

# Release the breaks and accelerate your Journey to AI



Use evolving trends and lessons  
learnt to reinvent your Data and AI  
Strategy to support your Digital  
Transformation Journey



Wolfgang Knupp  
Data and AI Architect  
IBM South Africa

# Data and AI Strategy

## Quote:

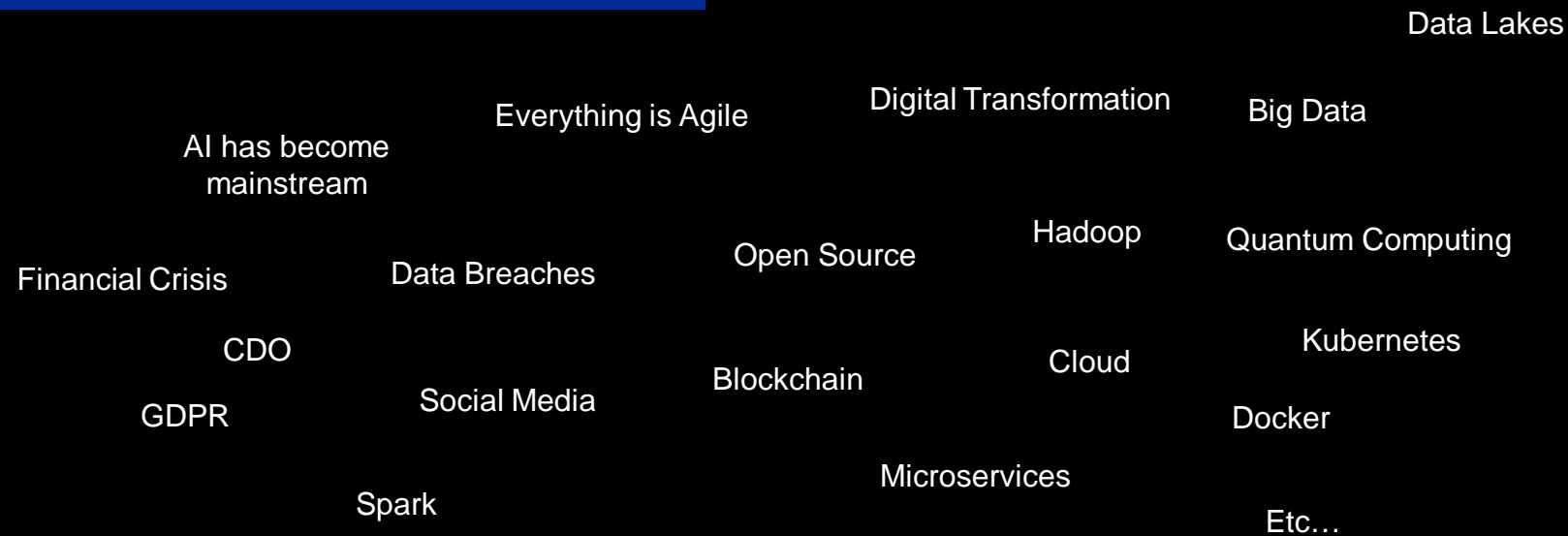
“Not having a data strategy is analogous to allowing each person within each department of your organization to develop their own chart of accounts and use their own numbering scheme. ”

Sid Adelman, Data Warehousing Expert (Adelman, Moss, & Abai, 2005)

# Data and AI Strategy

So, what has changed in the Data and AI world since 2005 ?

Not much ?... except...



# Data and AI Strategy

Your Data Strategy is intended to support your Business Initiatives.

and must :

- Be owned by Business (Chief Data Officer)
- Align with the Business Strategy
- Align with changes in technology
- Be revised regularly
- Include applicable Regulations
- Include Ethical Guidelines for AI
- Encourage ambitious and achievable goal setting (be testable / defensible)
- Include the standard elements as in 2005
  - Operational Model
  - Data Governance
    - Data Ownership
    - Common Business Glossary of Terms, Policies and Rules
    - Security, Data Privacy, Lifecycle, Data Quality, etc...
- CDO, Data and AI Reference Architecture, DevOps, Cloud

# Data and AI Strategy

Your Data Strategy is intended to support your Business Initiatives.

## Lessons learnt:

- Your data strategy must reflect your own Business Strategy
  - not be dictated by vendors or consulting houses
- Best of Breed technology is too expensive
  - time to value too long
  - cost of skills to integrate adjacent best of breed technologies
- Culture needs to adapt -> Change management is vital
- IT overload -> Business User Data / Insights Self Service to encourage innovation, fail fast
- And last but not least... There is No AI without IA

**There is no AI  
without an IA  
(information architecture)**

**80%**

of data is either  
inaccessible,  
untrusted or  
unanalyzed

**81%**

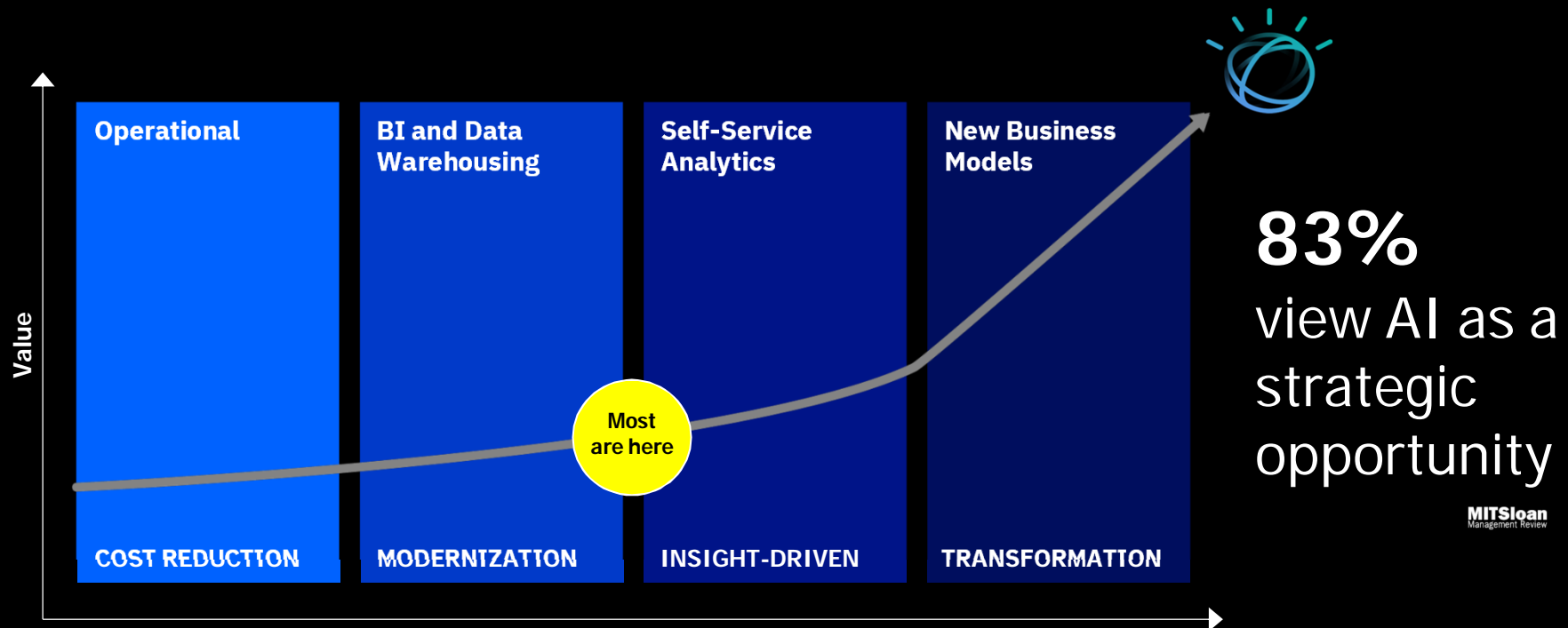
do not  
understand the  
data required  
for AI

