Teradata Loom® 2.5

Integrated Platform for Data and Metadata Management

HADOOP DATA MANAGEMENT

Take Control of Your Data Lakes

Organizations today are increasingly adopting Apache™ Hadoop-based data platforms because of low cost, scalability, data and processing flexibility to handle a variety of data sources and formats, disrupting traditional ways of information management. Along the way, Hadoop-based data lakes have emerged as a popular use case among early adopters—the vision of which is built on a novel idea of building a single, centralized place to store data from disparate sources in its native format and making it available to multiple analytical processing applications. Cognitive Augmentation is one of the key promises of the data lakes by allowing data from multiple sources like the internet, sensors, and machines to be manageable in Hadoop.

However, as with all technologies, the hype around Hadoop is rampant—and data professionals are overwhelmed by issues with data governance, data security, metadata management, and data quality that are increasingly prohibiting the wider adoption of Hadoop-based data lakes and other emerging data platforms in big data architectures.

Without employing a thoughtful approach to building data lakes, they can quickly turn into ungovernable, inaccessible data stores. This can result in substantial risks such as the inability to determine the data origins, data context, semantic consistency, and data lineage, making it very difficult for data analysts and the wider audience to find and work with the data in a seamless fashion. This results in organizations struggling to trust, understand, and use data effectively for their unique purposes. Therefore, it's very critical to have an integrated data management solution in place to ensure rapid, quick access to high-quality, high integrity data.

Teradata Loom enables data analysts and data scientists to easily find, access, and understand data in Hadoop and quickly start with data analysis to accelerate the time from data acquisition to delivering business insights.

Business Benefits

- Accelerate the time to value of your Hadoop investments.
- Democratize accessibility and data understanding across the organization.
- Rapidly improve the productivity of analysts and data scientists.
- Ease of integration with existing data platform architectures like EDW, Hadoop and NoSQL.

Technical Benefits

- Single, unified integrated platform from discovery to metadata management to data preparation.
- Automated source discovery, metadata generation, data cataloging, and data profiling.
- Self-service data discovery and user personalization.
- Maintain data lineage for data sets.
- Generic measurement model for external sources and statistics.
- Secured access to data and metadata; LDAP and Kerberos integration.
- Agile data preparation for faster insights.

"Without descriptive metadata, and a mechanism to maintain it, the data lake risks turning into a data swamp. Without metadata, every subsequent use of data means analysis start from scratch, like a form of data amnesia."

– Gartner, "The Data Lake Fallacy: All Water and Little Substance"
Niick Heudecker, Published: 23 July 2014

TERADATA®

Maximize ROI from Your Hadoop Investments

Apache Hadoop-based platforms lack rich features, enterprise readiness and is mostly an engineered system, which necessitates additional development efforts to fully realize their potential benefits. Most of the data that is loaded onto Hadoop is generally without a schema and in raw format. This results in data professionals spending an inordinate amount of time to find, understand, and prepare the data for actual data analysis.

Without repeatable processes for generating and maintaining descriptive metadata and data lineage, data must be collected, assembled, and refined by each user separately and independently to be able to drive meaningful business insights. This cycle gets repeated each time. In many cases, organizations experience steeper learning curves as they need to acquire new people, special skills, and new technologies. While so many diverse datasets and data operations increase complexity, the time spent from the data acquisition to insights delivery is too high for organizations to justify the ROI from Hadoop-based data platforms. Taking effective data and metadata management steps in Hadoop reduces complexities and enables analysts and other business users to quickly start analyzing the data, thereby accelerating their productivity.

A Truly Integrated Platform for Data and Metadata Management

Teradata Loom is an integrated big data solution for effective data and metadata management on Hadoop, enabling rapid analyst productivity by making it easy to find, access, and understand data. With deep integration capabilities, Teradata Loom captures all kinds of metadata about data and processes across Hadoop, and provides a single, robust platform for data management, maintaining data lineage, and data preparation through metadata management. Open RESTful APIs provide robust data access to the data and metadata on the Teradata Loom repository from external systems, bringing in an additional level of flexibility and extensibility to the already evolving big data architectures.

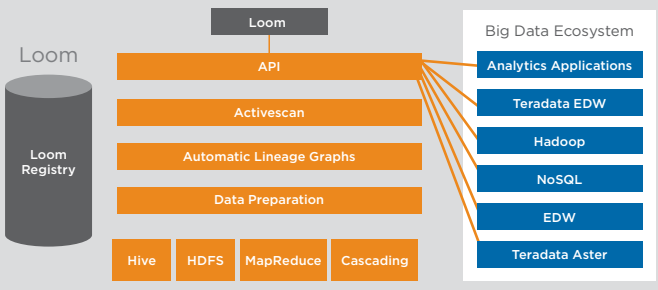

Figure 1. Loom architecture and big data ecosystems.

