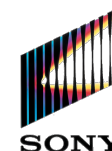


Overview

Automated Data Engineering for Rapid Analytics & Business Agility

Company Overview

- Founded in 2014
- HQ in Palo Alto, CA
- The only end-to-end automated software platform for Agile Data Engineering
- Proven in production at large enterprises



BNY MELLON

EAGLE
INVESTMENT SYSTEMS

DIAMOND
Resorts

Select list of current customers
(confidential)

Our Mission

To empower every company to harness data as effectively as Google, Facebook, and Amazon

Infoworks is v3.0 for our Founder, Amar Arsikere

Google



zynga



Infoworks

Our Solution

Infoworks ADE is a data engineering platform that enables large enterprises to:

- *Easily consolidate and utilize internal & external data sources*
- *Automate complex data engineering processes*
- *Launch unprecedented numbers of new analytic and data science use-cases without extensive teams of developers & data scientists*

Infoworks Provides the Agility Required for Successful Digital Transformation

The Infoworks Agile Data Engineering Platform

Agility is Key Requirement

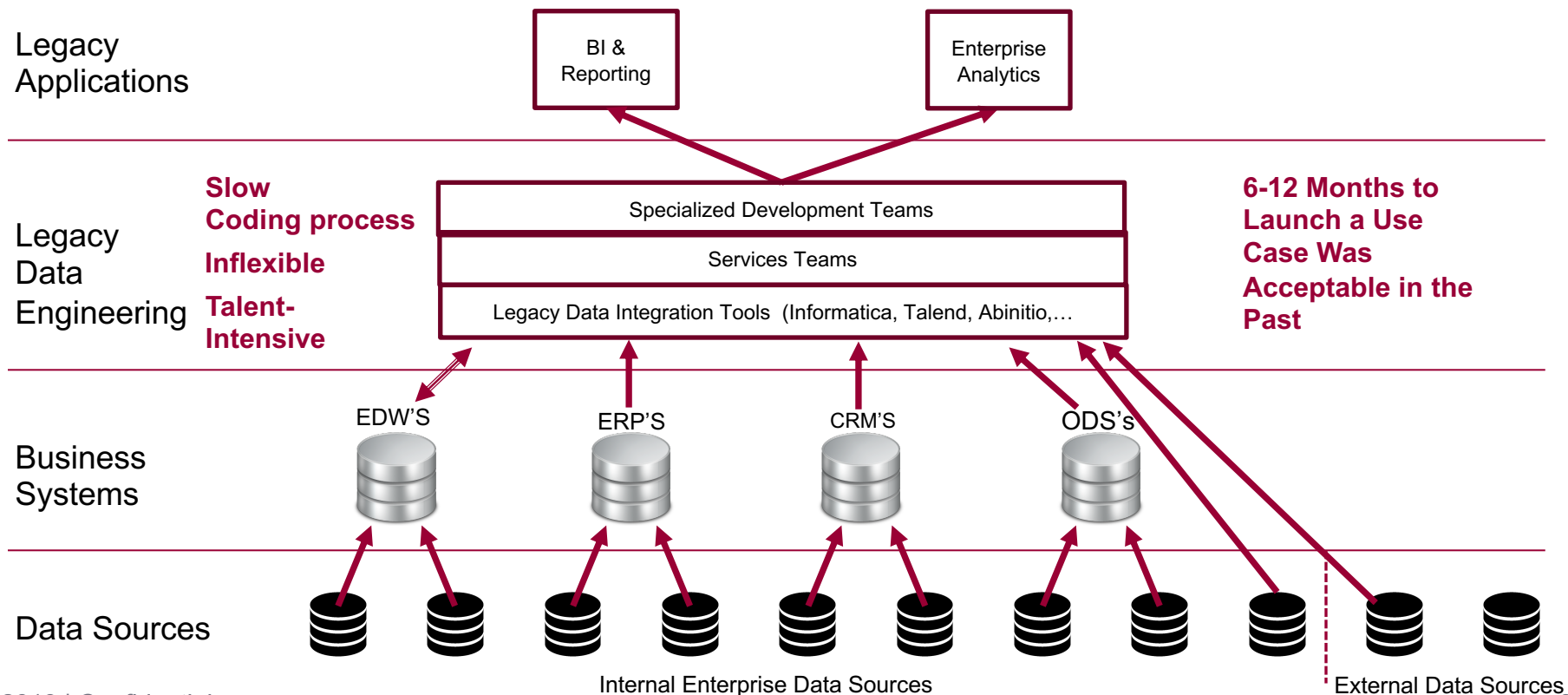
- Fast onboarding of new data sources
- Rapid onboarding of new analytics use cases
- Self service implementation of analytics use cases
- Minimal dependence on specialized talent
- Rapid adaptation to new business requirements
- End-to-end governance and lineage tracking
- Reusability of data and data artifacts
- Seamless extension to cloud, hybrid, and multi-cloud environments
- Resilience to easily adapt to changes in underlying technologies and infrastructure

