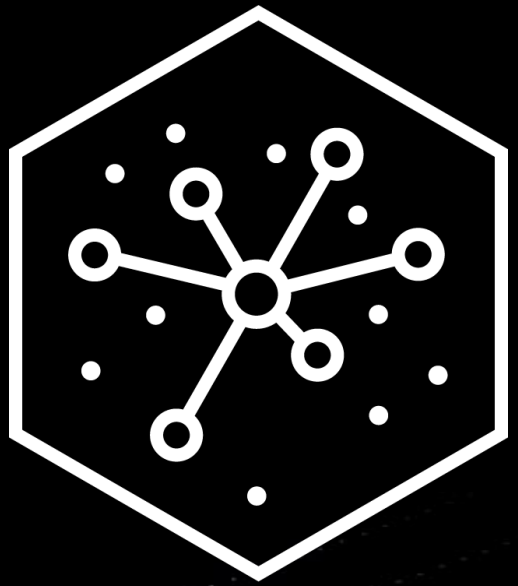
IBM Journey to Cloud and AI Analytics Modernization Workshop



Featuring: Cloud Pak for Data 3.0.1

Starts at 9:00am EST

August 27, 2020

IBM Analytics Modernization Workshop
Agenda

Part 1	• Introduction • Business Use Case	• Lab 01 • Lab 02
Part 2	• Collect: Connect • Organize • Collect: Virtualize	• Lab 03 • Lab 13 • Lab 05
Part 3	• Analyze • Deploy • Infuse – OpenScale • Infuse – Cognos Analytics • Wrap up	• Lab 06 • Lab 07 • Lab 08 • Lab 09 • Lab 10

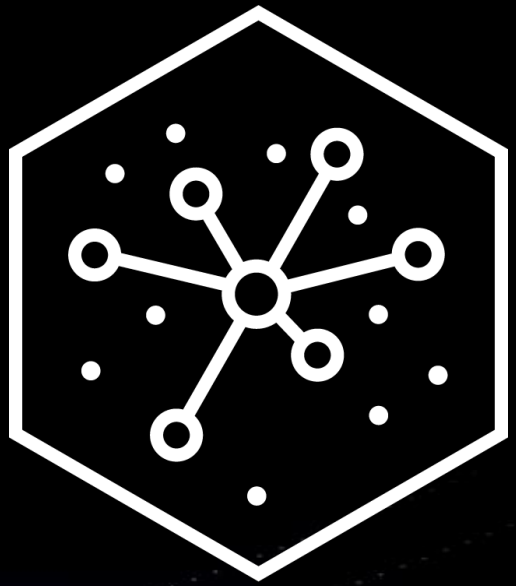
IBM Analytics Modernization Workshop

Part 1

• Introduction • Business Use Case	• Lab 01 • Lab 02
• Collect: Connect • Organize • Collect: Virtualize	• Lab 03 • Lab 13 • Lab 05
• Analyze • Deploy • Infuse – OpenScale • Infuse – Cognos Analytics • Wrap up	• Lab 06 • Lab 07 • Lab 08 • Lab 09 • Lab 10

Introduction

Lab 01 – Getting Started

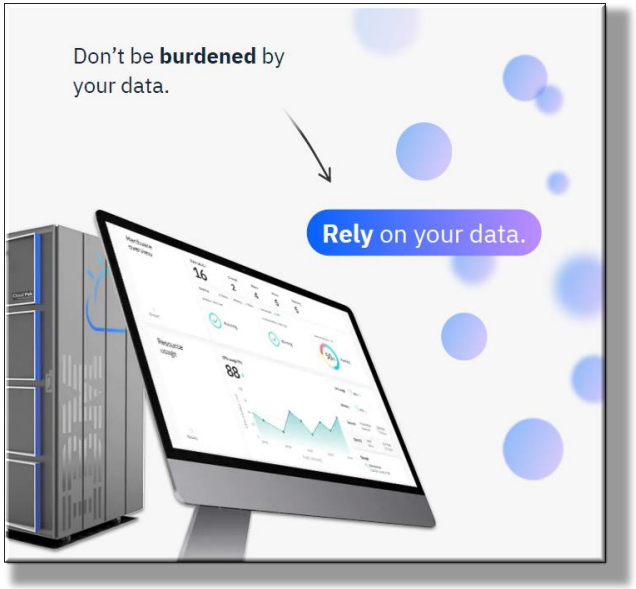


IBM Cloud Pak for Data

IBM Cloud Pak for Data is a single unified, integrated platform which helps to simplify the collection, organization and analysis of data.

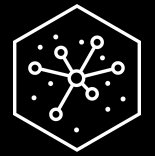
With it, enterprises can turn data into insights through an integrated cloud-native architecture.

IBM Cloud Pak for Data is extensible and easily customized to unique client data and AI landscapes through an integrated catalog of IBM, open source, and third-party microservices.



Why IBM Cloud Pak for Data?

Customer Challenges in creating value from data



"We have been accumulating data at amazing pace for years, but are doing nothing with it."

"We can't understand why it is so hard to deliver measurable business insight?"

"Over 60% of our time is spent finding the right data, obtaining access to it, verifying it and massaging it so that it can be utilized by the LOB – a bottleneck LOBs can no longer afford."

"We have an increased demand for self-service – for increased number of Data Scientists and Business Analysts that are all using their own tools with no oversight – efforts are duplicated, costs are driven up!"

"We need the capabilities to automate production tasks into a single experience for DevOps, Data Management and Data Scientists."

"How can we provide a true 'Shop-for-Data' and 'Self-Service' experience?"

"How do we monitor the success of what models and dashboards we have deployed?"

"How do we connect our disparate data sources without having to make more and more copies of the data?"

"We have no way to measure the quality of our data!"

"We are not doing well when it comes to putting models into production and managing the lifecycle of those models."

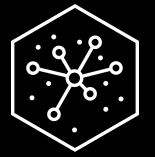
"We have a huge backlog of analytic projects that are not making it into production."

IBM: Companies have to innovate to ward-off potential disruptions and keep pace with the competitors – Data Scientists are everywhere – but are not being provided the data they need!

IBM: Governance is not only important due to security concerns – but lack of governance can lead to faulty data fed to the models – that can be a disaster to projects, reputation, and the business as a whole.

IBM: Productionalizing models and the model management lifecycle are critical to long term success.

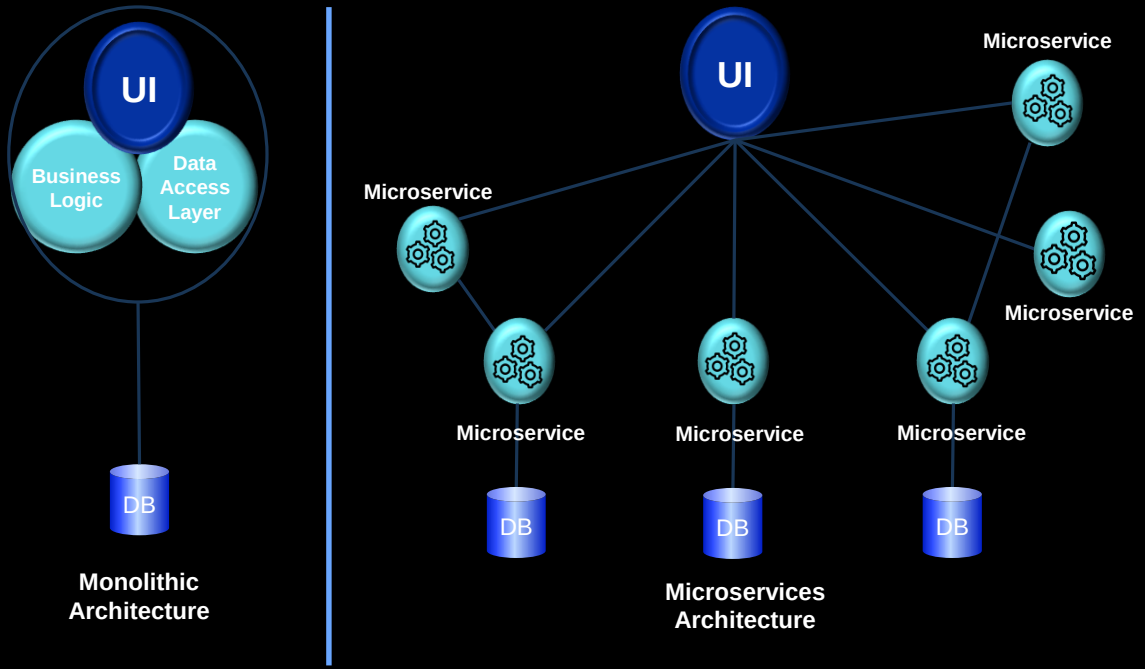
Considerations for Cloud Pak for Data



- Integrated Multi-modal platform
  - Use tool of choice and collaborate via project entities
  - Code/Click Options
  - All Analytics – Dashboard, Predictive, Prescriptive
  - All Data
  - Seamless user experience
- Hybrid Cloud
  - Cloud native architecture
  - Cloud agnostic – any vendor cloud or data center
  - Scalable data and analytic services
  - Flexibility to move data science to the data.
- Operationalize Machine Learning
  - Ease and flexibility of deployment at enterprise scale
  - Advanced model management capabilities.
  - Monitoring model performance
- Governance
  - Omnipresent, yet invisible – infused throughout
  - Data automatically integrated with governance capability for auto-discovery, catalog, and search subject to policies and rules
- Automate, Automate, Automate

Microservices – the first key to cloud native applications

Making development & deployment more efficient



Microservices benefits *

- ***Improved fault isolation:***

  Larger applications can remain largely unaffected by the failure of a single module

- ***Technological flexibility:***

  Try out a new technology stack on an individual service and roll it back if required

- ***Easier development:***

  A new developer can more easily understand the functionality of a service

- ***Optimized deployment:***

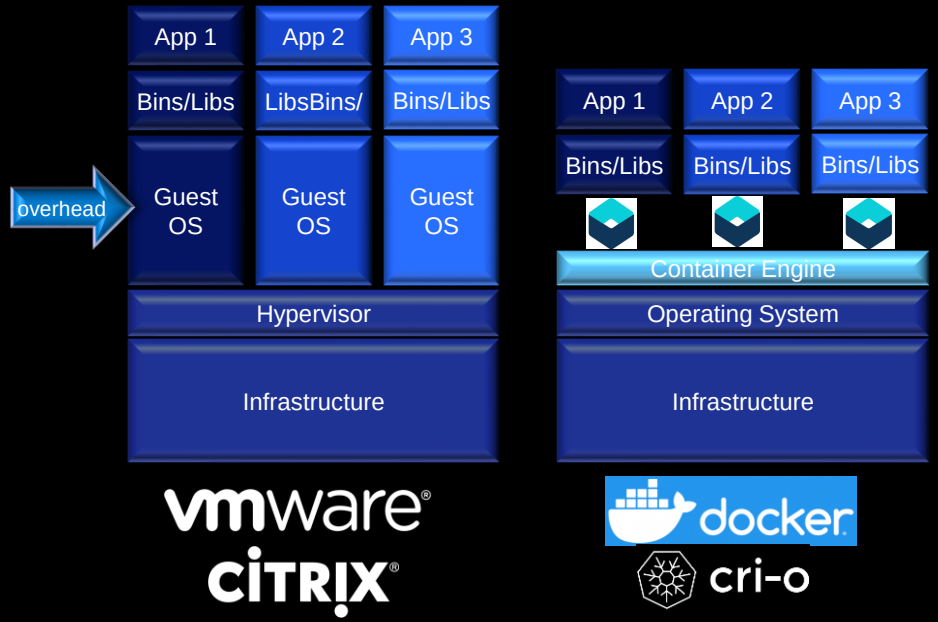
  Auto provision, auto scale and provide auto-redundancy

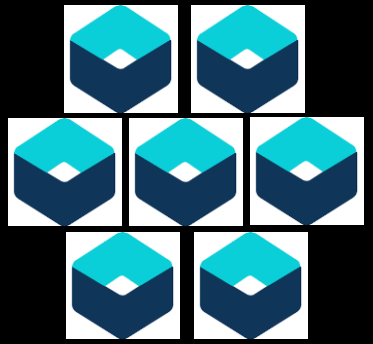
* This is not a claim that a microservice-based application approach is always better for every use case scenario

Containers – the second key to cloud native applications

Reducing operational and development costs

Virtual machines vs. Containers *





Containers can be 2 – 3 times more resource efficient than virtual machines

On average Docker developers ship software 7x more frequently

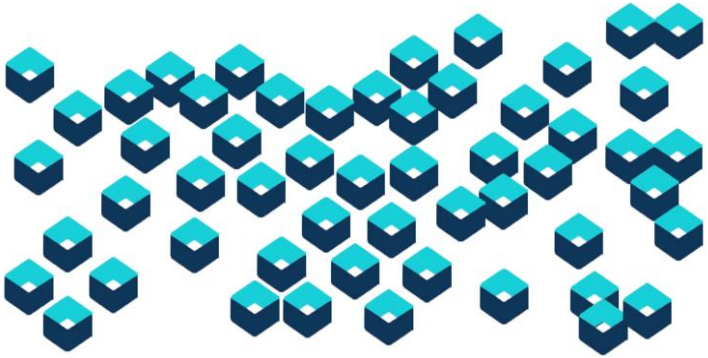
* Containers virtual software in the way that virtual machines have virtualized hardware

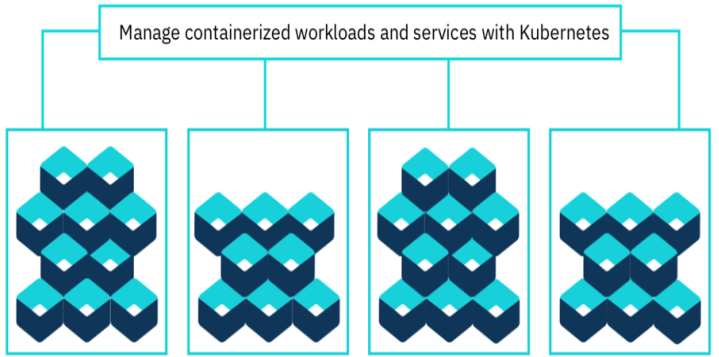
Container automation and orchestration is essential
Enter: Kubernetes

kubernetes

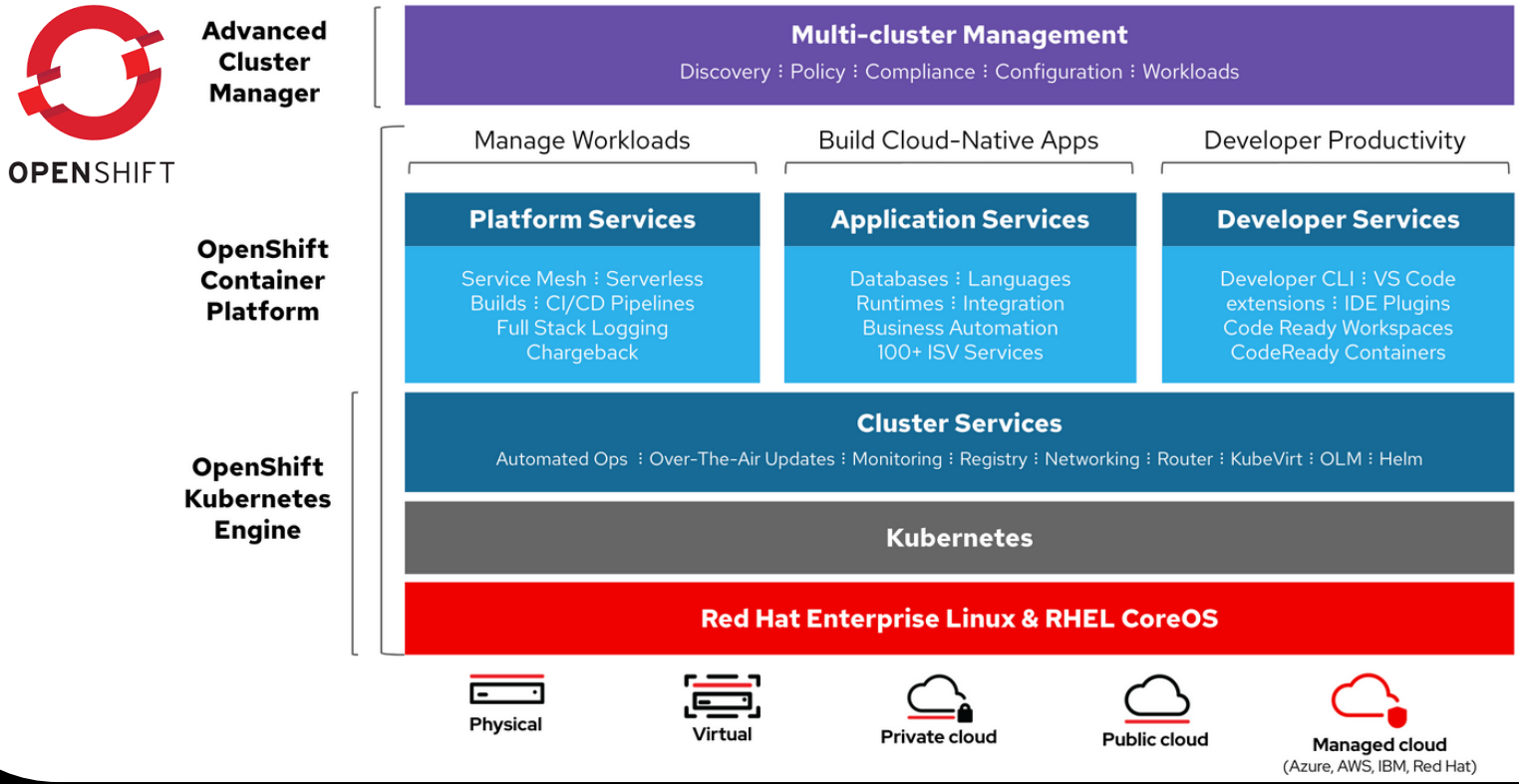
Containers are revolutionizing IT
But they require orchestration

Kubernetes - κυβερνήτης
Means “helmsman” or “pilot”





Red Hat OpenShift
Enterprise Kubernetes Platform



Cloud Pak for Data (CPD)
Make your data ready for AI

There is no **AI** without **IA**

Infuse – Deploy trusted AI-driven business processes

Analyze – Scale insights with ML everywhere

Organize – Create a trusted analytics foundation

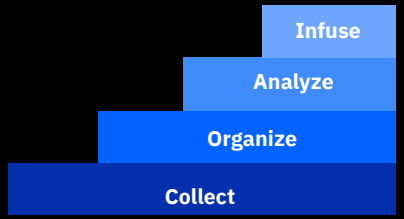
Collect – Make data simple & accessible

Strong Foundation – Built on "Cloud native architecture"

IBM Cloud Pak for Data

Unified, modular, deployable anywhere

The AI Ladder



App Developers and SREs| Business Partners | Data Engineers | Data Stewards | Data Scientists | Business Users

Integrated User Experience

Extensible: APIs, partner ecosystem, and solutions

Collect

- Provision SQL and NoSQL databases
- Create Connections
- Data virtualization
- ...

