Wednesday, 29th April 2020

Accelerate AI/ML and Data Science Projects with AI-Powered Modern Data Prep

Informatica

Today's Speakers

Preetam Kumar

Product Marketing Manager
Enterprise Data Preparation

Paul Siddall

Next Gen Analytics Product Specialist

© Informatica. Proprietary and Confidential.

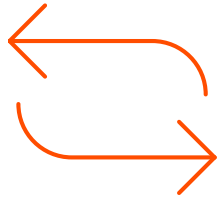
Informatica

Cloud Is the Key Enabler of the New Digital Era

83%

of enterprise **workloads** will be in the **cloud** by 2020

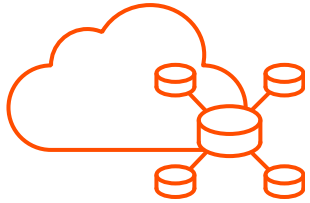
Source: Forbes, Logic Monitor Cloud Adoption



83%

of organizations expect **workloads** to **move freely** between **clouds**

Source: Turbonomic's 2019 State of Multicloud



75%

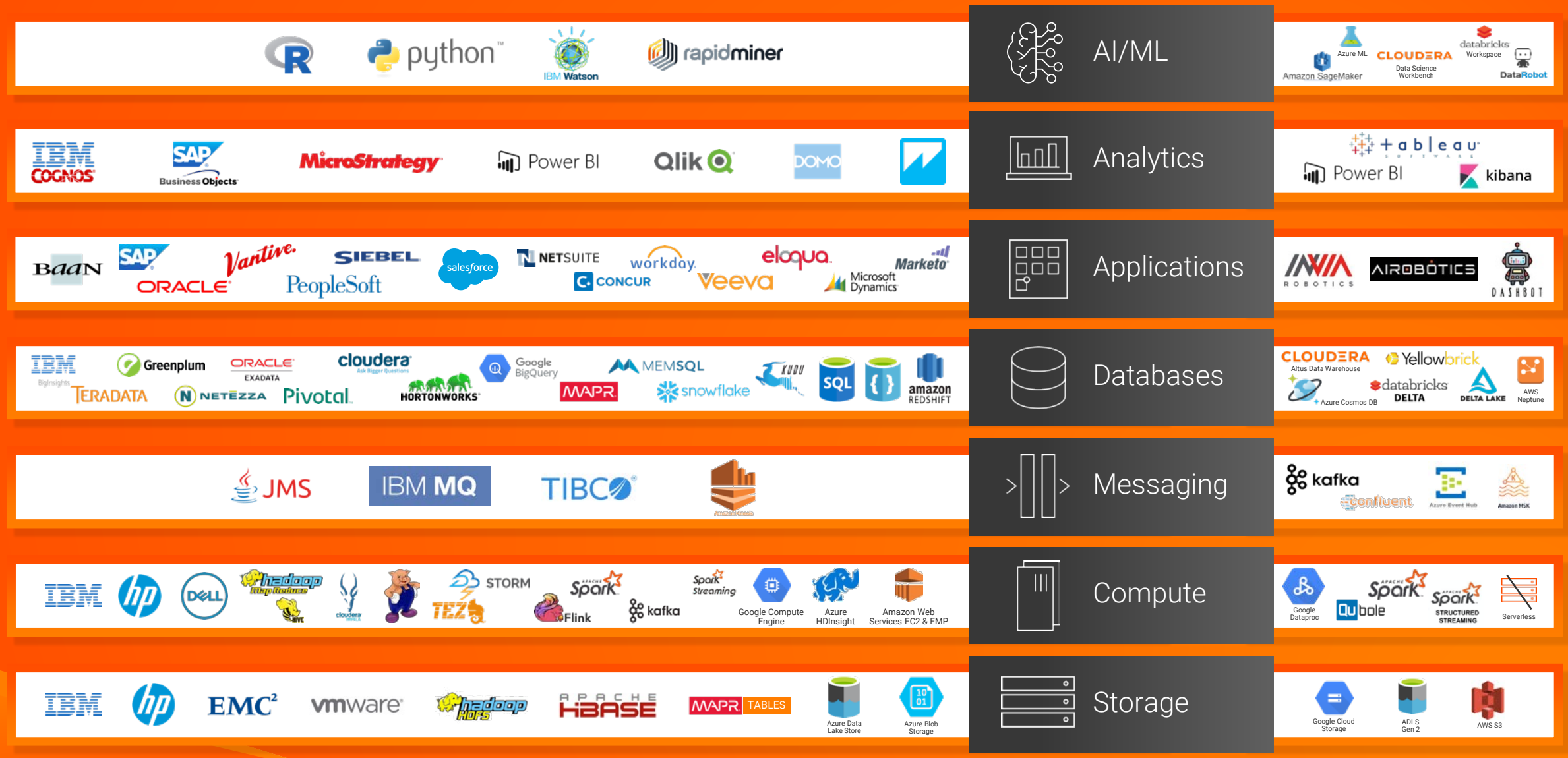
of all **databases** will be deployed or migrated to a **cloud** platform by 2022, with only 5% ever considered for repatriation to on-premises

Source: Gartner, The Future of the DBMS Market Is Cloud by Adam Ronthal, Donald Feinberg, Merv Adrian | June 23, 2019