Top Customer Use Cases

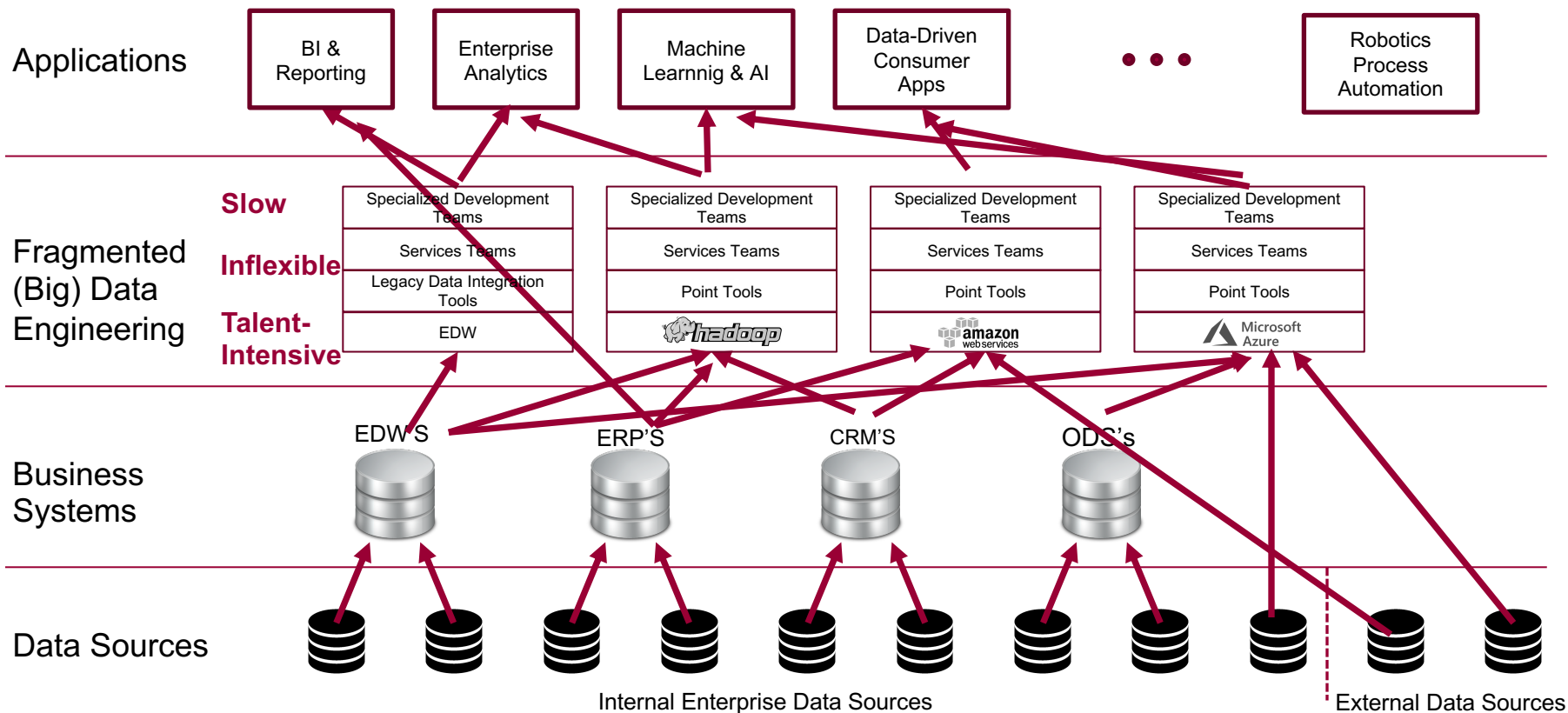
- Customer 360
- M&A Integration
- AI&ML / Data Science
- Operational and Financial 360
- EDW Migration
- Hybrid Cloud and Multi-Cloud Analytics