Organize

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

Analyze and Infuse

- Business reporting
- Data science and visualization
- AI lifecycle automation
- AI Apps
- Industry accelerators
- ...

Core Services

- User access management
- Security contexts and RBAC
- Volume management
- Monitoring and metering
- Service provisioning
- Operators
- Diagnostics
- Backup and migrate

Red Hat OpenShift

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

CPD Services Ecosystem

1. Cloud Pak for Data Base Services

Collect		Organize	Analyze		Deploy / Infuse
Data Virtualization		Watson Knowledge Catalog (With IA, IGC, Refinery, InstaScan)	Watson Studio	Analytics Engine	Data Science: Model Design & Deployment
Db2 Warehouse	PostgreSQL		Dashboards	IBM Streams	Watson OpenScale
Db2 Event Store	IBM Streams	Open Source Management	Industry Accelerators (many)		Watson Machine Learning
Db2 Big SQL	NPS	Infosphere Regulatory Accelerator	Watson Machine Learning		

OpenShift / Control Plane (Lite)

2. Premium Services *(purchase license or BYOL)*

Collect	Organize	Analyze	Deploy / Infuse
Db2 AESE Virtual Data Pipeline	Infosphere DataStage Edition Infosphere multi-cloud Data Mvmnt Master Data Management	Cognos Analytics SPSS Modeler Decision Optimization Planning Analytics Hadoop Execution Engine	Watson Assistant / Discovery Watson API Kit (Speech to Text, Text to Speech, Natural Language Understanding) Watson Financial Crimes Insights Planning Analytics

3. Third Party Extension Services

mongoDB. · Datameer · accelerite · Prolifics Global Technology Solutions Provider · figure eight · Lightbend · Cockroach LAB · NetApp portworx · EDB POSTGRES · wand · OA4ERP Operational Analytics for ERP · PlanetScale · DR · Senzing · knowis · ANALYTICS ZOO · OA4ERP

Cloud Pak for Data
Supports the AI Lifecycle

Connect and **discover** content from multiple data sources across your organization.

Provision databases and **virtualize** data access.

Grant user access levels and enforce **business policies**.

Index for search, visualize assets with **lineage, metrics, and quality profiles**.

Find data and analytics assets in the **Enterprise Catalog**.

Understand, cleanse and prepare your data to create data preparation pipelines visually.

Use popular open source libraries to prepare structured and unstructured data

Create Machine Learning, Deep Learning, Optimization, and other advanced mathematical models.

Design your models **programmatically** or **visually** with popular **open source** tooling and IBM frameworks.

Train at scale with support for **distributed** compute and **GPUs**.

Manage your models across **non-prod and prod**.

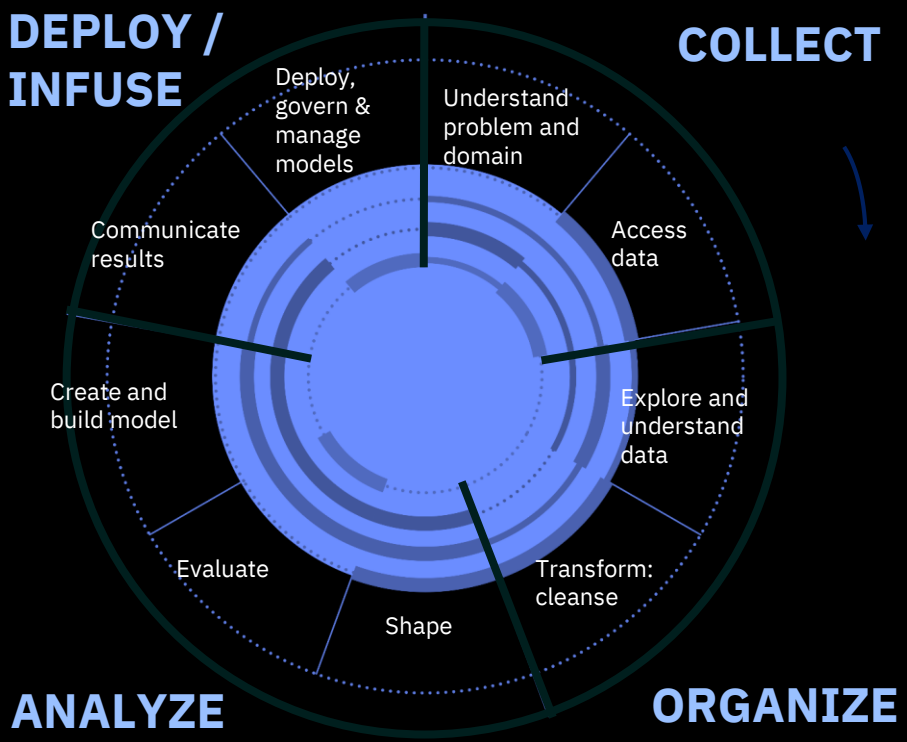
Deploy your models and scale automatically for online, batch or streaming use cases with **SLAs**.

Monitor model **performance** and **automatically trigger retraining** and redeployment as rolling upgrades.

Incorporate **trusted and governed models** into applications, dashboards, and operational systems

Cloud Pak for Data (CPD)

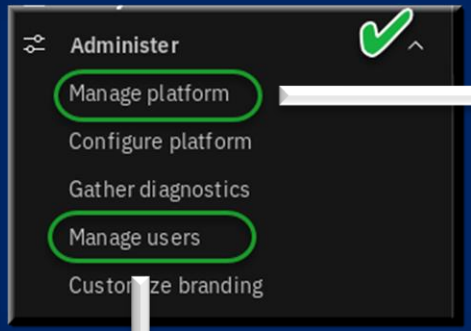
Increases workforce productivity across the analytics lifecycle



Administrator / Architect	Data Engineer	Data Steward
Ensures the usability of the compute, network, storage, etc.	Architects data pipelines & ensures operability	Governs data and ensures regulatory compliance
		

Business Analyst	Data Scientist	Application Developer
Works with data to apply insights to business strategy	Dives deep into the data to draw insights for the business	Plugs into analysis and code to build applications
		

CPD Administration
Administer and manage the platform



Deployments

See the deployments running in each Red Hat OpenShift project on the platform. You can see how many resources each deployment is currently using compared to the number of resources that the deployment has reserved.

Filter by: All types | Clear all

Find deployments

Name	Type	Installed on	Service instances	vCPU	Memory (GB)
Total			2	6.01 of 75.08	101.20 of 201.70
IBM Cloud Pak for Data Cognos Analytics Advanced	Premium	5/28/2020 4:32 PM	0	0.00 of 0.60	0.11 of 0.11
SPSS Modeler for Watson Studio	Premium	5/28/2020 4:32 PM	0	1.45 of 2.10	0.42 of 8.40
Db2 Warehouse	Base	5/28/2020 4:32 PM	1	0.67 of 10.15	16.80 of 44.10

Name	User ID	Username
admin	1000330999	admin
Business Analyst	1000331009	businessanalyst
CPD User	1000331002	cpduser
Data Engineer	1000331003	dataengineer

Name	Pods	vCPU	Memory (GB)
Total	8	0.14 of 1.65	1.50 of 2.80
db2wh-1590588027600-ibm-unified-console-api	5	0.10 of 1.00	1.30 of 2.00
db2wh-1590588027600-ibm-unified-console-influxdb	1	0.00 of 0.10	0.08 of 0.25
db2wh-1590588027600-ibm-unified-console-ucgoapi	1	0.00 of 0.25	0.07 of 0.10

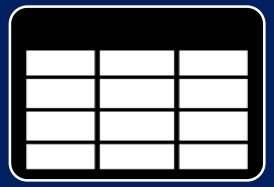
Business Use Case

Lab 02 – Business Use Case: Customer Churn

Trade Co. Challenges

- Customer retention problem leading to declining revenue
- Underperforming rules-based system to identify separation (churn) risk
- Lack of centralized, vetted, and reliable data to ensure accuracy of analytics
- Disparate analytical tools for reporting and model development
- No simple way to infuse machine learning models into the customer facing Stock Trader Application

Separation (Churn) Risk: Current Rules Based System



Built Using Limited Data

Rules are developed using a single source of data that contains customer demographic information.

Manual Process to Develop Rules

Rules are manually developed based on the past experience of the marketing team. Rules are only updated once a year.

Low Overall Predictive Accuracy

Low overall predictive accuracy. We are both missing identifying customers who ultimately separate and incorrectly assigning high risk to customers who ultimately stay.

Separation (Churn) Risk: New Data Driven Approach

Incorporate Multiple Data Sources

Use vetted centralized transactional data along with customer demographics to understand separation behavior. Also, include the outcomes of the rules-based system for each customer where an accurate prediction was rendered.

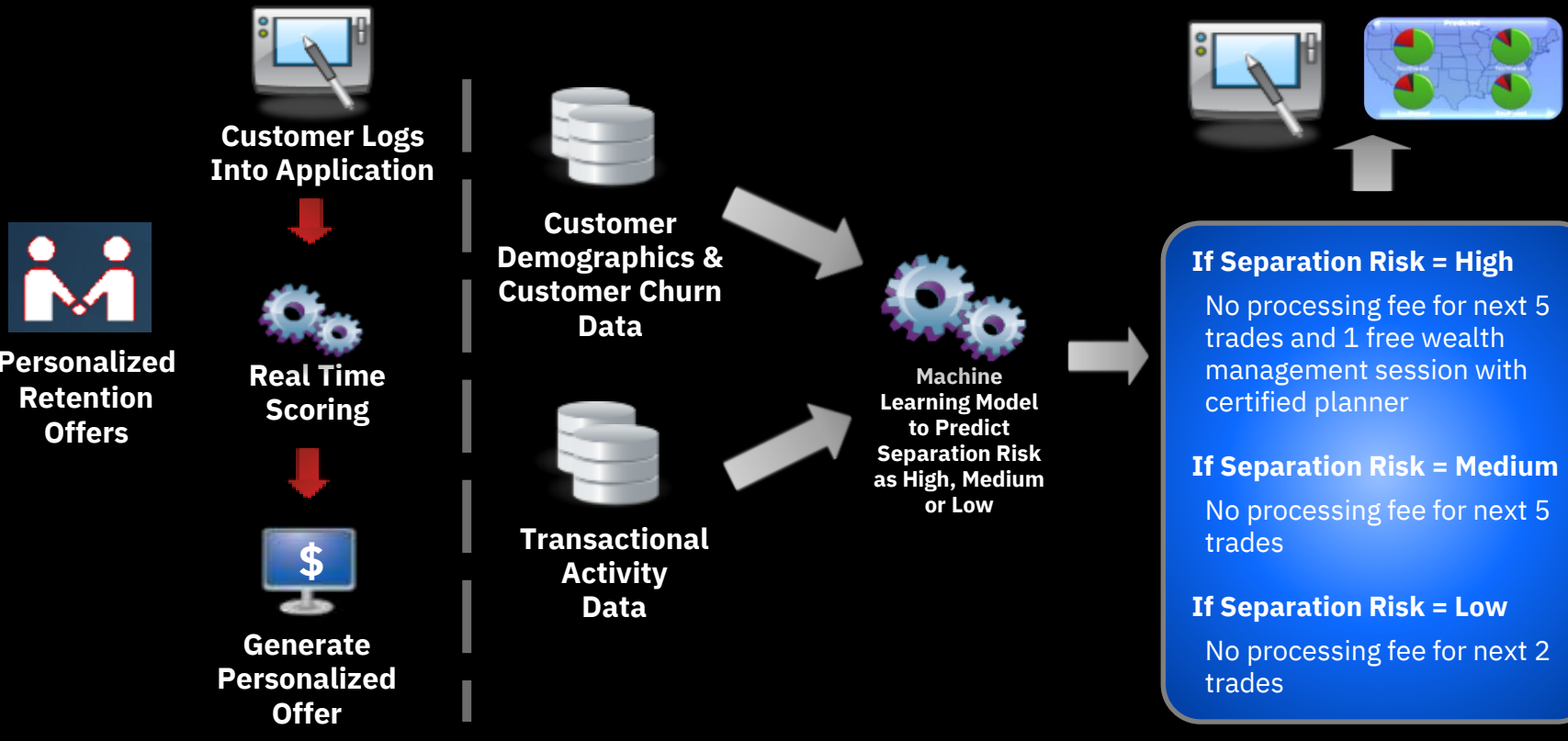
Data Driven Process to Develop Machine Learning Models

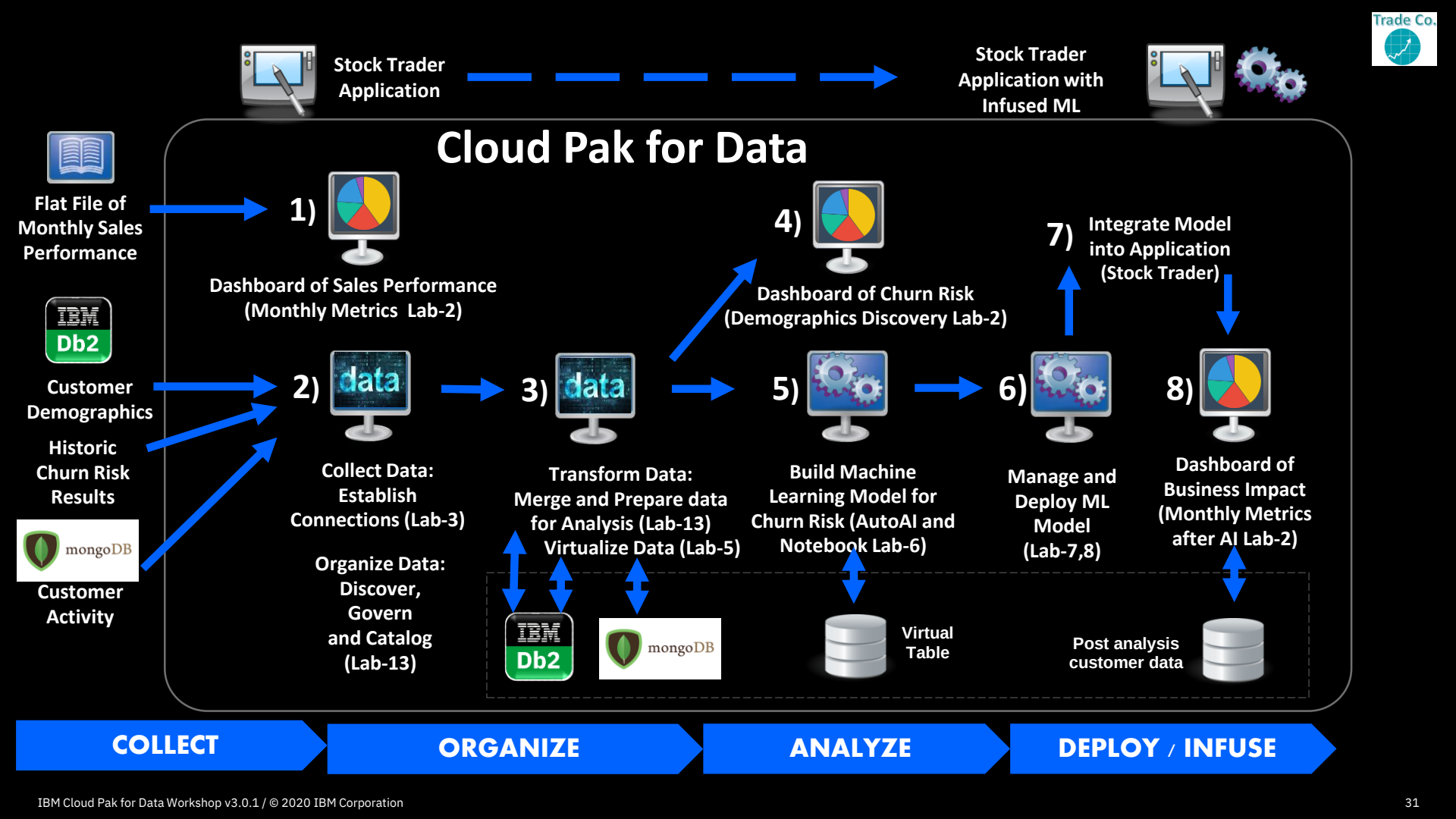
Develop predictive models for separation risk that automatically discover and incorporate all the patterns in the data including interactions and contingencies.

