NOVETTA
ENTITY
ANALYTICS

EMPOWER YOUR ANALYSTS. GO BEYOND BIG DATA.

Delivering Unparalleled Clarity of Entity Data

White Paper

DELIVERING UNPARALLELED CLARITY OF ENTITY DATA

The ability to bring together data from multiple sources to produce complete views of entities for analysis will dramatically transform how most organizations perform analytics. Real-world entities – people, organizations, locations, assets, events or anything else that can be described by data attributes – are much more complex than just a name, address and basic transactional and customer service data. Once analytics tools and applications can access advanced 360-degree entity views and perform analytics on the data, organizations are able to garner richer and more contextual insights than ever before.

Although the term "360-degree view" has been in use for years, it typically refers to gaining a complete view of just a subset of data about a person, usually data gathered from internal structured sources and presented by master data management software (MDM) or stored in a data warehouse.

Today's new technologies can create comprehensive views of all available entity information from structured, unstructured and semi-structured data in internal and external sources. We call this an advanced 360-degree entity view.

Novetta's team of data and computer scientists began developing entity analytics technologies more than 15 years ago to help federal government agencies better understand their data. These government agencies have many large, diverse, rapidly growing data sources and need to be able to quickly and accurately use new data for analysis without requiring significant time or human interaction. They turned to Novetta's team to help them accurately combine data from many sources and create complete entity views. Novetta's experts spent years measuring and analyzing many diverse data types, and used this knowledge to create proven repeatable methods for characterizing, processing, combining and creating resolved entities from data in many sources and formats. The source-agnostic methods Novetta's team developed allow federal government agencies to build advanced 360-degree entity views and are now built into Novetta Entity Analytics software.

Novetta designed the commercial version of Novetta Entity Analytics to run on Hadoop and to quickly and easily create complete entity records about people, organizations, locations, events, and other real-world entities from many sources. These rich entity records, which also include details about relationships and hierarchies between entities, enable organizations to perform new types of analytics that weren't possible before.

This white paper provides:

- An overview of how organizations can use advanced 360-degree entity views to achieve superior business results.

- A summary of the current challenges organizations face when attempting to combine data sources and resolve entities within Hadoop.

- A description of the capabilities Novetta Entity Analytics provides to deliver the most accurate entity information possible.

ADVANCED 360-DEGREE ENTITY VIEWS YIELD GREATER BUSINESS RESULTS

Advanced 360-degree entity views created from data of many sources and types provide much richer details about each entity's specific actions and behaviors, but this is just the first step in delivering more actionable business decisions. The data can then be used by analytics tools and applications, such as data warehouses, enterprise search, link analysis visualization, online transaction processing (OLTP) and online analytics processing (OLAP) to gain a clearer picture from the data and enable more accurate predictions.

Organizations that are able to create true advanced 360-degree entity views of information from multiple internal and external sources can use this data in a variety of ways, including:

- **Sales and marketing targeting** – Novetta Entity Analytics reconciles vast amounts of customer behavior and fragmented customer identity data to provide organizations with more complete customer intelligence they can leverage to create more effective marketing campaigns and programs, and use for more effective strategic planning.

- **Risk, fraud and abuse detection** – Novetta Entity Analytics connects the dots between people and their behavior patterns, affiliations, organizations, locations and events across multiple data sources and provides complete profiles of suspicious actors that enable organizations to more easily detect potential risk exposure, fraudulent activity and abusive transactions.

- **Security threat assessment** – Novetta Entity Analytics automatically collects and organizes data from multiple sources so organizations can better assess, predict and mitigate risks and threats posed to their facilities, assets, and personnel before they become a reality.

Today's new source-agnostic entity resolution technologies quickly and easily combine, resolve and analyze data to provide an advanced 360-degree entity views from multiple internal and external data sources.

CHALLENGES OF COMBINING DATA SOURCES IN HADOOP

Hadoop is the most cost-effective platform for quickly creating a data lake, but the platform lacks tools to easily combine and analyze data from internal and external data sources. Combining disparate data types from various sources is a challenge because each has a unique structure that makes it difficult to connect and map them all together.

Organizations that want to combine and analyze internal data sources within Hadoop usually employ one of two methods, but each is problematic:

• Data discovery or data science applications require too much guesswork. Users have to determine the best foreign keys to use to connect data sources and make assumptions when creating data model overlays. These guesses are hard to test, often incorrect when applied at scale, and lead to faulty data analysis and mistrust of sources.

