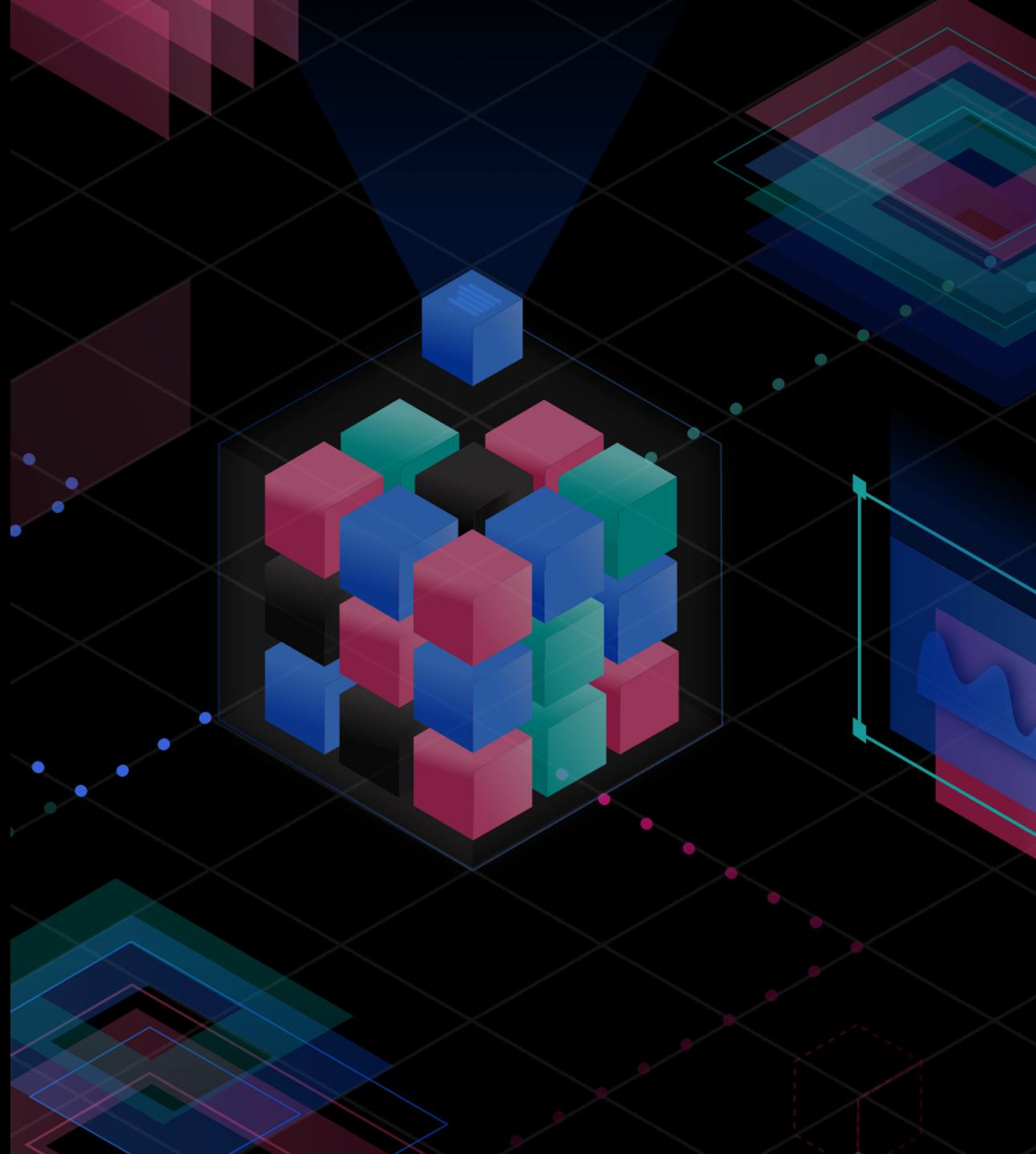


IBM Cloud Pak for Data

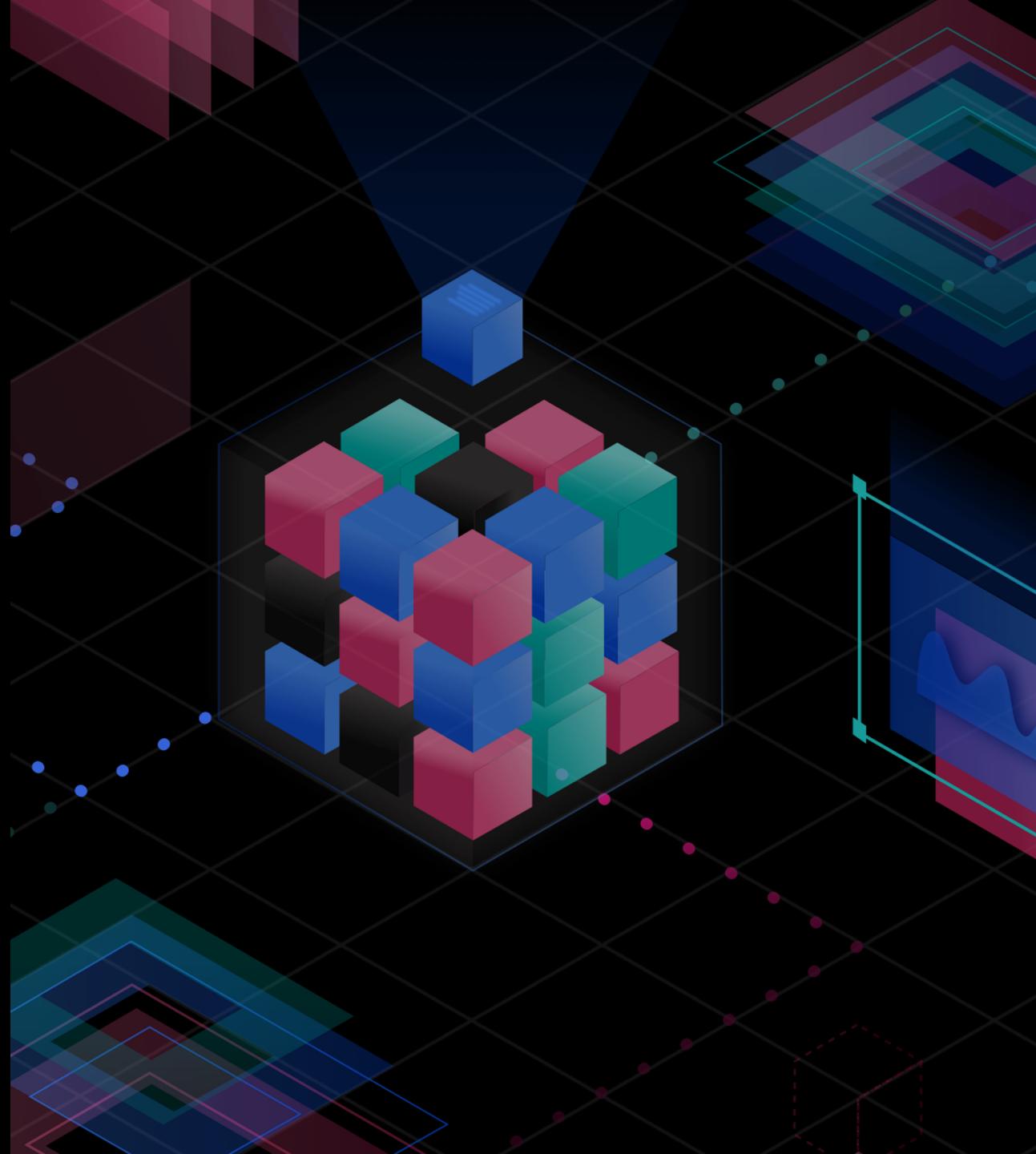
Any Data. Any Cloud. Anywhere.

Rachana Vishwanathula
Data Scientist
Hybrid Cloud Build Team



Agenda

- Becoming a data-driven organization
- The ladder to AI
- What is Cloud Pak for Data?
- Deployment patterns
- Use Cases





You're a data
company now

What it means to deliver a data-centric customer experience

"A critical component of continuous customer experience includes evaluating where AI resources can extend or support your customer-facing associates, like chat bots, virtual customer assistants, or real-time analytics at key points in the customer journey."