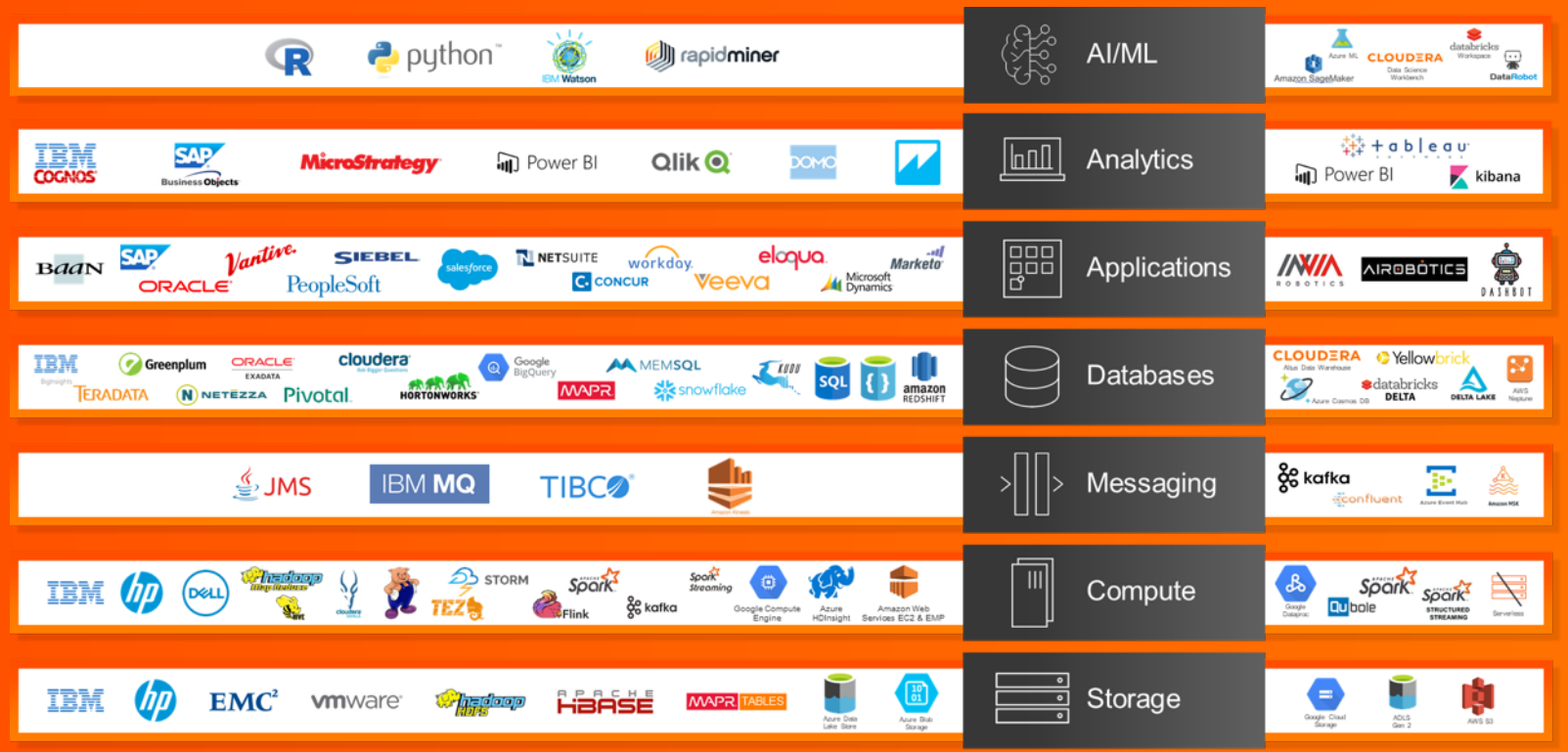
© Informatica. Proprietary and Confidential.

Informatica

Technology Shift



The Bold New World of AI and Analytics



AI & MACHINE LEARNING

ANALYTICS AND APPS

CLOUD DATA WAREHOUSE AND DATA LAKES

Data Science Projects Rarely Make It into Production

80%

of data science labs that have failed to establish measurable operationalization practices will be outsourced by 2021

Source: Follow 4 Data Science Best Practices to Achieve Project Success; Afraz Jaffri, Alexander Linden, Carlie Idoine, Jim Hare, Erick Brethenoux, 6 January 2020, Gartner

Main barriers include:

- Difficulty deploying into business processes and applications
- Lack of DevOps skills and right tools
- Unable to secure or govern data and analytics
- Poor planning and unreasonable expectations
- Data quality and integrity challenges

Innovations of "high" benefit, such as autoML, MLOps and edge analytics, will have a significant impact on data science programs during the next 2-5 years.

Source: Hype Cycle for Data Science and Machine Learning, 2019
Shubhangi Vashisth, Alexander Linden, Jim Hare, Peter Krensky
6 August 2019, Gartner

© Informatica. Proprietary and Confidential.

Informatica

Data Quality Problems Keep Users Awake at Night

60%

of organizations **challenged** by data quality

Source: IDC

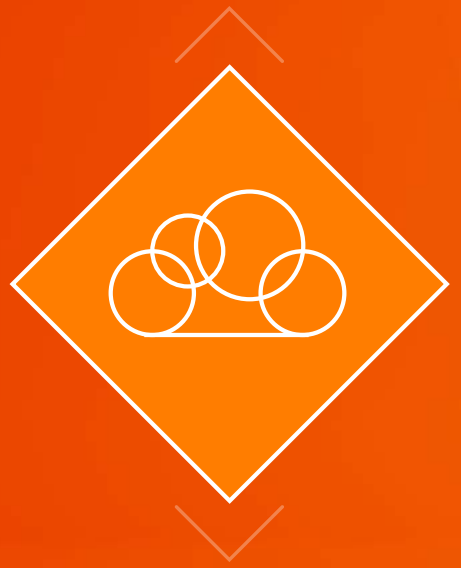
While some data quality issues are unfathomably complex, many yield quickly and produce outsize gains. Eliminating a single root cause can prevent thousands of future errors, save millions, and make things better for all involved.

