



What's Missing from Your Big Data Plans - Novetta Entity Analytics

Actionable Insight Through Big Data

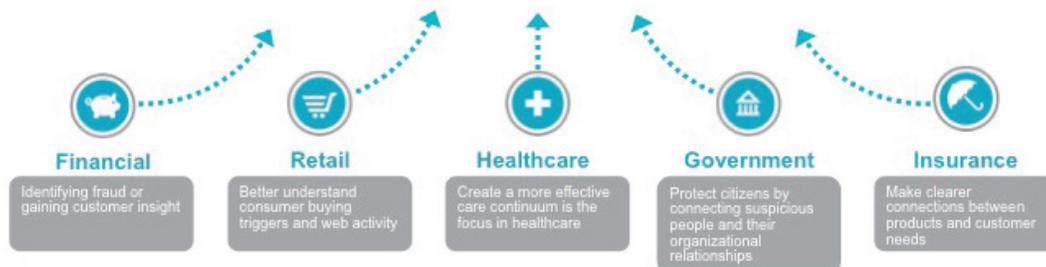
The term "Big Data" is dominating conversations in most organizations and IT departments from large established Fortune 100 companies to the small startups. Organizations are either in the midst of planning how they will leverage Big Data platforms or are in the midst of their deployment. These organizations recognize the value that Big Data technologies offer in allowing organizations to economically ingest and process sources and volumes of data that were previously either not contemplated or unrealistic. However, organizations often miss an important piece of responsible data management. While Big Data platforms like Hadoop offer an ability to quickly ingest and persist billions of records from a multitude of sources of unstructured, semi-structured and structured data, into one place, this is only the first step. The key to driving real business value is to leverage that data into actionable insights. In order to create any meaningful and actionable plans the "big data" needs to be integrated and refined.

Governance and mastering of Big Data sources is challenging. By definition, the minimal investment required to create a single Big Data platform to store this broad array of sources has lowered the bar and provided an opportunity for the business to inexpensively ingest and evaluate data to determine its value. While software exists to help characterize Big Data repositories, in order to assess the value and organize the data into actionable information; organizations need an equally cost effective and scalable solution to refine and integrate the centralized sources of data into complete business entities such as customer, vendor, product, location, event, et cetera. Novetta Entity Analytics provides the solution to address the organization, integration, and refinement of Big Data natively to Hadoop offering similar economics and value of the platform. Additionally Novetta Entity Analytics expands the insights and value that Big Data sources offer by not only integrating data about customers, vendors, suppliers, products, et cetera but also by creating the relationships amongst this data. It enables organizations to understand how customers are



Leverage ALL data to enhance business intelligence and data discovery

1. Increase Operational Efficiency
2. Improve Business Decisions
3. Predictive Analysis



connected to which products, channels, vendors, et cetera. Let's take a look at the much-discussed Big Data example below:

Company XYZ plans to create an "Enhanced 360-degree view of the customer" - extending the existing customer views by incorporating additional internal and external sources of customer identity and behavior. Specifically, Company XYZ plans to leverage Hadoop to ingest click stream data, online customer profiles, the past year's worth of transactional warehouse data, and 3rd party marketing data for a more complete view of the customer and their purchases combined with their 3rd party demographics. These insights will drive improved customer segmentation and predictive analytics on which customers should be offered which products and what is the target audience for specific offers.

Before the advent of Big Data technology, Company XYZ could leverage traditional ETL or MDM processes to refine and integrate the data into useful representations of the parties and the products. However, within Hadoop these technologies are either not natively available, nor cost effective relative to scalability and performance or cannot provide the level of sophistication necessary to combine performance with measures of accuracy and quality. Company XYZ has identified the sources of data, the process for ingestion, and the desired output. However, without a tested data integration and refinement solution native to a Big Data platform like

