled using Spectrum's Metadata Insights Profiling capability referencing the Metadata Insights Modeling physical customer data models to identify any data quality issues related to data duplication, accuracy, timeliness, completeness, and consistency.
- A Data Quality Assessment is prepared based on the Metadata Insights data profiling report, to outline data quality issues, data quality rules to address each issue and implementation of data quality rules.
- A Logical customer data model is defined and mapped to the physical model data sources for the data needed to populate each entity using Spectrum's Metadata Insights Modeling graphical user interface and insights into data quality provided by the Data Quality Assessment. The logical model represents the customer business entities and relationships that the business wants to understand for each physical model/data source. In this example, the model is based on the business needs to create a single customer view and the connected multiple views of customers that arise in different contexts.
- Spectrum's Metadata Insights Lineage and Impact Analysis feature can also be used to view the dependencies between data sources, destinations and the processes that use the data.

Figure 04

DQ Profile Column	Inferred Datatype	Length	Uniques(%)	Duplicates(%)	Nulls(%)	Blanks(%)	Patterns	Percentage	DQ Issue	DQ Issue Category	Sample Values	Data Quality Rule	Spectrum Data Quality Resolution
Phone	Text	16	84%	3%	0	2%	(999) 999-9999 99.99.99.99 (9) 999-9999 (99) 999-9999 9999999999	3%	Non Standard Phone Format	Non Standard Format	6175550000	Identify and reformat invalid phone number formats	Parse phone numbers into component parts (CountryCode, AreaCode, Exchange, Number). Validate using Open Parser if the PhoneNumber field is populated Reconstruct and reformat phone numbers using Transform
Zip	Text	10	45%	46%	0	9%	99999 9999 99999-9999 z9z 9z9 zz99 9zz	9%	blank zip codes when address, city and state are populated	blank Values	-	Validate and standardize & correct address	Validate, standardize and correct address to populate all relevant address elements according to postal authority standards using UAM Validate Address

Example data quality issue records from a Data Quality Assessment

Figure 05



Example: Logical Customer Data model from Spectrum Metadata Insights

Customer data cleaning and enrichment

Spectrum Platform data flows can be configured to read customer data from the Metadata Insights logical customer model exposed through a model store. The Spectrum Platform data flows can then be configured to parse, normalize, standardize, validate and enrich customer information based on the Spectrum implementation for each data quality rule outlined in the Data Quality Assessment report. This improves data quality of the mastered customer data and improves the effectiveness of customer information exact and fuzzy matching, in order to determine which customer records represent the same customer. Examples of the batch data quality and enrichment, functionality in FCC examples of our SCV accelerator implementations includes:

- Identify and move customer information entered in incorrect fields to the correct fields.
- Validate and format National Identity Card numbers (e.g. SSN) using valid patterns for each demonstrated country.
- Validate and format phone numbers using valid patterns for each demonstrated country
- Validate and format email address.
- Validate and format dates (e.g. date of birth).
- Validate and format bank account numbers and bank codes according to bank specific and international standards.
- Parse and normalize personal names according to business-specific rules for handling naming conventions, such as conjoined names and natural or reverse order personal names.
- Parse and standardize company names against the Pitney Bowes World Premium Point of Interest dataset (WPPOI) and enrich the business customer records with company registered name, trade name and brand name, segmentation fields for defining type of organization (SIC1, SIC2, SIC8, MiCode ect), address, phone email, relevant business information fields (e.g. sales volume) and business organization information (e.g. parent company). Fuzzy matching is used to compare the parsed company name, phone and address in the customer record to similarly parsed registered, trade and brand company name, address and phone information in the WPPOI dataset. This is done to match suspect company information against a candidate group of company information selected to have likely matches from the WPPOI dataset using a Match Key. Parsed and standardized versions of company phone, email, address as well registered, trade and brand names are used to more accurately match and consolidate customer information. The advanced fuzzy matching process maximizes matches and reduces false positives based on matching individual elements such as company name and suffix. The WPPOI data set contains business data on over 100 Million business and non-business POIs covering more than 220 countries.
- Parse, standardize and validate addresses according to postal authority standards.
- Geocode the standardized and validated address to the best level of location precision available for each address. In most cases this level of precision is the exact property location depending on the geocoding data available for the country. The customer record is enriched with the returned latitude/longitude coordinates and geocoding result quality indicators in addition to other optional return attributes (e.g. assessor's parcel number identifier) depending on the geocoding datasets used. In the U.S., the Pitney Bowes Master Location Data (MLD) dataset is available to not only provide the best level of point location precision available for each address but also provide the MLD PBKey unique and persistent location key for each address in the U.S.. An exact match on the MLD pbKey is used to more accurately and efficiently match and consolidate customer information versus matching on parsed standardized address components using a combination of fuzzy and exact matching. The MLD dataset uses every possible source of address and corresponding coordinate to create a single Master list of address Locations applying complex algorithms to work out the definitive version of every address ("Golden Address") and the most precise coordinate for that address ("Golden Location"). The Pitney Bowes Global Geocoding module provides geocoding support for over 120 countries for non U.S. specific implementations

Customer data consolidation

The customer data is ready for matching and consolidation once it is cleaned and enriched. Two matching and consolidation phases are typically implemented via Spectrum data flows if the input customer data is a batch versus one record input a time (e.g. web service). In the first phase, the customer records in the input batch are matched against themselves to determine if there are any duplicates. Customizable exact and fuzzy matching rules are used that match on attributes like account number, bank code, personal name, registered/trade/brand company name, address (e.g. PBKey), DOB, email, phone, and/or national Identity Card #. The fuzzy matching uses customized matching rules with Phonetic and Edit Distance-based algorithms among others (e.g. syllable alignment, acronym, exact match, character frequency and edit distance). The solution assigns a universally unique 32-digit identifier key unique to each record representing the same customer. A 'Best of Breed' customer record can be created from all matching input customer records by evaluating customer attributes (e.g. name, account number, bank code, DOB, address, phone, email, national id etc) from each record according for completeness, accuracy and/or timeliness to determine which attributes from which record should go into making up the master customer record.

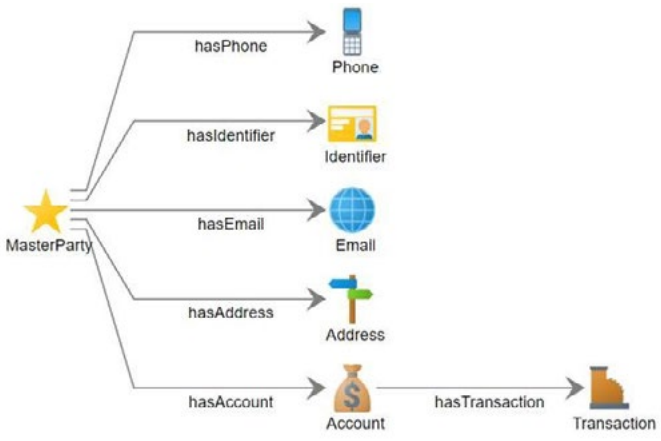
In the second phase, the output records from the first consolidation phase are matched against the existing mastered customer records stored in the Spectrum Graph database based Master Data Management repository to see if a master record for that customer already exists. The existing unique customer id from the repository replaces the newly assigned unique id if a repository match is found. Similar implementation specific exact and fuzzy match rules are used in this phase of processing as the first phase. Again, a 'Best of Breed' customer record is created to update the existing master record by evaluating the same customer attributes from the input record and the matching master record for completeness, accuracy and/or timeliness to determine which attributes from which record should go into updating the master customer record. The Best of Breed input record becomes a new master customer record in the repository if no match is found in the existing master repository. These two phases may be reversed depending on the implementation.

