



# Infoworks

## THE AUTONOMOUS DATA ENGINE

Solution Brief:

Automated Data Warehouse Offload and Migration

Migrate from legacy data warehouses to the cloud  
and big data in days

## The Challenge

Migrating off your legacy data warehouse infrastructure is harder than you think

For years, the big data world has been talking about how Hadoop and cloud would create a new cost curve that would eliminate “old-fashioned” data warehouse appliances. For those companies who have made the move, the results are amazing. Analytics that use to take 30 hours covering a month’s worth of historical data now take 30 minutes analyzing an entire decade’s worth of historical data. Unfortunately, very few companies have fully made the move to “big data.” In fact, according to Gartner, over 80% of big data projects fail to deploy to production.

The reason this mass migration never occurred is that big data, whether on premise or in the cloud, is still too hard, and the expertise is expensive and extremely hard to find.

To make matters worse, migrating from your existing DW environment is even harder. You have to rewrite the workflow logic to run in the new environment which means you need an expert who knows both the old and the technology. You then need another expert who knows how to import data in bulk and then make sure that the data stays in synch while you complete the transition from the old system to the new system. Yet another expert is required to convert and optimize your old transformation logic to the new system. The finally, all of this work has written in a manner so it is easily manageable while running on top of a distributed infrastructure.

## 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 10 Retailer: EDW Offload 8x Faster  
26 day migration with Infoworks vs  
8 month migration without**

- Migrated from Teradata to Hadoop cluster
- 300 Tables of data and metadata migrated
- 70 SQLs to automated and visual pipelines
- 5 BTEQ scripts to graphical workflows