High Accuracy from Adaptive Machine Learning

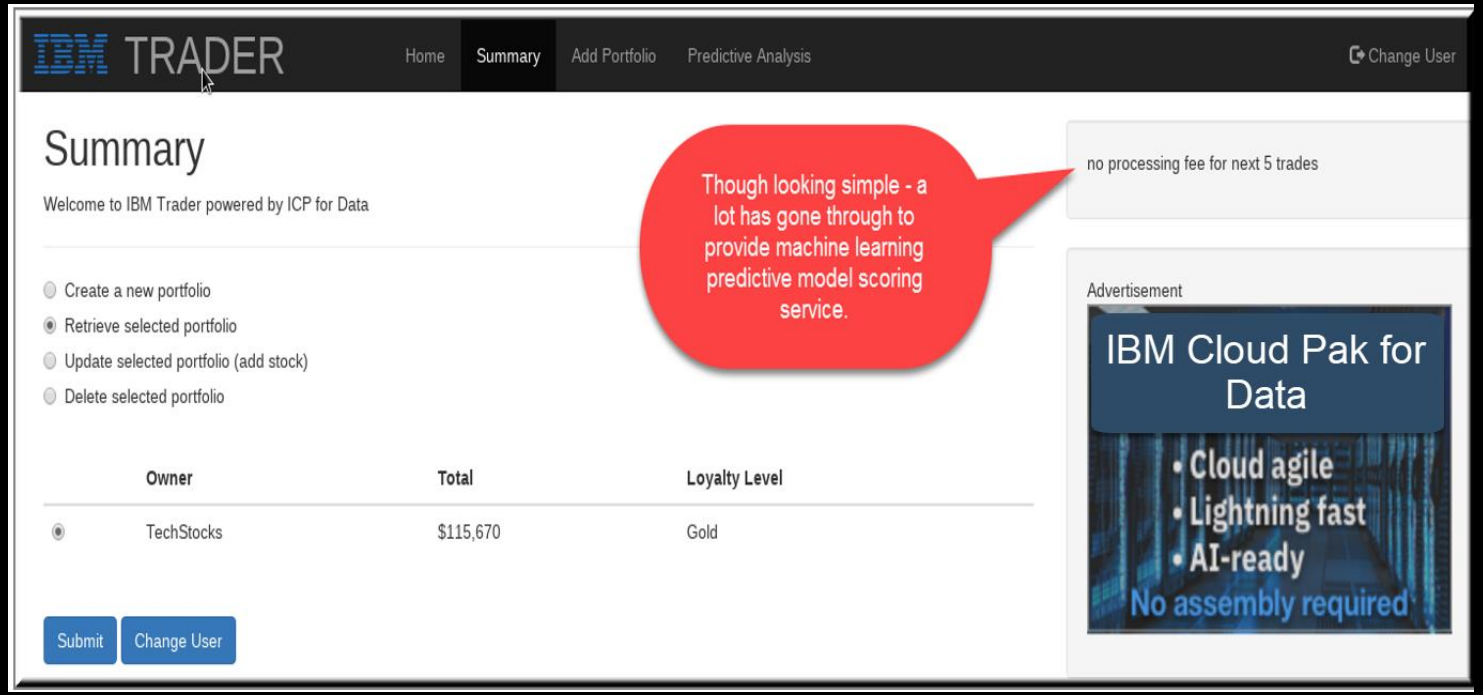
Models will classify separation risk with a higher overall accuracy and will adapt to changing patterns in risk to maintain that accuracy. Machine Learning models will incorporate all the understanding from the rules-based system and build on that to develop highly complex set of predictive conditions.

Deployment: Stock Trader App. Integrated with AI




Stock Trader Application
Stock Trader Application with Infused ML
Cloud Pak for Data
Flat File of Monthly Sales Performance
1)
Dashboard of Sales Performance (Monthly Metrics Lab-2)
4)
Dashboard of Churn Risk (Demographics Discovery Lab-2)
7)
Integrate Model into Application (Stock Trader)
Customer Demographics
Historic Churn Risk Results
Customer Activity
2)
Collect Data: Establish Connections (Lab-3)
Organize Data: Discover, Govern and Catalog (Lab-13)
3)
Transform Data: Merge and Prepare data for Analysis (Lab-13) Virtualize Data (Lab-5)
5)
Build Machine Learning Model for Churn Risk (AutoAI and Notebook Lab-6)
6)
Manage and Deploy ML Model (Lab-7,8)
8)
Dashboard of Business Impact (Monthly Metrics after AI Lab-2)
Virtual Table
Post analysis customer data
COLLECT
ORGANIZE
ANALYZE
DEPLOY / INFUSE

Stock Trader – After
Monetizing the ML model

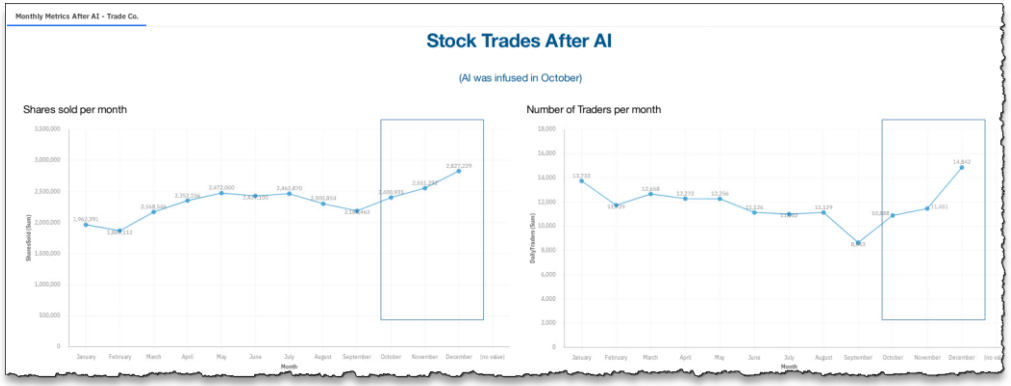


Trade Co. Dashboards

Before and After deploying the CPD developed ML model

Before AI

After AI



Lab-01: Getting Started

Review Home Page options

Manage Users
- ✓ Roles
- ✓ Configure LDAP

Manage Platform ** (not working)
- ✓ List of Deployments
- ✓ DB2 Service
    - ✓ Number of Pods
    - ✓ Number of Service Instances

Review Profile and Settings
- ✓ Git Integration

Review Services
- ✓ Watson OpenScale

Review My Instances
- ✓ Provisioned Instances
- ✓ Environments
- ✓ Jobs

Custom Branding

Lab-02: Business Case

Review Monthly Metrics

Review Demographics Discovery

Review Monthly Metrics After AI

IBM Analytics Modernization Workshop

Part 1

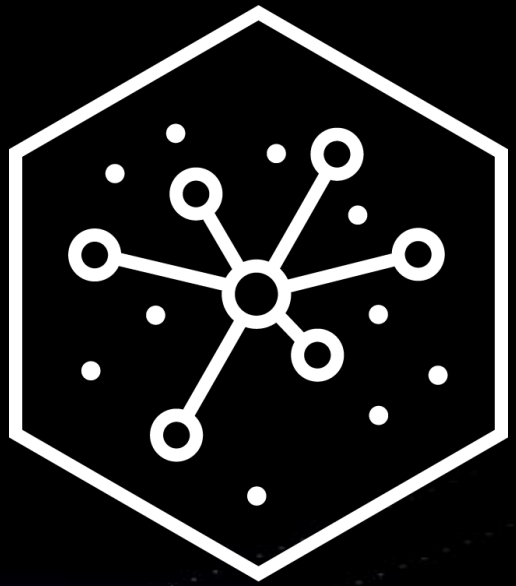
• Introduction • Business Use Case	• Lab 01 • Lab 02
• Collect: Connect • Organize • Collect: Virtualize	• Lab 03 • Lab 13 • Lab 05
• Analyze • Deploy • Infuse – OpenScale • Infuse – Cognos Analytics • Wrap up	• Lab 06 • Lab 07 • Lab 08 • Lab 09 • Lab 10

Cloud Pak for Data

We will return for review at 10:45 am. Please work on Lab-1 and Lab-2

Collect

Lab 03 – Collect: Connections



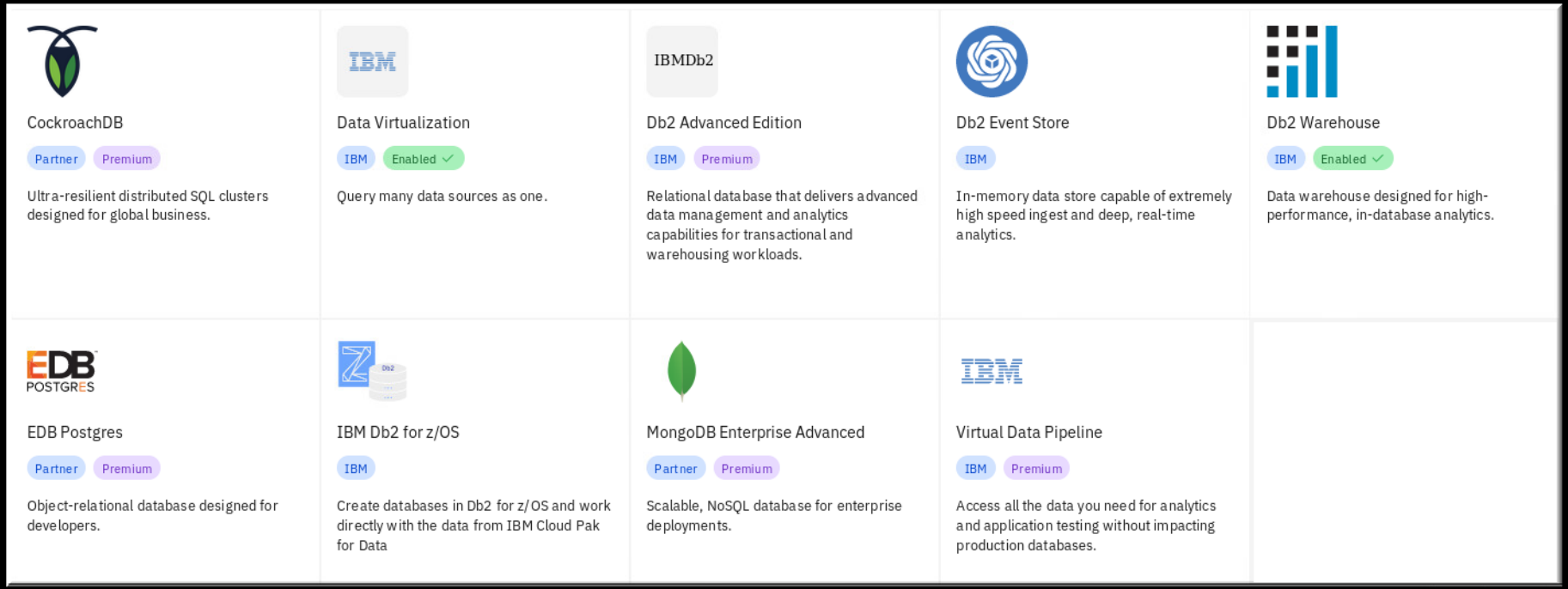
IBM Analytics Modernization Workshop

Part 2

• Introduction	• Lab 01
• Business Use Case	• Lab 02
• Collect: Connect	• Lab 03
• Organize	• Lab 04
• Collect: Virtualize	• Lab 05
• Analyze	• Lab 06
• Deploy	• Lab 07
• Infuse – OpenScale	• Lab 08
• Infuse – Cognos Analytics	• Lab 09
• Wrap up	• Lab 10

CPD Collect
1. Provision in-cluster databases

Provision, host, and manage these data sources directly on the CPD cluster



CPD Collect

2. Connect to existing data sources: IBM *

* Note: This list is constantly updated and shows what exists as of August 27, 2020.

Connect directly to these data sources and perform the CPD component functionality shown

IBM Data Sources	Cognos Dashboards	DataStage Edition	Data Virtualization	WKC	Watson Studio
Analytics Engine HDFS				✔	✔
Classic Federation		✔			
Cloud object storage (IBM)		✔		✔	✔
Cloud object storage (infra)				✔	✔
Cloudant				✔	✔
Cognos Analytics				✔	✔
Compose for MySQL				✔	✔
Data Set		✔			
Data Virtualization	✔			✔	✔
Data Virtualization Mgr z/OS			✔		
PostgreSQL databases				✔	✔
Db2		✔	✔	✔	✔
Db2 Big SQL			✔	✔	✔
Db2 Event Store			✔	✔	
Db2 for i			✔	✔	✔
Db2 for z/OS		✔	✔	✔	✔
Db2 Hosted				✔	✔
Db2 on Cloud		✔	✔	✔	✔
Db2 Warehouse	✔	✔	✔	✔	✔
Db2 Warehouse on Cloud	✔	✔	✔	✔	✔
Distributed Transactions		✔			
DRS		✔			

IBM Data Sources	Cognos Dashboards	DataStage Edition	Data Virtualization	WKC	Watson Studio
External Source		✔			
External Target		✔			
HDFS via Hadoop				✔	✔
Hierarchical		✔			
Hive via Hadoop				✔	✔
Impala via Engine for Hadoop				✔	✔
Informix		✔	✔	✔	✔
Informix Enterprise / Load		✔			
ISD Input / Output		✔			
Java Integration		✔			
Lookup File Set		✔			
Netezza		✔	✔		
Planning Analytics				✔	✔
Obj. Strg. OpenStack Swift				✔	✔
PureData for Analytics				✔	✔
WebSphere MQ		✔			
Z/os DVM sources (VSAM, IMS, Adabas, etc.)			✔		

CPD Collect

2. Connect to existing data sources: 3rd-party *

* Note: This list is constantly updated and shows what exists as of August 27, 2020.

Connect directly to these data sources and perform the CPD component functionality shown

Third-party Data Sources	Cognos Dashboards	DataStage Edition	Data Virtualization	WKC	Watson Studio
Amazon Redshift			✔	✔	✔
Amazon S3		✔		✔	✔
Apache Cassandra		✔			
Apache Derby			✔		
Apache Hbase		✔			
Apache HDFS				✔	✔
Apache Hive			✔	✔	✔
Apache Kafka		✔			
Azure Storage		✔			
BDFS		✔			
Cloudera Impala			✔	✔	✔
Dropbox				✔	✔
Filesystem		✔		✔	✔
FTP Enterprise		✔			
FTP		✔		✔	✔
Generic JDBC				✔	✔
Google BigQuery		✔	✔	✔	✔
Google Cloud Storage		✔		✔	✔
HDFS Generic web-HDFS				✔	
HDFS HttpFS				✔	
Hive JDBC		✔	✔		
Hive JDBC CDH		✔		✔	
Hive JDBC HDP		✔		✔	
Hortonworks HDFS				✔	✔

Third-party Data Sources	Cognos Dashboards	DataStage Edition	Data Virtualization	WKC	Watson Studio
Looker				✔	✔
MariaDB			✔		
MSFT Azure Blob and File		✔			
MSFT Azure Lake Store				✔	✔
MSFT Azure SQL DB				✔	✔
MSFT SQL Server	✔	✔	✔	✔	✔
Minio				✔	✔
Mongo			✔	✔	
MySQL			✔	✔	✔
ODBC		✔			
OData				✔	✔
Oracle		✔	✔	✔	✔
Pivotal Greenplum		✔		✔	✔
PostgreSQL	✔		✔	✔	✔
Salesforce.com		✔		✔	✔
SAP HANA			✔		
SAP Data Object		✔		✔	✔
Snowflake		✔	✔	✔	✔
Sybase Enterprise		✔	✔	✔	✔
Sybase IQ / OC		✔		✔	✔
Tableau				✔	✔
Teradata		✔	✔	✔	✔

CPD Collect
3. External Data Sets

The Weather Company

Historical Weather Data

- 3 Years of Historical Weather Data
- Current Weather Conditions
- Weather Forecast Data
- Location Look-up Services
- Industry Accelerators (Retail & Manufacturing)

90-day trial

Equifax

Demographic Consumer Data

- Ability to Pay
- Economic Cohorts
- Credit Styles Pro
- Income 360
- Wealth Complete Data
- IXI-Data

Premium

People Data Labs

People Data

- Dataset of ~1.5B profiles consisting of both b2b and b2c data on each person
- Data is accessible via an Enrichment API or Data License