Gartner

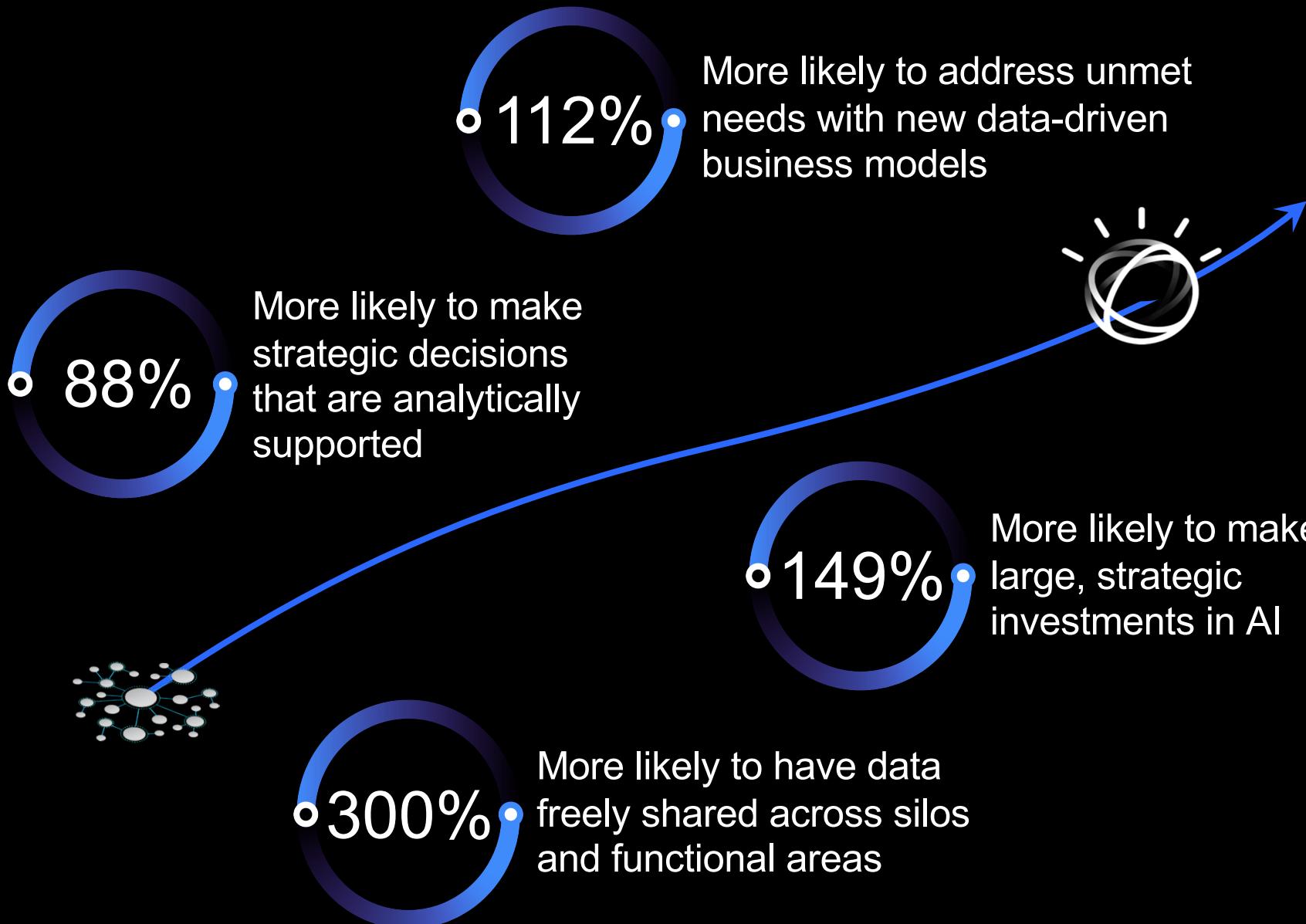
Smart data will take priority over big data... for too long, the industry relied on sheer quantity of data with no quality metrics for making key audience assumptions."

TC

"If you fail to keep up with your clients' increasing demands, the chances they will reach out to your competitors increases."

Forbes

The best performers are data-driven



Data-driven cultures
put data to work
across complex
digital ecosystems

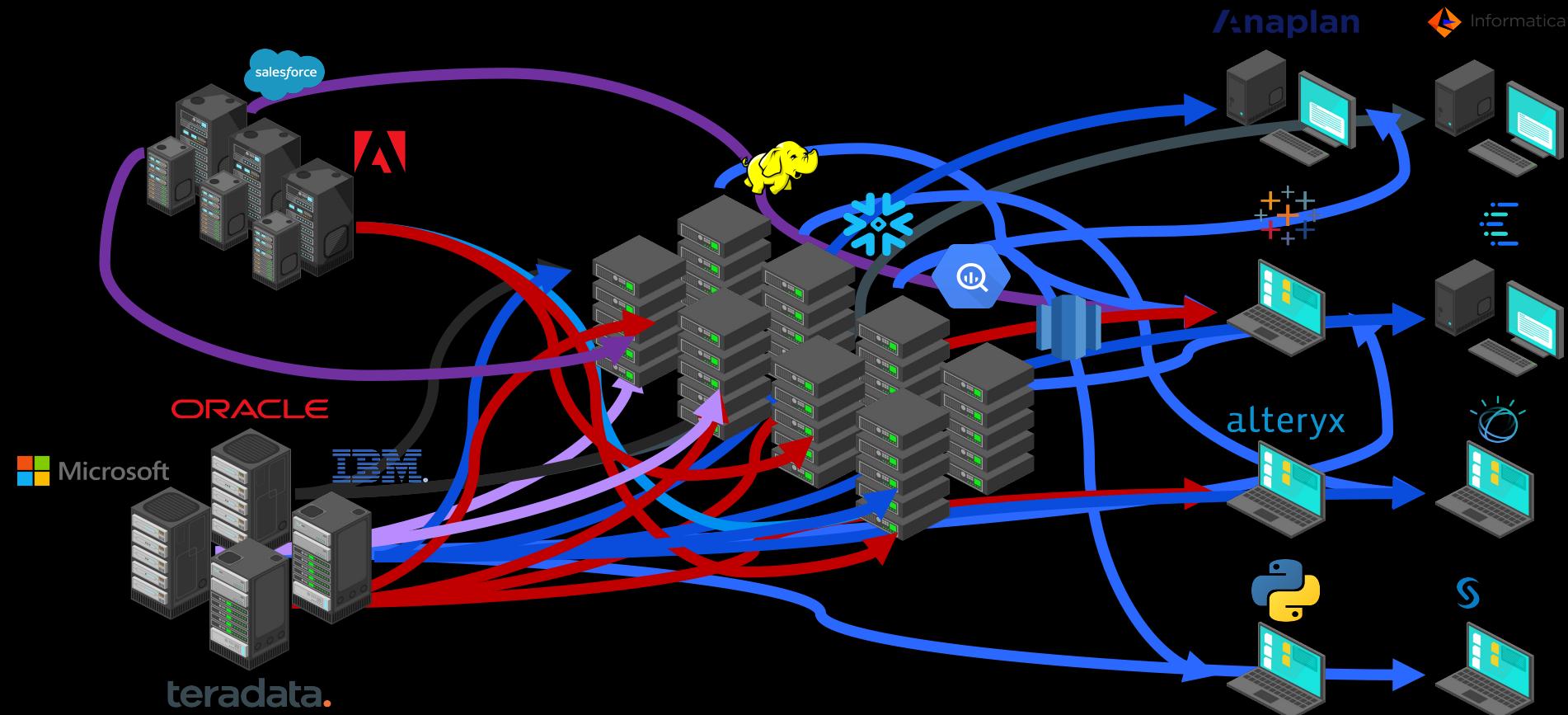
Data management complexity is the great inhibitor



- Data is spread across expansive digital ecosystems as cloud apps proliferate
- Overly manual efforts to move and unify data are slow, costly, and have limited impact
- Rapid growth in data with decreasing lifespans will only make things worse

42%
annual growth
in enterprise
data volumes

Common reality:
Tool oriented, little value, unnecessarily complex, and costly



The AI Ladder

IBM's prescriptive approach to the journey to AI

Infuse

Operationalize AI throughout the enterprise

Analyze

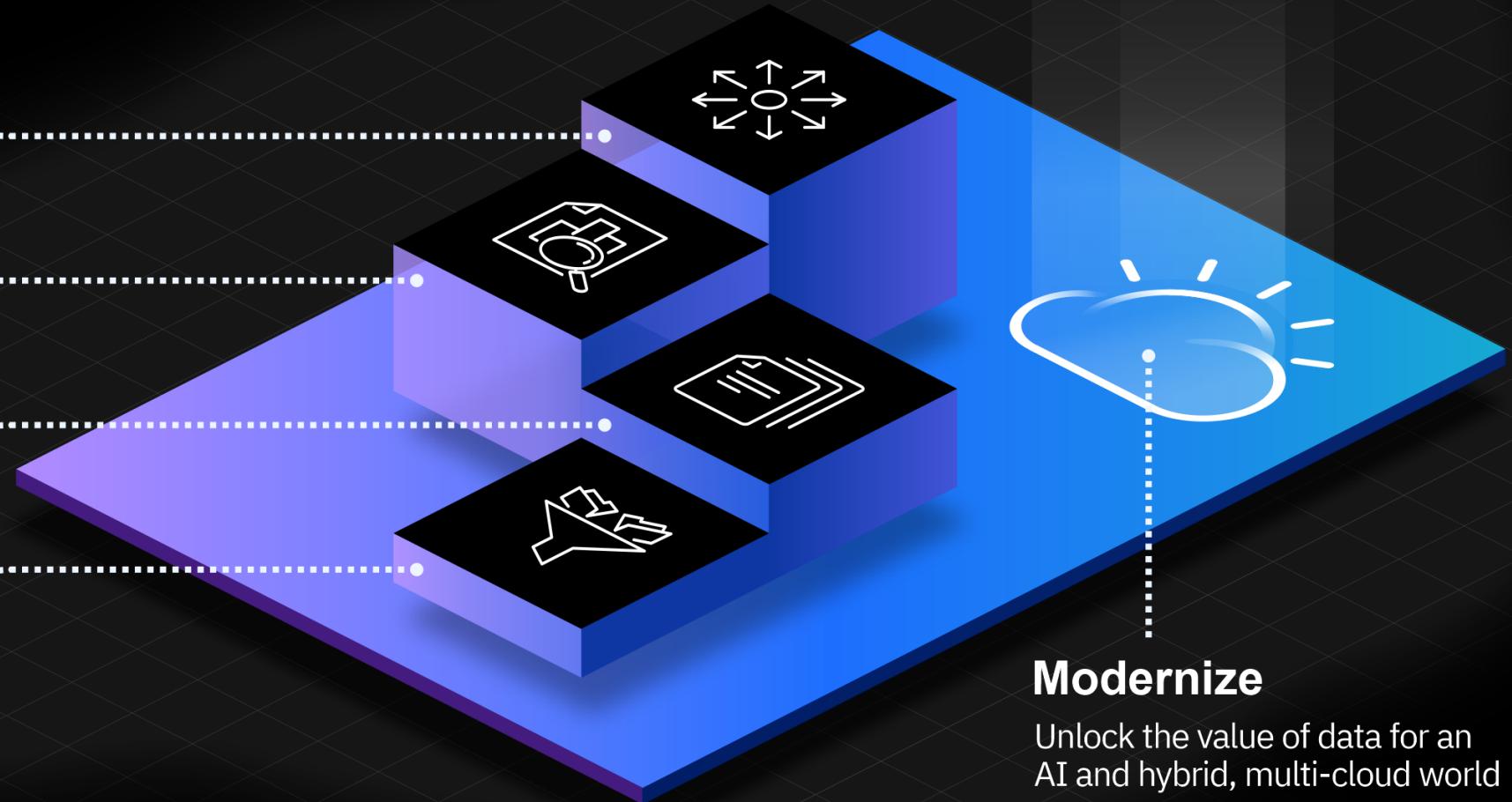
Build and scale AI with trust and transparency