End-to-End Automation Enables 10-100x Improvements in Time-to-Value with 1/10th of the Resources

Legacy Processes Are Simply Not Agile Enough



Today's Fragmented Environments are not Agile and are Talent-Intensive



Data Engineering is the Challenge

“Death by 1000 Paper Cuts”

Data Ingestion

- Change Data capture
- Parallelization of data load
- Slowly changing dimensions
- Conversion of source types to big data types

Data Synchronization

- Data Merge
- Data Synch
- History table creation

Data Transformation

- Building initial load data pipelines
- Building CDC pipelines
- Building SCD pipelines
- Pipeline change management
- End to end lineage creation

Data Models

- Building OLAP cubes
- Building in-memory models



Data Governance

- Data access control
- Change management tracking
- Enabling compliance reporting

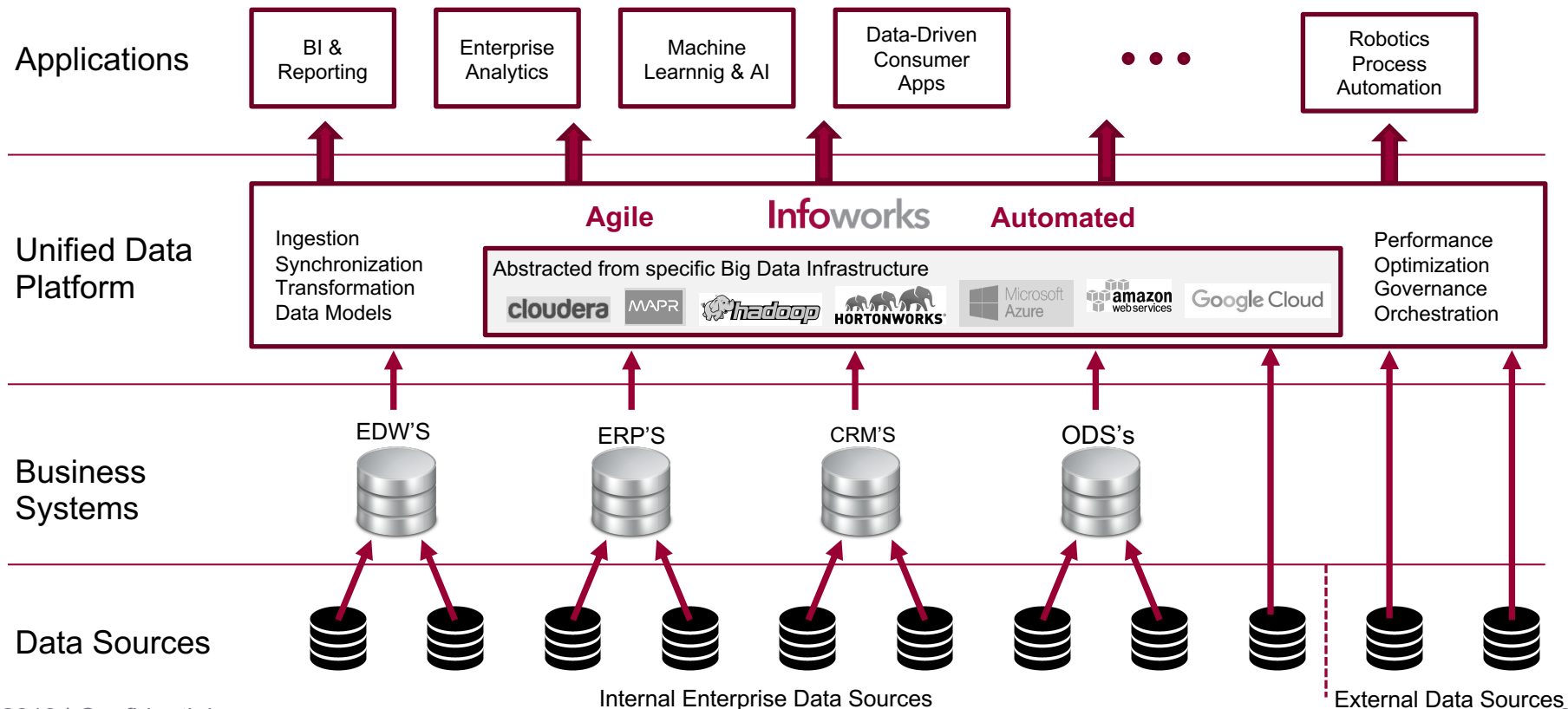
Performance Optimization

- Tuning of data load
- Tuning of data transformation
- Tuning of cube generation
- Tuning of in memory models

Production Orchestration

- Scaling jobs
- Migration from dev to production
- Operationalizing data science models
- Monitoring operational environment
- Identifying and restarting failed jobs

Achieving Any Data, at Anytime, for Any Application



Infoworks Use Case & Productivity Gain Examples



0 to Production-Ready in 14 Days

- **Use case:** Assets under management analytics
- Initial engagement started with a need to solve ingestion. Upgraded to end-to-end Infoworks platform deployment during the POC
- Customer implemented the complete use case using Infoworks in their production environment:
 - 1.2 Billion records base table