• Custom scripts merge data sets together, but do not scale. Hadoop experts write custom, source-specific scripts, using Pig or Java, to merge specific data sets together. However, scripts use hard and fast rules to combine structured data from specific sources, e.g. matching records based on an account number, so they have to be completely rewritten each time a new source is added.

When organizations want to add external sources to internal sources within Hadoop, the problems become even more technically challenging because external data is often fragmented and dirty, and can come from structured, semi-structured or unstructured sources. In addition, external data cannot be easily correlated to internal sources or added to existing analytical models, and most organizations do not want to govern it or add it to operational systems.
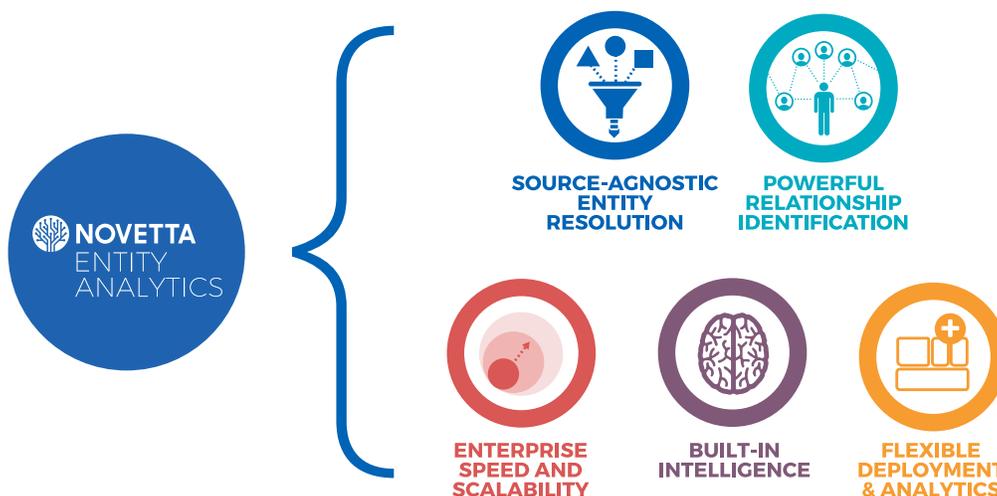
To overcome these challenges, organizations need the ability to easily combine and analyze data from any internal and external source or format within Hadoop.

SOURCE-AGNOSTIC METHODS DRIVE NEW ENTITY ANALYTICS

Novetta Entity Analytics gives organizations unparalleled clarity into their enterprise data and the ability to derive the actionable business intelligence they need to better understand and connect with customers, improve operational processes, and reduce fraud. The source-agnostic software is designed to combine and resolve multiple data sources in Hadoop, including fragmented, dirty, structured, semi-structured, and unstructured data, using empirically-based, statistically-sound, repeatable processes to produce advanced 360-degree entity views of data.

Novetta Entity Analytics includes highly accurate entity resolution and relationship awareness capabilities that produce the most complete entity views possible without impacting systems of record, workflows or other critical enterprise operations and applications. In addition, the software's built-in intelligence, best practices and automated processes allow it to be used effectively by a broad range of users. Novetta Entity Analytics is built for rapid performance and scalability, and also provides flexible deployment and analytics options.

Novetta Entity Analytics Delivers Critical Requirements for Advanced 360-degree Entity Views

Highly Accurate, Source-Agnostic Entity Resolution

Novetta Entity Analytics utilizes its highly accurate entity resolution processes to detect entity information and combine it into advanced 360-degree entity views, without the need for foreign keys or specific internal identifiers.

To resolve entities, Novetta Entity Analytics uniquely characterizes and analyzes data, uses the attributes within the data to resolve entities and combine entity records, reviews combined records to detect potential conflicts, and applies pre-built conflict resolution rules to correct potential issues. The result is precise, accurate and complete views of entities. Analytics tools and applications can then access the comprehensive entity data, perform analytics, and gain insights that weren't possible before.