Organize

Create a business-ready analytics foundation

Collect

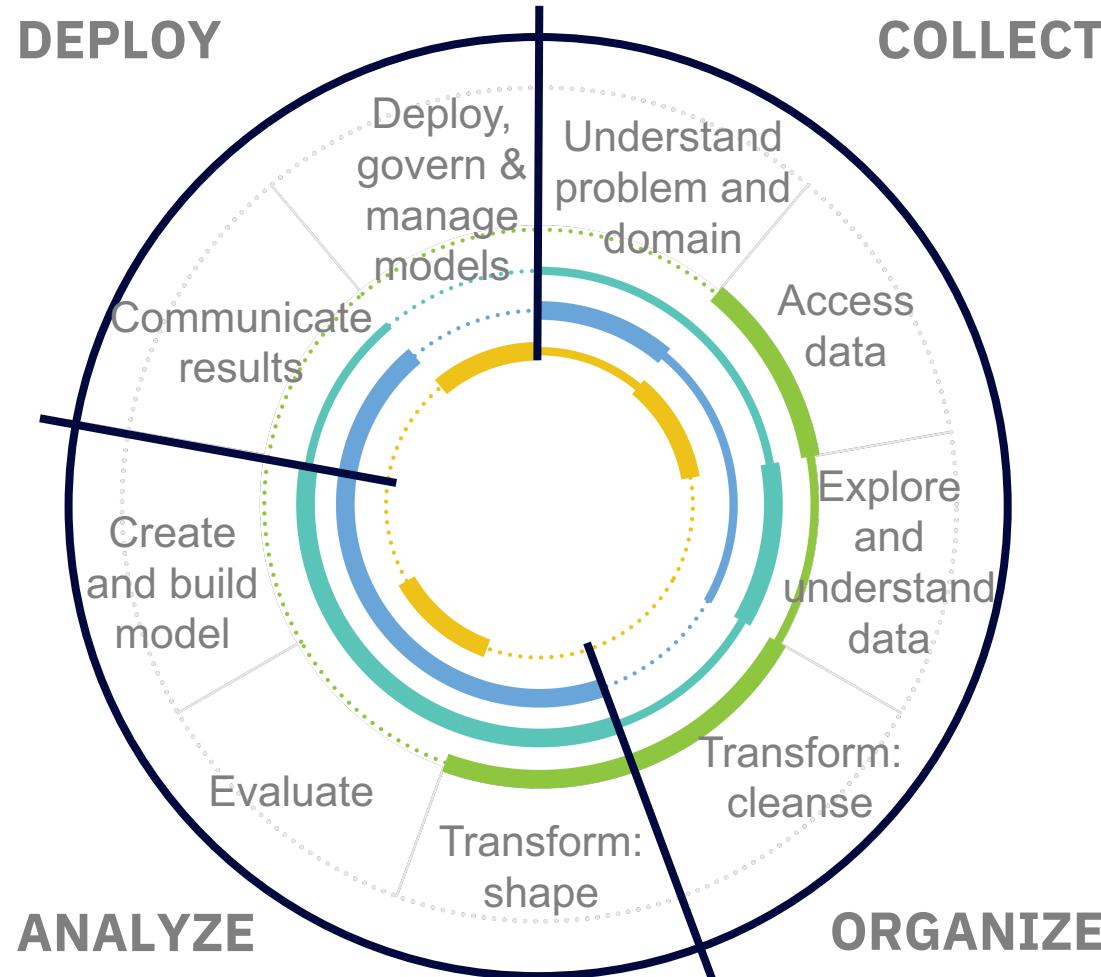
Make all data simple and accessible



Modernize

Unlock the value of data for an AI and hybrid, multi-cloud world

Methodology



Data Engineer

Architects data pipelines & ensures operability



Data Steward

Governs data and ensures regulatory compliance



Data Scientist

Gets deep into the data to draw insights for the business



Business Analyst

Works with data to apply insights to business strategy



App Developer

Plugs into analysis and code to build apps

Decision Maker(s)



CXO

CIO / CDO / CAO
LOB Leader

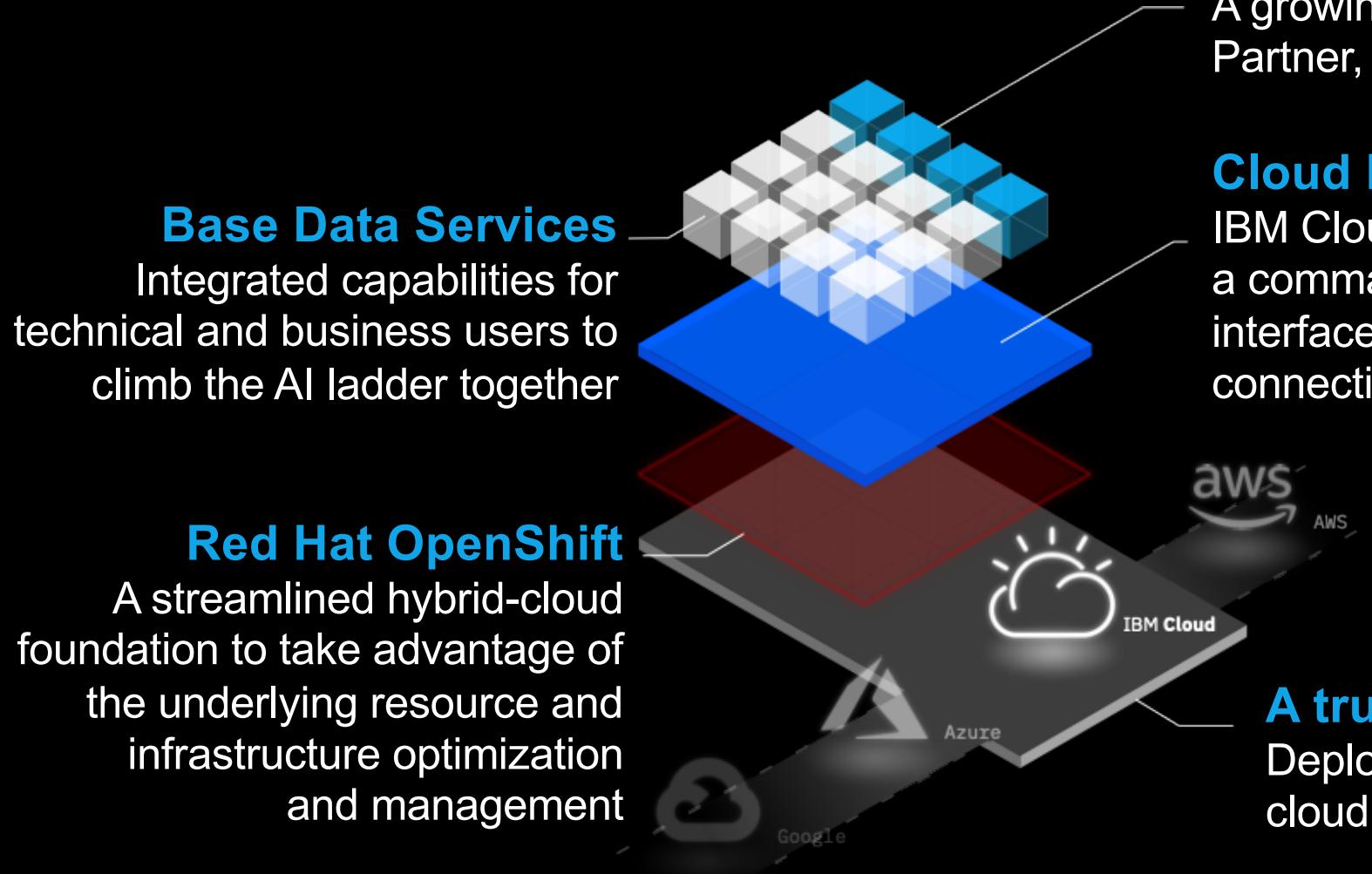


Sys Admin



Cloud Pak for Data

Simplifies, unifies, and automates your journey to AI



Extended Services

A growing ecosystem of Open Source, Partner, and IBM services

Cloud Pak Control Plane

IBM Cloud Pak Foundational Services provide a command-line interface (CLI), administration interface, services catalog, central list of connections, and the central user experience

A truly Hybrid-Cloud world

Deployable on-premises or on any vendor's cloud platform; also available as-a-Service

Cloud Pak for Data

Unified, modular, deployable anywhere

App Developers | Business Analysts | Data Engineers | Data Stewards | Data Scientists | Business Users

Integrated User Experience

Extensible: APIs, partner ecosystem, accelerators, and solutions

Collect

- Data virtualization
- SQL and NoSQL databases
- Event ingestion
- Streaming Analytics
- Apache Spark

Organize

- Data transformation
- Data quality and classification
- Policies and rules
- Data cataloging
- Self-service discovery and search

Analyze

- Data science and visualization
- AutoAI
- Model trust and explainability
- Model optimization

Infuse

- Business reporting and visualization
- Financial planning and analysis
- Cloud native AI services
- RegTech and Financial Crimes Insight

Cloud Pak core services
Security, Administration, Operations

Red Hat OpenShift

IBM Cloud | Amazon Web Services | Microsoft Azure | Google Cloud | Hyperconverged system

Cloud Pak for Data

Open, Extensible API Platform



App Developers



Business Partners



Data Engineers



Data Stewards



Data Scientists



Business Users

Personalized, Collaborative Team Platform



Instant, Pre-assembled Provisioning



Admin & Ops Dashboards



Cloud-native Data Micro Services

Collect Data

- Databases on-demand
- Data warehousing
- Fast data ingest / Streaming data
- Data Virtualization

Organize Data

- Data integration and Transformation
- Data curation
- Governance, privacy: policies & rules
- Data asset lifecycle management

Analyze Data

- Data visualization & exploration
- Machine learning & deep learning
- Model management & deployment
- Dashboards & business reporting