 - 54M incremental updates every day
 - 11 sources, 20 transformation pipelines, 1 orchestrator workflow
 - Customer was concerned about meeting business SLA of 9 hours. Infoworks actual performance: 5 hours
- The entire use case was implemented by the customer (1 engineer) **in 2 weeks** with no formal Infoworks training
- The use case is now in production

Without Infoworks	With Infoworks
~5 months 5 engs	10 days 1 engr

Infoworks Accelerates Time-to-Value

Recent Customer Use Cases

Advanced Analytics

(Fortune 10 Healthcare Co.)

Implemented a complex, machine learning, near-real-time, business process in 19 days

- Synchronized with Teradata every **10** mins
- **15 min** data-availability SLA
- Implemented by **2** engineers in **19** days from requirements to production

Without Infoworks	Infoworks
~6 months	19 days

9.5x Improvement

Data Lake Creation

(Fortune 100 Technology Co.)

Implemented Enterprise wide Data Lake involving 1500 data sources

- Synchronized data (CDC/Merge) from 1500 data sources
- **Serving reference data** for all Enterprise analytics
- Implemented by **2** engineers in < 2 months including a **Data shopping cart**

Without Infoworks	Infoworks
~2 years	60 days

10x Improvement

Cloud Portability

Migrated self-serve BI use case and dashboards from Azure to GCP in 1 day

- **7** data sources,
- **8** pipelines
- **8** optimized models
- **3** cubes
- **13** reports & dashboards
- Sub-second query response

Without Infoworks	Infoworks
~6 months	1 day

180x Improvement

Infoworks Accelerates Time to Value

Recent Customer Use Cases

EDW Offload (Leading Healthcare Co.)