Redman, Salmon and Nagle – Harvard Business Review

© Informatica. Proprietary and Confidential.

Informatica

Trends Driving Demand
for Enterprise Data Preparation at Scale

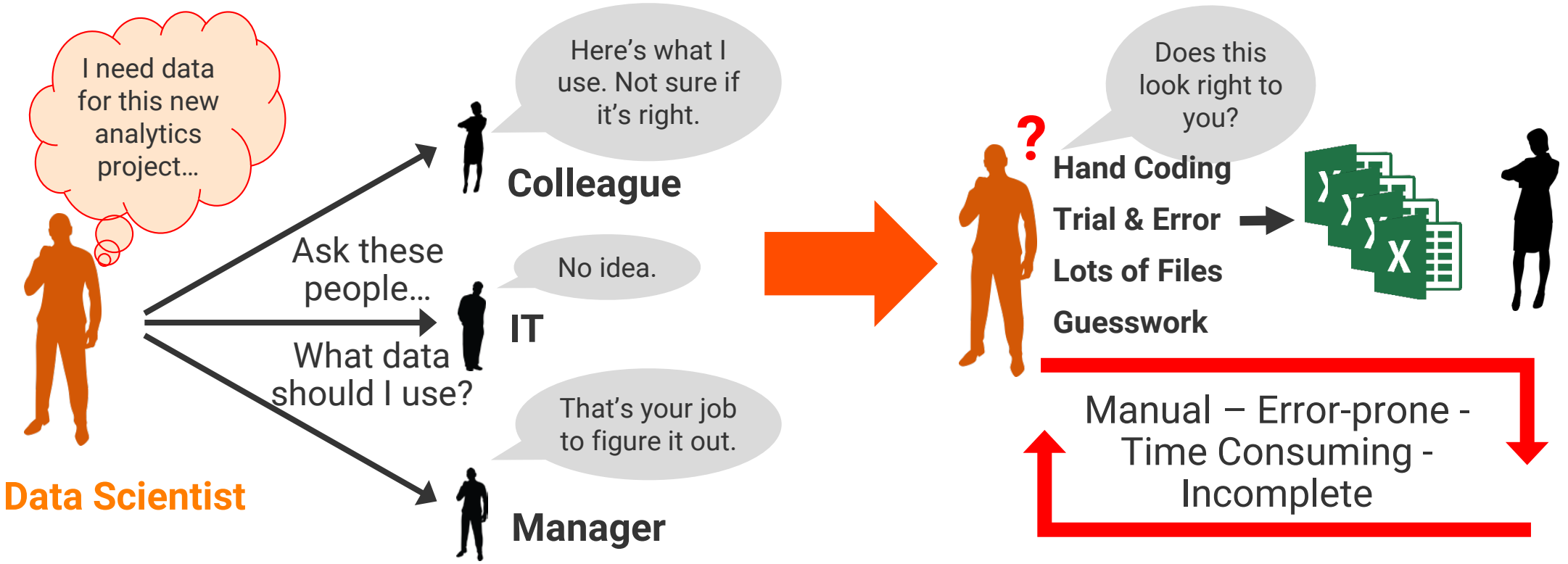


Hybrid and Multi-Cloud Data Lakes

Analytics, AI/Machine Learning Workloads

DataOps and MLOps

A Day in the Life of a Data Scientist



Struggle to Find Data

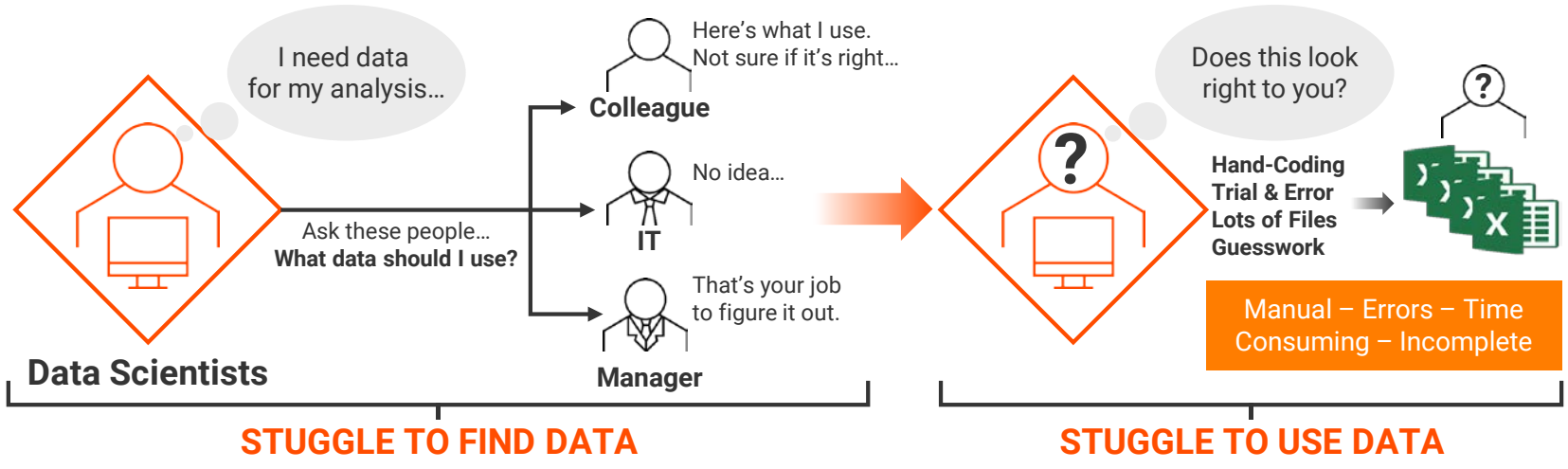
Struggle to Prepare Data

© Informatica. Proprietary and Confidential.