Enterprise Data Catalogs



Cloud Pak for Data



Kubernetes Platforms

✓ Elastic
Runtimes

✓ Product Ready Operations ✓ Built-in Continuous
Delivery

✓ Secure to the Core

Cloud Pak for Data v4.0 Packaging

AI Ladder Rung	Base Services	IBM Cartridges
Collect	Db2 Warehouse Data Virtualization Db2 BigSQL Netezza Performance Server IBM Streams Analytics Engine for Apache Spark Hadoop Execution Engine	Db2 AE/SE Informix
Organize	Watson Knowledge Catalog (including IGC) Information Analyzer (included in WKC) Data Management Console Data Privacy (Beta) IBM Match360 with Watson Guardium (Integration)	Master Data Management Product Master DataStage Information Server Knowledge Accelerators
Analyze	Watson Studio (includes Data Refinery) Watson Machine Learning (includes AutoAI) Watson Machine Learning Accelerator Watson OpenScale SPSS Modeler Decision Optimization	-
Infuse	Cognos Dashboards Embedded	Cognos Analytics Planning Analytics with Watson Watson Assistant Watson Discovery Watson Speech Services Financial Crimes Insights OpenPages with Watson Open Data for Industries Financial Services Workbench

Collect – Data management

Make data simple and accessible



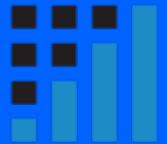
Multi-Cloud

- Fast and flexible data service management
- Ease of portability across private and public clouds



Unified Console

- One experience across all data services
- Coherent, efficient, scalable data services experience
- Ease of provisioning, monitoring, and more



Comprehensive Ecosystem

- Netezza, Db2 Warehouse, AESE, and Event Store
- Partnerships with MongoDB, EnterpriseDB, and more
- Data Virtualization



Data Governance and Self-Service Analytics

- Data automatically integrated with data governance capabilities for self-service data discovery

Collect Data

1. Provision, host & manage data sources on the platform

Data sources

 Db2 Event Store IBM Premium

In-memory data store capable of extremely high speed ingest and deep, real-time analytics.

 Db2 Advanced Enterprise Server Edition IBM Premium

Relational database that delivers advanced data management and analytics capabilities for transactional and warehousing workloads.

 Db2 Warehouse IBM

Data warehouse designed for high-performance, in-database analytics.

 IBM Db2 for z/OS IBM

Create databases in Db2 for z/OS and work directly with the data from IBM Cloud Private for Data

 Data Virtualization IBM Available

Query across many data sources as one.

 MongoDB Partner

Scalable, open source NoSQL database.

2. Connect to Existing data sources and/or upload your data

Sample Data Sources



Add data source

Data source name *
Type data source name here

Description
Type your description here

Data source type *
Big SQL

dashDB
Db2
Db2 for z/OS
Hive - HDP
HDFS - HDP

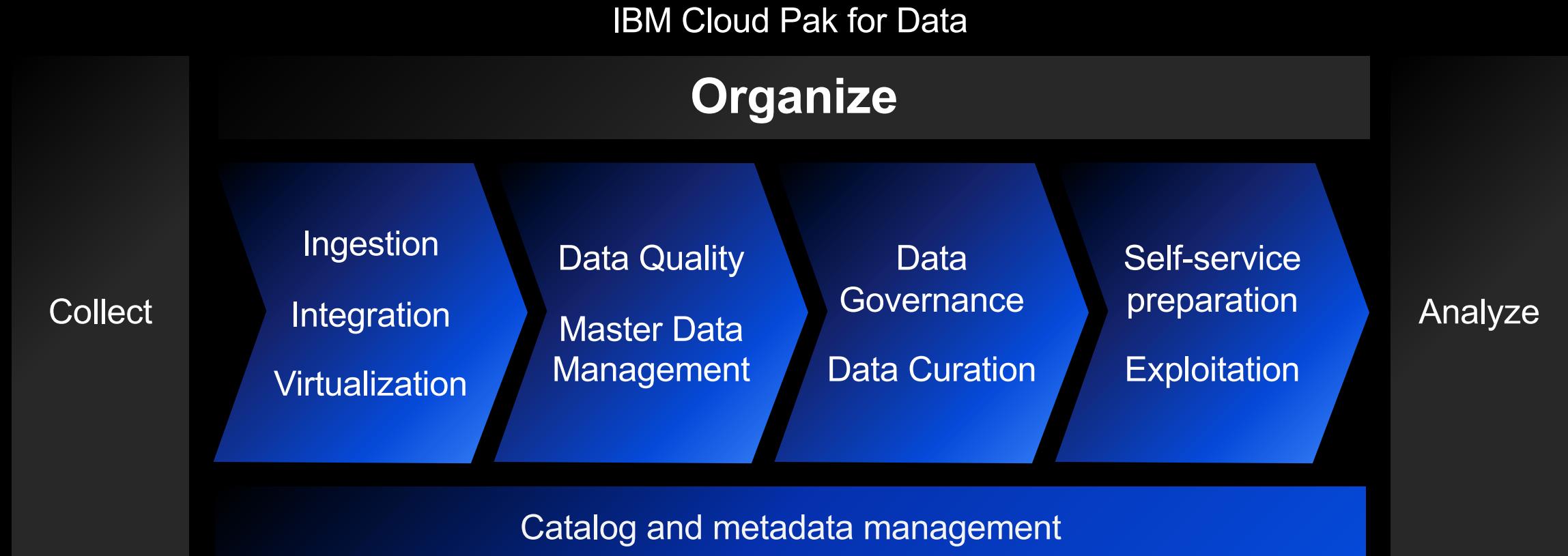
url here
here
here
here

Shared

(+) Add remote data set

Organize - DataOps

Create a business-ready analytics foundation



Private



IBM Cloud



Microsoft Azure



aws



Google Cloud

Organize Data

Auto-discover, catalog and publish meta data, track lineage etc.

Discover assets

Connection *
db3

Schemas ⓘ
Example schema[database_name/schema_name]

Browse

Select the tasks that you want to run

Profile and classify data Assign business terms
 Analyze data quality Publish to enterprise catalog

Explore Assets
In IBM Cloud Private for Data

ASSET TYPES

Search asset types

- Glossary and Governance
- Databases
- Data Files
- Unstructured Data Sources
- Data Science
- Logical Data Models
- Physical Data Models
- XML Schema Definitions
- Master Data Management
- Applications
- Files
- Stored Procedure Definitions
- Business Intelligence

Data exploration
In IBM Cloud Private for Data

Glossary and Governance

- Term (54)
- Category (4)
- Information Governance Rule (125)
- Information Governance Policy (80)
- Collection
- Label (6)

Steward

Data Class

Databases

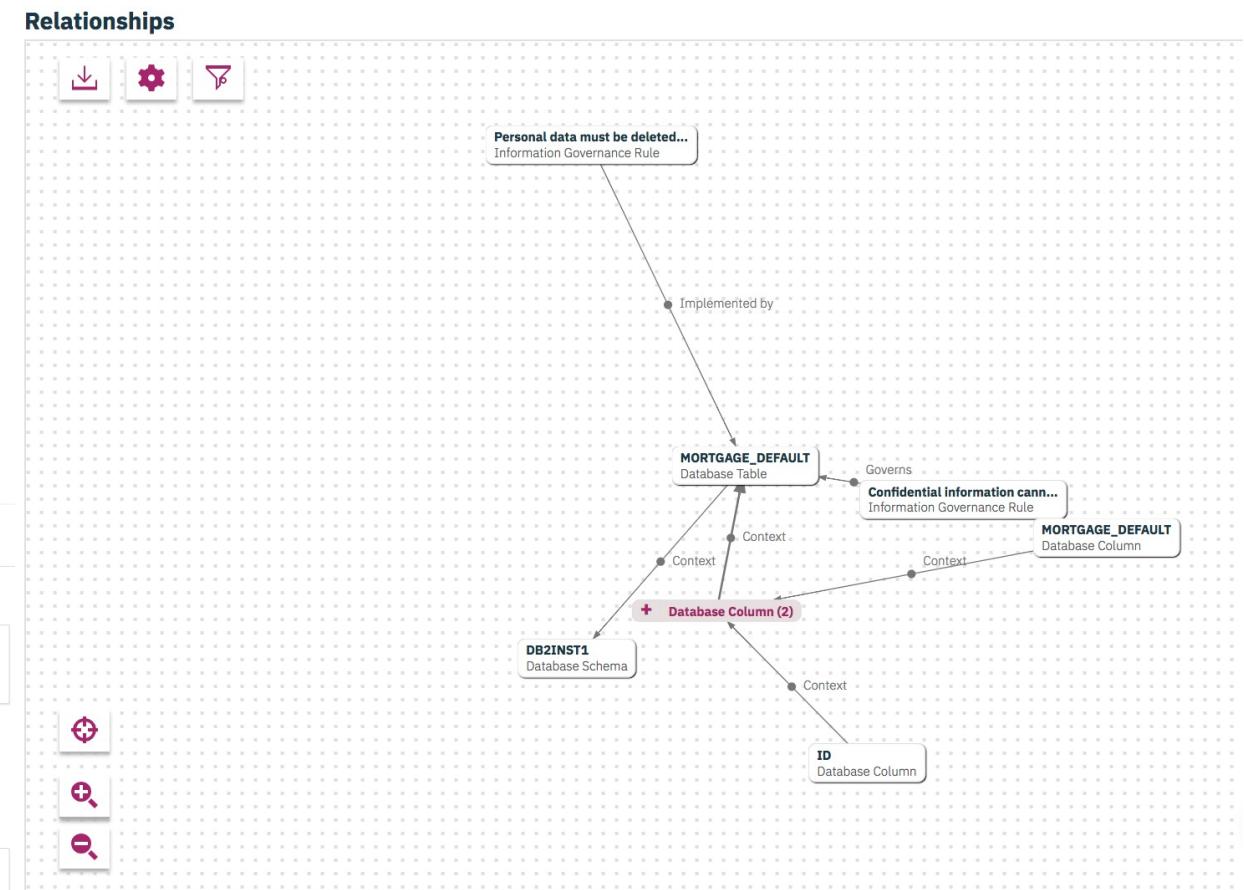
- Host (2)
- Database (2)
- Database Schema (7)
- Database Table (17)
- View (288)
- Database Column (4593)