//

*No amount of AI algorithmic sophistication will overcome a lack of data [architecture] ... bad data is simply paralyzing*

**MIT Sloan**  
Management Review

# Clients are advancing their use of data for business value



# Barriers to successful AI transformations

There is no Artificial Intelligence (AI) without Information Architecture (IA)

## Data Ecosystem

- Data in silos
- Difficult to access
- No lineage

## Workflow

- Not integrated
- Not governed
- Lack dev/prod parity

## Analytics Tools

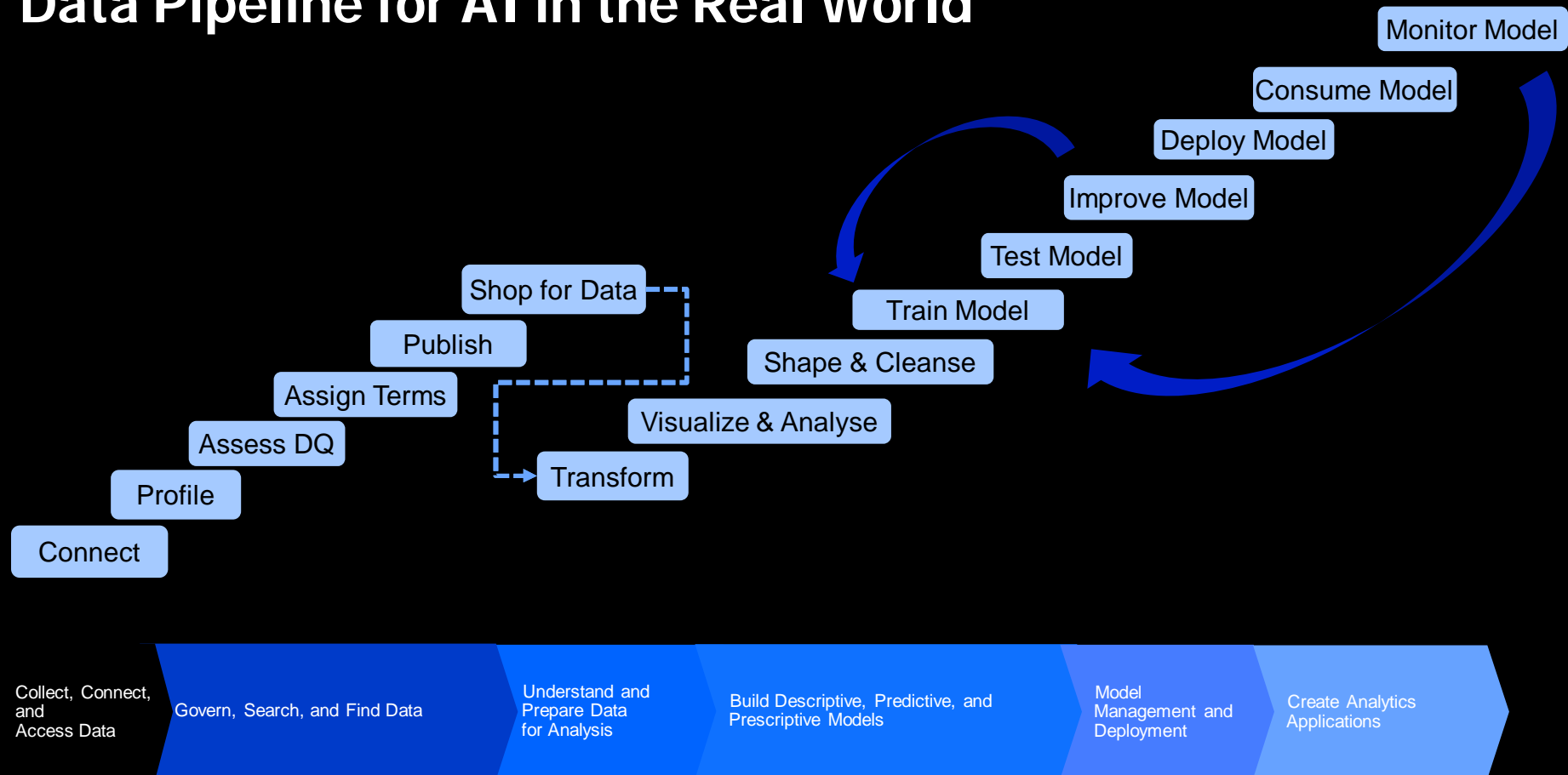
- Discrete tools
- Different preferences
- Difficult to manage

