



Infoworks

AGILE DATA ENGINEERING

Eliminate Big Data Complexity Through Automation

Solution Brief:

Agile Data Workflows for Business Intelligence & Analytics

Reduce the cycle time to develop, test and deploy business intelligence reports and analytics models into production

The Challenge

Self-Service analytics and BI is much more than just data visualization

An increasing number of companies have jumped onto the self-service bandwagon. As IT resources and technical talent have become more scarce, business people have taken advantage of new advances in commercial and open source tools to do much of the work themselves. The good news is that they can now do quite a bit of their own business intelligence reporting without the help of IT.

However, self-service tooling in the big data space has mainly focused on data visualization and interactive transformation of data as part of a data science discovery process. The goal has been to make it easy for data scientists and advanced business analysts to explore data to find a “recipe” that can then be used in a production process. The problem is that these tools don’t provide self-service capabilities to organize the data for efficient access. They also don’t enable production processes at scale.

Additionally, they don’t deal with production issues that occur as a result of constantly changing source data that can break production pipelines. Lastly, they don’t deal with 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 who can transform data science experiments into rock solid big data production workflows.

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.

Oil and Gas Exploration

Implemented a streaming ingestion, near-real-time, predictive analytics solution in 2 weeks

- Streams data from oil wells in near real-time
- Integrates with historical data
- Built a real time drilling dashboard with preventive maintenance analytics
- POC to production ready in 2 weeks

18x Improvement

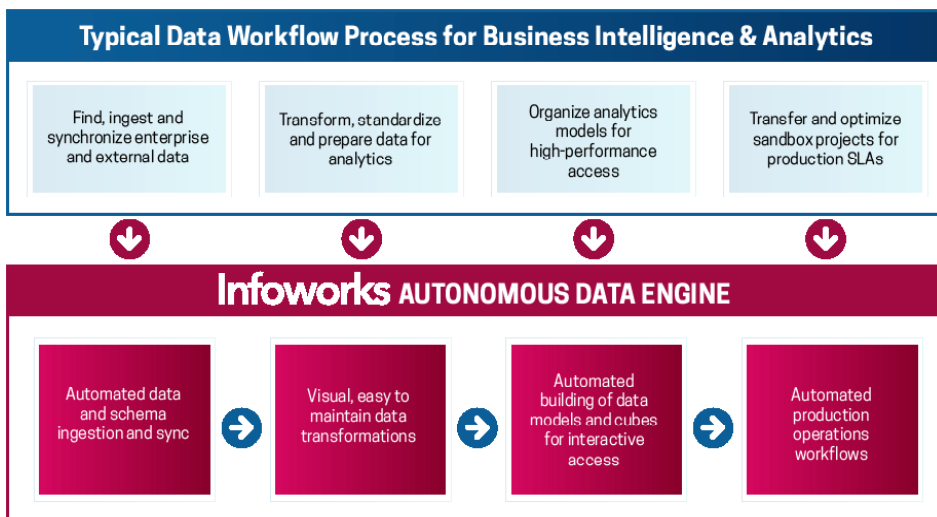
The Solution

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

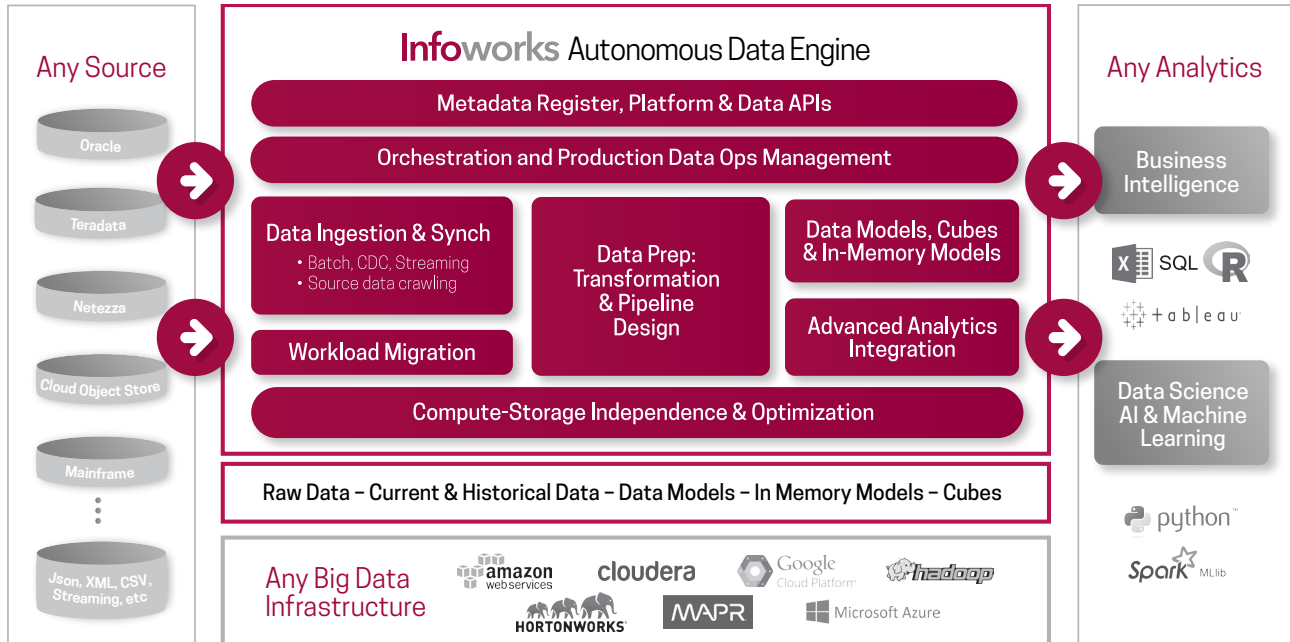
There are many great visualization and interactive transformation tools which are great for analytic discovery. They just can’t be used to build production workflows.

Infoworks addresses the challenge of building production big data workflows with an end-to-end solution that enables self-service from source to consumption. 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 applies an unprecedented level of automation to data workflows and data engineering to eliminate big data complexity. In addition, we provide an optimized operational environment to track & manage complex data pipelines in production.

Infoworks Automates the Data Workflow Process for BI and Analytics



Any Source, Any Big Data Platform, Any Analytics



End-to-End Solutions Capabilities

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.

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 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