Database Alias

Stored Procedure

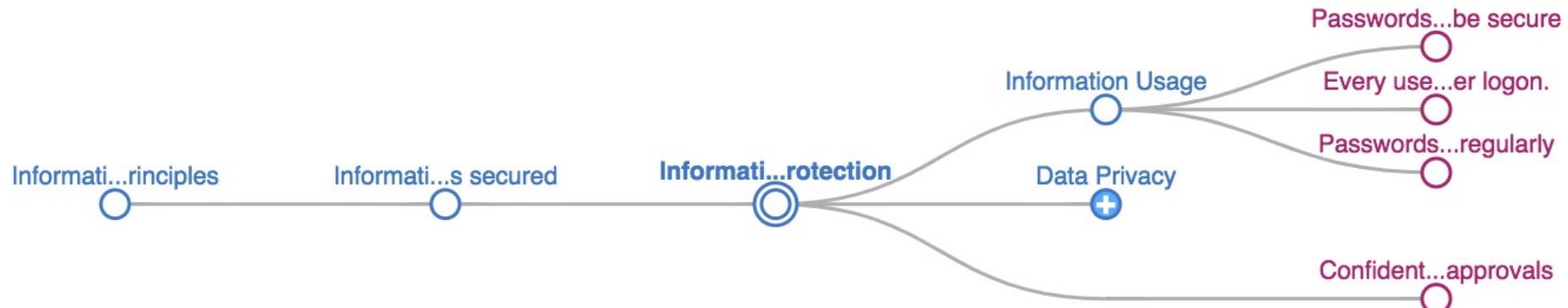
Stored Procedure Parameter

Analyze Data 4



Organize Data

Define & Enforce Governance Policies / Rules



Rules Policies

Search for Rules in the catalog

125 Rules available

Import Rules Create Rule

NAME	CREATED	MODIFIED	⋮
Mask an email address by using an auto generated method	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮ Details Delete
Mask a US Social Security Number by using a random method	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮ Details Delete
Mask a French INSEE Number by using a random method	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮ Details Delete
Computer addresses have a valid format	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮ Details Delete

Rules Policies

Search for Policies in the catalog

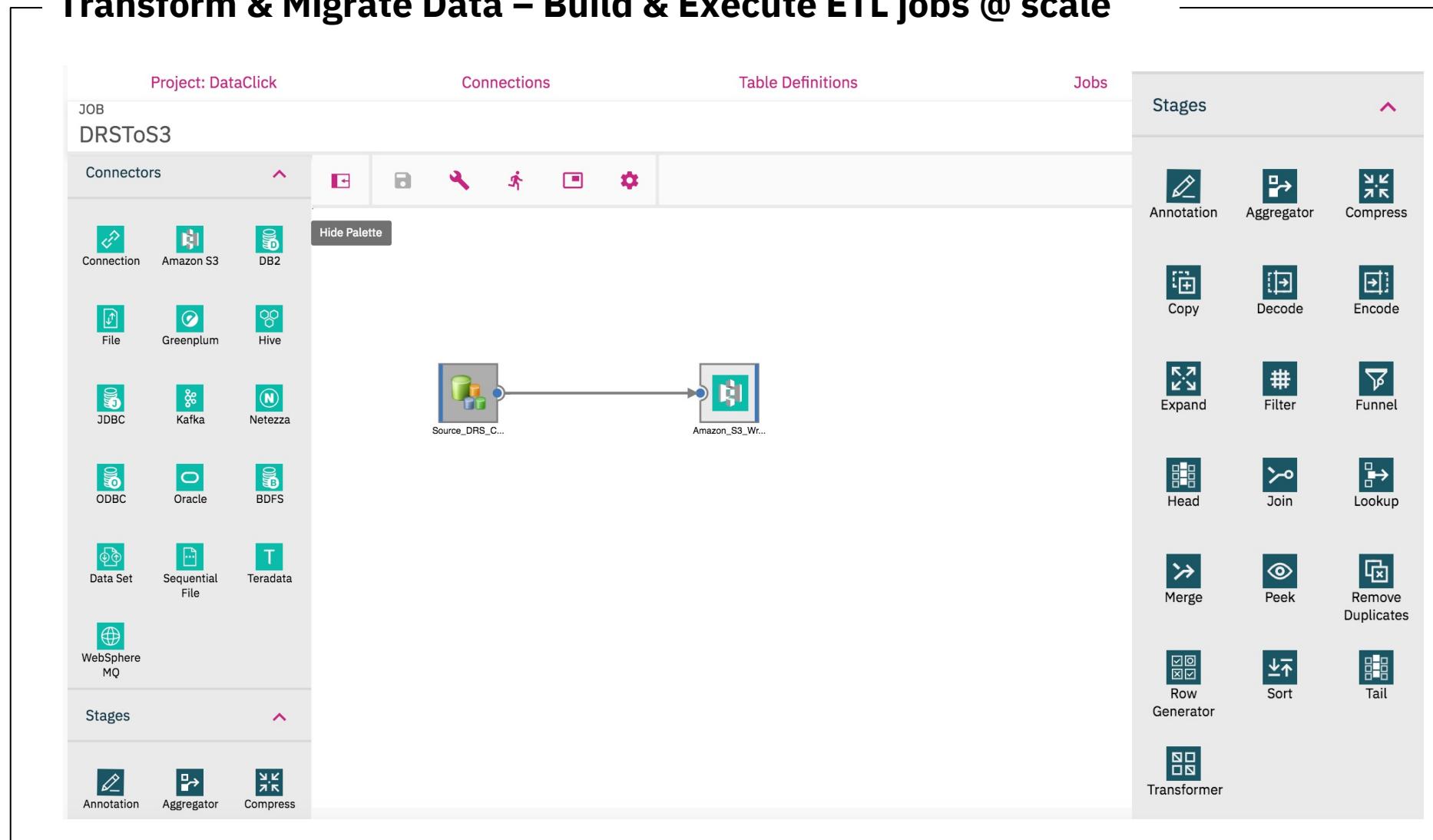
80 Policies available

Import Policies Create Policy

NAME	CREATED	MODIFIED	⋮
Information Protection	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮ Details Delete
Authoritative Sources provide the best information	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮ Details Delete
Policy Enforcement	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮ Details Delete
Information Supply Chain Integrity	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮ Details Delete

Organize Data

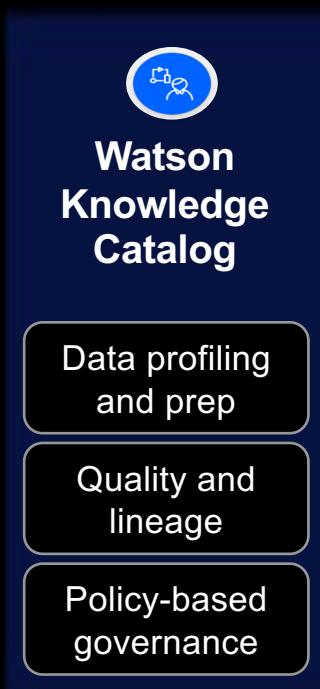
Transform & Migrate Data – Build & Execute ETL jobs @ scale