Powerful Relationship Identification

Novetta Entity Analytics also detects relationships between entities to provide greater context about entities and hierarchies, such as roles within formal or informal groups, or transactional relationships. To identify relationships, the software uses similar processes to those it uses for entity resolution, except it looks at payload attributes within data sources in addition to the attributes it already uses for entity resolution.

Novetta Entity Analytics uncovers both simple and complex relationships, which can be used to further improve the accuracy of resolved entities or prompt users to ask more interesting questions of the data. Simple relationships could include householding, such as two resolved entities that shared the same house at the same time and are either family members or roommates. Complex relationships could be detected by uncovering an individual's behavior patterns from information within transactional records or identifying their employer from information contained within hospital inpatient records.

Novetta Entity Analytics: Uncovering Relationships for Greater Clarity

TO WHOM OR WHAT ARE THEY CONNECTED?
Family
Employees
Coworkers

TO WHAT ORGANIZATIONS ARE THEY AFFILIATED?
Watch lists
Industry affiliations
Official records

WHO IS THIS?

WHAT PRODUCTS DO THEY USE?
Automobile
Smartphone
Computer

HOW DO THEY SHAPE EVENTS?
Accidents
Purchases
Role they play

Built-in Intelligence, Automated Processes and Best Practices

Novetta Entity Analytics contains much of the knowledge Novetta's team has gained over many years consulting on large data integration and entity analytics projects. This built-in intelligence includes the following automated processes and best practices:

- Discovers and extracts data from unstructured and semi-structured sources.

- Characterizes data and creates histograms to show users how data conforms to expected norms.

- Identifies data values that are appropriate or inappropriate for use in resolve entities.

- Generates strict, loose, and moderate threshold settings for users to select based on use cases and data sources.

- Selects resolution strategies, based on specific combinations of data sources, to streamline data integration and improve data matching.

- Detects potential conflicts within resolved entities and applies pre-built conflict rules to improve overall data accuracy.

Enterprise Speed and Scalability to Support All Workflows

Novetta Entity Analytics scales linearly across a Hadoop computing cluster and rapidly derives optimal resolution rule sets to deliver highly accurate resolved entities from large volumes of data. The software works on small clusters with two or three data sources and a few million records, to those with large clusters, many data source and billions of records.

Novetta Entity Analytics is built on a unique, high performance, distributed parallel processing model that partitions data into small groups of records that are likely to be related to the same real-world entity and performs entity resolution on each partition. The software then combines the data from each partition, performs further data analysis and processing to improve accuracy and gain insights into specific entities and groups of entities, and automatically organizes entity data into a multidimensional index.

Flexible Deployment and Analytic Options

Novetta Entity Analytics provides customers with flexible deployment and analytics options to make it easy to get started combining and analyzing different data sources and entity types. These options include:

- Deployment on premise or in the cloud on Cloudera, Hortonworks and Pivotal Hadoop distributions.

- Installation without disruption or modification of operational or transactional systems, changes to business processes or rework of data flows.

- Resolution of real-world entity types to support a variety of workflows, departments and use cases.

- Secure access controls and easy-to-use configuration console to support a broad range of users and projects.

- Easy access to entity and relationship data for other applications by importing data or accessing it as a web service.

ACHIEVE CLARITY AND NEW INSIGHTS FROM BIG DATA

Novetta Entity Analytics allows organizations to bring together multiple data sources within Hadoop and produce complete entity views of data. These advanced, comprehensive 360-degree entity views can dramatically transform how most organizations perform analytics to deliver more actionable insights than ever before.

With Novetta Entity Analytics, users don't have to be data integration or entity resolution experts because they can rely on the breadth and depth of experience and ease of use Novetta has built into its software. In fact, organizations that deploy Novetta Entity Analytics will realize immediate business results without having to hire a team of data scientists and data integration resources to get the job done.

NOVETTA
From Complexity to Clarity.

Headquartered in McLean, VA with over 700 employees across the US, Novetta has over two decades of experience solving problems of national significance through advanced analytics for government and commercial enterprises worldwide. Novetta's Cyber Analytics, Entity Analytics and Multi-INT Analytics capabilities enable customers to find clarity from the complexity of Big Data at the scale and speed needed to drive enterprise and mission success. Visit novetta.com or email contact@novetta.com for more information.

7921 Jones Branch Dr, Suite 500
McLean, VA 22102

(571) 282-3000
novetta.com

@novettasol

novetta

WP08242015