## Culture

- Not collaborative
- Slow provisioning
- Lack trust in AI

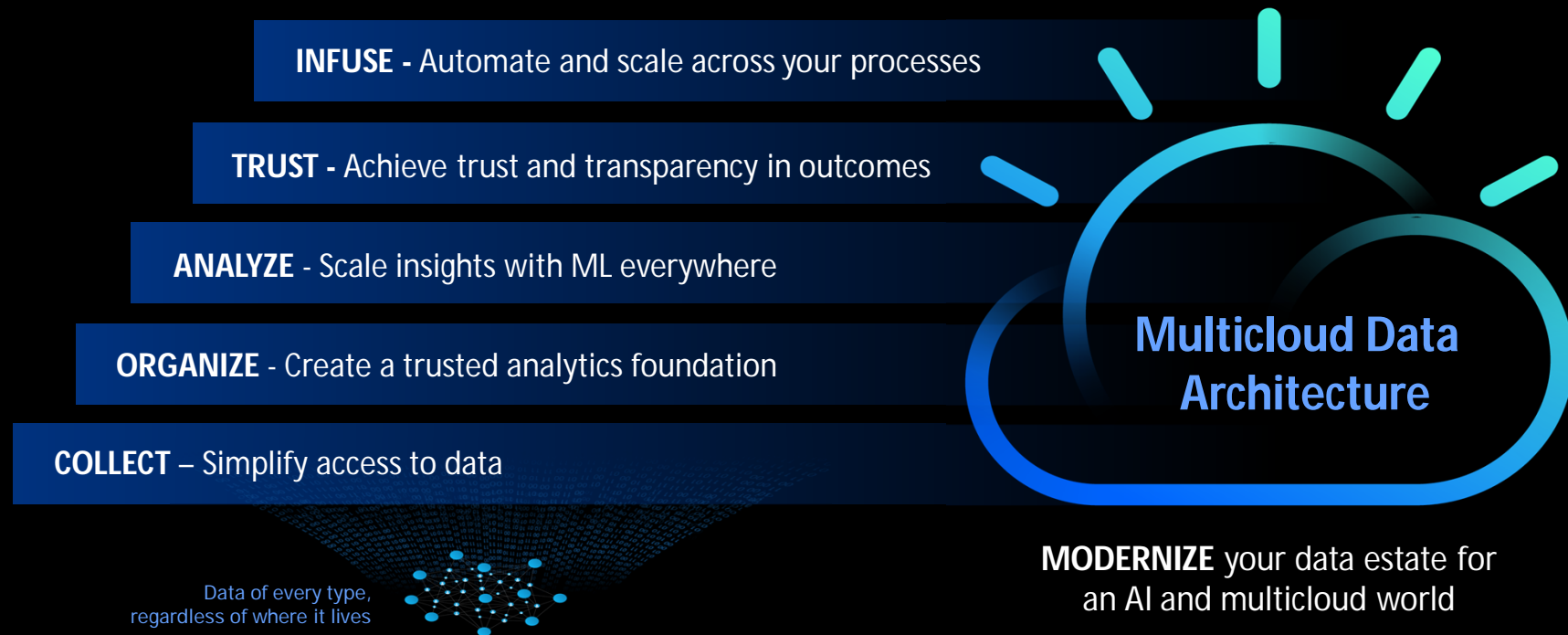


# Data Pipeline for AI in the Real World



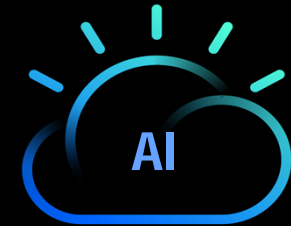
# The AI Ladder

A prescriptive approach to accelerating your journey to AI

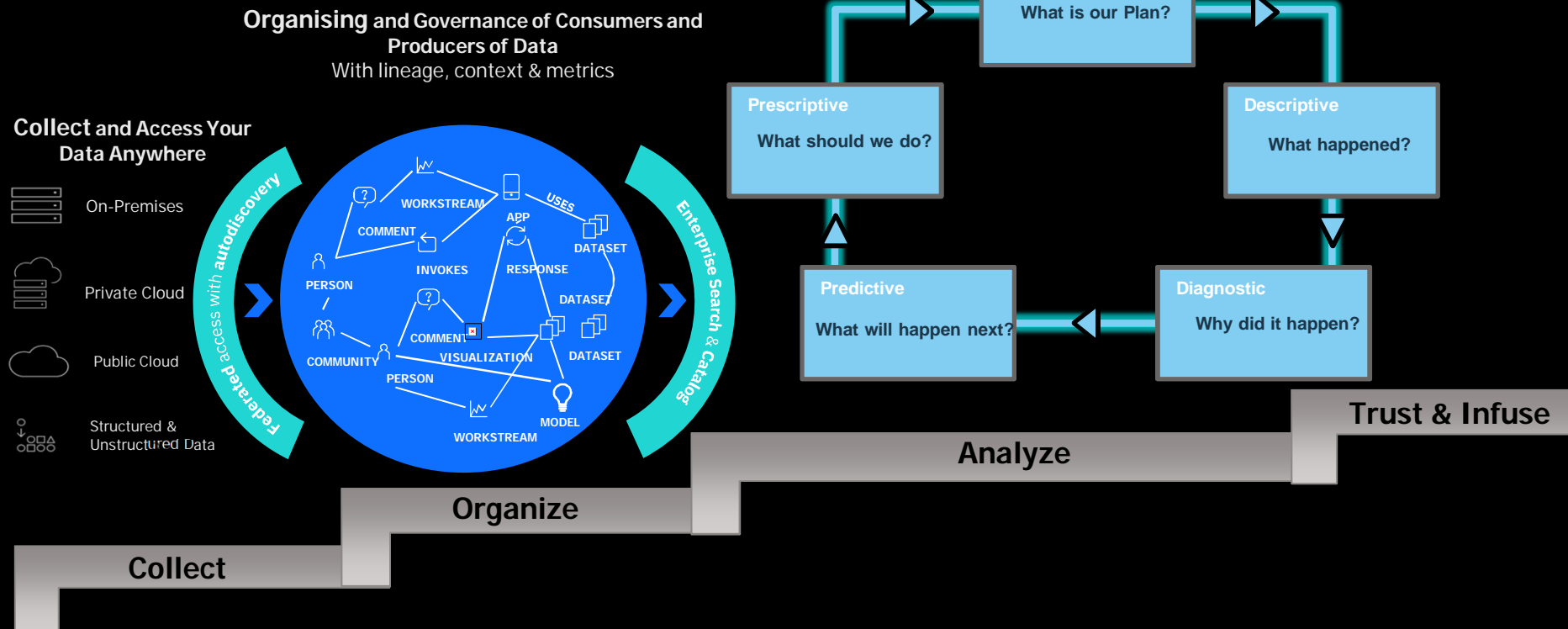


**MODERNIZE** your data estate for  
an AI and multicloud world

# Ladder to AI with IA



Apply Analytics Lifecycle to gain Actionable Insights  
With BI, Machine Learning and Optimisation



# IBM Solution to accelerate the Journey to AI: IBM Cloud Private for Data



App Developers



Business Partners



Data Engineers



Data Stewards



Data Scientists



Business Users



Instant, Pre-assembled Provisioning



Admin & Ops Dashboards



Cloud-native Data Micro Services

Collect Data

- Databases on-demand
- Data warehousing
- Fast data ingest / Streaming data
- Federated query