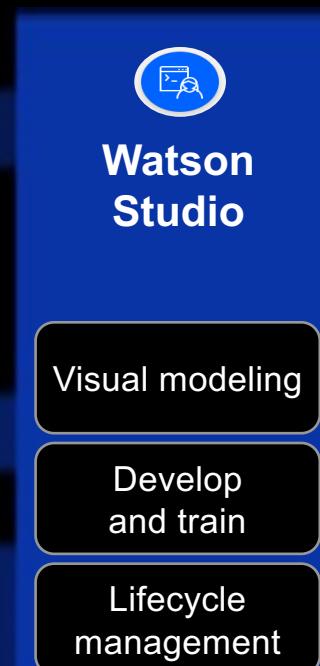
Analyze – Data science

Operationalizing AI with trust & transparency

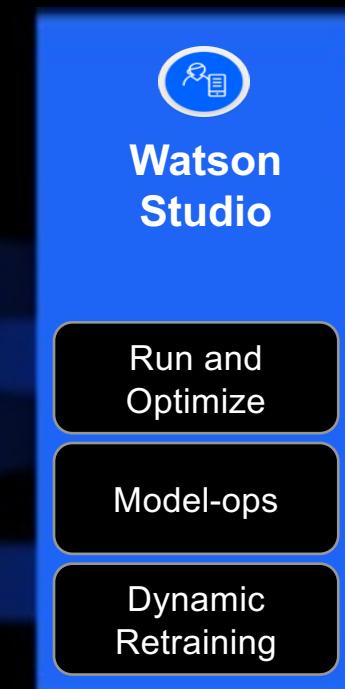
Prepare and Organize Data



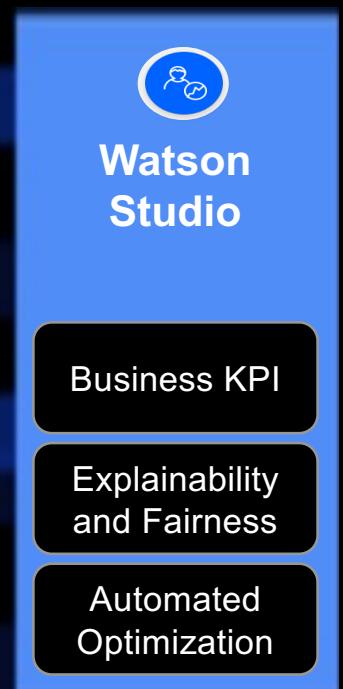
Build and Train AI Models



Deploy and Run AI Models



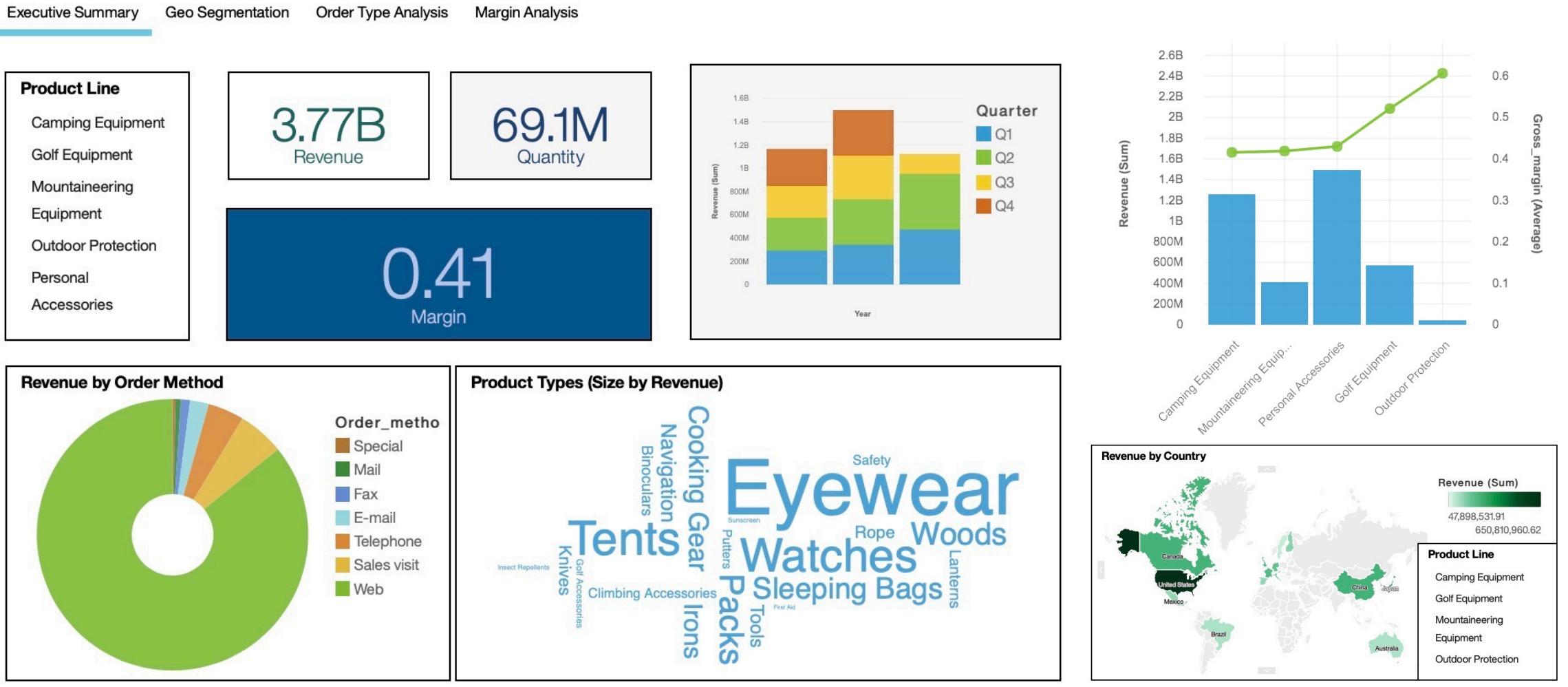
Manage and Operate Trusted AI



Automated AI Lifecycle

Analyze Data

Build Visualizations & Dashboards



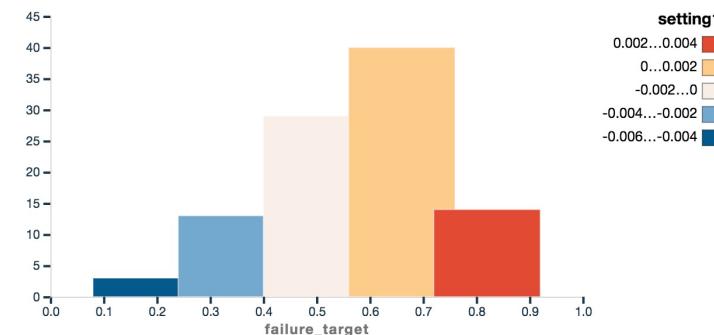
Analyze Data

Build , Deploy and manage Data Science & Machine Learning Models

Projects > Oil and Gas > Sensor LOS-model-evaluation.py

Sensor LOS-model-evaluation.py

```
34 threshold = {u'metric': 'areaUnderROC', u'min_value': 0.45, u'mid_value': 0.75}
35
36 # replace "label" below with the numeric representation of the label column that you defined while training the model
37 labelCol = "label"
38
39 # create evaluator
40 from pyspark.ml.evaluation import BinaryClassificationEvaluator
41 evaluator = BinaryClassificationEvaluator(labelCol=labelCol)
42
43 # compute evaluations
44 evaluation["metrics"]["accuracyScore"] = predictions.rdd.filter(lambda x: x[labelCol] == x["prediction"]).count() * 1.0 / predictions.count()
45 evaluation["metrics"]["areaUnderPR"] = evaluator.evaluate(predictions, {evaluator.metricName: "areaUnderPR"})
46 evaluation["metrics"]["areaUnderROC"] = evaluator.evaluate(predictions, {evaluator.metricName: "areaUnderROC"})
47 evaluation["metrics"]["threshold"] = threshold
48
49 if(evaluation["metrics"][threshold.get('metric', 'INVALID_METRIC')] >= threshold.get('mid_value', 0.70)):
50     evaluation["performance"] = "good"
51 elif(evaluation["metrics"][threshold.get('metric', 'INVALID_METRIC')] <= threshold.get('min_value', 0.25)):
52     evaluation["performance"] = "poor"
53 else:
```



Overview API

POST https://lkd-230-1-master-1.tyre.ibm.com:31843/dmodel/v1/sensor/pyscript/sensor-los/score Deployment token

sensor-los

No description available.

TYPE	ASSET	ALLOCATED CPU	ALLOCATED MEMORY	REQUESTS
Web service	Sensor LOS v1	Unallocated	Unallocated	0

Request

Function name * score

Body *

```
{"input_json_str": "[{"LoanAmount":12995,"Residence": "Owner Occupier","YearsCurrentEmployer":34,"AppliedOnline": "YES","NumberOfCards":2,"SalePrice":137000,"Income":4575,"YearCurrentAddress":8,"CCDebt":772,"Loans":1,"ID":100282,"Location": 100}]}
```

Response

Project releases 1

Sensor-IoT-Online

CREATOR



SOURCE
Oil and Gas D0model

DEPLOYMENTS
1

LAST MODIFIED
12 Sep 2018, 4:48 PM

Infuse - AI Applications

Accelerate time-to-value with pre-built AI apps

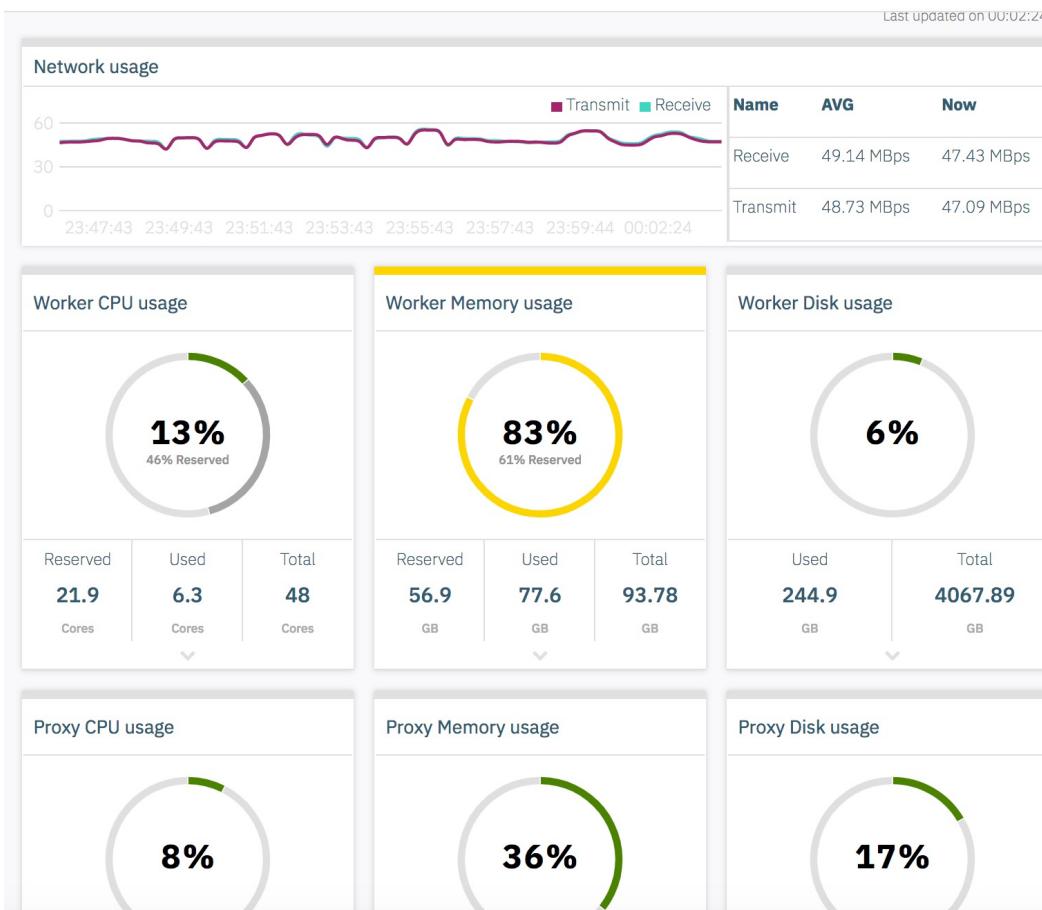


Hyperconverged
Private Cloud Systems

Administer

A single Pane to administer & monitor your data & analytics workloads

Admin Dashboard



Pod list

POD NAME	STATUS	NODE NAME	USED CPU	USED MEMORY	USED DISK
audit-trail-service-56c998198d-xgmf7	✓	172.16.204.201	0.00	286MB	0B
auth-apikeys-6mrfg	✓	172.16.204.11	0.01	604MB	0B
auth-apikeys-w7mpl	✓	172.16.203.175	0.01	455MB	0B
auth-apikeys-xs2g2	✓	172.16.203.116	0.01	442MB	0B
auth-idp-72zdh	✓	172.16.203.175	0.01	640MB	0B
auth-idp-8mgfb	✓	172.16.203.116	0.01	687MB	0B
auth-idp-sltk	✓	172.16.204.11	0.00	808MB	0B
auth-pap-4jdf2	✓	172.16.203.175	0.00	195MB	0B
auth-pap-bfnd2	✓	172.16.204.11	0.00	203MB	0B
auth-pap-worsn	✓	172.16.203.116	0.00	202MB	0B