## 8x Faster Migration

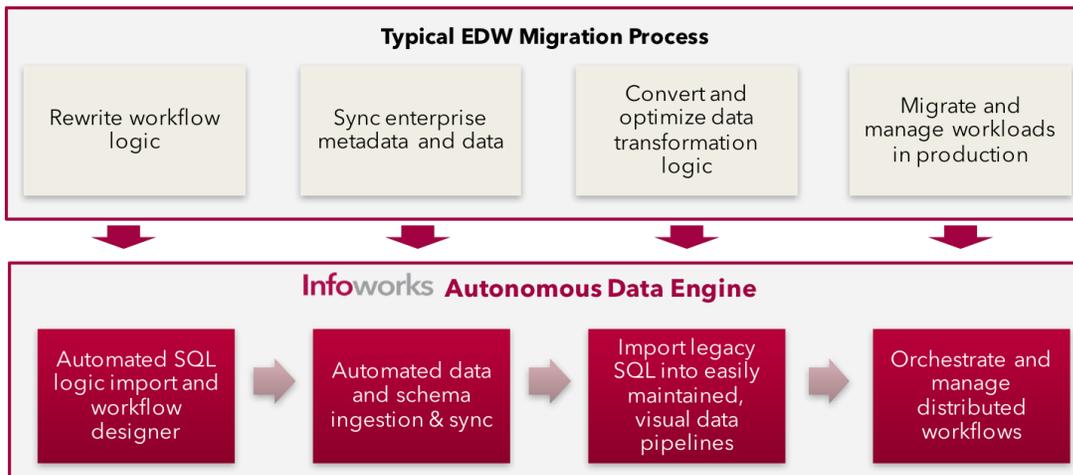
### The Solution

Automation is the key to eliminating complexity

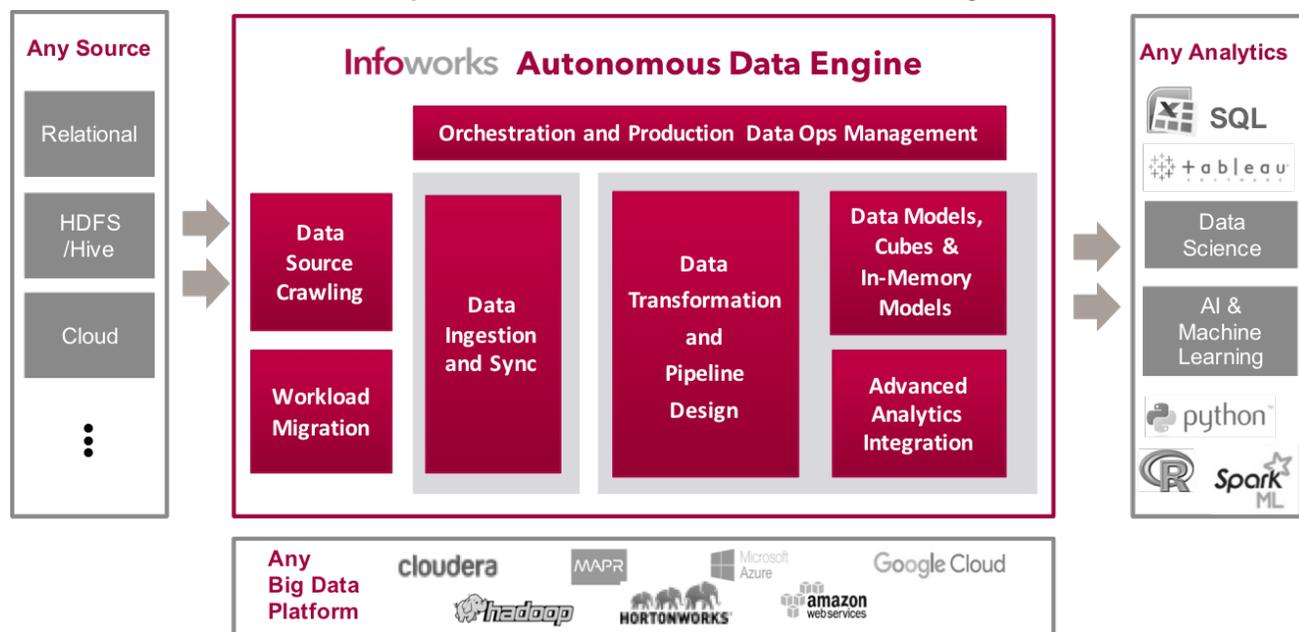
There are many visual coding tools available that provide pretty interfaces to solve point problems within a data warehouse offload or migration project. The issue is that the user must stitch these tools together to make it all work. There is no integration and little to no automation.

Infoworks addresses the end-to-end challenges with end-to-end solutions that are more than just a pretty user interface. Our Autonomous Data Engine automates most of the work for you, which is why our Fortune 500 customers are in production in a matter of days or weeks. Infoworks automatically converts legacy data warehouse environments to big data and the cloud in days, converting and optimizing your old transformation logic and workflows ...automatically. We then provide an optimized operational environment to track & manage production data pipelines.

### Infoworks Automates the Data Warehouse Migration Process



## Infoworks Components for Data Warehouse Offload and Migration



Infoworks provides a complete solution that makes migrating from a traditional data warehouse to a big data environment quick to complete and easy to maintain by delivering the following automated components as part of a single package:

### Automated Workload Migration

Automates migration of workloads (ETL logic, BTEQ in Teradata, SQL workloads, and other such programs) from legacy data warehouses to a big data environment. With automated data, schema and workload migration, the Autonomous Data Engine provides a comprehensive solution for data warehouse offload and migration.

### Automated Data Ingestion & Synchronization

#### Data Source Crawling and Ingestion

Automatically crawls data sources, ranging from flat files, XML, JSON to relational databases such as Teradata, Oracle, and SQL Server.

In the same way Google crawls the web to get web data, the Autonomous Data Engine crawls data sources and ingests source data in a high-performance, parallel process, while automatically preserving data precision.

#### Metadata Synchronization

Learns the metadata and infers data relationships for the data ingested from external data sources as well as data sets created using Infoworks. It also tracks end-to-end data lineage so that users can trace data elements back to the original source systems and perform downstream impact analysis.

### 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.
- Supports streaming, batch and incremental mode of data synchronization and export.

### Data 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.

## Data Models, Cubes and In Memory Models

Build target models including in-memory models, that are automatically optimized for fast access. Visually design star schemas, and automatically build high-performance OLAP cubes accessible from industry standard tools such as Tableau, Microsoft Power BI, MicroStrategy, etc..

## Advanced Analytics Integration

Integrate data pipelines with advanced analytics algorithms from libraries such as SparkML & R, with no need for coding. Build trained models or import pre-trained models into data pipelines.

## Orchestration and Production Operations Management

Design end-to-end work-flows 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.

## Portability

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 work-flows from Microsoft Azure to Google Cloud Platform in less than one day.

## Enterprise-Grade Security Integration

The Autonomous Data Engine provides security integration for user authentication and data security policies. It supports Single-sign-on/LDAP integration, Kerberos based authorization. It supports encryption for data in motion and at rest.

## 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

 (408) 899-4687

 sales@infoworks.io

 www.infoworks.io