Premium

BCC

Real Time Stock Data

- Real Time Stock Market Data

Premium

CPD Collect

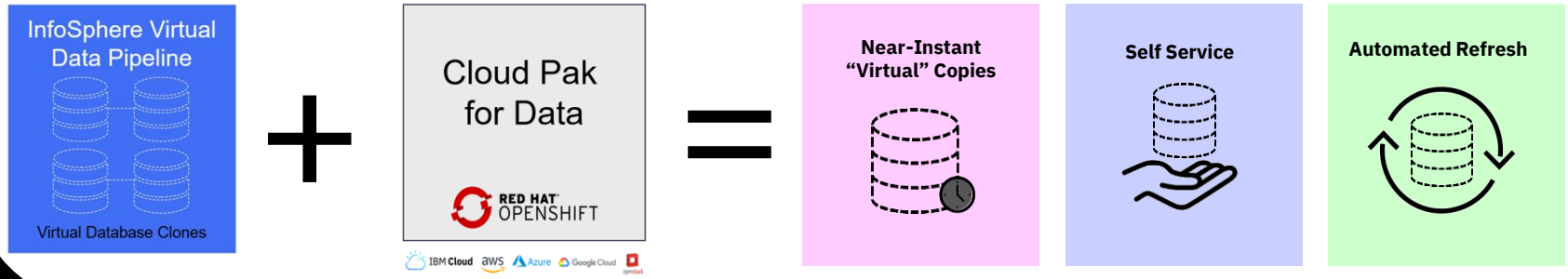
4. Premium Service: Virtual Data Pipeline – Overview

Provision and refresh analytics and test data in minutes

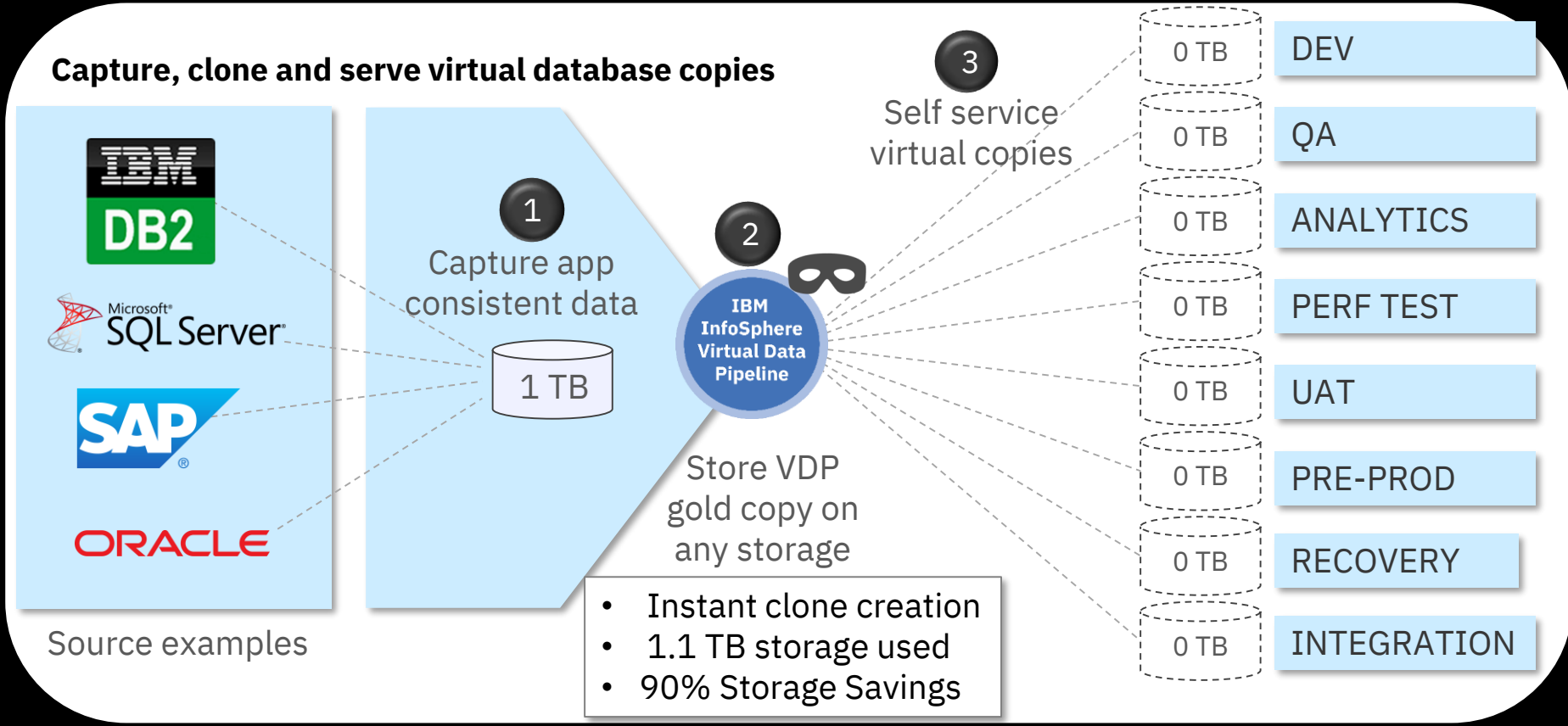
Virtual Data Pipeline in Cloud Pak for Data is a service that allows users to *instantly provision virtual database copies* to work with near real-time for data analytics and application testing, AI model training and testing, and data virtualization.

Accessing production data quickly and securely can be a challenge:

- *Risk*: Accessing Production data can impact operations and present a security and data privacy risk
- *Time Consuming*: Moving data to data warehouses and marts or creating duplicate copies can result in stale data
- *Cost*: Creating multiple copies and versions of data for each functional area uses a lot of storage and increases costs

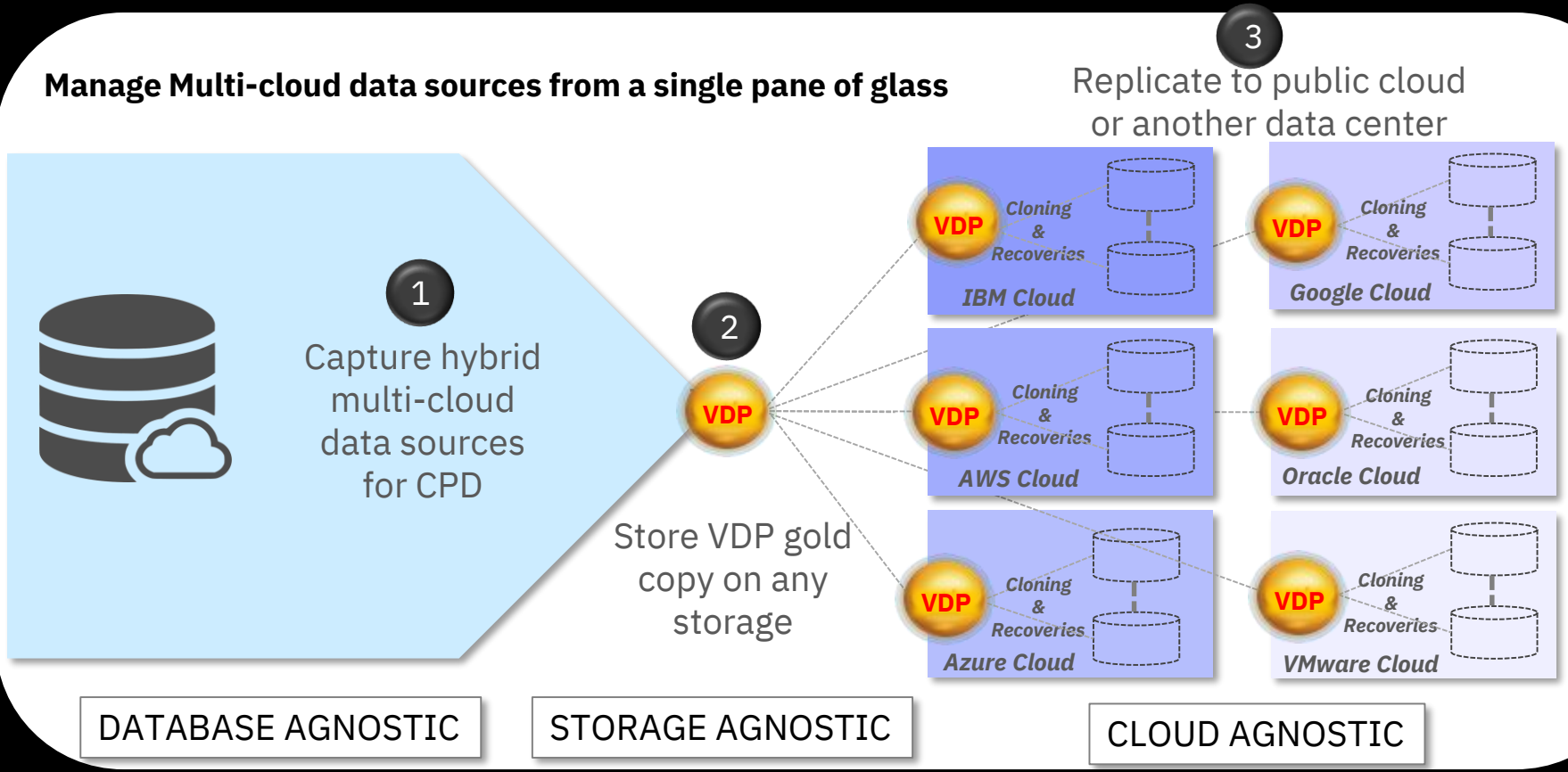


Virtual Data Pipeline
Cost savings example



Virtual Data Pipeline
Multi-cloud data management example

Manage Multi-cloud data sources from a single pane of glass

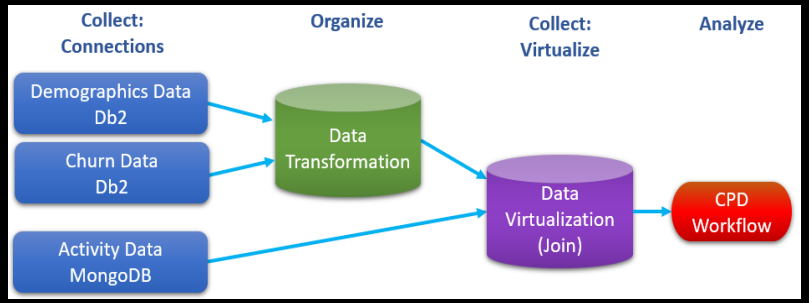


DATABASE AGNOSTIC

STORAGE AGNOSTIC

CLOUD AGNOSTIC

Lab-3 Collect: Connect



- Review DB2 Data
  - Customer Demographics
  - Customer Churn
  - Credentials
- Review DB2 Connection
  - Connection Parameters
- Review MongoDB Data
  - Customer Activity
  - Credentials
- Review MongoDB Connection
  - Connection Parameters

IBM Analytics Modernization Workshop

Part 2

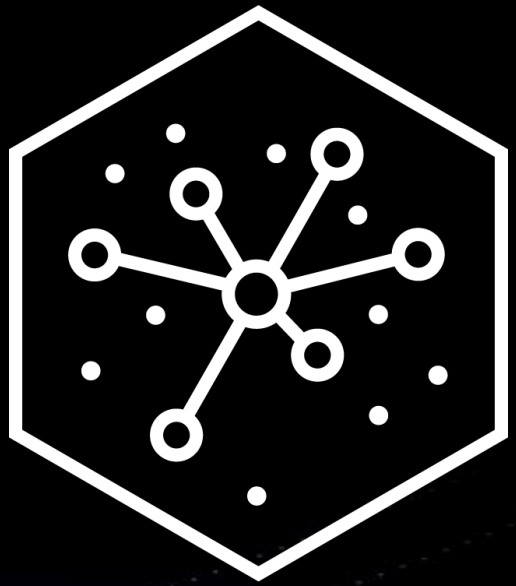
• Introduction • Business Use Case	• Lab 01 • Lab 02
• Collect: Connect • Organize • Collect: Virtualize	• Lab 03 • Lab 13 • Lab 05
• Analyze • Deploy • Infuse – OpenScale • Infuse – Cognos Analytics • Wrap up	• Lab 06 • Lab 07 • Lab 08 • Lab 09 • Lab 10

Cloud Pak for Data

We will return for lab review at 11:30 am. Please work on Lab-03.

Organize

Lab 13 – Organize



IBM Analytics Modernization Workshop

Part 2

- Introduction
- Business Use Case

- Collect: Connect
- Organize
- Collect: Virtualize

- Analyze
- Deploy
- Infuse – OpenScale
- Infuse – Cognos Analytics
- Wrap up

- Lab 01
- Lab 02

- Lab 03
- Lab 04
- Lab 05

- Lab 06
- Lab 07
- Lab 08
- Lab 09
- Lab 10

CPD Organize
DataOps

CPD Organize
Watson Knowledge Catalog (WKC)

Enterprise Data Governance

- Data governance—know your data
- To set up the foundation of a DataOps program, organizations need to comply with regulatory requirements, communicate and enforce policies and standards, and manage metadata.

CDO, Data Stewards

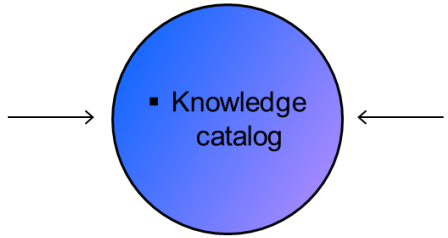
- Build/Import Business Glossary
- Manage Reference Data
- Manage Data Classes
- Auto-Discover metadata assets
- Classify data assets
- Define policies/rules
- Data governance workflow
- Data Lineage

Enterprise Data Quality

- Data quality—trust your data
- Data is useful only if its quality, content, and structure is well understood. Delivering reliable, quality, timely data for business consumption is a continuous process.

Data stewards and data quality analysts

- Profile data
- Understand, monitor and remediate data quality
- Apply validation rules



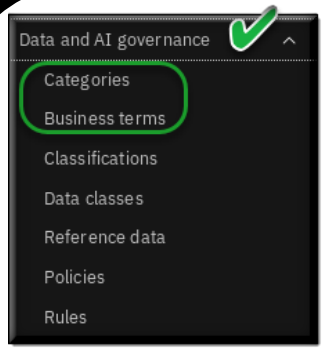
Enterprise Data Consumption

- Data consumption—use your data
- Enterprises need to surface business-ready data to consumers allowing them to deliver timely value to the business and make better decisions.

Data analysts, data scientists, business analysts

- Search and find relevant data
- Prepare data for consumption and analysis
- Policy Enforcement
- Tag, rate, comment on and share the data
- Data Lineage

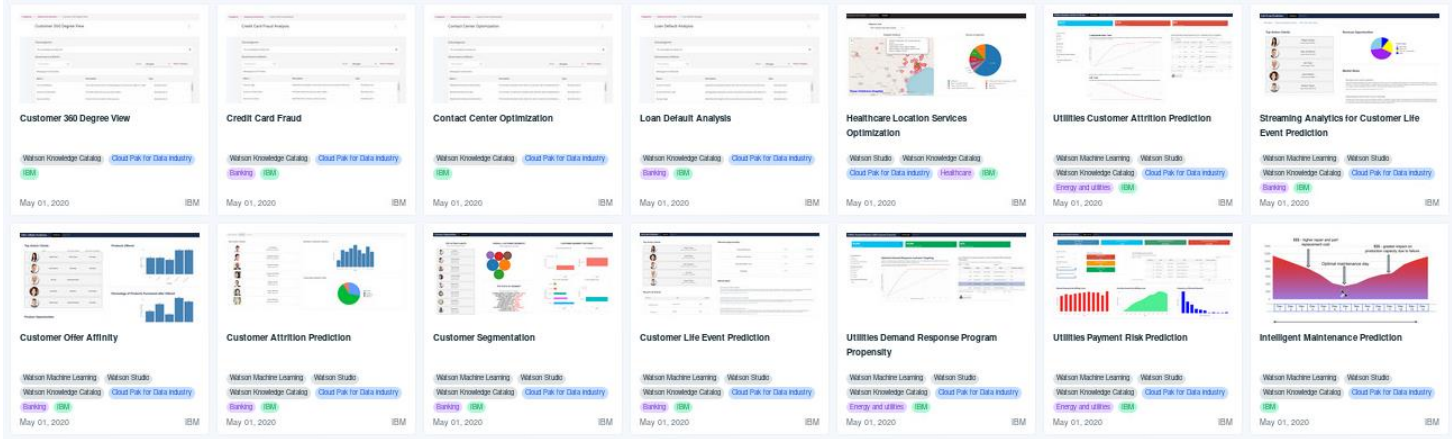
CPD Organize
Data and AI governance: Categories and Business terms



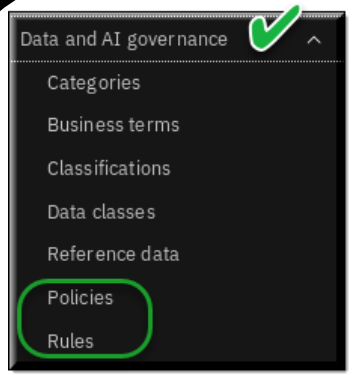
Categories provide logical structure to a business glossary