Runtimes 5

NAME	RUNTIME TYPE	USER	PROJECT	JOB NAME (RUN ID)	DATE STARTED	CPU (CORES)	GPUS	MEMORY (GB)	STATUS
Jupyter with Python 3.5 for GPU	Environment	admin	Oil and Gas	12 Sep 2018, 7:09 AM	—	—	—	—	●
RStudio with R 3.4.3	Environment	admin	Oil and Gas	12 Sep 2018, 8:54 AM	—	—	—	—	●
Data Refinery	Environment	admin	Oil and Gas	12 Sep 2018, 8:59 AM	—	—	—	—	●
Jupyter with Python 2.7, Scala 2.11, R 3.4.3, Spark 2.0.2	Environment	admin	Oil and Gas	12 Sep 2018, 10:14 AM	—	—	—	—	●
Jupyter with Python 2.7, Scala 2.11, R 3.4.3, Spark 2.0.2	Environment	admin	SF-SAS	14 Sep 2018, 8:54 AM	—	—	—	—	●

The newest generation of Cloud Pak for Data

Scale the value of ALL your data and accelerate your journey to AI with greater trust and productivity

Unlock new insights to more people

Democratize data discovery, understanding and relevancy

Deliver trusted, business-ready data

Dynamically govern data quality, privacy and usage policies

Unify data across hybrid clouds

Access and manage diverse data minimizing data movement

Unify and speed data-AI lifecycles

Empower collaborative, persona-centric workflows

Unleash data science productivity

Automate complex tasks, improve quality, lower skill barriers

Persona-centric Experiences (aaS)



Collect Data

Make Data Simple and Accessible



Organize Data

Create a Business-ready Analytics Foundation



Analyze Data

Build and Scale Trusted AI and BI



Infuse AI

Operationalize AI Across the Business

Watson AI and Openly Extensible Ecosystem



Intelligent Data Fabric

AutoCatalog
Discovery and Understanding

AutoSQL
Distributed Query

AutoAI
AI Model Lifecycle

AutoPrivacy
Privacy and Usage Enforcement



Any Data. Any Cloud. Anywhere.

Data Warehouses

Data Lakes

Data Clouds

Data Streams

Databases



Automation in Cloud Pak for Data

AutoAI

Automate training, predictive accuracy and explainability for models that reply on siloed data sets and time-variant conditions

AutoPrivacy

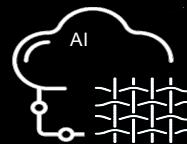
Automate enforcement of universal privacy and usage policies across the ecosystem of sources and consumers

AutoCatalog

Automate data discovery, understanding and enrichment by maintaining real-time knowledge of business-ready data

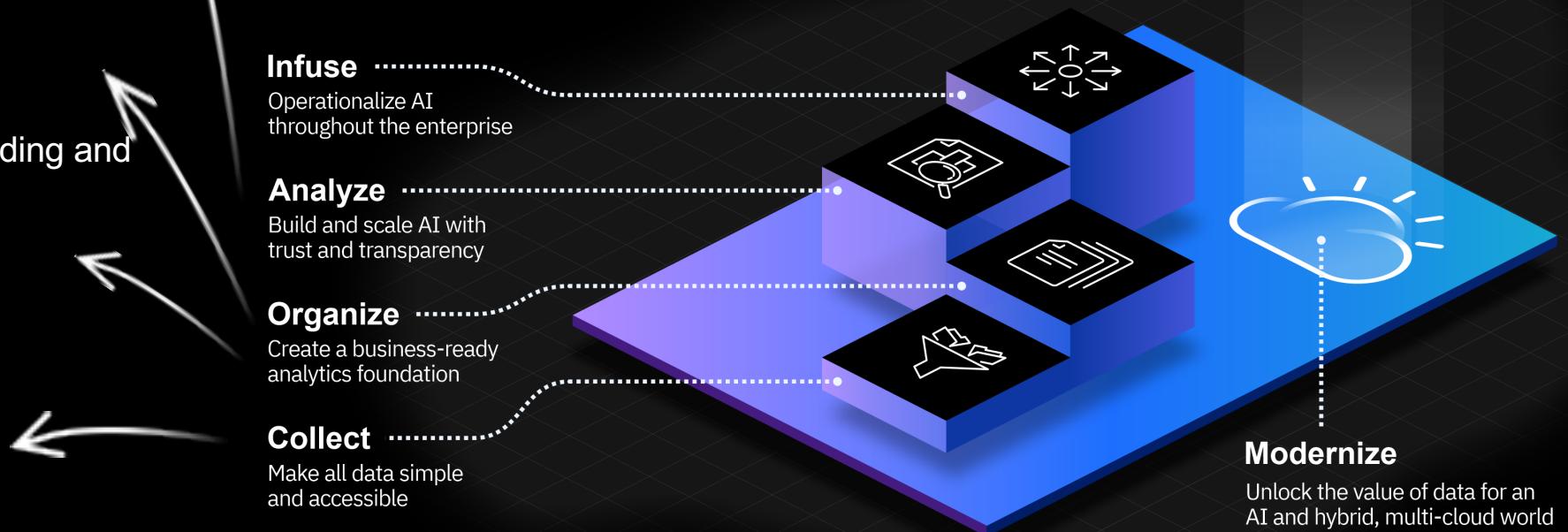
AutoSQL

Automate access to unify data across distributed data/cloud landscapes without data movement



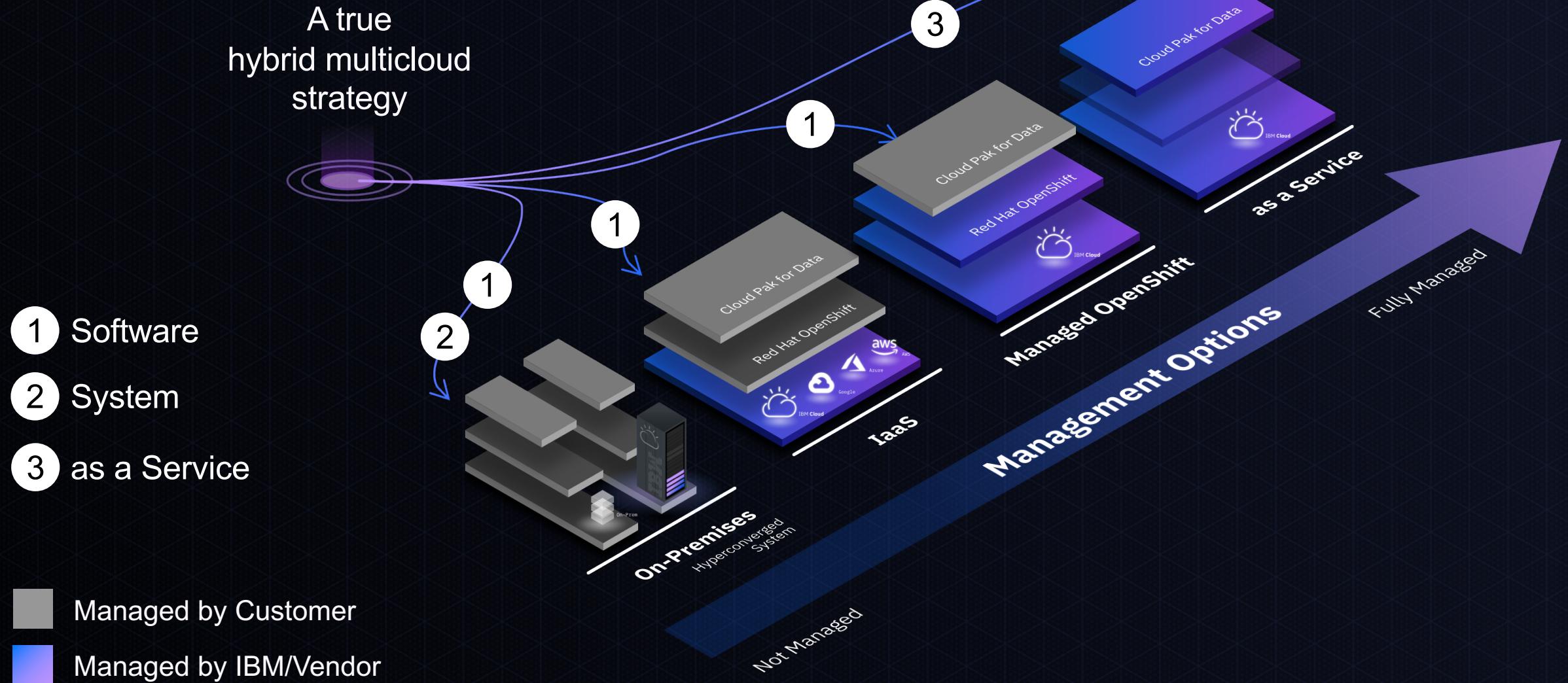
An intelligent data fabric

- Weave together new insights from unified data
- Democratize data usage across more people
- Unleash productivity and reduce cost/risk
- Make AI business easier and faster



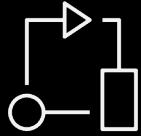
Deploy anywhere

Including as a managed service on IBM Cloud



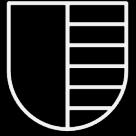
Cloud Pak for Data as a Service

Fully integrated data and AI platform, delivered as a service



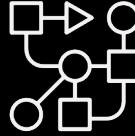
UNIFIED

Simplify the usage and management of essential data and AI capabilities with an integrated and fully managed platform.



GOVERNED

Ensure trusted and secure insights with leading governance capabilities on the cloud for enterprise.



ADAPTABLE

Increase speed and optimize costs with an as-a-Service solution that flexibly adapts to your changing business needs.



AI INFUSED

Empower your team with AI and automation to help augment their skills and accelerate AI development.