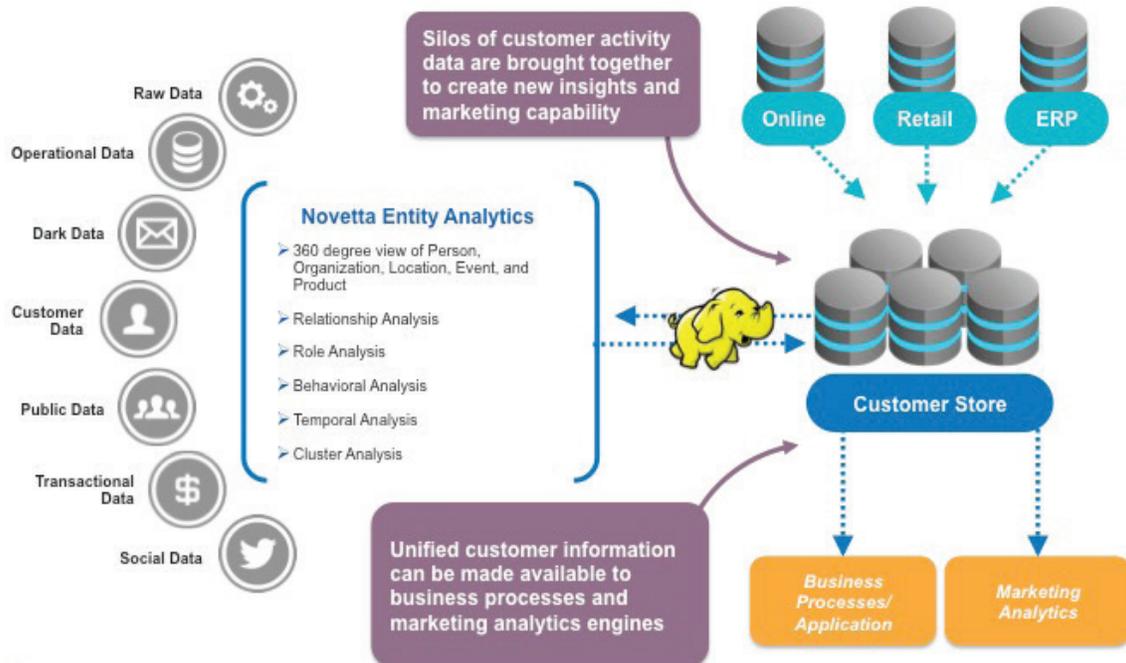
Hadoop, that extends beyond traditional ELT/ETL processes, the "complete view" of the customer will be suspect and likely provide inaccurate results.

Introducing Novetta Entity Analytics

Novetta Entity Analytics offers the missing piece to Company XYZ's plans; data integration and refinement native to Big Data. Novetta Entity Analytics has been developed, deployed and tested widely within a natural proving ground of Big Data technology, the Federal government. Since 9/11, US Government Agencies have been tasked with converting ungoverned Big Data into a manageable, refined and trustworthy source of insight. Multiple agencies leverage Novetta Entity Analytics to provide the required data integration, refinement, and analysis of billions of "records" from hundreds of sources from a variety of structured, semi-structured and unstructured data. Novetta is now bringing this technology to the commercial markets including retail, insurance, healthcare, financial services and oil & gas .

With Novetta Entity Analytics organizations can:

- Accelerate operational insights by constructing complete profiles of customer, organization, location, product, citizen, et cetera. These profiles are constructed from any volume of structured, semi-structured, or unstructured data originating from any application



- Improve customer service and retention by identifying dissatisfied customers and service problems found in call details, transactions and other volumes of interaction data and documents that previously could not be connected to an enterprise view of a customer due to volume, structure, and/or availability
- Increase revenues by creating unified customer profiles from a combination of now available and manageable sources of data including mastered data, transactional information, 3rd party data, and internet behavior data, that expose the relationships to existing products and services improving the effectiveness of cross-sell/up-sell efforts
- Detect threat and fraud by connecting the dots between people, organizations and events across data sources including the integration of transactional history with current and historical master data which was previously not possible with current fraud analysis solutions

Lower costs by solving large complex data integration, management, and analysis problems using a cost effective, predictable, linearly scalable platform for consumption of billions of data points.

Data Exploration and Characterization

As discussed previously, Hadoop offers the ability to centralize data without the development and maintenance of costly and time-consuming data ingestion and transformation processes. However, the result is a centralized environment without consistent structure, naming convention, format, and/or content.

- Provide a true unified view across multiple systems and sources.
- Uncover hidden insights buried in your data – including unstructured, semi-structured and structured data
- Create and manage complex view, relationships and hierarchies
- Reveal patterns, trends, and relationships across enterprise data between people locations, events, and products.