Automated Data Source Scanning

Teradata Loom's extensible metadata registry is automatically populated as new files are automatically scanned while they get loaded onto Hadoop. File formats, such as JSON, CSV, Apache Log, Apache Parquet, and Apache Avro are determined, and critical metadata, metadata statistics, and schema are automatically generated and maintained in tabular structures. Critical properties including data size, location, number of columns, number of rows, originating system, and date of transformation are automatically generated. Automatic transformations can be customized to be applied to each file type as the data is made available for analysis.

Data and Metadata Management

Teradata Loom's flexible architectural model allows for tracking all of the data assets across Hadoop. Loom maintains the relations between original datasets and transformed datasets, creating data lineage and provenance for all datasets. Teradata Loom also captures metadata on partitioned tables and containers in hive for efficient processing. This basic abstraction to organize and manage the diverse datasets greatly simplifies the tasks and amplifies the productivity of anyone who wants to begin analyzing the data while creating an audit trail for all datasets for verifiable data provenance.

Loom Data Wrangling for Agile Data Preparation

Teradata Loom's data-wrangling capabilities significantly simplify data preparation. Loom enables highly exploratory, iterative interactions with the datasets to quickly prepare the data for meaningful statistical analysis. Analysts and data scientists today spend 80 percent of their time in finding and preparing the data—time that ideally should have been spent on the analysis itself. With Teradata Loom, professionals can spend more time analyzing the data rather than preparing it, thereby dramatically increasing their productivity.

Self-Service Data Discovery

Built on a collaborative platform, the Loom workbench is a simple browser-based, intuitive user interface accessible in a self-service fashion by multiple users in the organization for maintaining and managing the

TERADATA.

data throughout its lifecycle—from automated scanning, metadata generation and management, data lineage, and data preparation stages.

Loom users can browse and search for data based on multiple criteria. Users can also personalize the entities by flagging them as favorites for faster discovery of relevant data.

Metadata Level Permissions

Teradata Loom's fine grain security model automatically cascades authorizations to the metadata level. This is based on underlying file and directory level permissions to allow data and metadata access to only users with required authorizations. Teradata Loom is tightly integrated with authentication mechanisms like Kerberos and LDAP on Hadoop to ensure secured access to data.

Generic Measurement Model

The 'new' generic measurement model in Teradata Loom allows external statistics and measurements to be captured and registered in Loom. These custom statistics from external partner systems, in addition to the ActiveScan statistics, help users to understand the data better and enable users to extract insights from data faster.

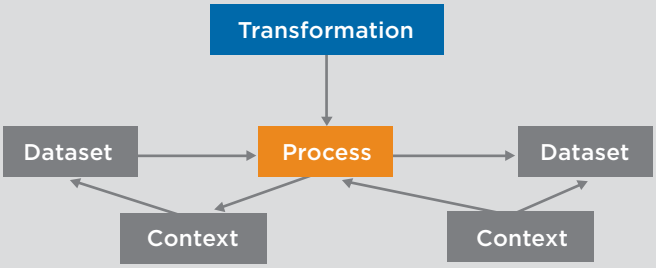

Figure 2. Data transformation with Teradata Loom.

JSON Support

JavaScript Object Notation (JSON) is the primary language that powers the Internet of Things (IoT), a global collection of millions of sensors, embedded microprocessors, and social media platforms. Teradata Loom automatically discovers and parses JSON data, representing it in simple table structures for easy data manipulation and preparation activities by users across the organization.

Supported Data Platforms

Teradata Loom is certified on the following Hadoop-based data platforms in the market today. This is not an exhaustive list.

- Teradata Open Distribution for Hadoop (TDH)
- Hortonworks Hadoop Data Platform (HDP)
- Cloudera CDH
- MapR Distribution including Hadoop

Try Teradata Loom Community Edition for Free

Teradata Loom Community Edition is a free-to-download version on a virtual machine that can be run on a laptop. It's a simple, fast way to interactively experience Teradata Loom and its powerful features for data and metadata management, data lineage, and data preparation. Get started working with your data today.

Download the free Teradata Loom Community Edition here **Teradata.com/tryloom**

Why Teradata?

Teradata helps companies get more value from data than any other company. Our big data analytic solutions, integrated marketing applications, and team of experts can help your company gain a sustainable competitive advantage with data.

For More Information

To find out more about Teradata Loom's data and metadata management capabilities and how Teradata can help you drive more value from your Hadoop investments, contact your local Teradata representative or visit **Teradata.com/loom**.

10000 Innovation Drive, Dayton, OH 45342 **Teradata.com**

06.15 EB8444