Cloud Pak for Data

Top 4 Use Cases



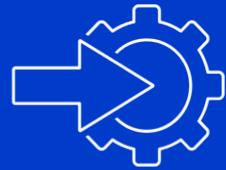
Manage your Data Anywhere

1. Manage all your enterprise data regardless of where it lives

(Data Virtualization)

2. Gain control & leverage your data from connected devices

(Fast data & Streaming analytics)



Operationalize Data Science & AI

Build, deploy, manage & govern models & data @ scale to improve business outcomes

(see following slides for details)

- e.g. a. Customer Churn
- b. Cross Sell / Up Sell
- c. Predictive Maintenance



Shift to Next-Gen workloads

Shift to Cloud Native

- a. Provision & scale in minutes
- b. Build once, deploy anywhere – multi cloud support
- c. Built in automation & collaboration to increase productivity



Smarter Governance

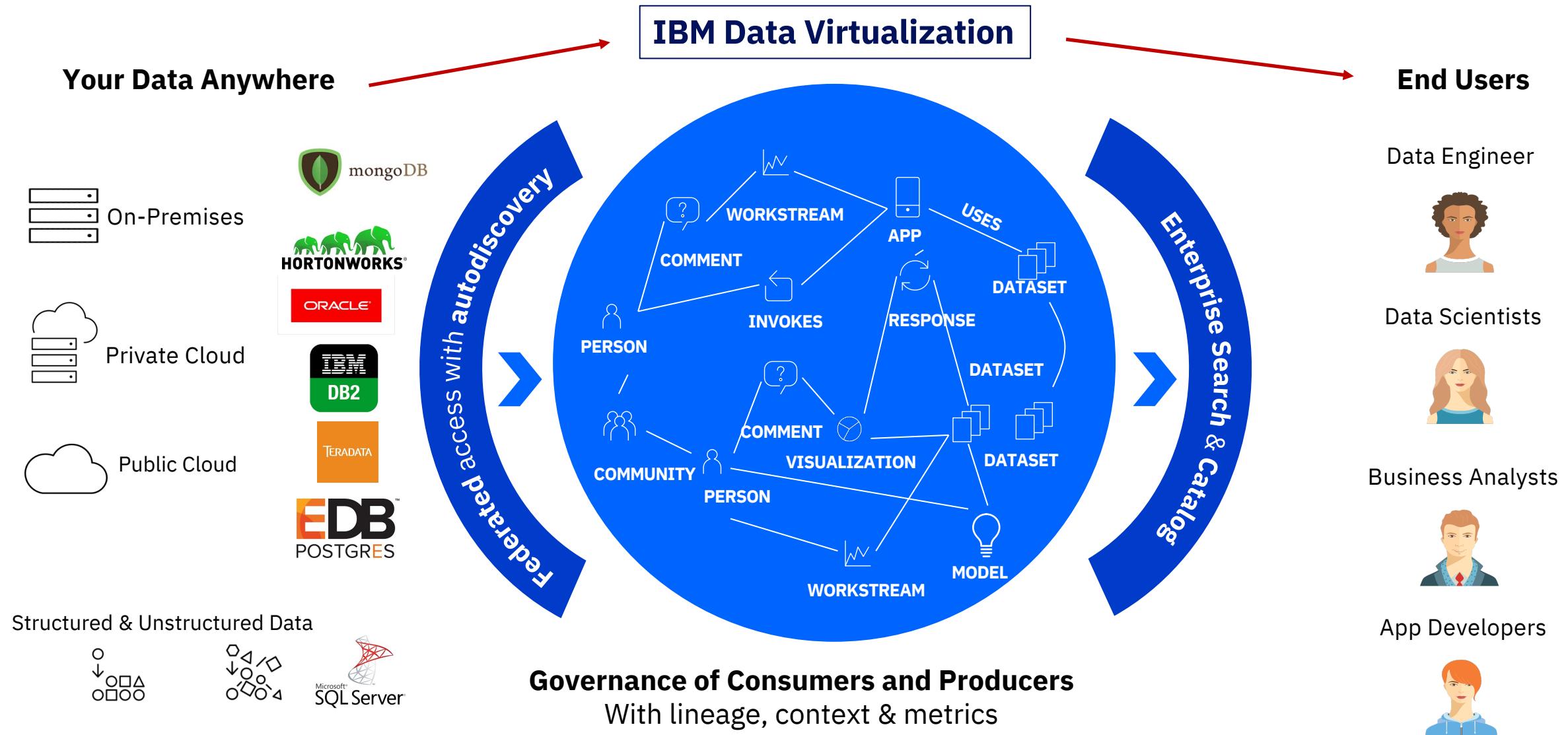
Governance to enable self service analytics

Auto-discover meta data, manage governance rules & policies, enforce privacy etc. to mitigate risk & ensure compliance

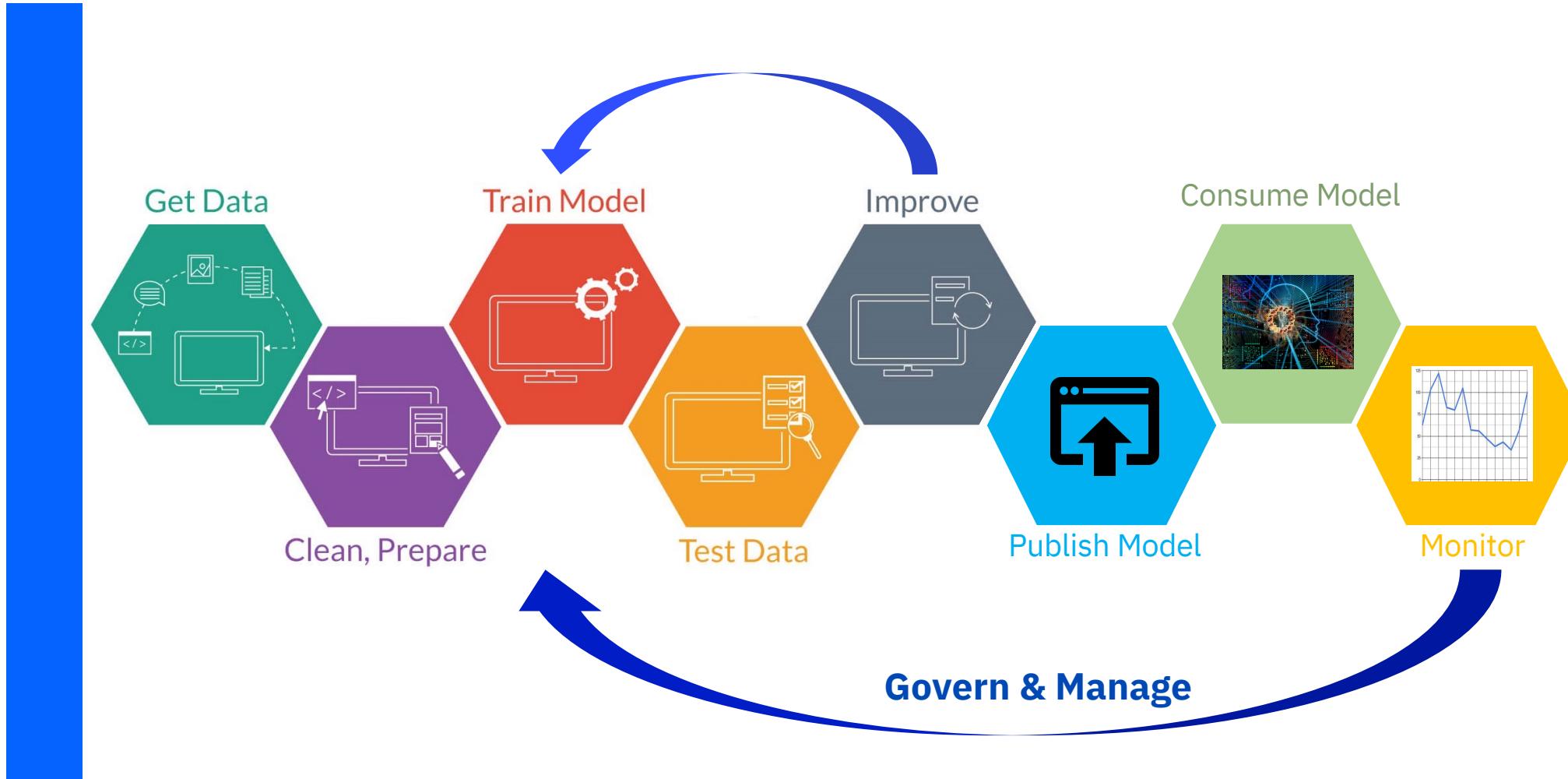
e.g. accelerate GDPR Compliance

Cloud Pak for Data

Use Case: Manage All Your Data – regardless of where it lives



Operationalizing Datascience & AI



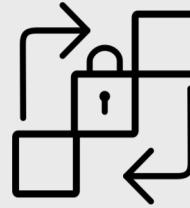
Cloud Pak for Data

Use Case: Modernize your Data & Analytics Workloads – monolithic to microservices



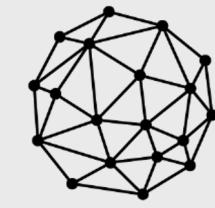
Reduce Costs

Run your **data & analytics services more efficiently** – single product stack and better resource utilization



Ship Faster

Increase agility and respond to changing business requirements through **DevOps** and **continuous delivery**



Increase Productivity

Enable **governed self-service**, collaboration, and provisioning, **across teams** on an integrated platform

Shift from Monolithic to Microservices for agility and efficiency

Cloud Pak for Data

Use Case: Compliance Readiness – accelerate transformation for compliance

Regulatory Compliance GDPR



- GDPR **applies globally**
- Potential **cost of not being compliant** is up to 4% of your organization's global revenue
- GDPR is in effect starting **May 25, 2018**

Cloud Pak for Data Out-of-the-box services



- Sensitive data discovery
- Data cataloging
- Data protection
- Compliance dashboards
- Record of processing

Integrate Optionally Thorough ecosystem



- Regulatory readiness dashboards
- Purpose & consent management
- Subject access rights
- Incident response