Business terms standardize definitions of business concepts

1. Manually create Categories and Business Terms
2. Import Categories and Business Terms from CSV or XML files
3. Import a Glossary from an **industry accelerator**

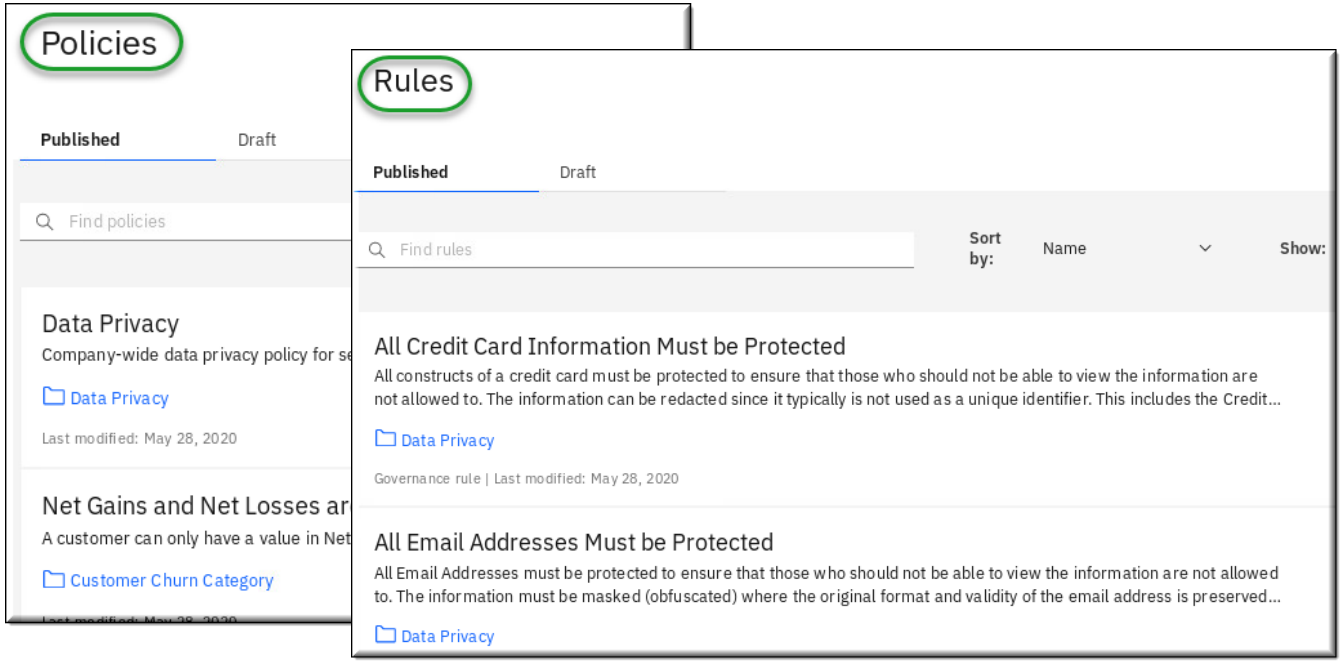


CPD Organize
Data and AI governance: Policies and Rules

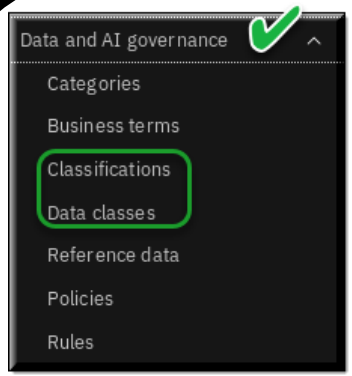


Policies describe how to control data and consist of one or more rules

Rules describe the criteria for compliance with business objectives

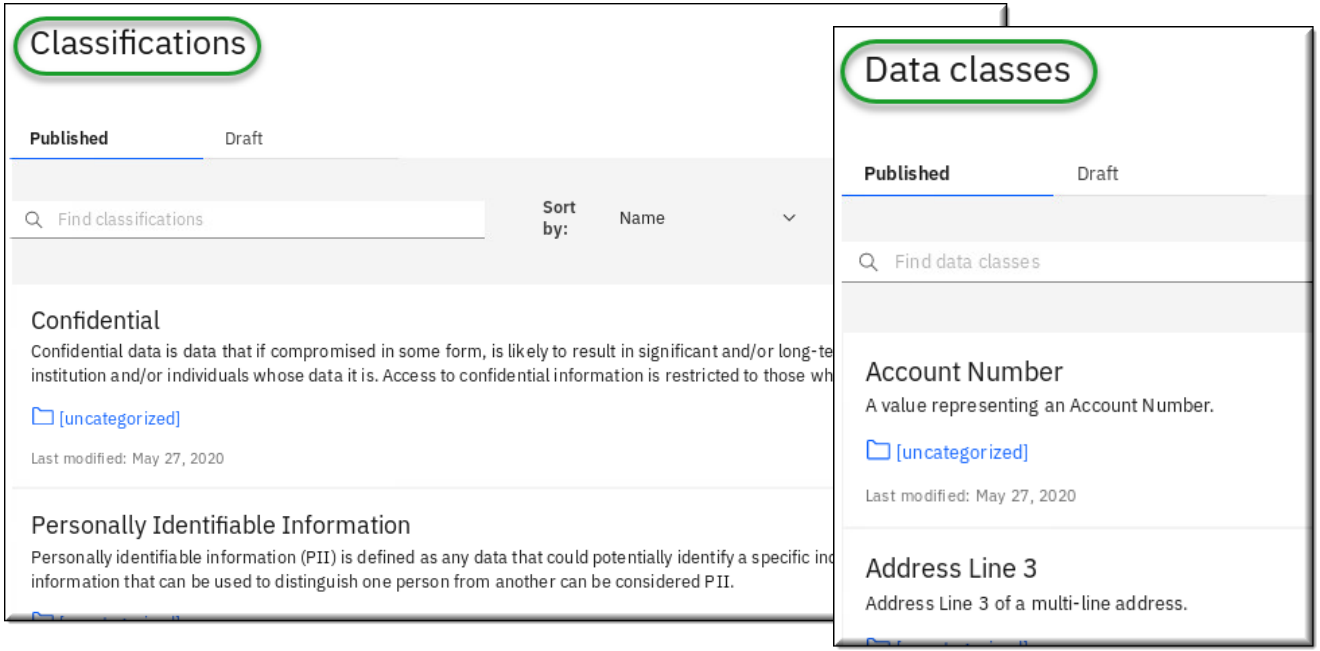


CPD Organize
Data and AI governance: Classifications and Data classes

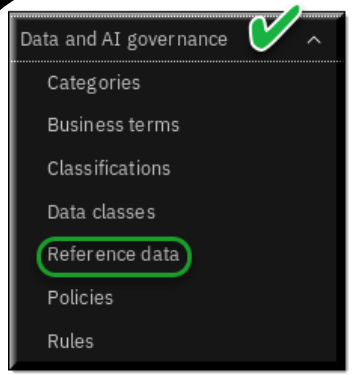


Classifications describes the sensitivity level of data

Data classes describe the contents of data in a column in a structured data set

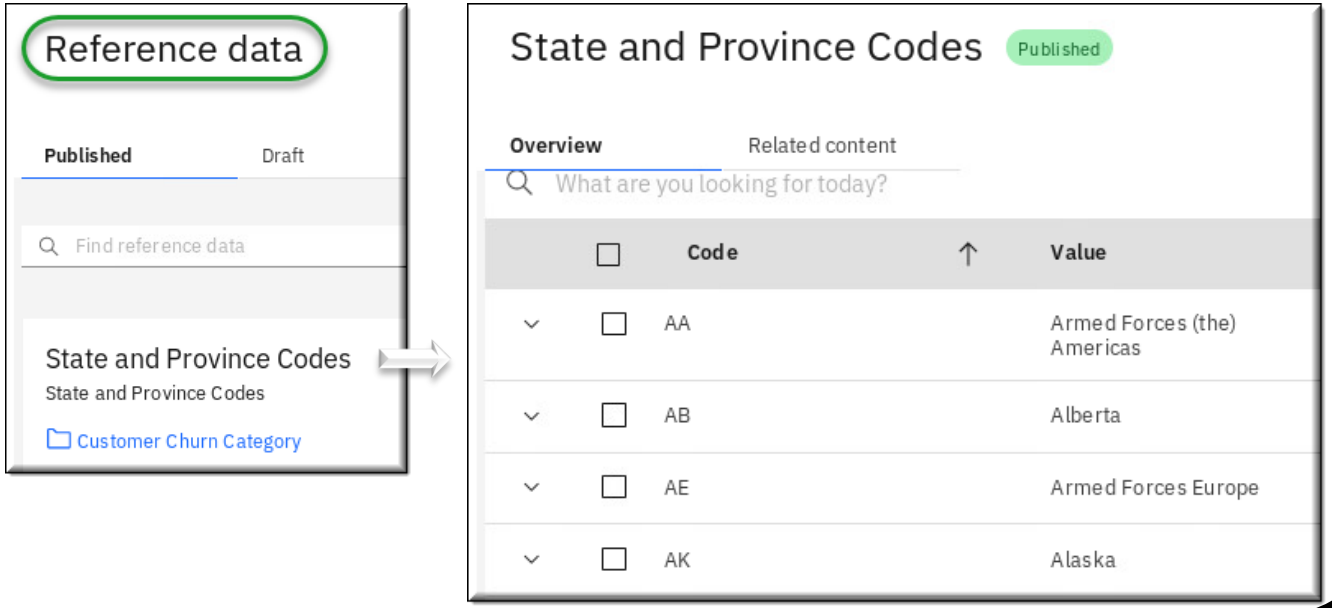


CPD Organize
Data and AI governance: Reference Data



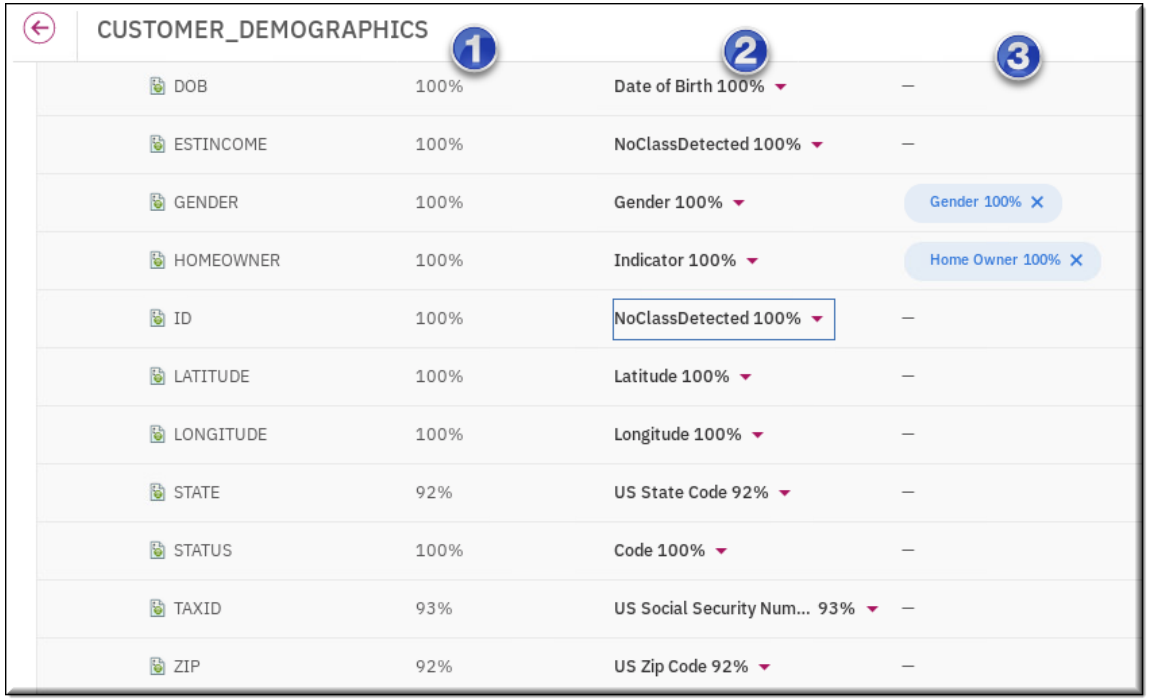
Reference Data Sets define list of permissible values that are allowed for use within a data field.

May be referenced by Business Terms, Policies, Rules and Data Classes



CPD Organize
Auto-discover assets

Data Discovery



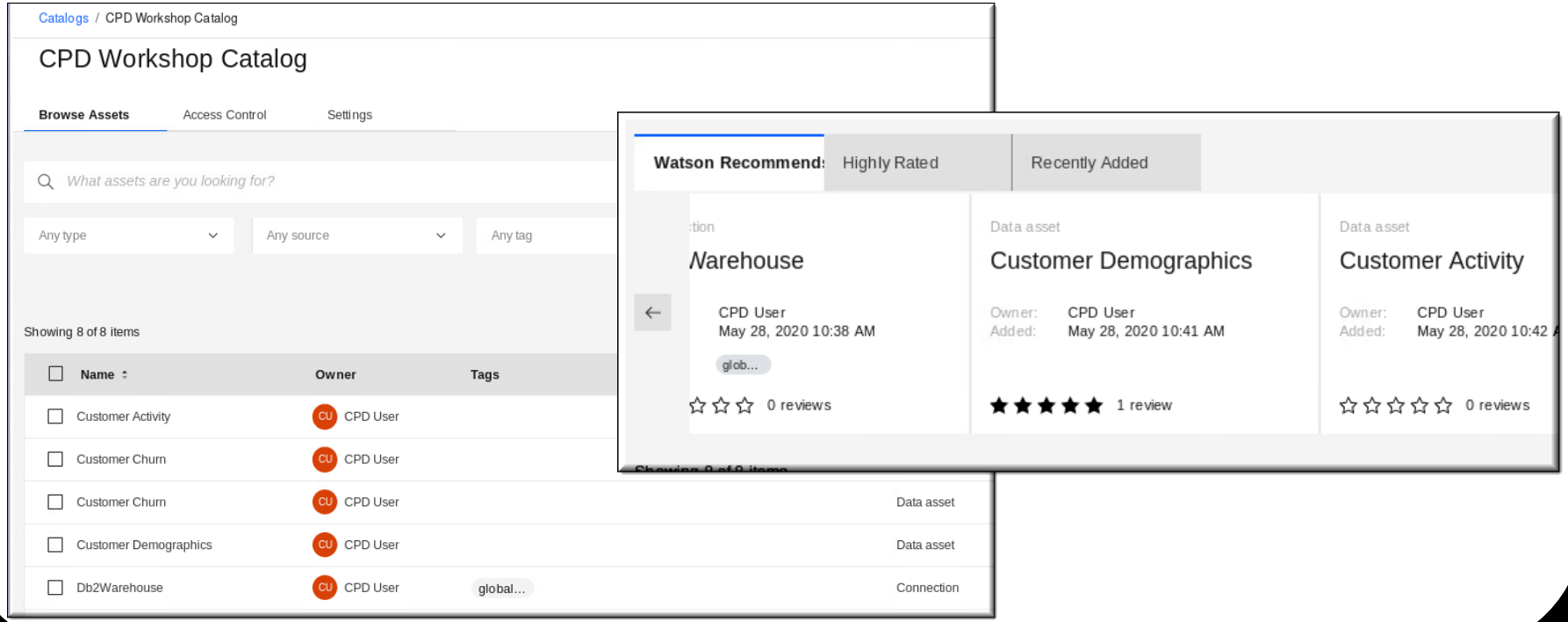
Use machine learning based auto-discovery to:

1. Analyze data quality
2. Analyze columns (Classify data)
3. Assign Business terms

You can perform discovery with data sampling to allow for self-service data access with a search.