Migrated Teradata workload - data, schema and logic in 26 days

- Migrated **5 control flows** in BTEQ to big data pipelines
- Migrated **70 Teradata SQLs** to big data pipelines
- Ingested and synchronized **300 tables, Production migration**

Without Infoworks	Infoworks
~9 months	26 days

9x Improvement

IoT Ingestion & Analytics (Oil & Gas Company)

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

- Stream data from oil wells. Integrate with historical data
- Build Real time drilling dashboard, Preventive maintenance analytics
- POC to production ready in 2 weeks

Without Infoworks	Infoworks
~6 months	10 days

18x Improvement

Self-Service BI (Leading Media Co.)

Complete implementation of self-serve BI use case for Spend analysis on Azure in 5 days

- **6** data sources – Oracle Financial data, SFTP data sources, General ledger
- **7** pipelines, **7** data models
- **5** cubes
- **5** reports & dashboards
- Sub-second query response

Without Infoworks	Infoworks
~6 months	5 days

35x Improvement

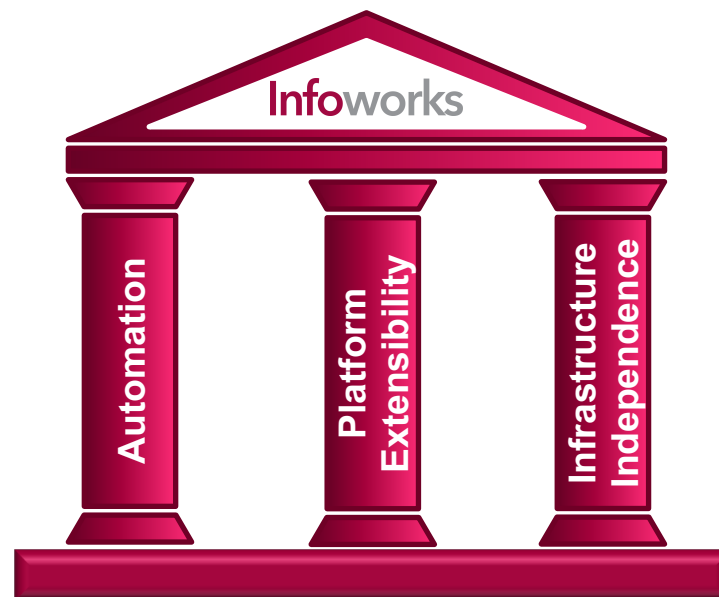
Infoworks Architecture



A Complete Agile Data Engineering Platform

- **Automation**
 - Complete code-free automation from data source to point of consumption by analytics
- **Platform Extensibility**
 - Supports customer or 3rd party applications
- **Infrastructure Independence**
 - Portable between and across environments on premise or in the cloud

Three Pillars of an Agile Data Engineering Platform



Any Source, Any Big Data Platform, Any Analytics

Infoworks

Any Source



Enterprise
Cloud
Bridge on
premise

Data
Crawlers

Data
Replicator

Cloud deployment: Azure, AWS or GCP

Infoworks Autonomous Data Engine

Platform, Metadata & Data APIs

Orchestration and Production Data Ops Management

Data Source
Crawling
(CDC, Merge)

Data
Ingestion
& Sync

Data
Transformation
& Pipeline
Design

Data Models,
Cubes &
In-Memory Models

Workload
Migration

Advanced Analytics
Integration

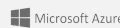
Compute/Storage Independence & Optimization