Customer records that were questionably matched in the automated processing can automatically be sent to the Spectrum Business Steward Portal exception repository for manual review as one option for handling exception records. The web based Business Steward Portal client enables a business steward to review, correct and re-process records that failed automated processing or were not processed with a sufficient level of confidence.

An Apache Elastic Search index which contains the master customers candidates used for matching can be dynamically updated with a unique key to improve throughput and avoid matching the same duplicate record more than once using fuzzy matching for large batch consolidation processing.

New customer and existing customer updates are loaded into the Data Hub graph database repository once consolidation is complete, unique customer ids have been assigned and 'Best of Breed'/'Golden' records have been created. One example model displayed below shows the Master Party, Party, National identifier, Phone, Email, Address and Account entities and their relationships. Graph models such as the one shown below can be easily created and modified graphically in the Spectrum Browser based Relationship Analysis Client (RAC) or in a dataflow. In addition to updating the MDM repository, other downstream systems are typically updated with the mastered and linked customer data using Spectrums data integration capabilities needed for those systems.

Figure 06



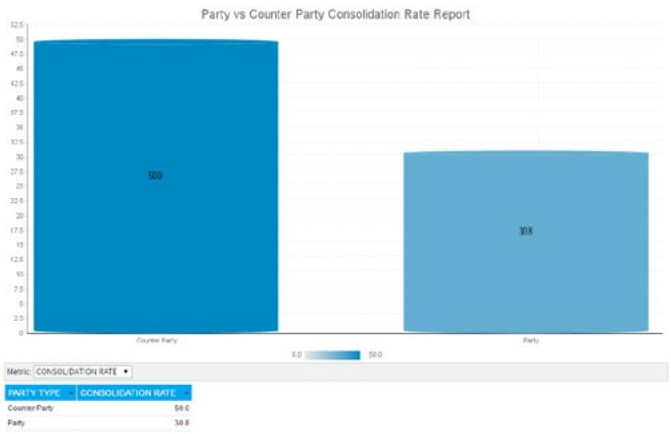
Example: Data Hub Graph Customer Model

Customer graph database and Location Intelligence visualization and analysis

Once the graph database and MDM repository are populated, Spectrum Brower based clients (Visual Insights, Data Hub Visualization and Spatial Analyst Connect) can be used to view, search, analyse and edit customer entities and relationships in graph database, map and dashboard forms. This helps identify relationships and trends.

The Spectrum Visual Insights Business Intelligence Dashboard provide data visualization and data analysis capabilities to represent Key Performance Indicators (KPIs) for customer data using dashboard, reporting and charting tools. This reporting tool provides useful summary information in interactive (e.g. drill down reports) formats for decision makers. Consolidation rate reports, match rule usage reports, segmentation type reports (e.g. how may parties versus counter parties, how many mastered parties versus non-mastered parties, how many business versus individual customers, how many business of each type) are a few of the more common types of reports implemented.

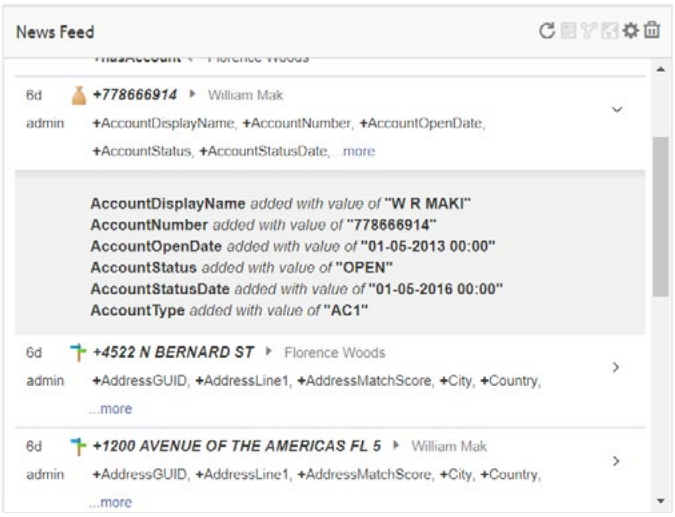
Figure 07



Example: Consolidation Rate Report shows to what degree parties and external parties were consolidated from individual customer records into master golden records