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



IBM Cloud Private



Kubernetes Platforms

✓ Elastic  
Runtimes

✓ Product Ready  
Operations

✓ Built-in Continuous  
Delivery

✓ Secure to the Core

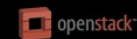
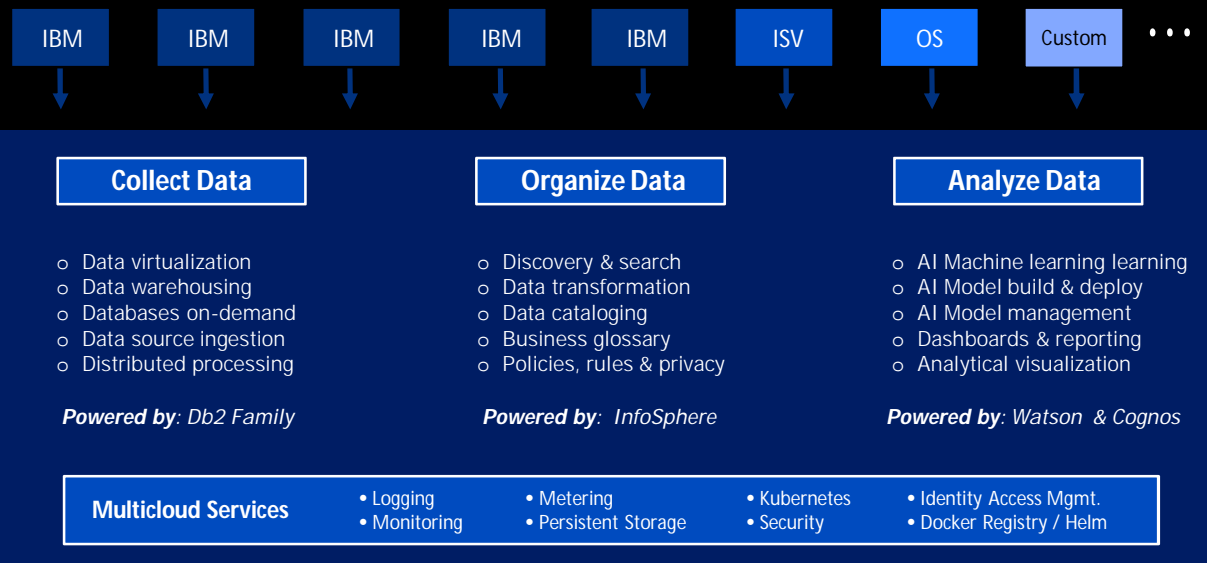
# IBM Cloud Private for Data

Unify on an open, multicloud Data & AI platform

## The Ladder to AI



Extensible "add-ons" from IBM's Data & AI Portfolio & 3<sup>rd</sup> Party Services



# 1) Deploys an Open Information Architecture for AI



Custom  
Extensions



App  
Developers



Data  
Engineers



Data  
Stewards



Data  
Scientists



Business  
Users & Analysts

## Data & AI Microservices

Collect Data

Organize Data

Analyze Data

Trust AI

Infuse AI



Enterprise Cloud  
Microservices



Containerized  
Workloads

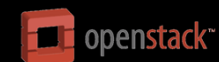
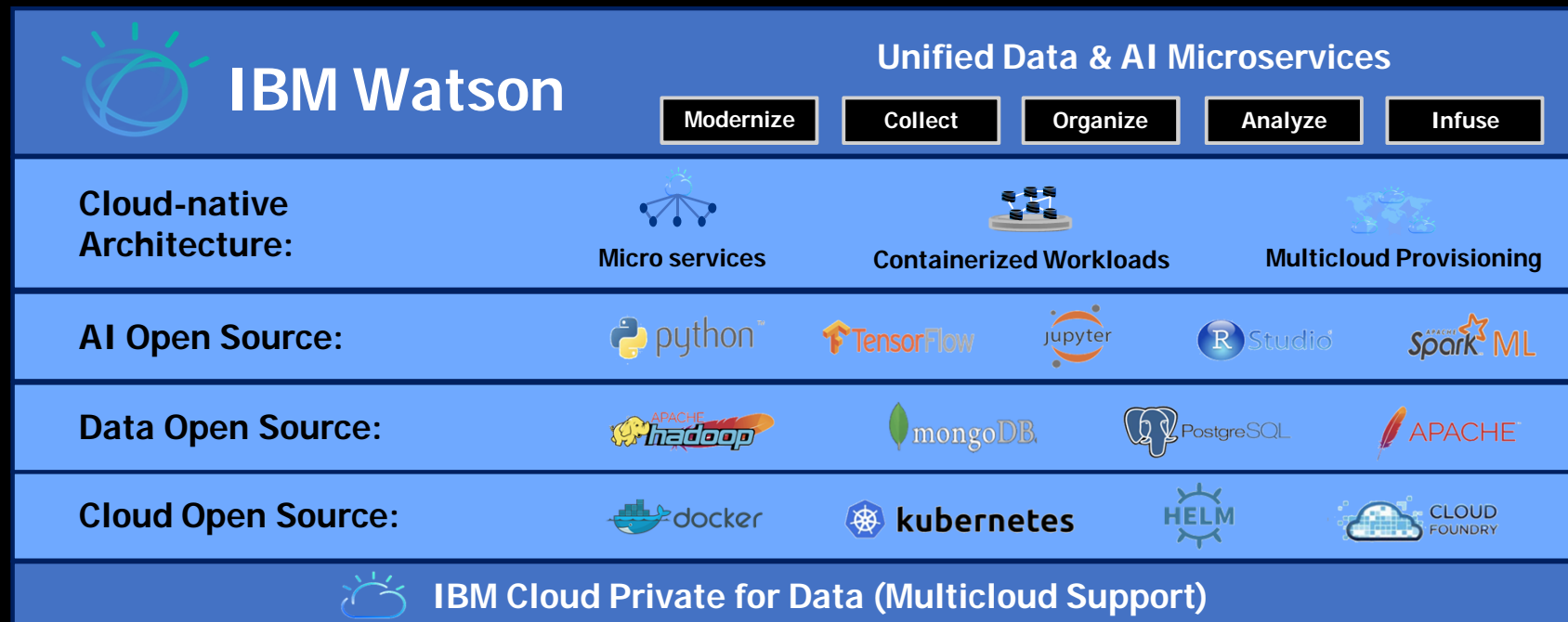


Multicloud  
Provisioning



IBM Cloud Private for Data (Multicloud)

# Open Source Meets Multicloud, Working as One



## 2) Makes your data ready for AI

**Deploy a governed platform unifying team workflows and AI models with the data they rely on**

- Automates the steps on the AI ladder
- Prepares and governs business-ready data for use in ML & AI models
- Built-in machine learning design, creation, train and deployment



