



# Infoworks

## AGILE DATA ENGINEERING

Eliminate Big Data Complexity Through Automation

### **Solution Brief:**

Agile Data Lake Creation and Management

**Reduce the cycle time to develop, test and deploy an enterprise data lake into operation**

## The Challenge

### Building a data lake is harder than you think

In an effort to take advantage of the flexibility of big data and cloud architectures an increasing number of companies have jumped onto the data lake bandwagon. However, implementing a data lake is more complicated than most people realize. In fact, simply populating the data lake can pose significant challenges as out of the box open source ingestion tools aren't especially scalable or robust. When you have large amounts of data, ingestion requires parallelization. However, not everyone has the in-house skills to write custom ingest code and even then, hand-coding is inflexible and difficult to repair.

Additionally, big data and cloud platforms that are used as the basis for data lakes don't deal with production issues that occur as a result of constantly changing source data that can break production pipelines. Data lakes don't typically support the merging of changed data directly. The result of not properly managing merged data is that access to fresh data becomes unnecessarily complex, and very slow. Lastly, big data platforms don't deal with enterprise production issues like the need to orchestrate potentially conflicting data pipelines or deal with issues like restarting jobs if a process fails. In the end, running big data workflows in production still requires experts that are either hard to find or extremely expensive to hire.

This is precisely why, according to a Gartner survey, over 80% of big data projects fail to deploy to production.

## Infoworks in Action

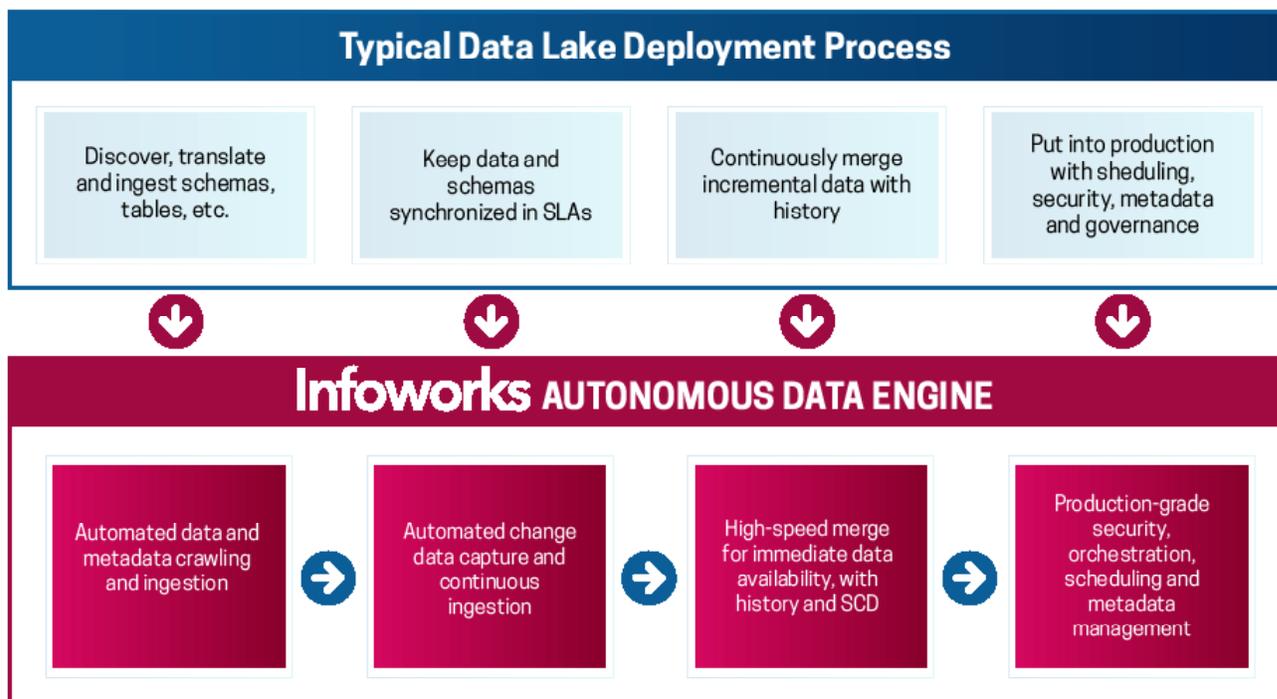
The Autonomous Data Engine has been deployed in production by large enterprises to run business critical applications. Infoworks' customers have successfully implemented complex, large-scale use-cases in days instead of months with minimal resources.