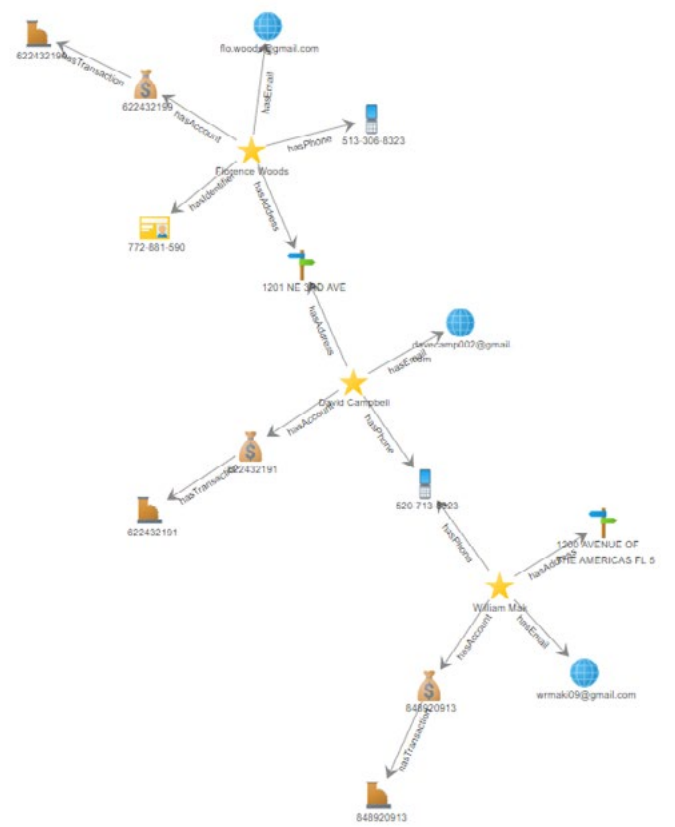
The Spectrum Data Hub Visualization and Relationship Analysis clients enable users to search, view, analyze and edit customer related entities and relationships in a graph database in graphical, map and dashboard forms to update information and identify relationships and trends. Relationship Analysis is useful in illustrating a network of people associated by a common theme like related financial transactions, accounts and/or locations of doing business. To enable this analysis, the Data Hub Visualization Canvas also enables users to write and run a Gremlin-based natural language to search for and display entities and/or their relationships that meet specified criteria. For example, queries can be written using fuzzy search to find business customers with similar names who have a specified type of account (e.g. Party PartyCompanyName is similar To Pitney && Party hasAccount Account AccountType Contains ac1). In addition, connected entities and relationships to selected entities and relationships can be loaded or unloaded as necessary to highlight hidden relationships and trends. The Data Hub Visualization dashboard also provides a news feed widget that displays updates to watched customer related entities if something needs to be tracked more closely. The dashboard also provides counts and other details for entities, relationships, properties, and connections. For example, the dashboard can display the top 20 master customers with the most transactions of a specified type. The Data Hub Visualization Map shows the individual location of selected customer related addresses with latitude/ longitude data on a base map providing spatial context. The map also shows the density of customers in any given region using Heat and Cluster maps.