### Collect Data

Collect data of every type, regardless of where it lives,



### Organize Data

Organize your data into a trusted, business-aligned analytics foundation



### Analyze Data

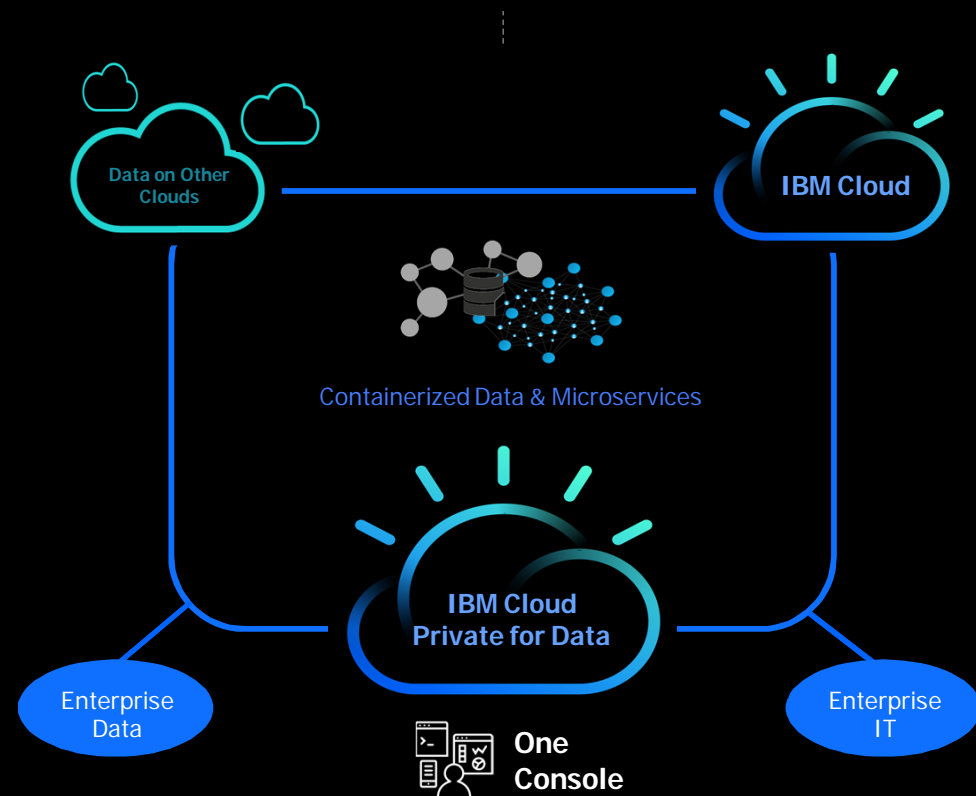
Analyze your data in smarter ways, with Machine Learning everywhere



### 3) Modernizes data for a multi-cloud world

**Leverage Cloud-native services to modernize and manage data in multicloud environments**

- One integrated console for collecting, organizing and analyzing data
- Virtualize all your data of every type, regardless of where it lives
- Build once, run anywhere across dynamically changing environments



# Watson on IBM Cloud Private for Data

## Extensible Add-on AI Services

### Watson Studio & Machine Learning

**Data Science Premium**  
Tools to design, build  
and deploy AI models



### Watson OpenScale

▪ Operational optimization services

### Watson Assistant

▪ Conversational AI services

### Watson APIs

▪ Interactive AI services (e.g., speech)

## Foundational Data & AI Services

### Cloud Private for Data

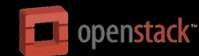
Open Source frameworks to build, train  
and deploy AI models powered by  
**Watson Studio** and **Machine Learning**



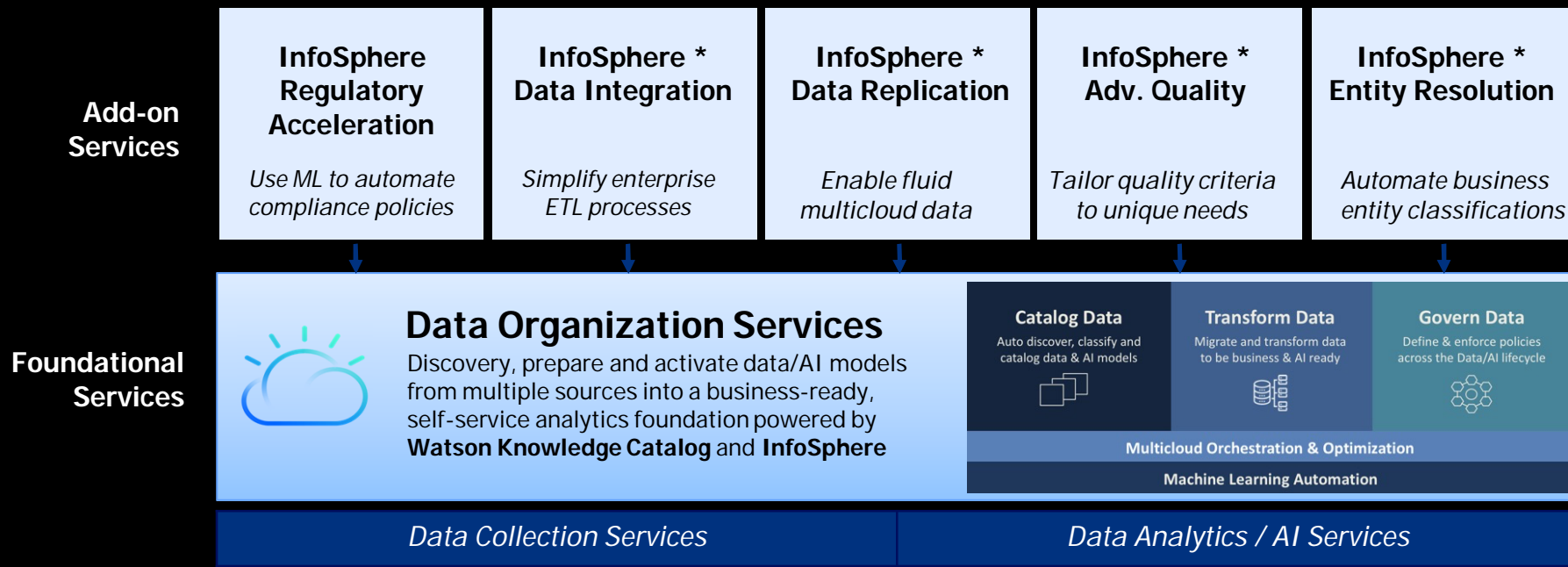
- Data visualization
- Machine learning learning
- Model build & deploy
- Model management
- Dashboards & reporting