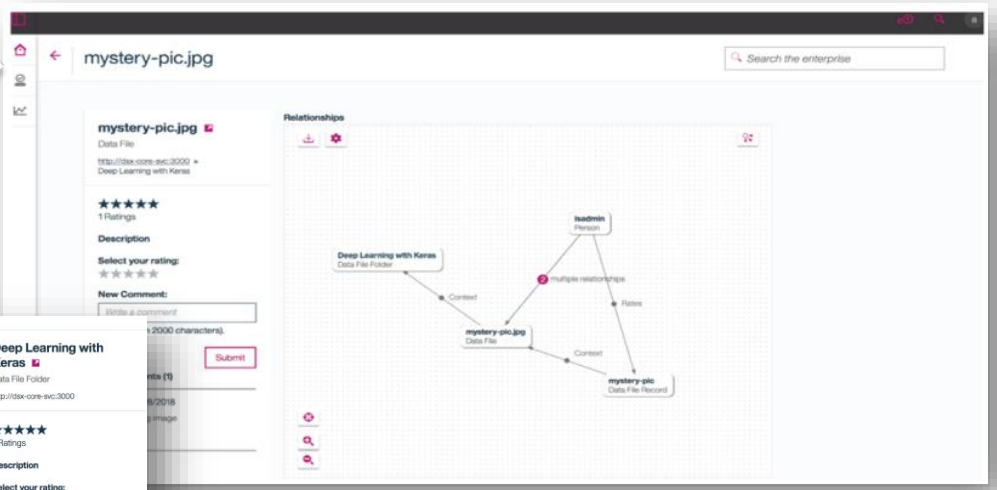
CPD Organize
Publish to a catalog

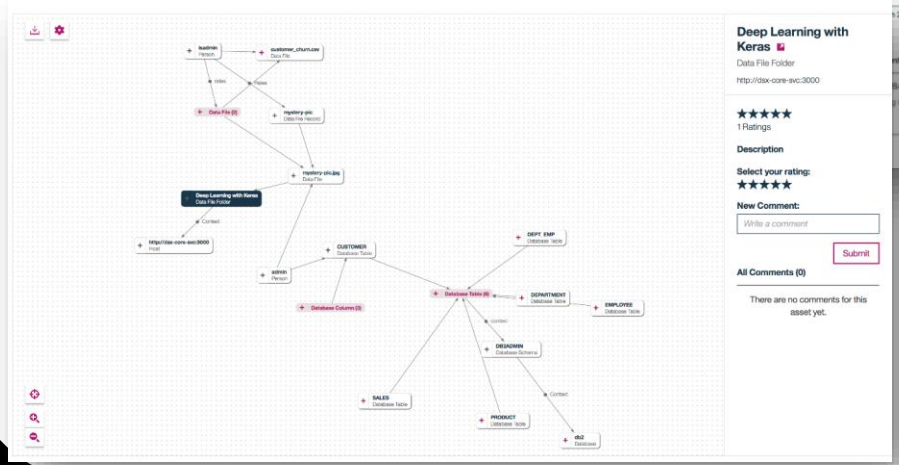
Catalog and govern your assets



CPD Organize
Relationship graph with explorer

- Explore relationships between data assets, terms, analytic assets, users, etc.
- Gain in-depth understanding of metadata through crowdsourcing (e.g., ratings, comments) and machine learning





Explore deeper to understand context and usage patterns

CPD Organize
Profile data

The **Profile** of a data asset includes generated metadata and statistics about the data.

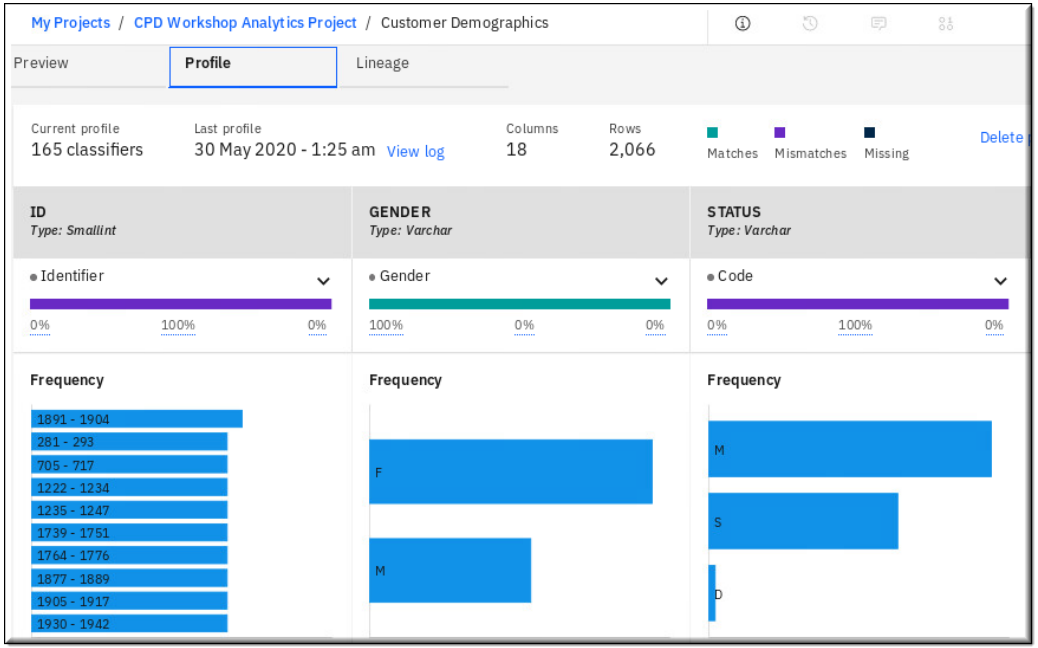
All catalog or project members can see data asset profiles.

Profiles are automatically created:

- In catalogs, profiles for unstructured data assets are created automatically, regardless of whether policies are enforced
- In governed catalogs, profiles for structured data assets are created automatically

Profiles can be manually created:

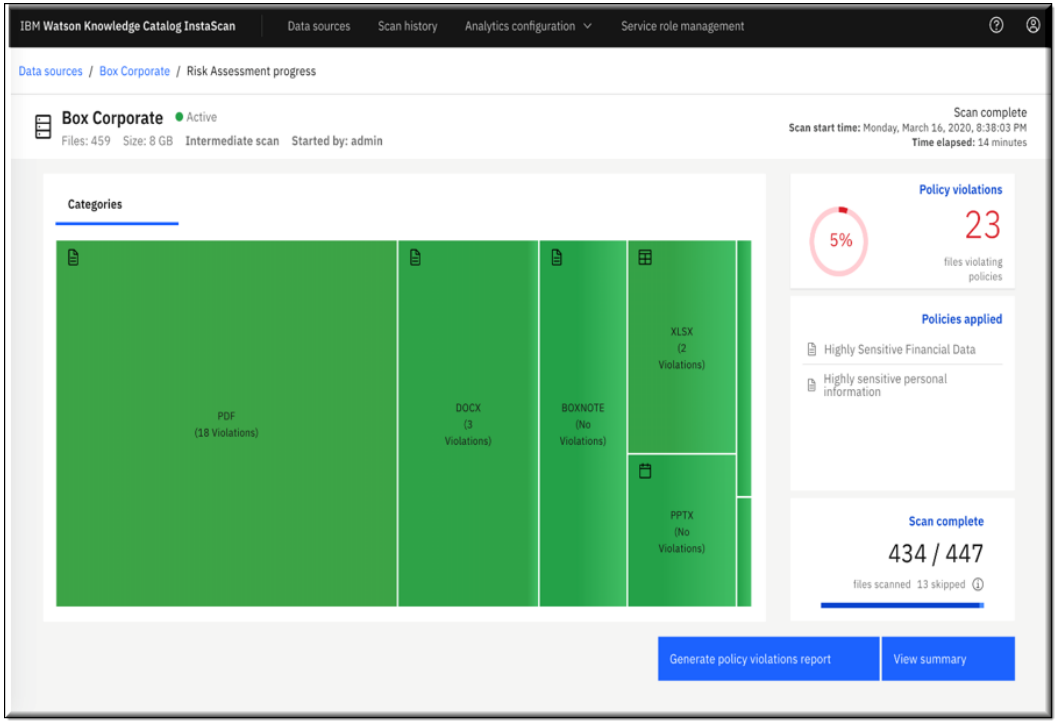
- In ungoverned catalogs for structured data assets
- In projects for both structured and unstructured data assets



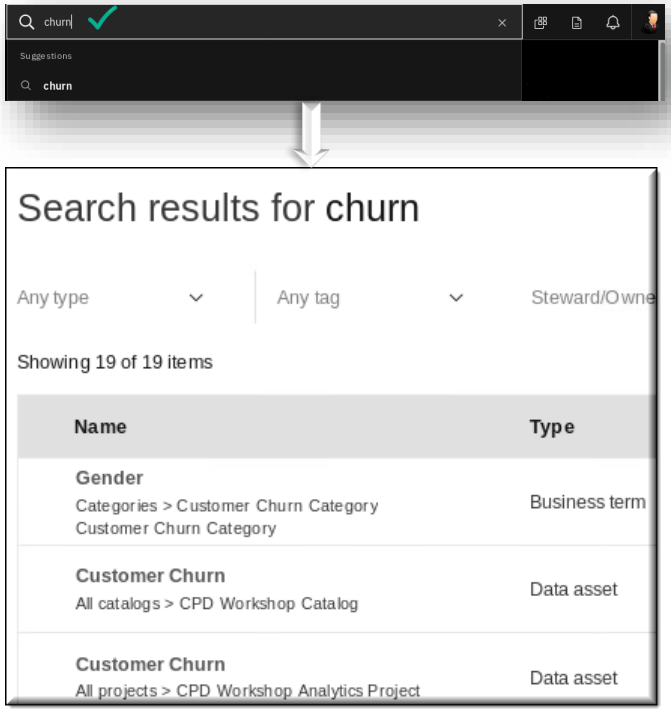
CPD Organize
InstaScan

InstaScan can perform risk assessments and compliance checks of unstructured data

- Scan email, PDFs, word processor documents and images
- Quickly determine which areas have high concentration of sensitive information and prioritize hot spots
- Automatically apply classification labels to files
- Create and share ongoing compliance reports with CISO or regulators
- Integrates with Box Shield to provides a comprehensive data privacy solution for unstructured data in Box



CPD Organize
Search for data

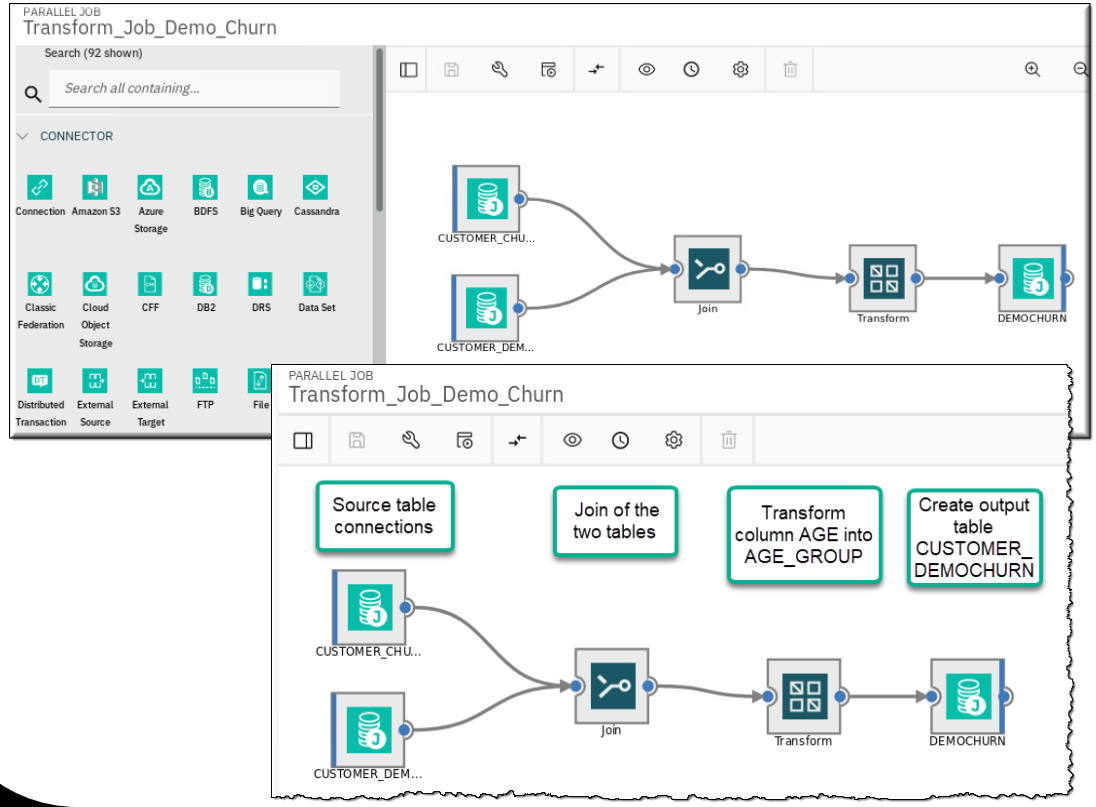


Relevancy of search results factors:

Text match	The provided text is searched in the asset name & description, where name contributes more to the higher place in the result list.
Asset rating	The higher the average rating the asset has, the higher it is on the results list.
Comments	The higher the number of comments, the higher the asset is on the results list.
Context match	The search results list might contain the closest neighbors of assets that are returned based on the text match.
Modification date	The assets that were modified recently are more likely to be returned in the search results.
Quality score	The higher the score, the higher the asset is on the results list. Quality score applies to database tables, views & columns, design tables, views & columns, data file records & fields.
Usage	The more relationships of type uses an asset has, the higher it is on the results list.

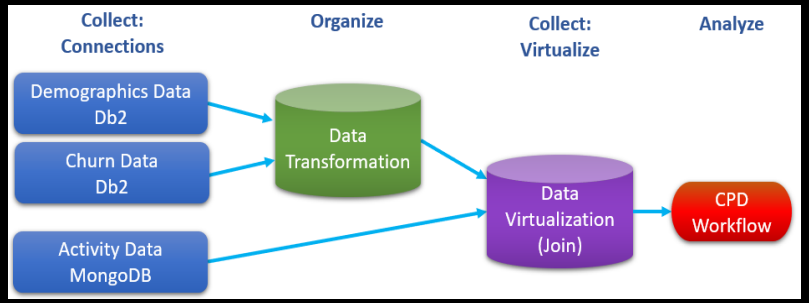
CPD Organize

DataStage - Transform and migrate data, build and execute ETL jobs at scale



- Use powerful data transformation capabilities
- ML infused Smart Job Clustering, Smart Job Assist and automated job sequencing
- Design once, run on any-cloud using Kafka
- Remotely execute job for co-located access to data, satisfy geo-political requirements and save costs
- In-line data quality and governance to build trusted data when the data is being delivered to a target environment such as a data lake
- Create CI/CD pipelines with GitHub

Lab 13 – Organize- Deeper Dive



Create a Project
- ✓ Add a Global DB2 Connection
- ✓ Add Connected Data
  - ✓ Customer Demographics Table
  - ✓ Customer Activity Table
- ✓ Preview Customer Demographics
- ✓ Profile Customer Demographics
  - ✓ Data Column Frequencies
  - ✓ Data Column Classification (e.g. TaxID-> US SSN)

Review Governance Artifacts
- ✓ Categories and Business Terms
- ✓ Classifications
- ✓ Data Classes
- ✓ Reference Data

Data Protection Rules and Masking
- ✓ Create Policy
- ✓ Define Protection Rule
  - ✓ Data Class DOB masked for user==‘developer’
- ✓ Assign the rule to the policy
- ✓ Demonstrate masked DOB on display for ‘developer’

Data Discovery Automation
- ✓ Crawl 2 DB2 Tables
- ✓ Classify Data Columns
- ✓ Score Data Columns, Data Table
- ✓ Map Business Terms to Data Columns
- ✓ Analysis Results in Workspace

Data Transformation
- ✓ Review a Transformation Job
- ✓ Create a Transformation Job
- ✓ Use the Visual Palette
- ✓ Compile, Run Job, View Output

Data Lineage
- ✓ Information Assets
- ✓ View Customer Demographics Data Lineage

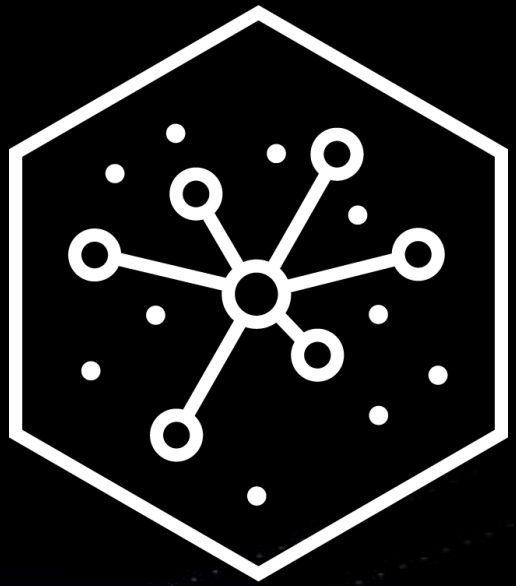
IBM Analytics Modernization Workshop

Part 2

• Introduction • Business Use Case	• Lab 01 • Lab 02
• Collect: Connect • Organize • Collect: Virtualize	• Lab 03 • Lab 13 • Lab 05
• Analyze • Deploy • Infuse – OpenScale • Infuse – Cognos Analytics • Wrap up	• Lab 06 • Lab 07 • Lab 08 • Lab 09 • Lab 10

Cloud Pak for Data

We will return for lab review at 2:00 pm. Please work on Lab-13.