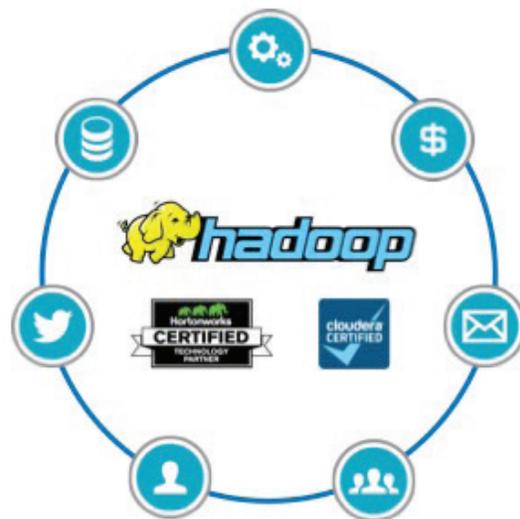
Novetta Entity Analytics provides a solution to this problem with capabilities and technology to:

- Inventory and record the data and sources that are present. The process uses Hadoop based technology to catalog the sources and the attribution of each of the sources.
- Evaluate the content of the sources and attribution to confirm the type and validity of the data.
- Annotate the unstructured sources to extract the relevant business entities and data that will be leveraged as part of the unstructured correlation process in linking to structured data

The Novetta Entity Analytics “Exploration and Characterization” process populates data back into the Hadoop cluster to be leveraged in the subsequent steps within the Novetta configuration console and within the cluster processing.

Data Assessment

Once Novetta Entity Analytics “Explores and Characterizes” the contents of your Hadoop cluster. The solution then measures and profiles the data as a means of assessing the value of the content for refinement, integration, matching, and analysis. The Novetta user interface console provides key metrics and graphics that help determine the strategies and tolerances for integrating the underlying data. The process understands what possible strategies and content can be useful for matching and correlation. It then builds uniqueness measures and statistical



data profiles to help determine optimal threshold and matching values. For instance, understanding that not all values or all attributes about a customer are relevant alone for determining if two records are referring to the same person and at what point should it be combined with another attribute.

Because Novetta Entity Analytics profiles and measures the underlying data, it is able to deal with the statistical densities and variation to provide both precision and accuracy to matching otherwise untrustworthy data. Novetta Entity Analytics provides these same measurements and profiling when it assesses what data and content is meaningful to what business entity and also to analysis and construction of relationships. All of these measurements are provided within the easy to use configuration user interface putting the power of data science into the hands of business and data integration analysts.

Entity Resolution and Summarization

Lastly, Novetta Entity Analytics comes with templates and workflows that are pre-loaded to address common business entities like customer, product, location, organization, et cetera. In addition, Novetta Entity Analytics provides templates and integrated workflows to create different types of relationships such as household, hierarchy, geography, and others. These templates and workflows reflect knowledge of the types of errors typically associated with data about the business entities, attributes, and relationships. Users of the Novetta Entity Analytics configuration console can simply select these templates for specific matching patterns and threshold tolerances for each attribute and corresponding business entity. This allows business analysts and data integrators to perform complex integration functions with simple to use workflows and user interfaces. Once completed, these strategies are executed and results of the entity grouping and linking jobs are available for review and publishing either back to the cluster or search engine. Business Analysts and data integrators can then leverage the analytical capability of Novetta Entity Analytics by constructing segmentation and summarization rules that are applied to the records that are grouped into a single entity and its relationships to other entities. For instance, users of the configuration console can create a segmentation that will identify the group of customers that are solely represented multiple times with the online store and not with retail locations. This population of resolved customer's entities can then be fed to other business

processes or further assessed by the Novetta Entity Analytics solution for iteration on the matching process or further segmentation.

The result of this processing is a rationalized view of a customer, vendor, supplier, provider, product, et cetera. This rationalized view including the relationships a entity has with other entities like suppliers, products, locations, is published back into the Hadoop cluster and exposed via a common integration layer. Likewise the resolved entities can be published to other systems for continued processing like campaign management, customer loyalty or even into your analytics platforms that could not originally deal with the scale and variety of the original data.

Conclusion

As organizations plan and evaluate the value of Big Data technology, the role of a data integration and refinement solution should not be overlooked. While the platform and other tooling are important, the translation to valuable business centered use cases requires context and refinement of the actual data into the entities that are meaningful to the business (customer, product, prospect, vendor, supplier, product, claim, et cetera). By constructing these entities, organizations have the context to not only identify their unique customers but also uncover the touch point for where customers connect with the organization and the relationships these customers have with the company's products, channels and locations. This data represents context to Big Data and is key to unlocking business value. Without a means to provide context, companies adopting Big Data strategies will not identify and realize the value of Big Data sources of information and will struggle to rationalize the economics of Big Data technology like Hadoop. Novetta Entity Analytics offers a unique set of capabilities to quickly and easily assess, integrate and analyze data within your Big Data environment.