Informatica

Now Multiply that Across All Data Scientists….

Enterprise Data Preparation (EDP) Flips the 80/20 Rule



80% of time spent on data prep

20% on analytics

Months / Weeks

VS

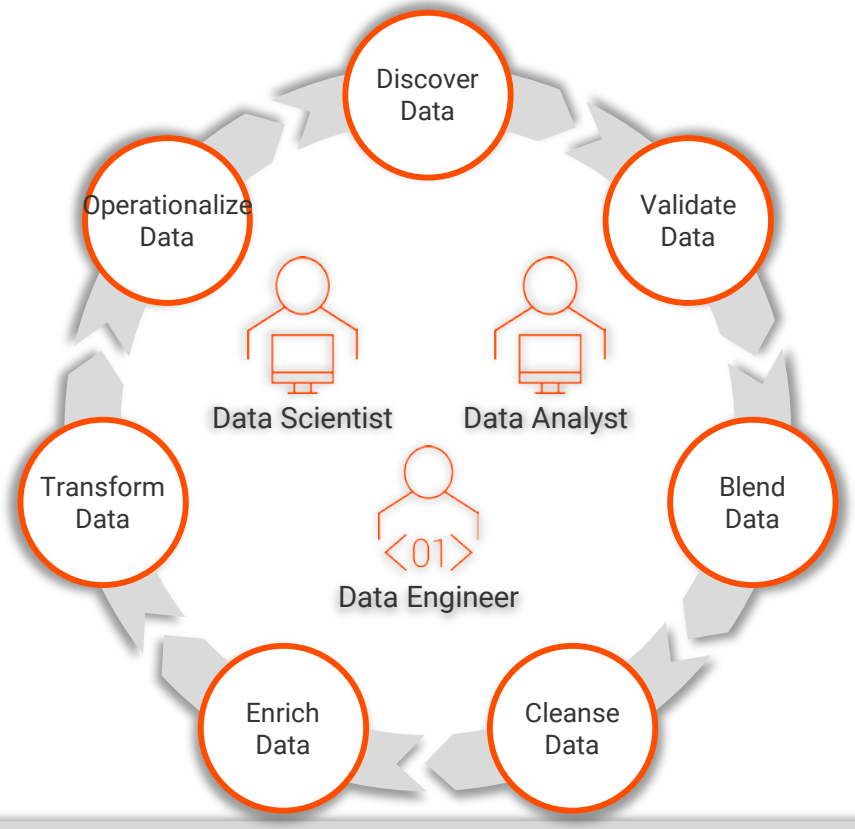
Hours / Minutes

20% of time spent on data prep

80% on analytics

Modern Enterprise Data Preparation

Agile, Iterative and Collaborative Process



Data Ingestion & Profiling | Data Catalog & Metadata Management | Data Transformation | Data Security | Data Quality & Governance | Data Enrichment | Data Modeling

AI-powered end-to-end data preparation for hybrid and multi-cloud data lakes

© Informatica. Proprietary and Confidential.

Informatica

Enterprise Data Preparation Empowers DataOps Teams

Democratize data pipelines for analytics and AI/ML workloads at scale

Data Engineer

How can I quickly prepare data for advanced analytics and operationalize data pipelines for reliable data delivery?

Data Analyst

How can I quickly and easily find, explore and prepare data for my analysis and create a repeatable process without depending on IT?

Data Preparation Process

Agile, Iterative, Collaborative

Data Scientist

How can I extract completely new patterns and models that detect or predict behavior?