Data Collection Services

Data Organization Services



# Unified Governance & Integration on Cloud Private for Data



\* On Roadmap Q2/Q3 2019



IBM Cloud



Google Cloud







# Thank you !

**Please visit the IBM Stand to see how  
IBM Cloud Private for Data  
Can help you to accelerate your Journey to AI**

# IBM Cloud Private for Data

## Top 4 Use Cases

### Use Cases

 <b>Manage your Data Anywhere</b>	 <b>Operationalize Data Science &amp; AI</b>	 <b>Shift to Next-Gen workloads</b>	 <b>Smarter Governance</b>
<p><b>1. Manage all your enterprise data regardless of where it lives</b> <i>(Data Virtualization)</i></p> <p><b>2. Gain control &amp; leverage your data from connected devices</b> <i>(Fast data &amp; Streaming analytics)</i></p>	<p><b>Build, deploy, manage &amp; govern models &amp; data @ scale to improve business outcomes</b> <b>(see following slides for details)</b></p> <p>e.g.</p> <ul style="list-style-type: none"><li>a. Customer Churn</li><li>b. Cross Sell / Up Sell</li><li>c. Predictive Maintenance</li></ul>	<p><b>Shift to Cloud Native</b></p> <ul style="list-style-type: none"><li>a. Provision &amp; scale in minutes</li><li>b. Build once, deploy anywhere – multi cloud support</li><li>c. Built in automation &amp; collaboration to increase productivity</li></ul>	<p><b>Governance to enable self service analytics</b></p> <p>Auto-discover meta data, manage governance rules &amp; policies, enforce privacy etc. to mitigate risk &amp; ensure compliance</p> <p>e.g. accelerate GDPR Compliance</p>

Collect Data


Solution Overview

## Solution Overview

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

**Data sources**


Data Source	Status	Description
Db2 Event Store	Premium	In-memory data store capable of extremely high speed ingest and deep, real-time analytics.
Db2 Advanced Enterprise Server Edition	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	Available	Query across many data sources as one.
MongoDB	Partner	Scalable, open source NoSQL database.



## Db2 Event Store

IBM Premium


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



IBM 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

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



IBM Data Virtualization

Available

Query across many data sources as one.

MongoDB

mongoDB

Partner

Scalable, open source NoSQL database.

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

*Sample Data Sources*

Db2

HDFS

Custom JDBC

Cloudera

Oracle

Teradata

Hortonworks

Netezza

### 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

☐ Shared

[Add remote data set](#)

Db2	HDFS	Custom JDBC	Cloudera
Oracle	Teradata	Hortonworks	Netezza

# Add data source

Data source name \*

Type data source name here

Description

Type your description here

Data source type \*

Big SQL

dashDB

D2

D2 for z/OS

Hive - HDP

HDFS - HDP

uri here

here

here

Shared

Add remote data set

# Organize Data

## Solution Overview

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

#### Discover assets

Connection \*

db3

Schemas <sup>i</sup>

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)

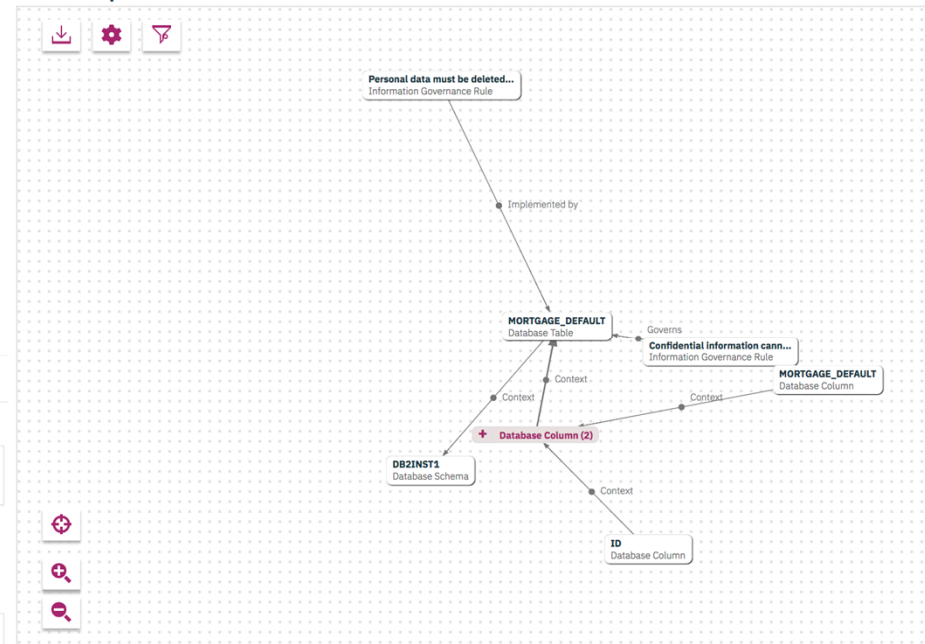
- Steward
- Data Class

##### 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 <sup>4</sup>

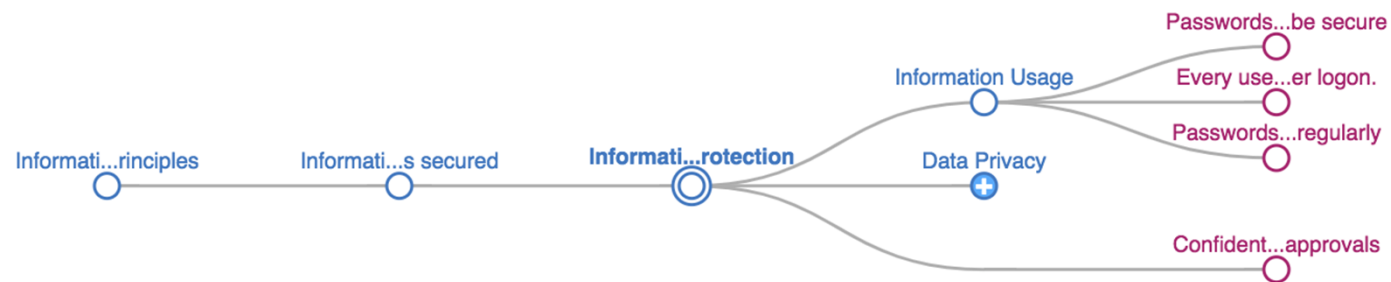
#### Relationships



# Organize Data

## 2. Solution Overview

### 2. 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	⋮
Mask a US Social Security Number by using a random method	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮
Mask a French INSEE Number by using a random method	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮
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	⋮
Authoritative Sources provide the best information	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮
Policy Enforcement	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮
Information Supply Chain Integrity	11 Sep 2018, 1:06 PM	11 Sep 2018, 1:06 PM	⋮