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

Any Big Data
Platform



cloudera



Any Analytics

Business
Intelligence



+ a b l e a u

Power BI

Data Science
AI & Machine
Learning

python™

Spark MLlib

Infoworks Automates Complex Tasks at Every Stage



ENTERPRISE IT

Configure New
Data Sources &
Authorize
Access

Provision &
Manage
Platform
Infrastructure

Infoworks Autonomous Data Engine

DATA INGESTION & SYNC

Automatic
Ingestion / CDC

Automatic Data
Type Conversion

Auto Crawling

Automatic
Schema Change

Automatic Merge

DATA TRANSFORMATION

Automated
Incremental
Pipelines

Automated Data
Validation

Automated
Dependency
Management

Suggest
New Data
Connections

HI-PERF MODELS

Automatically
optimize data
models

Auto create
OLAP cubes

Automatically
maintain
time axis

Automated
metadata lineage
to source

PRODUCTION OPERATIONS

Automated Fault
tolerance

Restartability

Monitor/
Debug

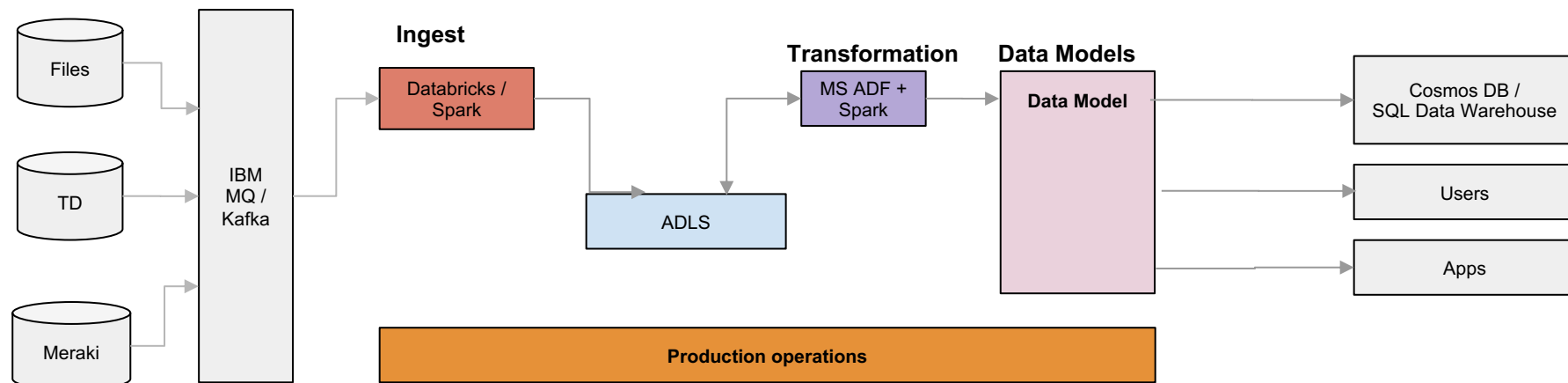


BUSINESS ANALYSTS

Define and
Implement
Analytics

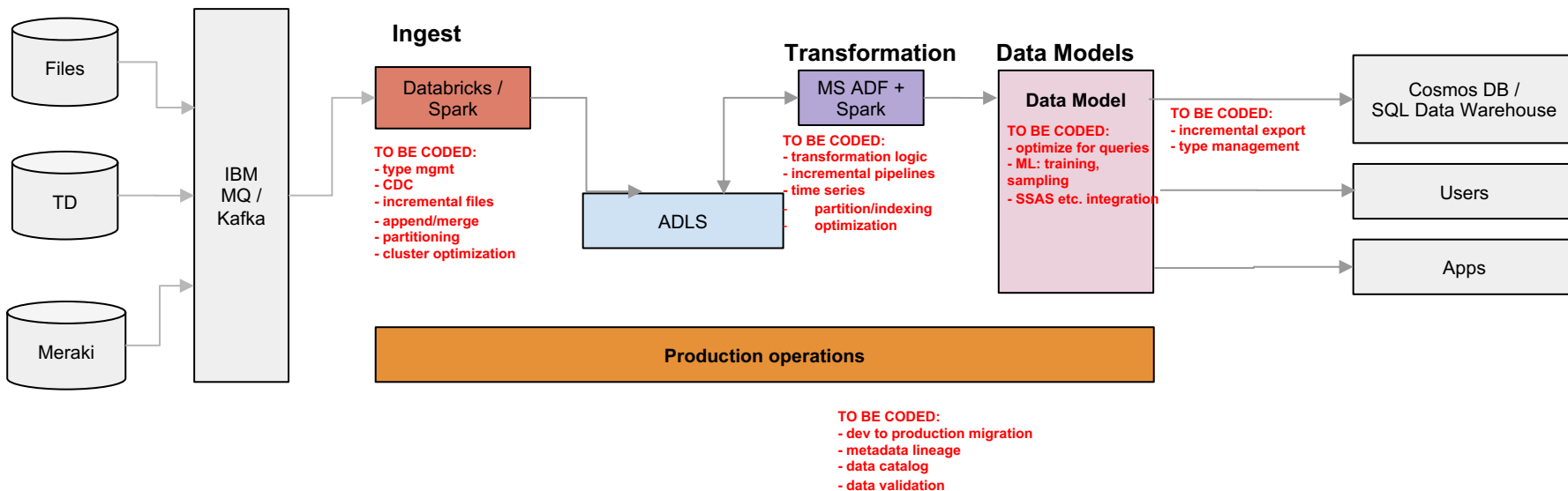
- Eliminates the need for specialized talent and consultants
- Enables new use cases to be launched 10x faster with fewer resources