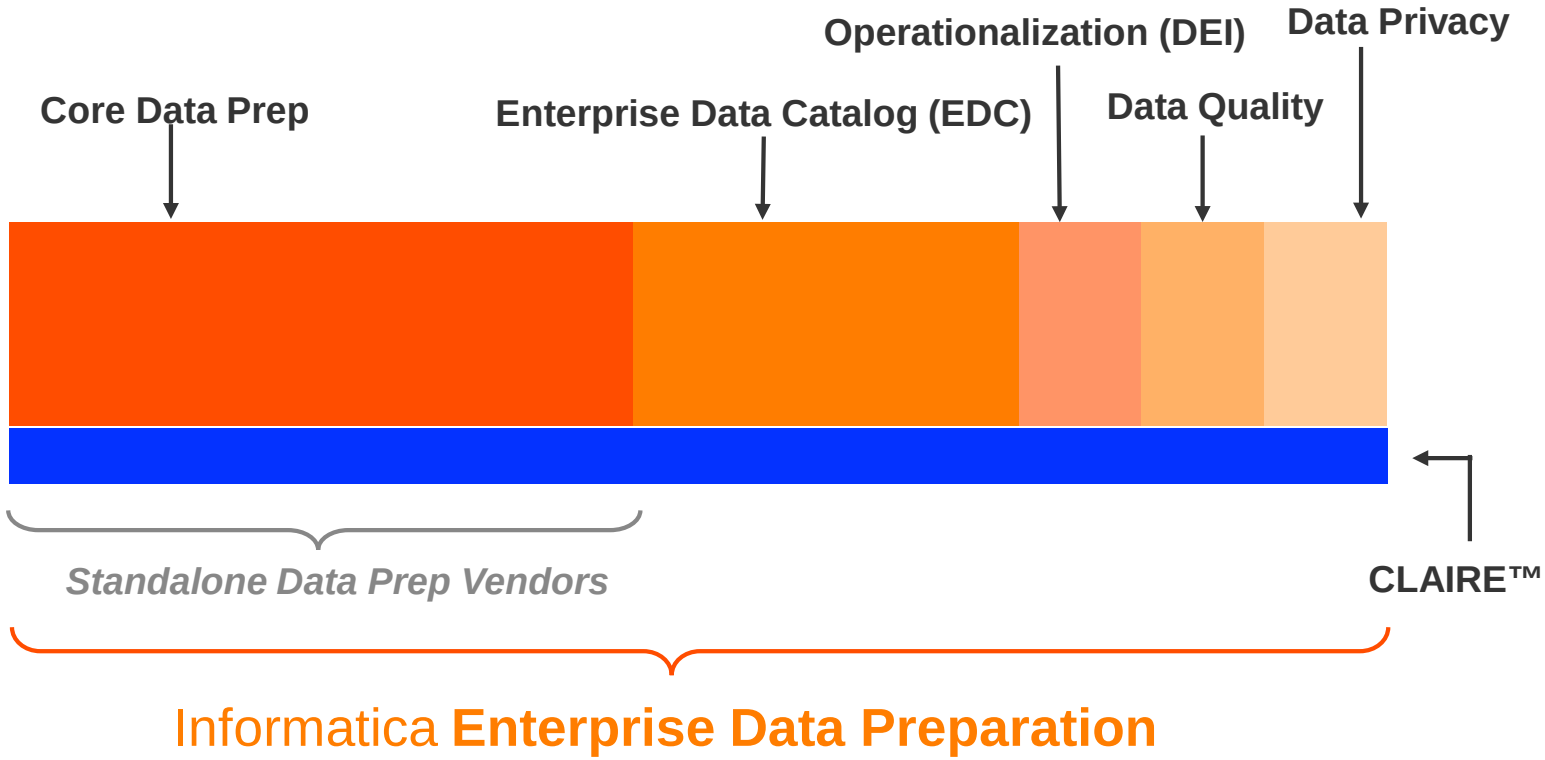
Data Steward

How can I enable self-service and ensure right data is available to right people in trusted and secure manner?

© Informatica. Proprietary and Confidential.

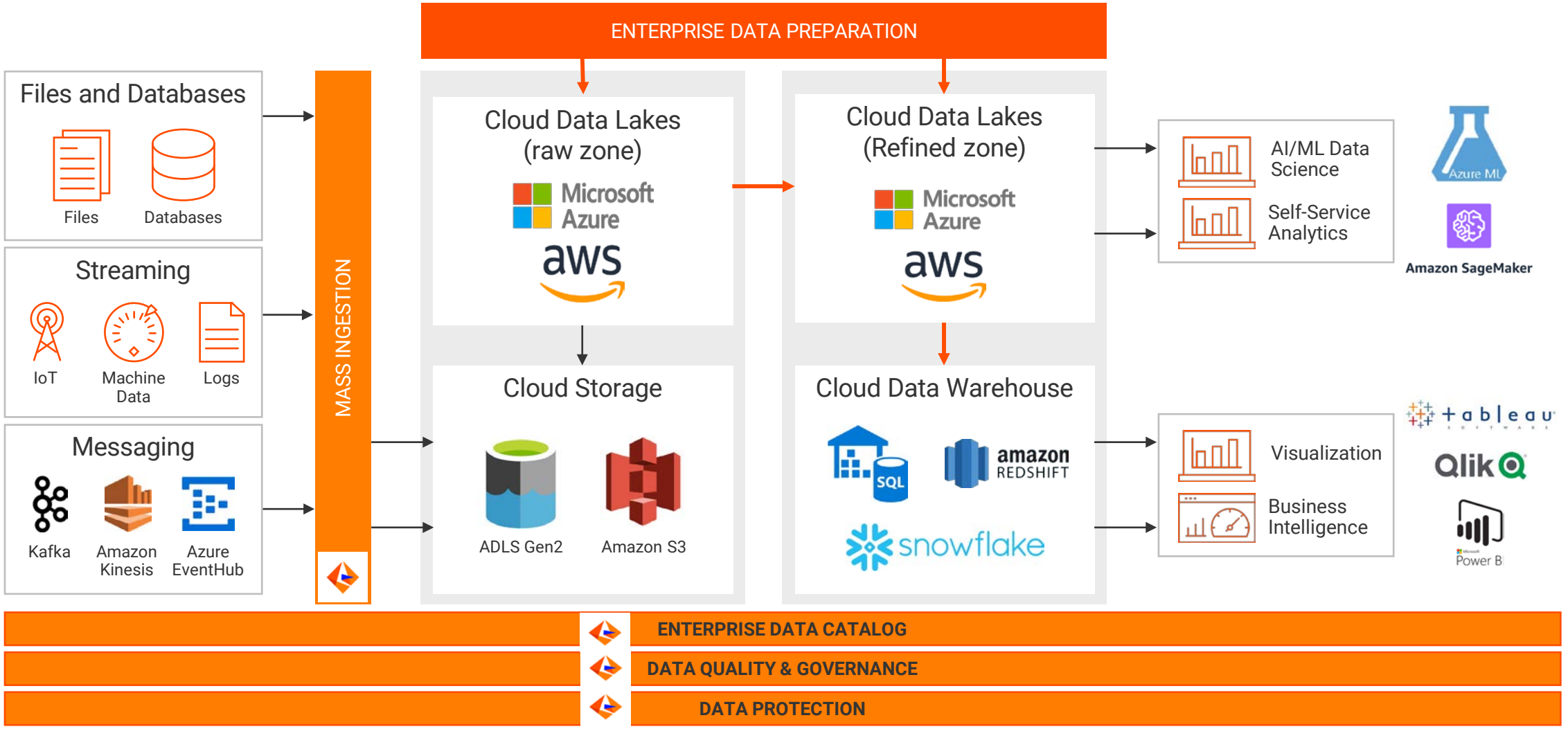
Informatica

Informatica Enterprise Data Preparation



© Informatica. Proprietary and Confidential.