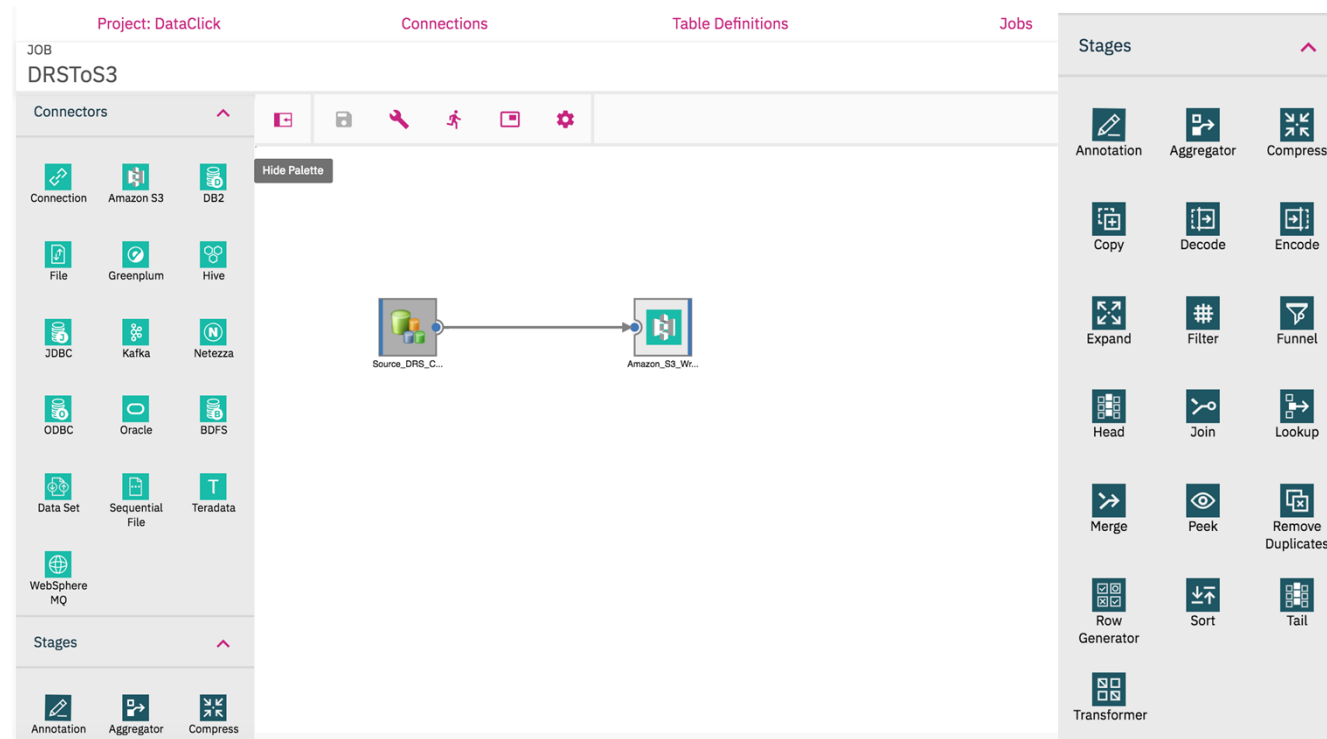
Details  
Delete



# Organize Data

## Solution Overview

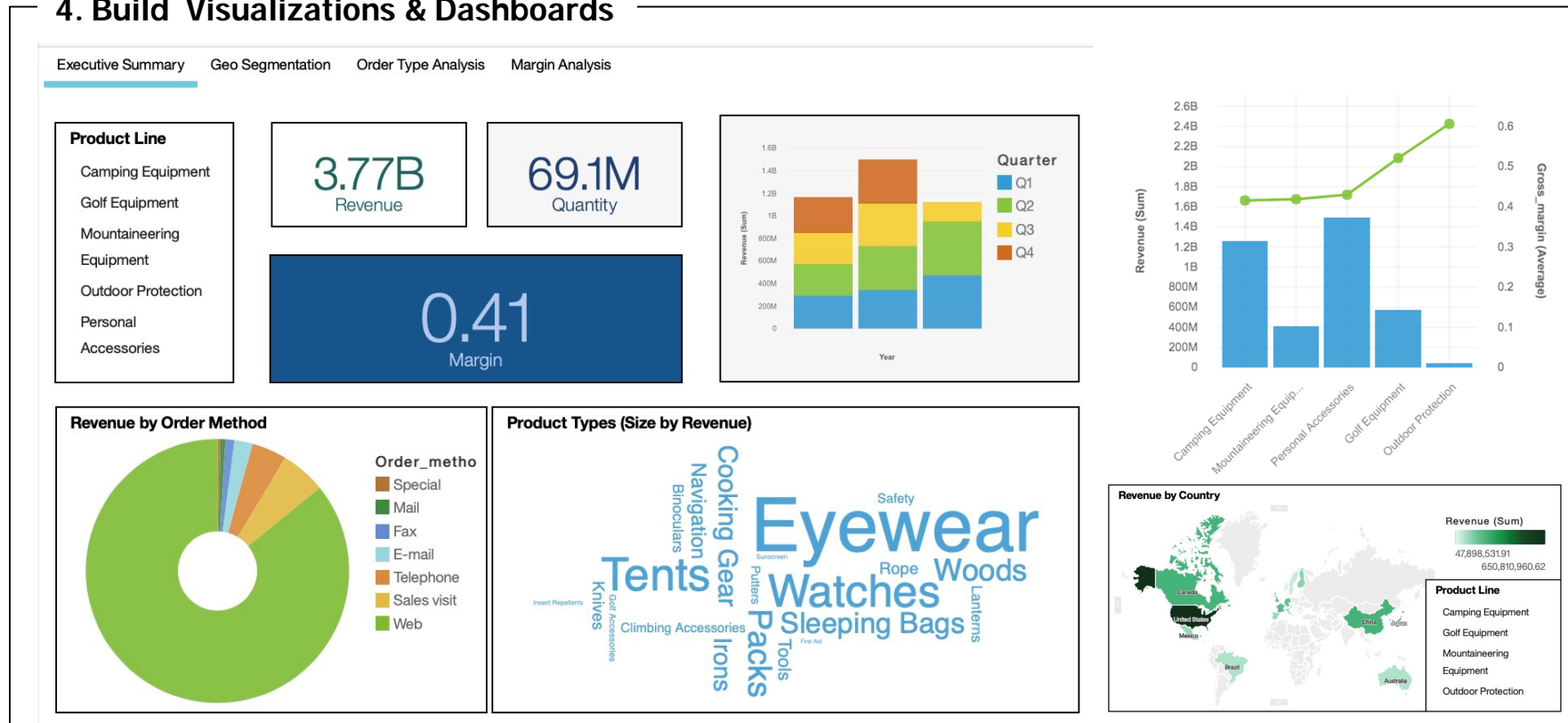
### 3. Transform & Migrate Data – Build & Execute ETL jobs @ scale