Example Data Engineering Implementation in Azure



Looks great!... except

....this requires a lot of coding

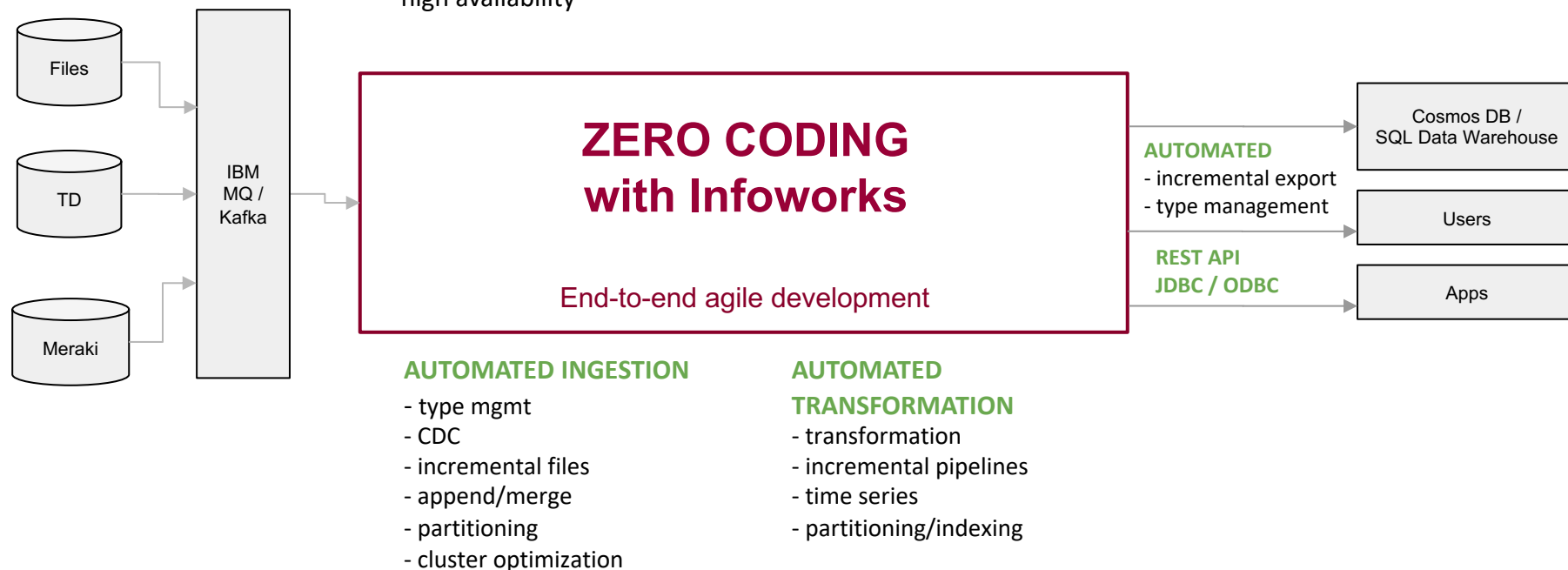


Time consuming, resource intensive, expensive, it is **not agile**

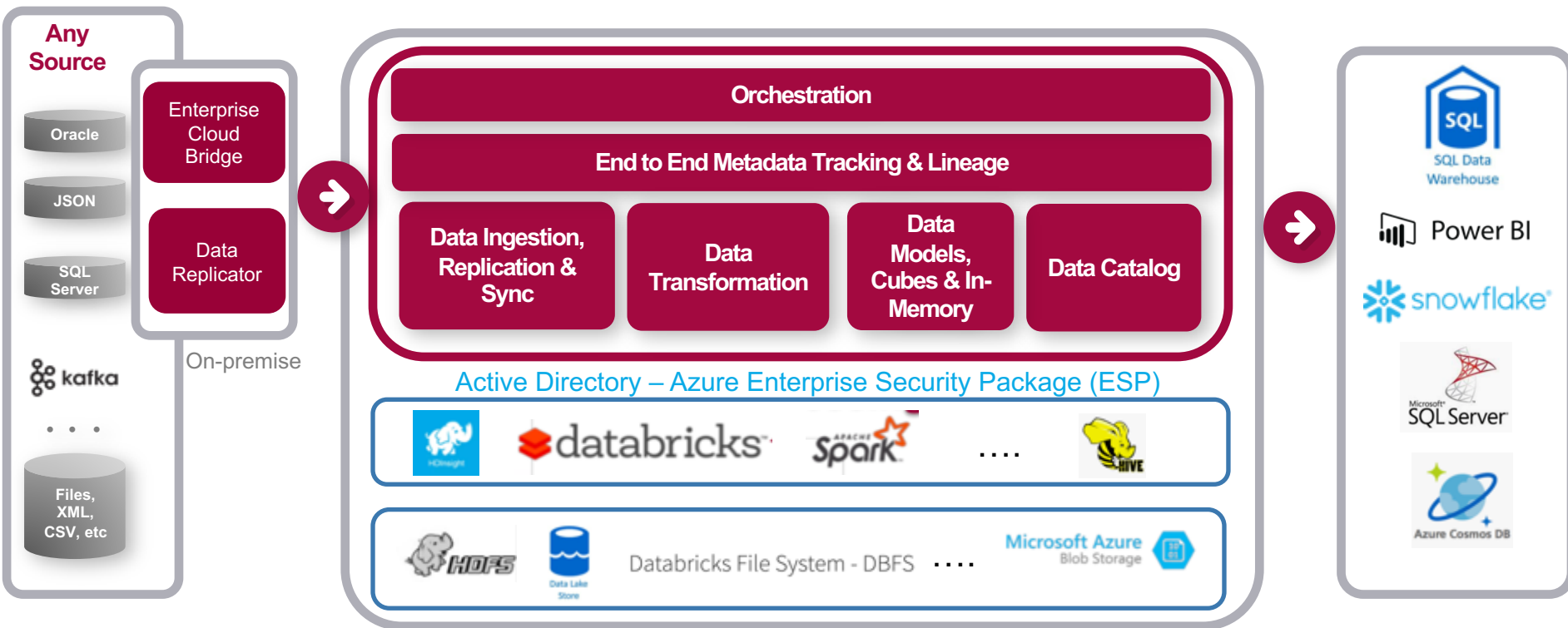
Infoworks Delivers Agility Through Automation

AUTOMATED END TO END PLATFORM

- dev to production migration
- metadata lineage
- data catalog
- data validation
- high availability
- self-service
- collaborative interface
- Not locked in



Infoworks & Azure Architecture



Infoworks accelerates the customer journey on Azure by automating data engineering

Thank You!