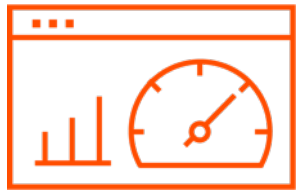
Enterprise Data Preparation Reference Architecture



© Informatica. Proprietary and Confidential.

Informatica

Enterprise Data Preparation Primary Use cases



Data Preparation for Self-service Analytics on Data Lakes

Find, explore and prepare data quickly in the data lake to support ad-hoc analytics followed by operationalization in a governed environment

Data Preparation for Advanced Analytics (AI/ML Projects)

Find, explore and prepare data that quickly so that it is ready for Data Science projects followed by operationalization of Data Science process

© Informatica. Proprietary and Confidential.

Informatica

Customer Case Studies

Accelerate Analytics, AI/ML Workloads Across Verticals

Democratize data and insights at enterprise scale

Financial Services	Healthcare	Insurance	Retail	Manufacturing	Government	Transportation
AML	Precision Medicine	Customer 360	Customer 360	Supply Chain Optimization	Cybersecurity	Customer 360
Fraud	Clinical Trials	Risk Management	Contextual Experiences	Asset Management	Citizen Experiences	Fleet Optimization
Risk Management	Drug Discovery	Underwriting	Order Management	Operational Intelligence	Case Management	IoT

© Informatica. Proprietary and Confidential.

Informatica

AXA XL, insurance company

Global commercial insurance and re-insurance

Business Drivers

- Build a Data Ecosystem Enterprise Platform (DEEP) on MSFT Azure Data Lake
- Democratize data discovery and extend use of EDP to data engineers, actuaries and data analysts

Solution

- Multiple use cases for data prep incl. Customer 360, Risk management and Underwriting
- Faster discovery of curated datasets and prepping of fully-governed data pipelines
- Able to perform robust analyses and reduce time to insights

"Informatica helps us **establish a single version of the truth** by bringing data into one governed repository, cataloging it, and making self-service data prep available."

John Mulvaney
Senior CoE Lead
Data Analytics Workbench

© Informatica. Proprietary and Confidential.

Avis Budget Group

Business Drivers

- Analyze terabytes of connected vehicle data from a fleet of 650,000 vehicles to drive innovation, increase efficiencies and improve customer experiences
- Support global vehicle analytics with end-to-end data pipelines and speed time-to-insights

Solution

- Able to prep and analyze data at scale for a plethora of use cases including customer 360, fleet management, mileage optimization and preventative repairs
- Able to increase productivity with rapid data discovery, understanding data assets and lineage to create fully-governed data pipelines at scale

"Informatica lets us use real-time data to optimize fleet management and telematics so that we can save money and drive our bottom line.

***Informatica Enterprise Data Preparation gives us easy, visual access** into the data in our data lake."*

Christopher Cerruto
VP of Global Enterprise Architecture and Analytics

AVIS Budget

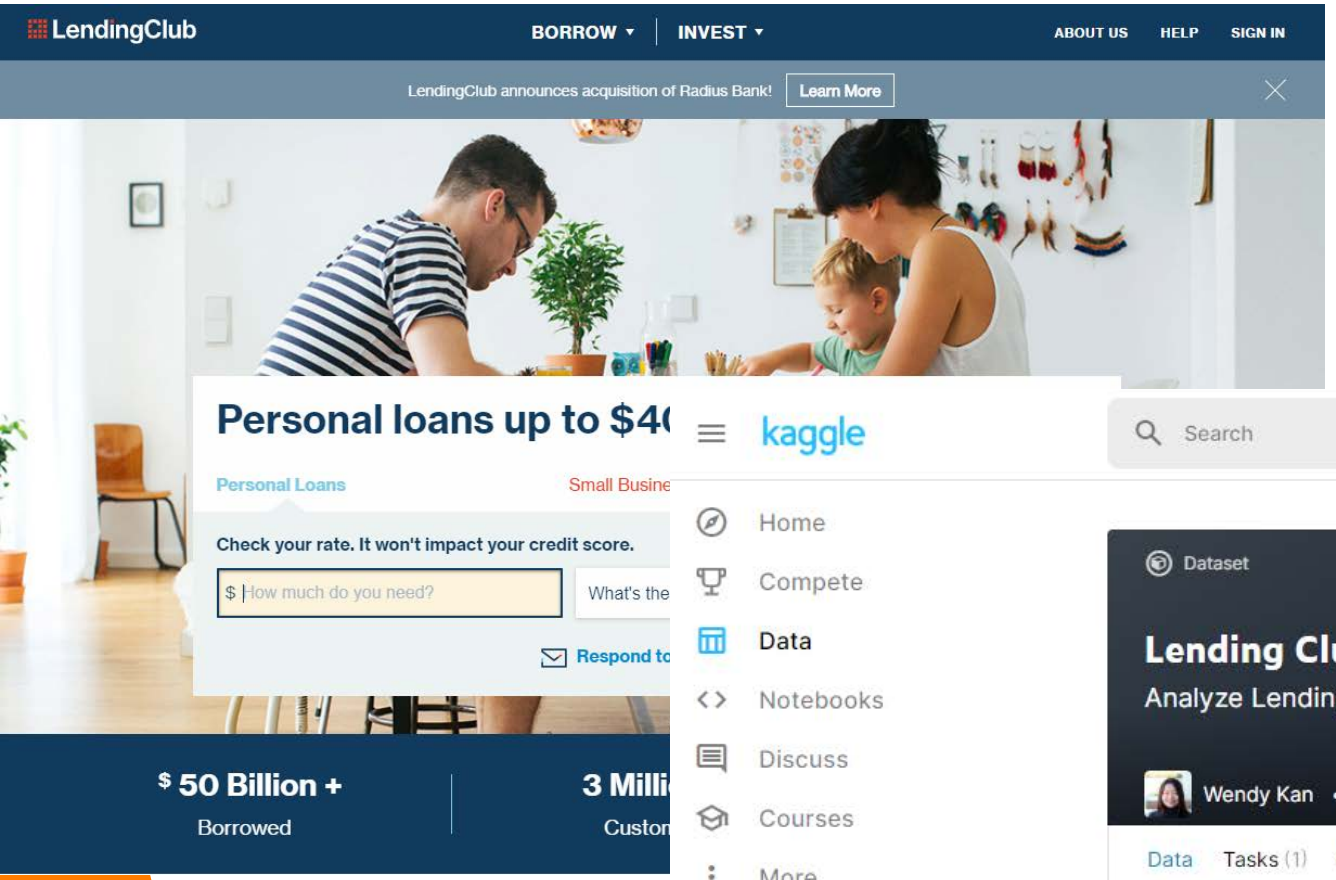
© Informatica. Proprietary and Confidential.

