



End-to-End Data Integration for Hadoop Data Lakes



Deliver timely, high-quality and well-governed transactional data to the business

Data Lakes enable enterprises to process vast data volumes and address use cases that range from batch analysis to streaming analytics and machine learning. Whether on premises or in the cloud, Data Lakes provide an efficient, scalable and centralized foundation for modern analytics.

But traditional tools for integrating this data are neither efficient nor scalable for Data Lake implementations. IT organizations often struggle to ingest data from hundreds or even thousands of sources that require custom coding and intrusive triggers and agents, tying up your most talented programmers with repetitive and error prone work.

A related challenge is efficiently transforming data into accurate, consistent and analytics-ready systems of record. Scarce programming resources are one obstacle. Another is the lack of metadata and lineage views, which forces users to individually collect, assemble and refine data for analytics.

Attunity solutions remove these obstacles and create an efficient, automated data pipeline that reduces time to analytics.

Data Lake Ingestion with Attunity



Attunity Replicate is a simple, universal and real-time data ingestion solution that delivers data efficiently to any major Hadoop/Data Lake platform. With Attunity Replicate, architects and database administrators can eliminate manual coding with a 100% automated interface that quickly and easily configures, controls and monitors bulk loads as well as real-time updates. You can ingest data across hundreds or thousands of end points – including any major RDBMS, legacy system, data warehouse, Data Lake distribution or streaming platform – through a single pane of glass. Attunity Replicate also minimizes production impact and administrative burden by copying source updates from transaction logs, with no need for agents.

Data Lake Transformation with Attunity



Attunity Compose for Data Lakes automates the creation and loading of Hadoop Hive structures, as well as the transformation of enterprise data within them. Our solution fully automates the pipeline of BI ready data into Hive, enabling you to automatically create both Operational Data Stores (ODS) and Historical Data Stores (HDS). And we leverage the latest innovations in Hadoop such as the new ACID Merge SQL capabilities, available today in Apache Hive (part of the Hortonworks 2.6 distribution), to automatically and efficiently process data insertions, updates and deletions.

Attunity Replicate integrates with Attunity Compose for Data Lakes to simplify and accelerate data ingestion, data landing, SQL schema creation, data transformation and ODS and HDS creation/updates. Here is a sample architecture and description of how a combined Attunity solution can manage data flows at each stage of a data lake pipeline.

CUSTOMER SUCCESS

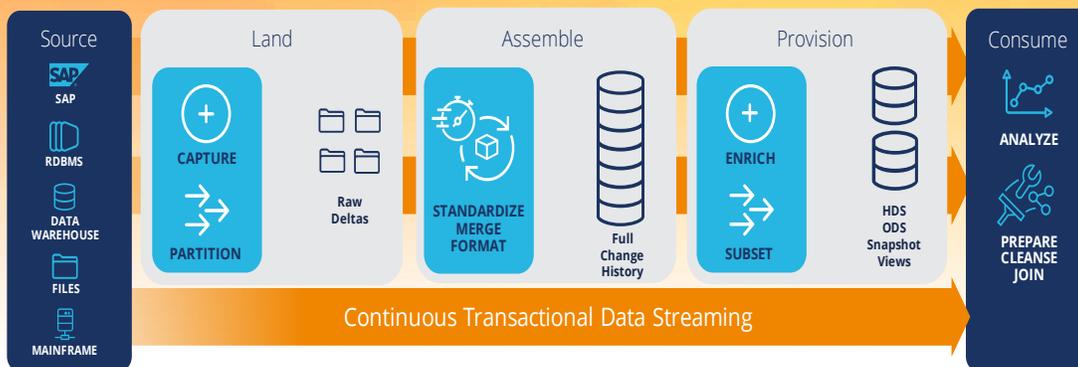


“Using Attunity, we were able to create our strategic analytical platform, insights analytics, which allows us to make important operational decisions that benefit our staff and students.”

JUERGEN STEGMAIR, LEAD FOR DATABASE ADMIN, UNIVERSITY OF NORTH TEXAS



Your Data Lake Pipeline



• Landing Zone



First Attunity Replicate copies data, often from traditional sources such as Oracle, SAP and mainframe, then lands it in raw form in the Hadoop File System (or cloud equivalent). This process enjoys all the advantages of Attunity Replicate, including full load/CDC capabilities, time-based partitioning for transactional consistency and auto-propagation of source DDL changes. Data is now ingested and available as change tables, but not yet ready for analytics.

BUSINESS BENEFITS

- Faster Data Lake operational readiness
- Reduced development time
- Reduced reliance on Hadoop skills
- Easier compliance

• Assembly Zone



Next Attunity Compose standardizes and combines change streams into a single transformation-ready data store. It automatically merges the multi-table and/or multi-sourced data into a flexible format and structure, retaining full history to rewind and identify/remediate bugs if needed. The resulting persisted history provides consumers with rapid access to trusted data, with no need to understand or execute the structuring that has taken place. Data managers and architects, meanwhile, maintain central control of the entire process.

• Provisioning Zone



Finally, data managers and architects provision an enriched data subset to a target, potentially a structured data warehouse, for consumption (curation, preparation, visualization, modeling and analytics) by data scientists and analysts. Data can be continuously updated to these targets to maintain fresh data.

• Metadata Integration and Management



Attunity provides automated metadata management capabilities to help enterprise users better understand, utilize and trust their data as it flows into and is transformed within their data lake pipeline. With Attunity Replicate and Attunity Compose you can add, view and edit entities (e.g., tables) and attributes (i.e., columns). Attunity Enterprise Manager centralizes all this technical metadata so the lineage of any piece of data can be tracked from source to target, and the potential impact of table/column changes across data zones can be assessed.

In addition, Attunity Enterprise Manager collects and shares operational metadata from Attunity Replicate with third-party reporting tools for enterprise-wide discovery and reporting. Attunity continues to enrich its metadata management capabilities and contribute to industry initiatives such as ODPI to help simplify and standardize Big Data ecosystems with common reference specifications.

Contact Attunity today to learn how we can help you streamline your Data Lake pipeline and speed your analytics readiness.