### Fortune 500 High-Tech Manufacturer 3 month implementation with Infoworks vs 3 year implementation without

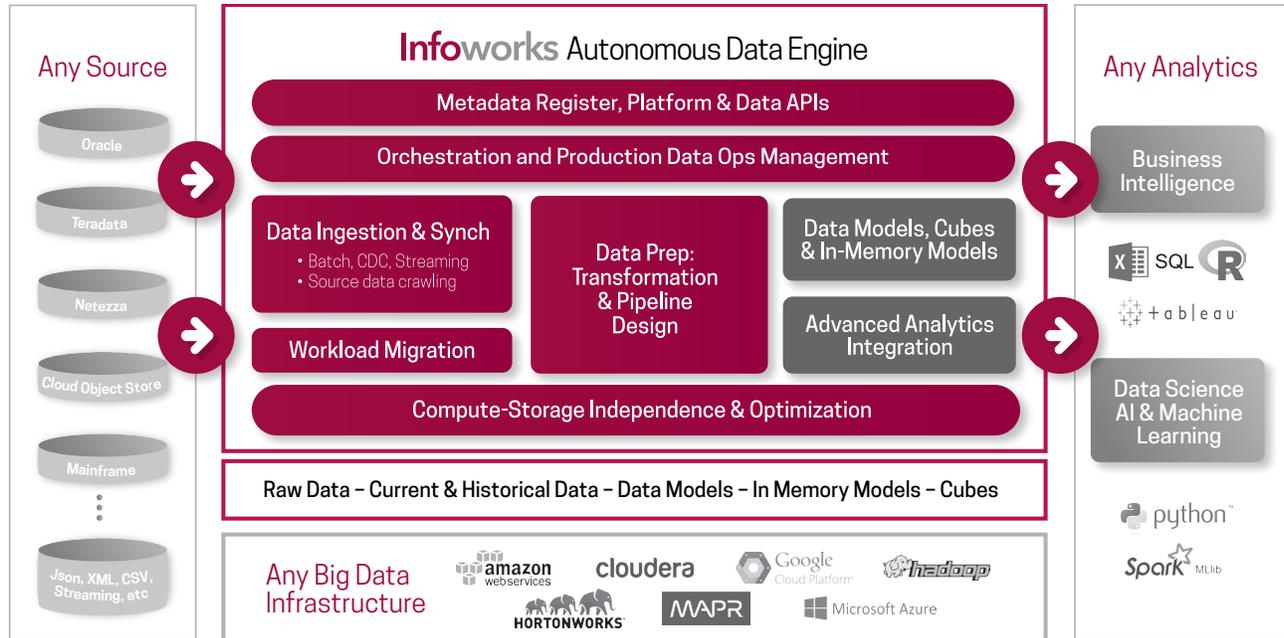
- Replaced an in-house data ingestion solution that had been developed over 3 years
- Ingestion framework for 1500+ data sources across the enterprise
- Change data capture & high-speed replication for all sources
- Production ready environment with full resource optimization
- Reduced the support team from five data engineers to one

# 9x Improvement

## Infoworks Automates Data Lake Creation and Management Processes



## Any Source, Any Big Data Platform, Any Analytics



### The Solution

#### End-to-end automation is key to successful deployment of big data projects

Infoworks provides a complete solution that makes creating and then managing a data lake, either on premises or in the cloud, a simpler process by delivering the following automated components as part of a single package:

### Data Ingestion and Synchronization

#### Data Ingestion

Ingests source data in a high-performance, parallel process, while automatically preserving data precision. Infoworks provides a no-code environment for configuring the ingestion of data into your data lake via batch, change data capture (CDC), and data streaming.

#### Metadata Synchronization

Automatically crawls data sources, ranging from flat files, XML, JSON to relational databases. Learns the metadata and infers data relationships for the ingested data from external data sources as well as for data sets created using Infoworks, making metadata searchable via a metadata repository.

#### Data Synchronization

Continuously synchronizes source data from enterprise databases, data warehouses, and file sources. Changing data is captured from the source systems using log-based and query-

based methods. The changed data is merged with the base data in a high-performance continuous merge process.

- Automatically handles slow-changing-data and schema changes and creates current and historical tables.
- Supports export functionality to other enterprise operational and data warehouse systems.

### Workload Migration

Automates migration of existing SQL workloads to a big data environment. With automated data, schema and workload migration, the Autonomous Data Engine makes it easy to migrate existing SQL workloads and also allows customers with in-house SQL expertise to write their pipelines in ANSI SQL and automatically migrate them onto the Infoworks platform.

### Data Prep: Transformation and Pipeline Design

Provides self-service data preparation using an interactive, drag-and-drop data transformation capability with support for SQL-based and other transformations. Users work with data in a collaborative, suggestion-based interface that reduces or eliminates dependence on IT skills.

## Orchestration and Production Data Operations Management

Design end-to-end workflows and orchestrate in production with fault-tolerant, distributed execution. Migrate from development environments to production across big data or cloud platforms with single-click operations.

## Compute-Storage Independence and Optimization

Infoworks automation also makes it easy to move from an on-premise Hadoop platform to the cloud, or from one cloud environment to another. One Infoworks customer moved an entire set of production workflows from Microsoft Azure to Google Cloud Platform in less than one day.

## Metadata Register, Platform and Data APIs

Provides the ability to search metadata about data sources, pipelines and workflows. Additionally, all platform services and data are accessible via a wide variety of API access points. Infoworks also tracks end-to-end data lineage so users can trace data elements back to the original source systems and perform downstream impact analysis.

## Enterprise-Grade Security Integration

Infoworks provides security integration for user authentication and role-based data security policies. It supports single-sign-on/LDAP integration, Kerberos based authorization as well as encryption for data in motion and at rest. Infoworks also provides role and domain management for controlling access to data managed within the data fabric.

“

## What Our Customers are Saying:

**“With Infoworks we were able to complete our project plan for the entire year, in a few days!”**

- Lead Enterprise Architect, Fortune 10 Retailer

**“Infoworks reduces our time to introduce new end-to-end analytics models from 6 months to a week, without IT involvement. This allows our analytics teams to quickly meet fast-changing business requirements and directly enables growth”**

- Director of Analytics, Leading CPG company

**“With Infoworks we can quickly execute on our large backlog of data projects. If we were to custom build our ideal data platform it would be like Infoworks”**

- Data Architect at a Fortune 100 technology company

**“Prior to Infoworks we had long turn-arounds for data requests to IT. Now we have self-service reporting end-to-end”**

- Business Analyst

# Infoworks

 (650) 391-9306

 sales@infoworks.io

 www.infoworks.io