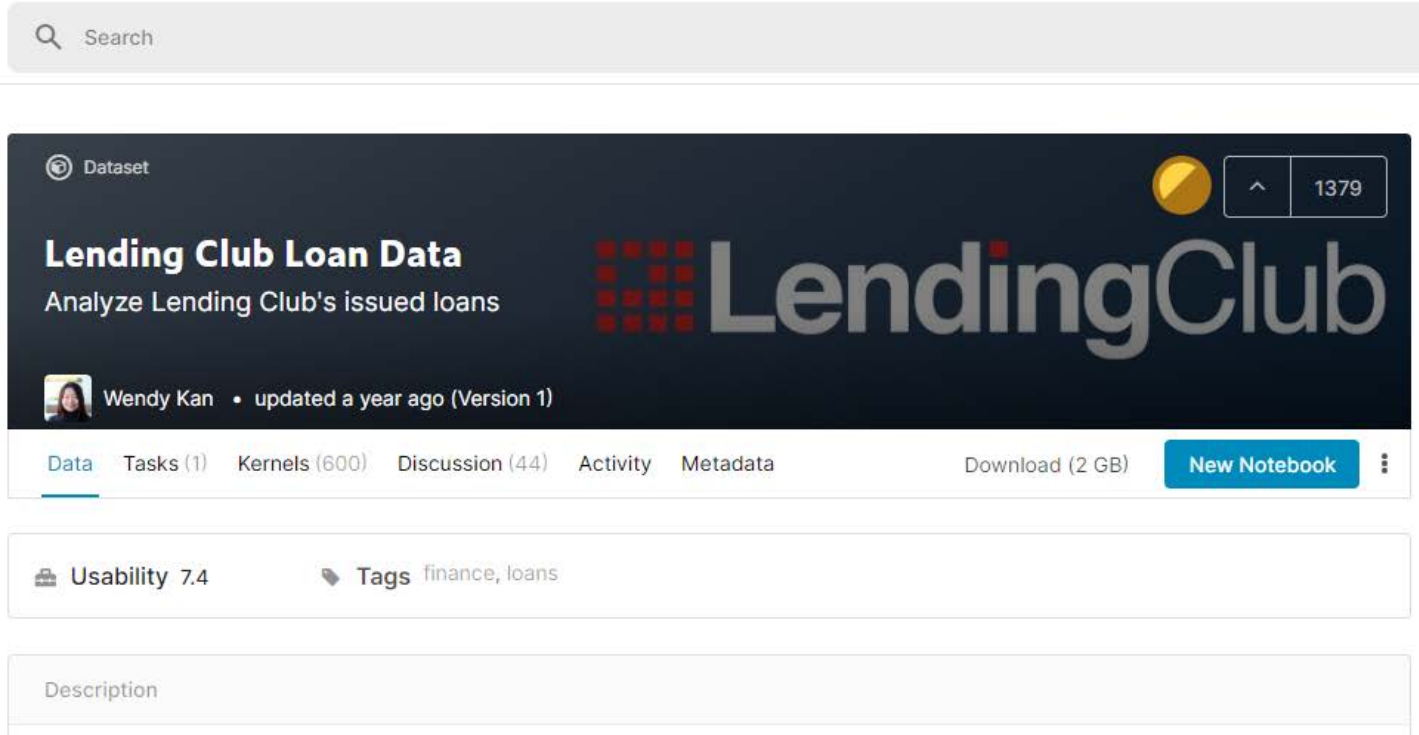
Informatica

Democratize Data Science

Informatica™

Data





© Informatica. Proprietary and Con

Use case

- Predict default
- Less defaults -> less money lost
- Higher interest rates = more money

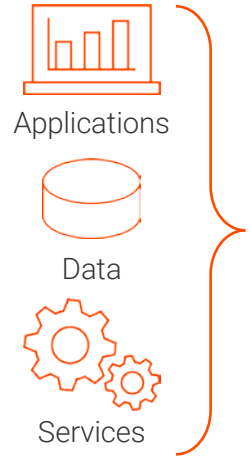
© Informatica. Proprietary and Confidential.