Collect

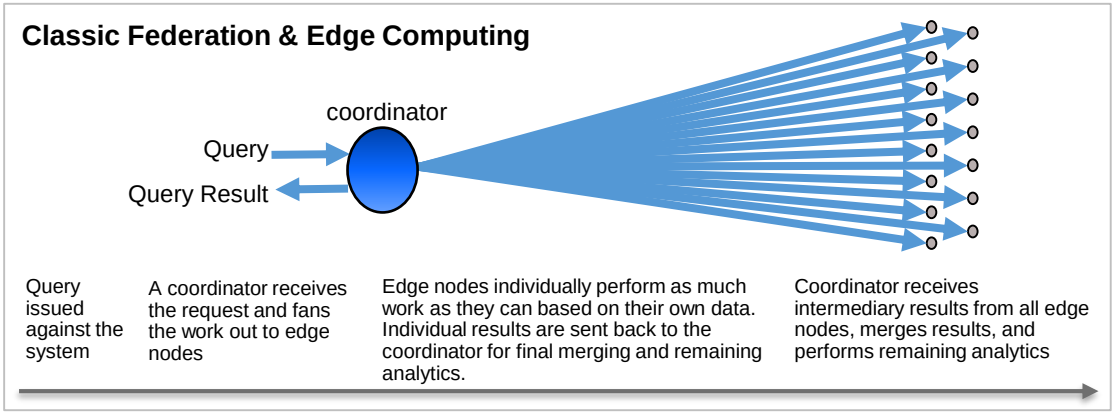
Lab 05 – Collect: Virtualize

IBM Analytics Modernization Workshop

Part 2

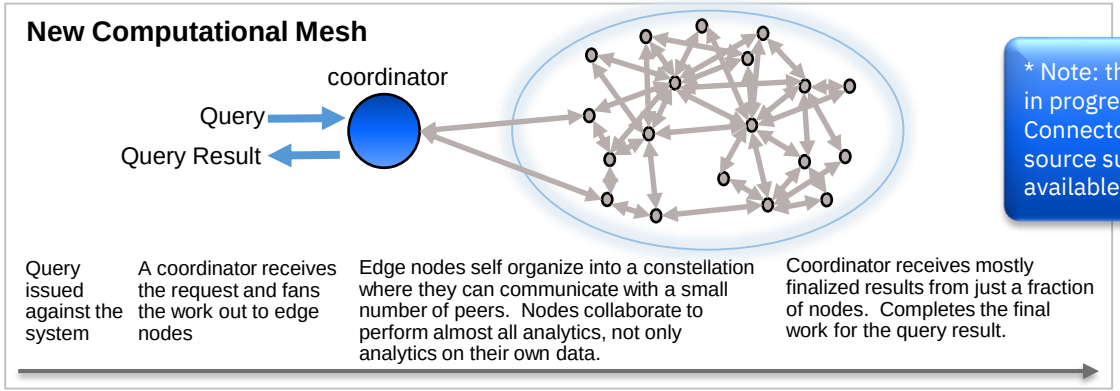
• Introduction	• Lab 01
• Business Use Case	• Lab 02
• Collect: Connect	• Lab 03
• Organize	• Lab 04
• Collect: Virtualize	• Lab 05
• Analyze	• Lab 06
• Deploy	• Lab 07
• Infuse – OpenScale	• Lab 08
• Infuse – Cognos Analytics	• Lab 09
• Wrap up	• Lab 10

CPD Data Virtualization
Constellation “Computational Mesh” benefit



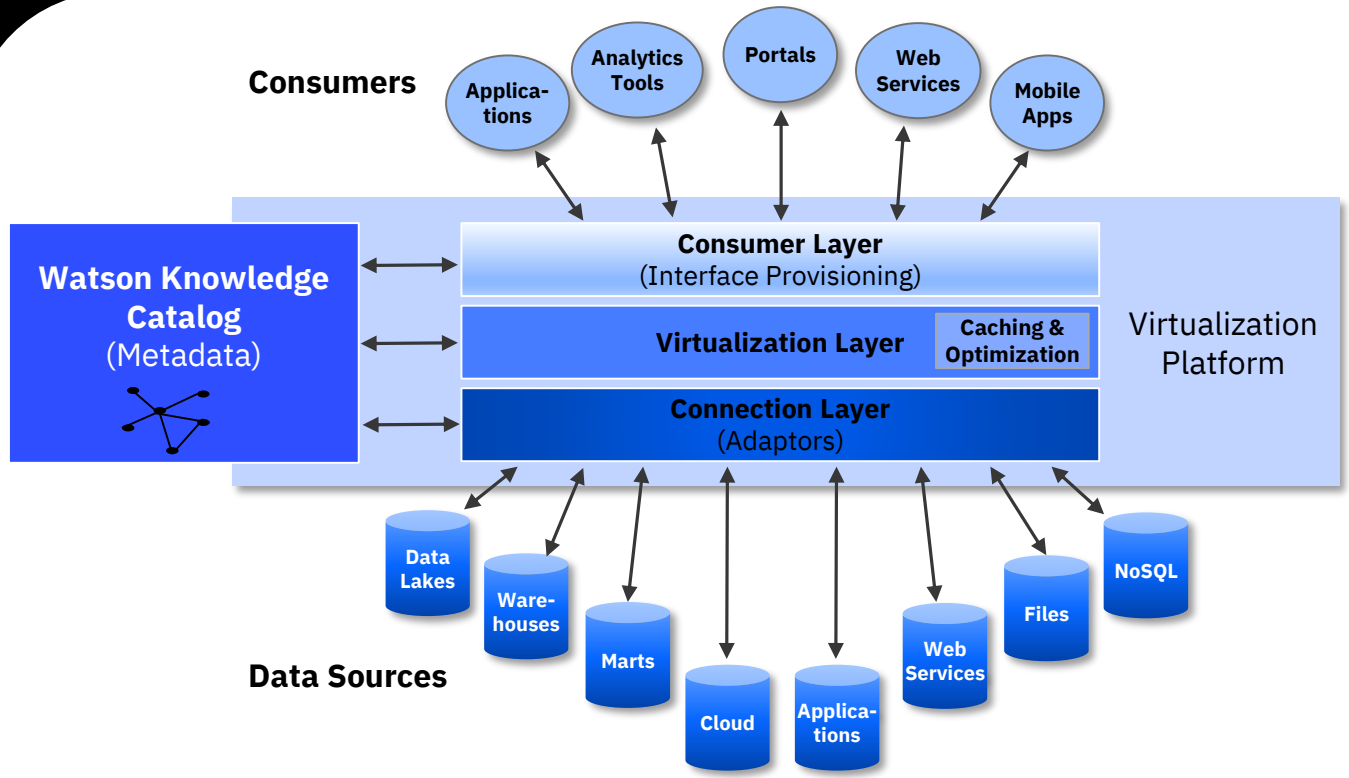
To be clear: Federation is a form of Data Virtualization and has been used successfully for many years in IBM products like Db2

CPD Data Virtualization uses a new Computational Mesh * approach which meets the performance demands of today’s modern data access requirements



* Note: this is a work in progress. Remote Connectors with data source support is available today.

CPD Data Virtualization
With Watson Knowledge Catalog (WKC) built in



- **Provides the ability to search, view, access, manipulate, and analyze data**
- **No need to know or understand its physical format or location**
- **No need to move or copy it**

CPD Data Virtualization
Benefits and use cases

Benefits

Simple:
- Self-discovering, self-organizing cluster
- Joins provide a one source input to analytics

Flexible:
- Once established, it is easy to add new sources to the constellation
- Integrates disparate data assets with simple automation, providing seamless access to data as one

Scalable:
- Can access thousands of sources, IOT and edge devices

Cost Effective:
- Leverages the compute resources of source systems to execute the SQL

Secure:
- Inherits privileges & masking policies of the data sources
- Built in governance, security, and access control

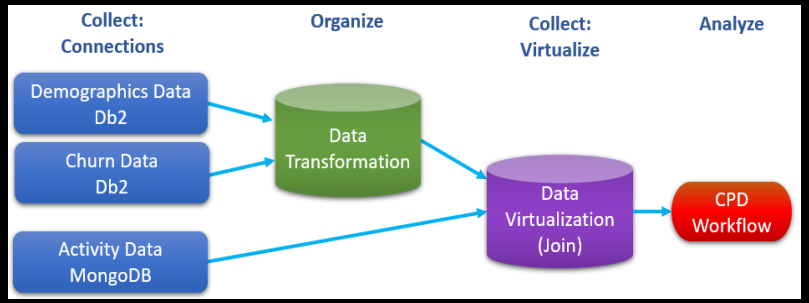
Use Cases

Data Scientists:
- Significant productivity increase getting access to sources discovery and assembly of data sets

Current State answer requirement:
- Current state required for up-to-date analytics
- One time access to data, then throw it away
  - e.g. “How much cash is ‘on hand’ across our branches worldwide?” “What is our current ‘claims’ liability?”

ETL and/or Data Governance saturation
- Self-service – In the event that Data Engineers cannot keep up with business demands for access to data

Lab-5 Collect: Virtualize Review



Add Data Sources to Virtualization Facility
- ✓ DB2, MongoDB

Virtualize CUSTOMER_DEMOCHURN
- ✓ Select Table
- ✓ Add to Cart
- ✓ Virtualize in Project

Virtualize ACTIVITY01
- ✓ Select Table
- ✓ Add to Cart
- ✓ Virtualize in Project

Join Virtual Tables
- ✓ Based on ID Field
- ✓ Create View
- ✓ Preview View in Project

IBM Analytics Modernization Workshop

Part 2

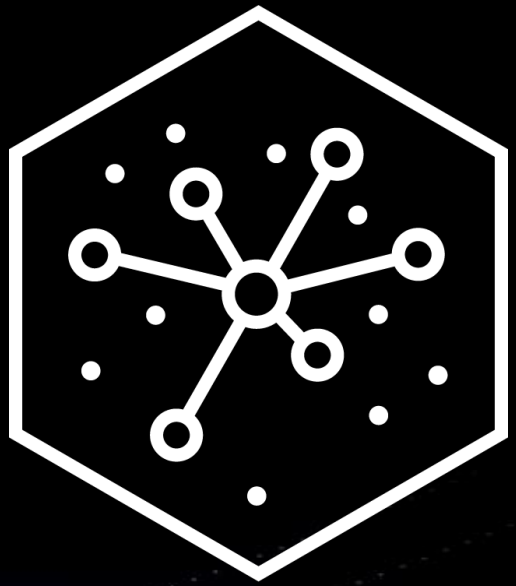
• Introduction • Business Use Case	• Lab 01 • Lab 02
• Collect: Connect • Organize • Collect: Virtualize	• Lab 03 • Lab 13 • Lab 05
• Analyze • Deploy • Infuse – OpenScale • Infuse – Cognos Analytics • Wrap up	• Lab 06 • Lab 07 • Lab 08 • Lab 09 • Lab 10

Cloud Pak for Data

We will return for lab review at 3:15 pm. Please work on Lab-5.

IBM Analytics Modernization Workshop

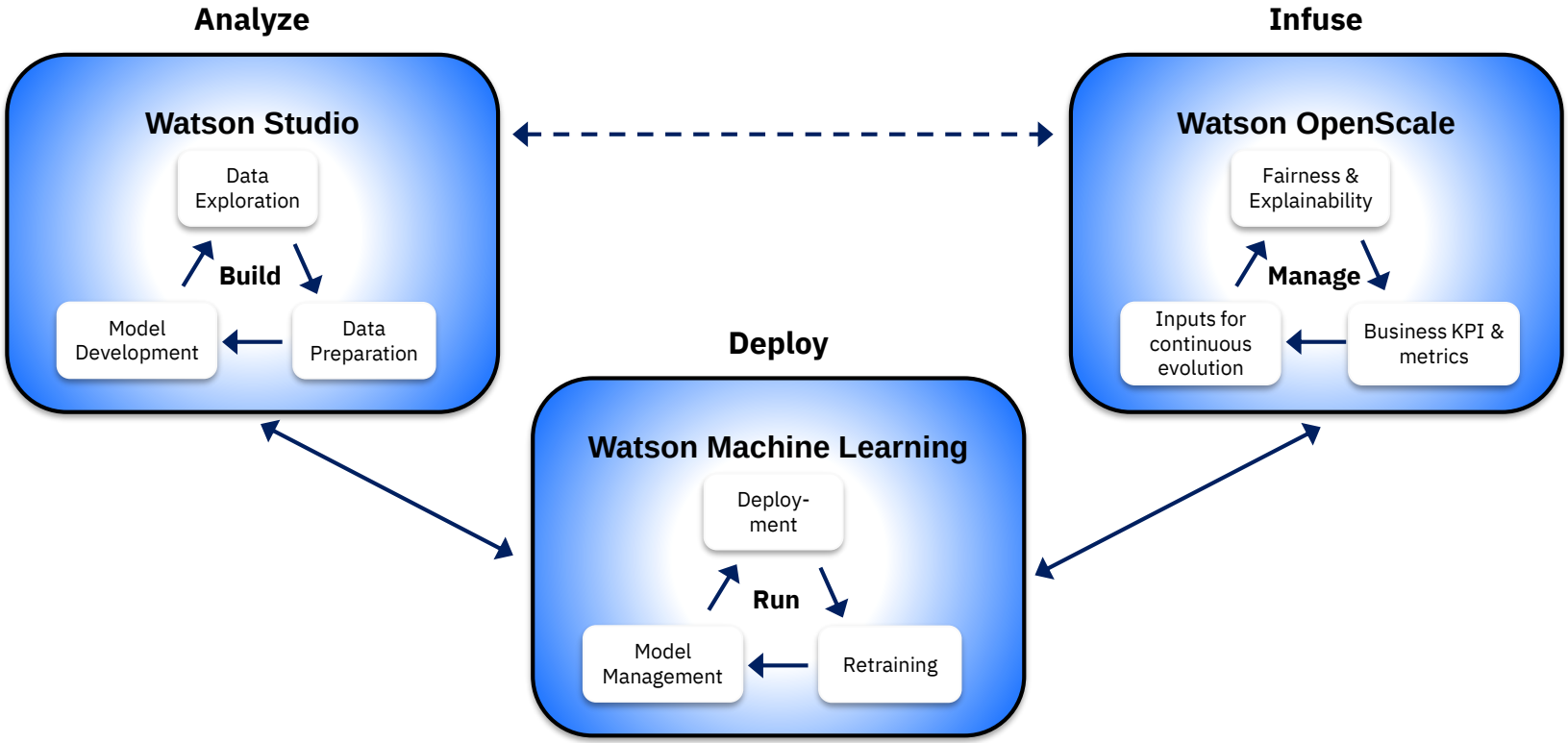
	• Introduction • Business Use Case	• Lab 01 • Lab 02
	• Collect: Connect • Organize • Collect: Virtualize	• Lab 03 • Lab 04 • Lab 05
Part 3	• Analyze • Deploy	• Lab 06 • Lab 07
	• Infuse – OpenScale • Infuse – Cognos Analytics • Wrap up	• Lab 08 • Lab 09 • Lab 10



Analyze

Lab 06 – Analyze: AutoAI & Notebooks

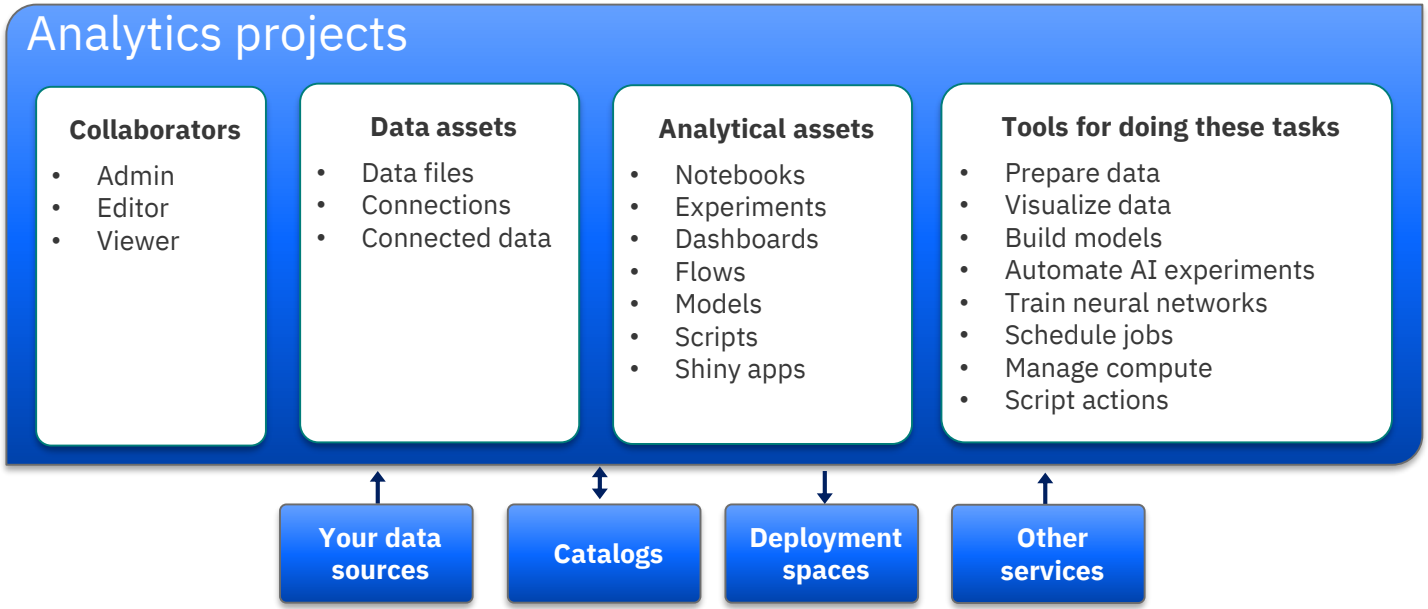
Analyze
The Data Science Lifecycle: Overview



Analyze
Watson Studio: Collaborating with Analytics Projects

Watson Studio provides the environment and tools to collaborate on business problems.

Watson Studio is centered around the *Analytics Project.* Data scientists and business analysts use analytics projects to organize resources and analyze data with various tools.