# Analyze Data

## Solution Overview

### 4. Build Visualizations & Dashboards



# Analyze Data

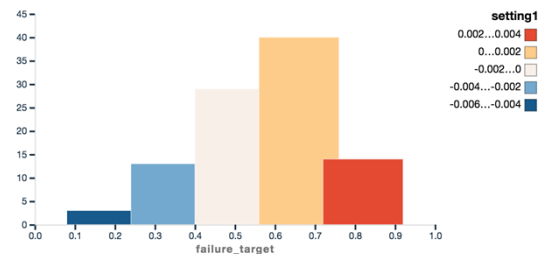
## Solution Overview

### 5. 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"] = predictions.rdd.filter(lambda x: x[labelCol] == x["prediction"]).count() * 1.0 / predictions.count()
45 evaluation["metrics"]["accuracyScore"] = evaluator.evaluate(predictions, {evaluator.metricName: "areaUnderPR"})
46 evaluation["metrics"]["areaUnderPR"] = evaluator.evaluate(predictions, {evaluator.metricName: "areaUnderPR"})
47 evaluation["metrics"]["areaUnderROC"] = evaluator.evaluate(predictions, {evaluator.metricName: "areaUnderROC"})
48 evaluation["metrics"]["threshold"] = threshold
49
50 if(evaluation["metrics"][threshold.get('metric', 'INVALID_METRIC')] >= threshold.get('mid_value', 0.70)):
51     evaluation["performance"] = "good"
52 elif(evaluation["metrics"][threshold.get('metric', 'INVALID_METRIC')] <= threshold.get('min_value', 0.25)):
53     evaluation["performance"] = "poor"
54 else:
```



Overview API

POST https://ad-130-1-master-1.sny.sbn.com:31843/dmodel/v1/sensor/predict/sensor-los/score Deployment token

### sensor-los

No description available.

TYPE	ASSET	ALLOCATED CPU	ALLOCATED MEMORY	REQUESTS
Web service	Sensor LOS v4.1	Unallocated	Unallocated	0

Request Response generate code

Function name \*  
score

Body \*

```
{
  "Input_json_str": "{"LoanAmount":12985,"Residence":"Owner","Occupier":"WentCurrentEmployer","AppliedOnline":"YES","NumberOfCards":2,"SalePrice":137000,"Income":45715,"YearCurrentAddress":8,"CODes":["7732","Loan":12,"70","300262","Location":100]}",
  "score": 0.75
}
```

#### Project releases 1

● Sensor-IoT-Online

CREATOR  
admin

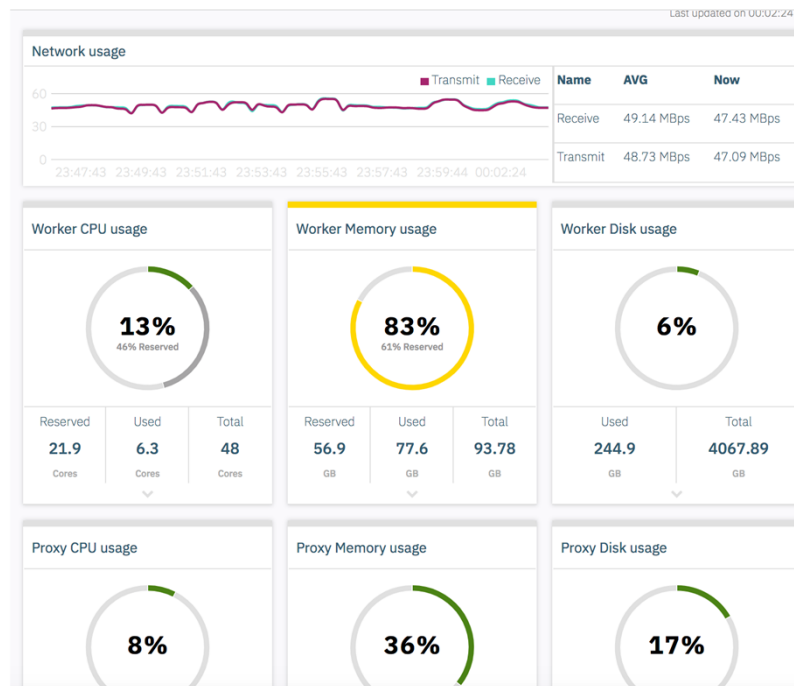
SOURCE  
Oil and Gas D0model

DEPLOYMENTS  
1

LAST MODIFIED  
12 Sep 2018, 4:48 PM

## 6. A single Pane to administer &amp; monitor your data &amp; analytics workloads

Admin Dashboard



Pod list

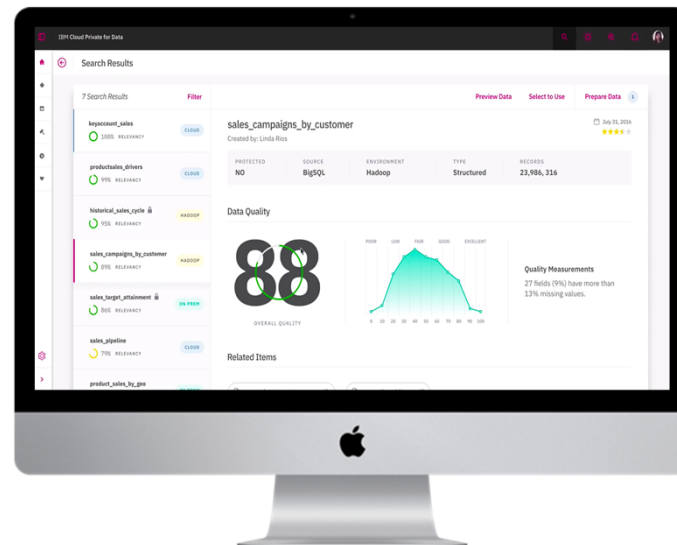
POD NAME	STATUS	NODE NAME	USED CPU	USED MEMORY	USED DISK
auth-trail-service-69c989f6d-vgm7f	Running	172.16.204.201	0.00	286MB	0B
auth-apkeys-6mrlg	Running	172.16.204.11	0.01	604MB	0B
auth-apkeys-w7mpt	Running	172.16.203.175	0.01	452MB	0B
auth-apkeys-xd2g2	Running	172.16.203.176	0.01	442MB	0B
auth-tdp-72zrh	Running	172.16.203.175	0.01	642MB	0B
auth-tdp-6mglb	Running	172.16.203.176	0.01	687MB	0B
auth-tdp-6tkk	Running	172.16.204.11	0.00	808MB	0B
auth-pap-4jy2	Running	172.16.203.175	0.00	195MB	0B
auth-pap-6fnd2	Running	172.16.204.11	0.00	203MB	0B
auth-pap-wcscs	Running	172.16.203.176	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	—	—	—	<div><div></div></div> ⋮
RStudio with R 3.4.3	Environment	admin	Oil and Gas		12 Sep 2018, 8:54 AM	—	—	—	<div><div></div></div> ⋮
Data Refinery	Environment	admin	Oil and Gas		12 Sep 2018, 8:59 AM	—	—	—	<div><div></div></div> ⋮
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	—	—	—	<div><div></div></div> ⋮
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	—	—	—	<div><div></div></div> ⋮

# Product Demo

## Solution Overview



<http://ibm.biz/experienceICP4D>