

Data Lake on AWS with Talend

Delivering Digital Transformation Through Advanced Analytics

CHALLENGE

Gaining actionable insights from legacy data sources is a common challenge faced by organizations of any scale. Scattered data sources, disparate data types, and inflexible data platforms can all compromise the ability to get what you need from your data.

In today's digital economy, organizations are constantly exploring new ways to harness the power of data to transform their business. Legacy data platforms and traditional data warehouses are not enough to manage the digital data deluge. Organizations with legacy platforms face significant challenges that affect their business. These include:

- Rigid data platforms with scattered information silos
- Complex integrations and point solutions
- Delayed business decision making due to delayed availability
- Legacy systems affecting data monetization initiatives
- Delayed onboarding of data
- Rising costs

These challenges result in:

- Missed business opportunities
- Revenue and cost leakage
- Governance and trust issues

DOING MORE WITH DATA LAKES IN THE AWS CLOUD

By having native connectivity to a variety of on-premises, cloud, and big data sources (including AWS), customers no longer need to hand-code integrations between data sources and can:

- Experience faster speeds and scale with Spark Streaming and Hadoop
- Allow any data workers to access and cleanse big data while governing its use
- Protect big data investments with a future-proof architecture

SOLUTION

An out-of-the-box open data lake solution with AWS, Talend, and Cognizant allows you to build, manage, and govern your cloud data lake in the AWS Cloud so that you can realize your digital transformation goals rapidly.

SOLUTION DELIVERS AN ACCELERATED APPROACH TO MODERNIZING DATA LAKES

Benefits include:

- Low-cost and elastic infrastructure with a flexible, pay-as-you-go, cloud model ideal for dealing with vast amounts of data.
- Rapid data ingestion for faster onboarding of data.
- Data modernization and data offload with native connectivity to almost any legacy system.
- Increased productivity by incorporating data integrity and data quality throughout the organization.

FILLING THE DATA LAKE

Talend's native connectivity to a myriad of data sources makes it easy to read and parse schemas, whether they are SaaS apps such as Salesforce.com, on-premises apps such as SAP, legacy apps and databases, SOAP or REST web services, or streaming data protocols such as Kafka, and MQTT.

Creating a data lake quickly requires the ability to ingest all types of data in either batch mode, or in real-time. Native code-generation from Talend, the ability to scale out job servers rapidly on AWS EC2 instances, combined with Cognizant's Data Ingestion Framework and Mainframe Offload capabilities ensures that filling an AWS S3 data lake without any data loss, and keeping all data integrity can be done so in a highly-performant way.



MANAGING AND USING DATA

Most data lakes consist of a large volume of extremely raw and unstructured data that would normally take a long time to process. With Talend's Big Data Integration platform, customers can use Apache Spark natively when creating jobs in Amazon EMR and achieve a higher throughput as a result. Data must be processed and prepared before it can be consumed by an analytics app such as Amazon Quicksight. Cognizant's QA automation framework ensures that business-critical metadata information is extracted and stored in an Amazon Redshift or Amazon S3 repository for reuse later, either for data discovery or reporting.

Ultimately, cleaned and prepared data can be used in analytics applications such as Amazon Quicksight for multi-dimensional business insights and also fed into Amazon Machine Learning to automate extraction of hard-to-discover trends.

MAKING IT EASIER TO PROCESS REAL-TIME DATA

Talend delivers a solution that empowers customers to architect a data lake quickly. They help companies improve productivity of their real-time big data projects in the AWS Cloud by incorporating data integrity and data quality throughout the organization.

With this solution, customers can:

- Leverage a low cost and elastic infrastructure with a flexible, pay-as-you-go, cloud model
- Quickly and easily ingest and integrate any type of data at scale
- Accelerate data modernization and data offload with native connectivity to almost any legacy system
- Deliver self-service data access and data preparation

Gaining greater insights

With Talend, users can get greater access and insights from the data they generate to gain a competitive advantage. Talend bridges the gap for companies who need real-time access to their data and the expertise to derive the right kind of insights from that data. If a company doesn't have the resources or bandwidth to deliver real-time analytics, decision making gets delayed and opportunities may be missed.

ABOUT AWS

For 11 years, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud platform. AWS offers over 100 fully-featured services for compute, storage, databases, analytics, mobile, Internet of Things (IoT) and enterprise applications from 44 Availability Zones (AZs) across 16 geographic regions in the U.S., Australia, Brazil, China, Germany, Ireland, Japan, Korea, Singapore, and India. AWS services are trusted by more than a million active customers around the world – including the fastest growing startups, largest enterprises, and leading government agencies – to power their infrastructure, make them more agile, and lower costs. To learn more about AWS, visit <http://aws.amazon.com>.

ABOUT TALEND

Talend (NASDAQ: TLND) is a next-generation leader in cloud and big data integration software that helps companies turn data into a strategic asset that delivers real-time, organization-wide insight into customers, partners, and operations. Through its open, native, and unified integration platform, Talend delivers the data agility required for companies to meet the constantly evolving demands of modern business. With Talend, companies can easily scale their data infrastructure and rapidly adopt the latest technology innovations in cloud and big data.

A leader in the 2017 Gartner Magic Quadrant for Data Integration Tools and Forrester Wave: Big Data Fabric Q4 2016, Talend's solutions support over 1500 global enterprise customers including Air France, GE, and Lenovo, across a range of industries. For more information, please visit www.talend.com and follow us on Twitter: @Talend.

ABOUT COGNIZANT

Cognizant (NASDAQ-100: CTSI) is an AWS Big Data Competency Partner and one of the world's leading professional services companies, transforming clients' business, operating and technology models for the digital era. Our unique industry-based, consultative approach helps clients envision, build and run more innovative and efficient businesses. Headquartered in the U.S., Cognizant is ranked 205 on the Fortune 500 and is consistently listed among the most admired companies in the world. Learn how Cognizant helps clients lead with digital at www.cognizant.com or follow us @Cognizant.