Figure 08



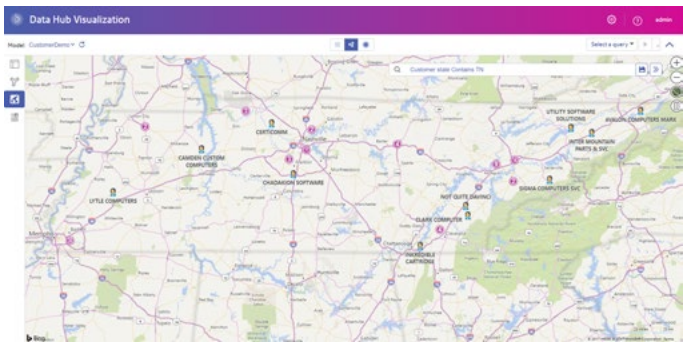
Example: News feed widget that displays updates to watched customer related entities

Figure 09



Example: Display of specified entities and/or relationships based on a Gremlin-based natural language query

Figure 10

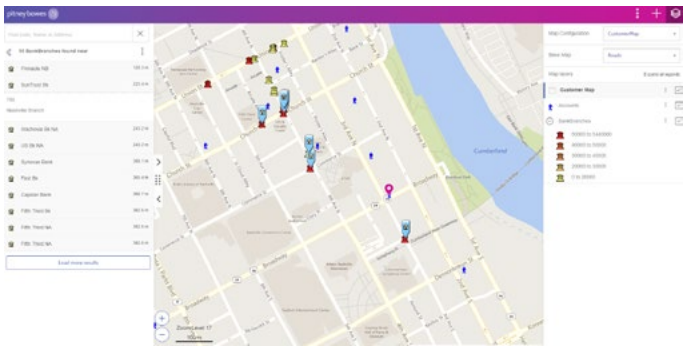


Example: Cluster map of customer locations

Spectrum Spatial Analyst Connect provides more advanced map based visualization, search, analysis and editing capabilities for customer information based on spatial context and relationships. Custom queries can be defined or pre-defined queries can be selected to search for customer profiles and locations that meet specified criteria. Selected customer and transaction locations (e.g. bank branch) can be displayed on a map to identify if there is anything unusual about the locations in regard to other related locations and the base map context (Bing roads/ aerial/hybrid maps are provided out-of-box). For example, the location of a credit card transaction can be mapped and checked if it is at a location the customer has never been before. Moreover, Range, Individual Value and/or Graduated Symbol Thematic maps can be dynamically created to color-code customer locations on the map by risk score, customer type, status or any other relevant attribute.

The Find Nearest and Drive Time/Distance functionality can be used to find other related customer and or transaction locations close by using straight line, drive distance or drive time calculations from a specified location as the first search criteria. The map view and the Data Hub Visualization Canvas graph view can be kept in sync so that if you edit one to correct or enhance customer related information the update can be seen in the other view. For example, the canvas graph view can show the relationship/link between a customer and a bank branch and all the other customers linked to that bank branch if you edit a customer record to assign a bank branch based on a distance search from the customer location.

Figure 11



Example: Find Nearest bank branch to customer to establish or validate relationship between a customer and an account that can then be viewed on the canvas graph view to show the relationship/link between a customer and a bank branch/account and all the other customers linked to that bank branch/account

Results and benefits

For many organizations, multiple inaccurate and inconsistent views of the customer across the organization leads to less effective financial fraud detection, regulatory fines, customers abandoning the process, poor customer experience and higher churn rates. With the right capabilities, however, organizations can lock-in customer relationships, tackle the high cost of fraud and cross-sell far more effectively. The successful implementation of a single customer view and the connected multiple views of customers that arise in different contexts can be achieved. Accurate, consistent and consolidated view of customer data across the entire enterprise can be provided using effective data integration, data cleaning, data enrichment, entity resolution, Master Data Management as well, as graph database and location intelligence visualization and analysis.

United Kingdom
Building 5, Trident Place
Hatfield Business Park
Mosquito Way
Hatfield
Hertfordshire AL10 9UJ
ukenquiries@pb.com

For more information, call 08444 992 992
or visit us online: pitneybowes.com/uk

17SWS05637_UK