ML stages

- Find the data
- Understand the data
- Clean it
- Enrich
- Feature engineer
- Model select
- Hyperparameter tune
- Deploy

Subject Matter Expert

Auto ML



Informatica
Data Catalog

Informatica
Informatica Prep

Data Discovery, Data Integration, Data Quality

© Informatica. Proprietary and Confidential.

Informatica

AutoML Benefits

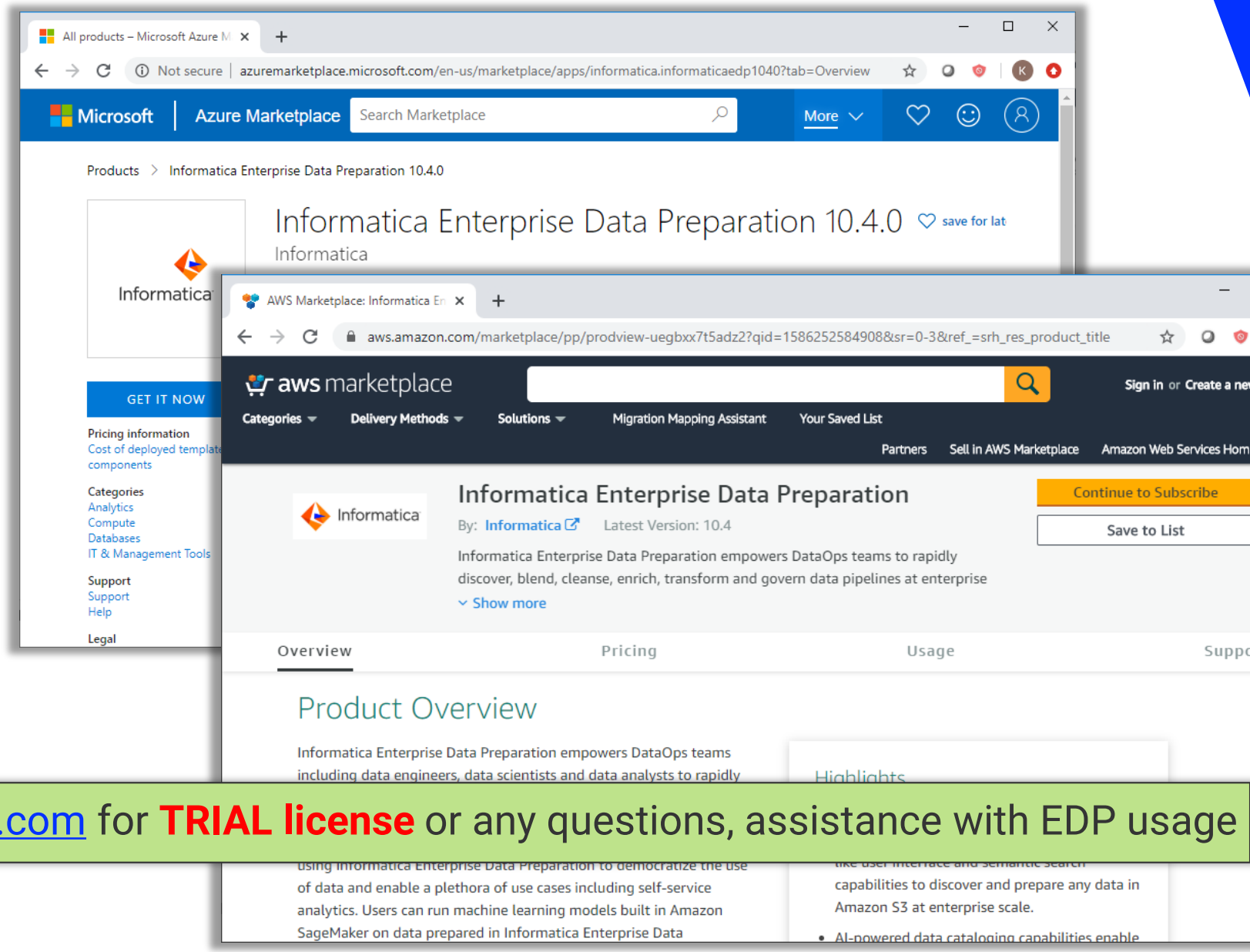
- Helps engineer features
- Picks an algorithm, hyperparameters and generates the model ready for deployment
- Ephemeral compute
- You can simply download the model and script
- Domain experts can rapidly build and deploy ML solutions

© Informatica. Proprietary and Confidential.

Demo

Try EDP 10.4 on Azure/AWS Marketplaces

- **Free TRIAL**
- EDP installed and configured with all pre-requisites
- Setup a Trial or POC or a PROD environment in a couple of hours
- Start preparing on your **ADLS or S3 data lakes** without any additional configuration
- Environment comes with sample dataset to get trained on EDP



Contact edpassist@informatica.com for **TRIAL license** or any questions, assistance with EDP usage

© Informatica. Proprietary and Confidential.

Learn More

- Watch a demo: Build ML Models with **DataRobot**, Informatica Enterprise Data Preparation and Databricks
- Learn about customer use cases: AXA XL, Avis Group
- Download the datasheet: Informatica Enterprise Data Preparation
- Solution Brief: Find and Prepare Any Data for Self-Service Analytics
- Read the Analyst Report: Constellation ShortList™ Self-Service Data Preparation
- Gartner Peer Insights Reviews

© Informatica. Proprietary and Confidential.

Thank you!