Infoworks

AGILE DATA ENGINEERING
Eliminate Big Data Complexity Through Automation

Agile Creation and Management of Data Lakes

Agile Data Warehouse Offload and Migration

Agile Data Workflows for Business Intelligence & Analytics

The Challenge
Complexity is limiting enterprises' ability to capture the strategic value of their data

There is a consistent "buzz" today about how leading companies are harnessing big data for competitive advantage. Your organization is striving to become one of those market leading companies. However, the reality is that over 80% of big data projects fail to deploy to production because project implementation is a complex, resource intensive effort that takes months or even years. The technology is complicated, and the people who have the necessary skills are either extremely expensive or impossible to find. In addition, the market is rapidly evolving from Hadoop to Spark to server-less technologies with organizations moving from on premise implementations to cloud implementations, very often managing their data pipelines across multiple environments.

The Solution
Automation is the key to ensuring big data agility

The Infoworks Autonomous Data Engine (ADE) applies an unprecedented level of automation to data engineering and data operations. Infoworks ADE eliminates big data complexity and delivers agility for your data analytics projects. The Infoworks Autonomous Data Engine:

- Automates the development of complete data workflows from source to consumption
- Automates delivery of data to BI and advanced analytics applications
- Automates migration of data and workloads from legacy data warehouse systems to big data platforms
- Automates orchestration and management of complex data pipelines in production
- Provides independence from the underlying big data platform. Automates migration of data pipelines from on-premise to cloud and from cloud to cloud

The Infoworks Autonomous Data Engine is the most efficient and agile solution to implement enterprise big data use cases including:

- Agile Creation and Management of Data Lakes
- Agile Data Warehouse Offload and Migration
- Agile Data Workflows for BI and Analytics

.

Agility
- Deploy new use cases in days….not months
- Add new data sources with a single click

Flexibility
- Choose any analytics tools, methods, or algorithms
- Choose any big data platform, on premise or cloud and migrate from platform to platform without re-coding

Operating Efficiency
- Focus resources on generating business value
- Decrease demand on IT through self-service
- Improve reliability in production operation of analytics use cases

Cost Efficiency
- Reduce the cost and time required to build and maintain analytics use cases in production
- Reduce cost associated with legacy DW's

Automation delivers tremendous value to your business across a wide variety of factors

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 analytics use-cases in days instead of months with minimal resources. Some examples of these successes are:
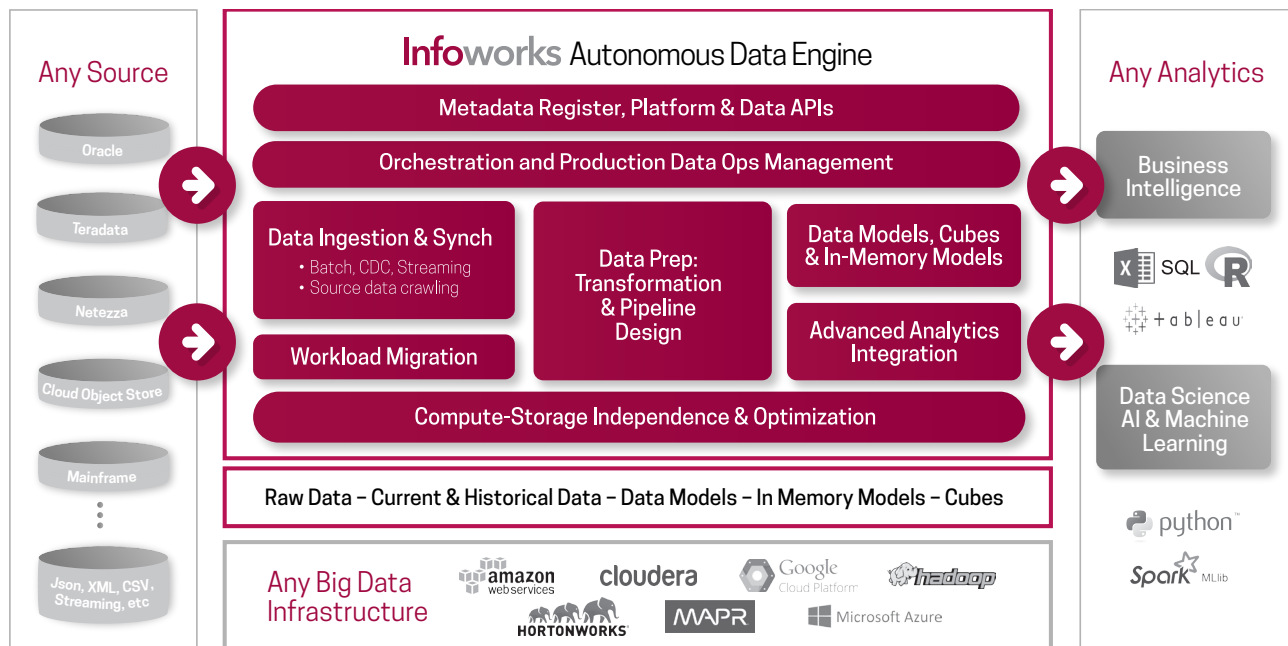
Fortune 10 Pharmacy: Advanced Data Application
Implemented near-real-time, machine learning claims processing data flow in 19 days
- Synchronized business process data from Teradata every 10 minutes
- Achieved the data load window SLA of 15 minutes
- Implemented by 2 engineers in 19 days from requirements to production

Oil & Gas Exploration: Self-Service BI and Analytics
Implemented a streaming ingestion, near-real-time, predictive analytics solution in 2 weeks
- Streams data from oil wells in near real time
- Integrates streaming and historical data
- Built a real-time drilling dashboard with preventive maintenance analytics

2

Any Source, Any Big Data Platform, Any Analytics

Any Source

- Oracle
- Teradata
- Netezza
- Cloud Object Store
- Mainframe
- Json, XML, CSV, Streaming, etc

Infoworks Autonomous Data Engine

- Metadata Register, Platform & Data APIs
- Orchestration and Production Data Ops Management

- Data Ingestion & Synch
  - Batch, CDC, Streaming
  - Source data crawling
- Workload Migration
- Data Prep: Transformation & Pipeline Design
- Data Models, Cubes & In-Memory Models
- Advanced Analytics Integration
- Compute-Storage Independence & Optimization

Raw Data – Current & Historical Data – Data Models – In Memory Models – Cubes

Any Big Data Infrastructure

amazon web services · cloudera · Google Cloud Platform · hadoop · HORTONWORKS · MAPR · Microsoft Azure

Any Analytics

- Business Intelligence
- SQL · R · tableau
- Data Science AI & Machine Learning
- python
- Spark MLlib

End-to-End Solutions Capabilities

Infoworks provides a complete solution that automates end-to-end data workflows from source to consumption as well as the on-going operational management of those workflows. Infoworks is platform independent and agnostic and delivers the following automated capabilities:

Workload Migration

Automates migration of workloads (ETL logic, BTEQ in Teradata, SQL workloads, etc.) from legacy data warehouses to a big data environment. With automated data, schema and workload migration, the Autonomous Data Engine accelerates data warehouse offload and migration and allows seemless migration across big data platforms, both on premise and in the cloud.

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.

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.

Data Models, In-Memory Models and Cubes

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 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