Explosive growth in the volume and variety of data, coupled with a rapidly increasing need for business analytics, means that demand far outpaces the ability to deliver data sets. The process of finding and unifying data sources is largely manual, driven by trial-and-error. Attivio Data Source Discovery reduces time spent cataloging and identifying data sources to accelerate the data discovery process.

• For structured data sets, Attivio evaluates patterns, ranges, sparseness, outliers, cardinality, primary foreign keys, and referential integrity. For large data sets, Attivio leverages a set of algorithms.

• Attivio adds an intelligent layer of metadata to help identify the information. For example, a column in a customer database containing all five-digit numbers is likely zip codes. Nine-digit numbers in the format xxx-xx-xxxx are probably Social Security numbers.

• Human tagging and active learning: experts can tag information in the catalog. This active learning approach refines the semantic understanding of the data as people use the system.

The Attivio Semantic Layer

Unify Data Across Silos

HOW ATTIVIO BUILDS A SEMANTIC METADATA CATALOG

Data Sources

The data discovery process begins with data capture and storage. In addition to the traditional EDW and database systems, most enterprises capture and store semi-structured and unstructured information to fuel analytics.

Semantic Layer

Attivio scans all sources and samples the contents to discover what they contain. It virtually extracts the information and appends additional metadata to build a semantic catalog. Governance and control remain at the source. This enables business users to discover, unify, and model the data they need in a self-service way, while IT ensures quality control.

Data Integration and Preparation

Data cleansing and ETL or ELT are optional steps in the process. Sometimes data is good enough to work with, but sometimes it is not. When this step is required, the relevant data sets are identified and unified in the Attivio semantic layer, and then sent through a cleansing or transformation process prior to the analysis.

Analytics Layer

The final step is to analyze the data that was captured, stored, catalogued and unified. Collecting data from the various source systems is time consuming, manual, and tedious. A semantic layer makes it possible to automate data cataloging and unify the data across silos, eliminating the manual steps and accelerating data discovery.