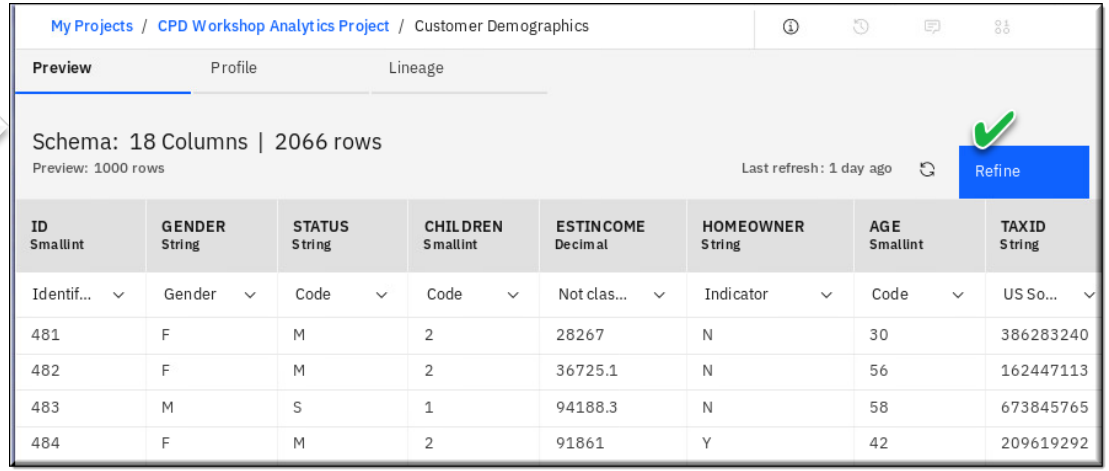


Analyze
Refine data with visualizations

Refine can cleanse and shape tabular data with a graphical flow editor using functions and logical operators.

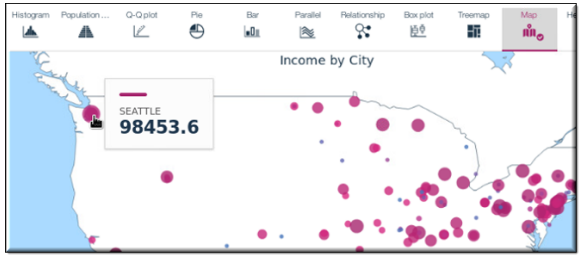
Use it to remove data that is incorrect, incomplete, improperly formatted, etc.

Shape the data by filtering, sorting, combining or removing columns. You can create a Data Refinery flow as a set of ordered operations on the data to run repeatedly any time.



ID Smallint	GENDER String	STATUS String	CHILDREN Smallint	ESTINCOME Decimal	HOMEOWNER String	AGE Smallint	TAXID String
Identif...	Gender	Code	Code	Not clas...	Indicator	Code	US So...
481	F	M	2	28267	N	30	386283240
482	F	M	2	36725.1	N	56	162447113
483	M	S	1	94188.3	N	58	673845765
484	F	M	2	91861	Y	42	209619292

Data Refinery also includes a graphical interface to profile data to validate it with 20+ customizable charts that give perspective and insights into the data.



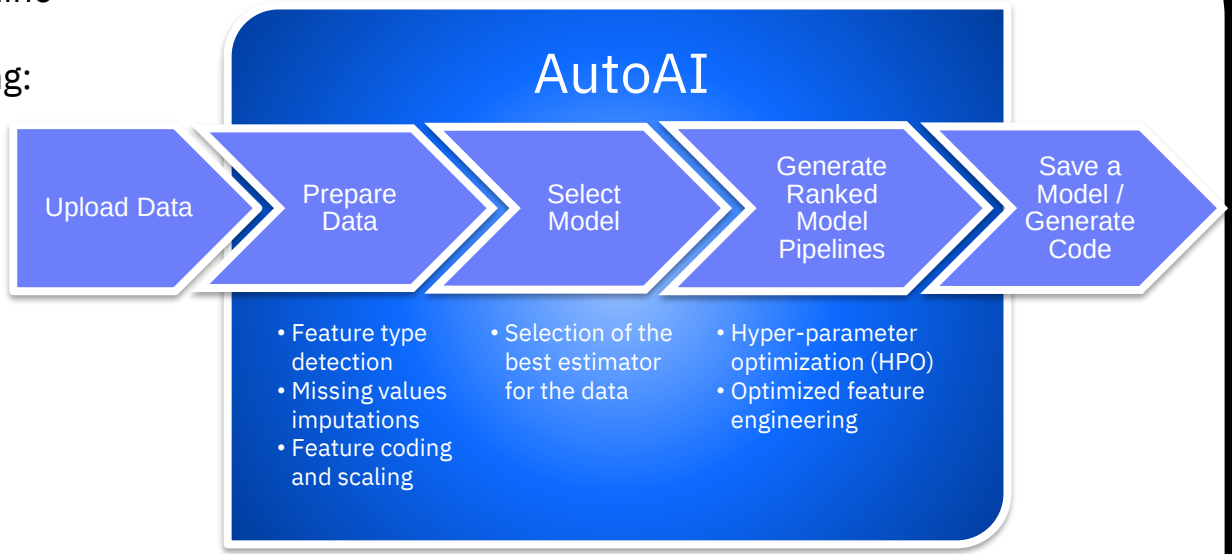
Analyze
AutoAI * – Overview

AutoAI is an award-winning technology that simplifies the Machine Learning model creation and AI lifecycle by automating the following:

- **Data preparation**
- **Model development**
- **Feature engineering**
- **Hyper-parameter optimization**

AutoAI delivers training feedback visualizations for real-time model performance results with:

- **Binary, Multiclass, and Regression support**
- **One-click model deployment**

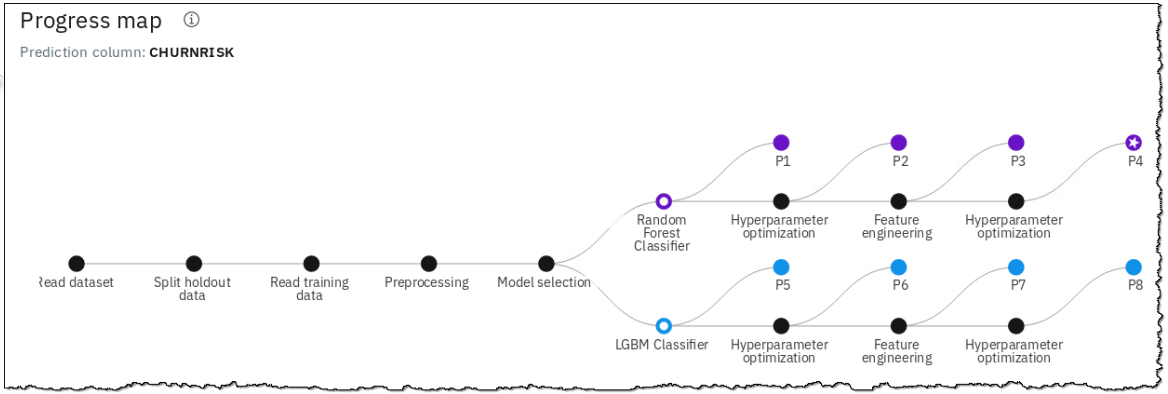


* AutoAI is enabled with the Watson Machine Learning service install, but it is driven through a Watson Studio Analytics Project

Analyze
AutoAI – Infographics

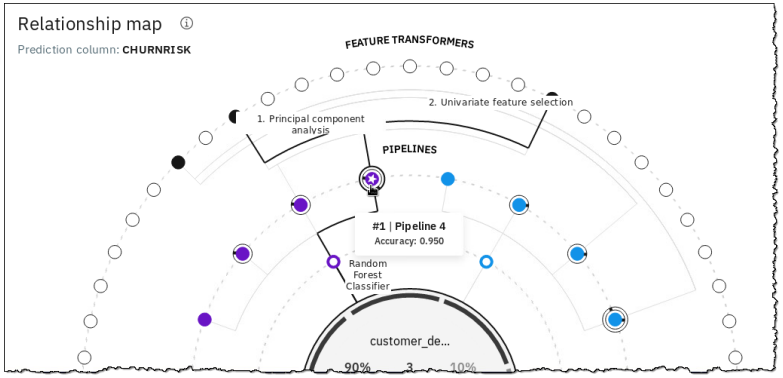
AutoAI Progress map

Displays a progress each step as it creates the best model for your data.



AutoAI Relationship map

Interactive infographic that shows the relationship of the pipelines, algorithms, and the feature transformers.



Analyze
AutoAI – Pipelines

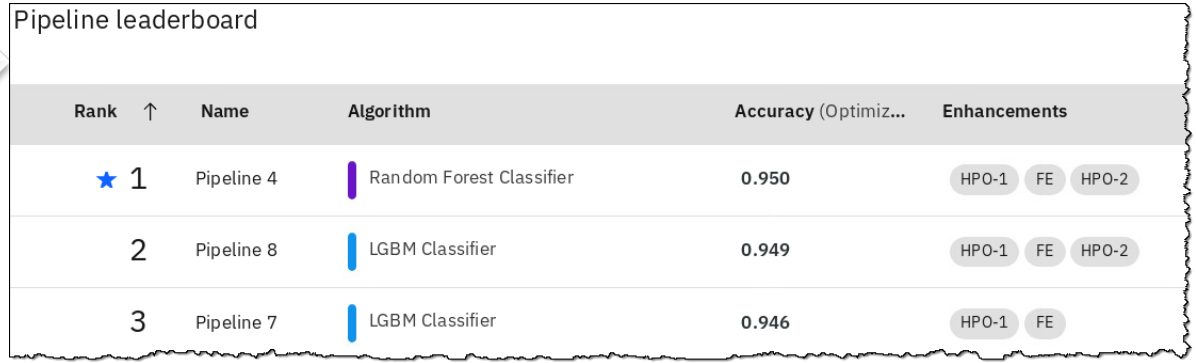
AutoAI pipeline leaderboard

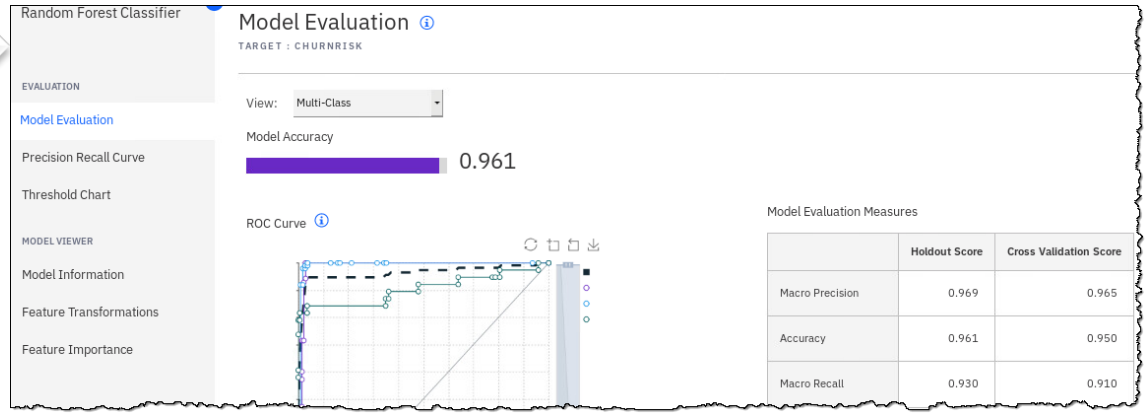
Shows the ranking of the pipelines for each potential model, the higher the better.

After AutoAI completes its model creation steps, you can drill into the pipeline(s) to understand how it came to its conclusion.

Save the pipeline in your project as a:

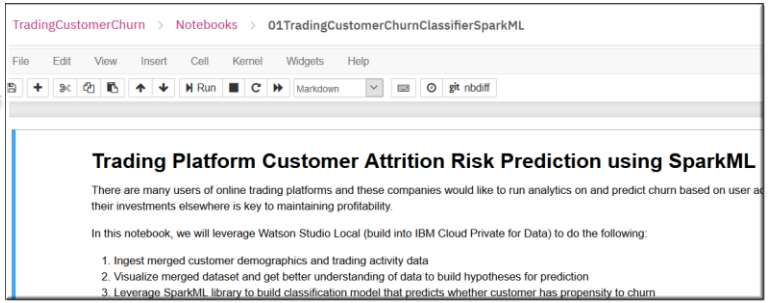
- **model**
- **notebook**



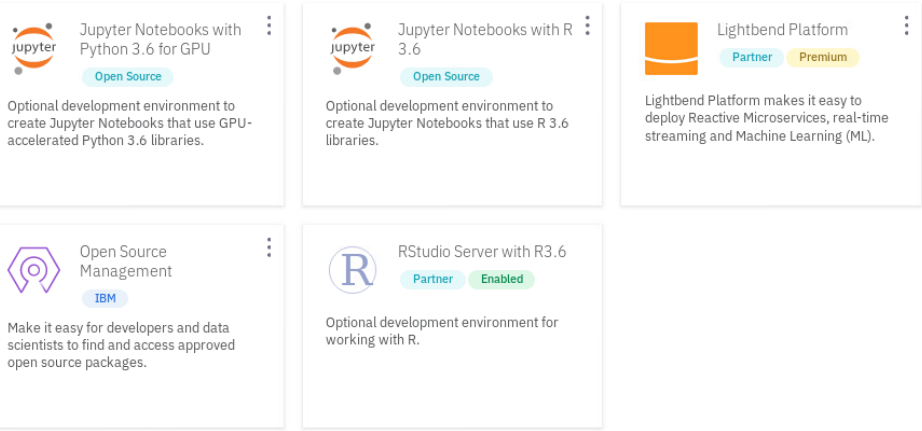


Analyze
Notebooks, RStudio and other tools

The default notebook environment:
Jupyter Notebook with Python 3.6



Developer tool services available:
- Jupyter Lab
- Jupyter Notebooks with Python 3.6 for GPU
- Jupyter Notebooks with R 3.6
- RStudio Server with R3.6
- Lightbend Platform
- OpenSource Management

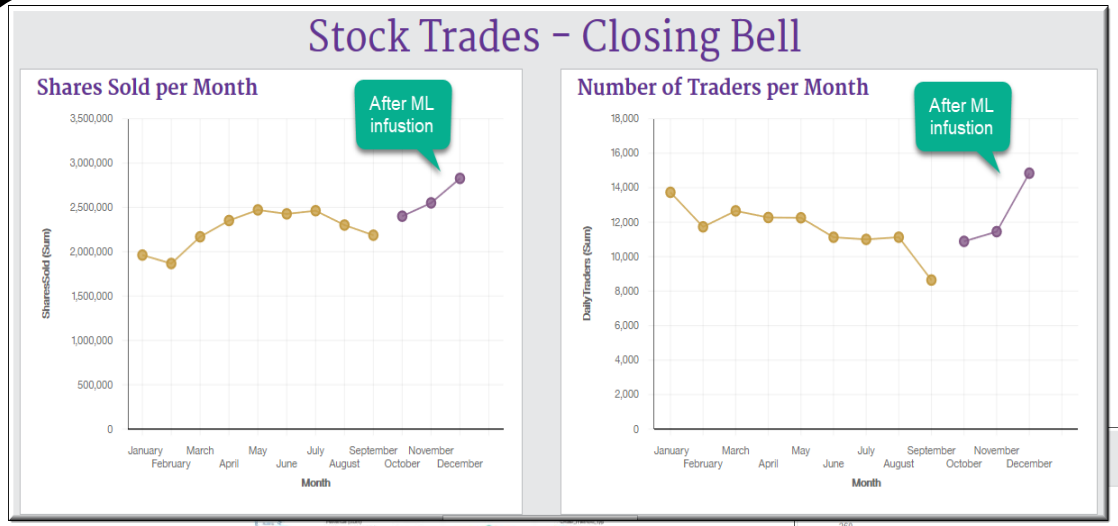


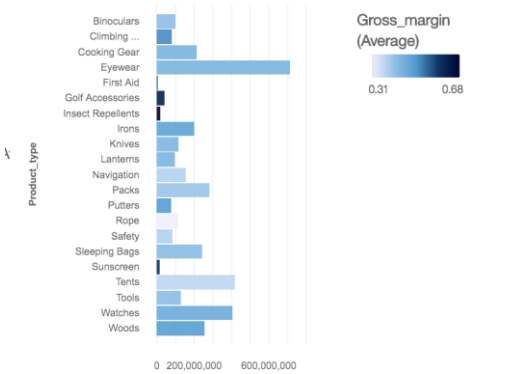
Analyze

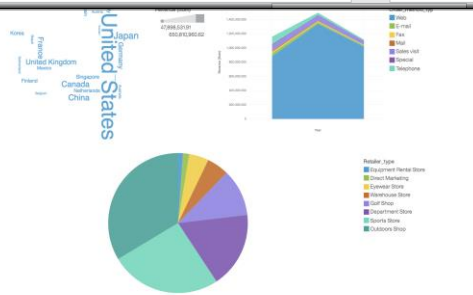
Watson Studio notebooks: Build Data Science & Machine Learning models

- Data Scientists and Data Engineers collaborate with each other in CPD platform – while still maintaining data governance
- Collaboration using GitHub or BitBucket is integrated into the platform, which brings a cohesiveness to the work culture and helps to automate CI/CD pipe line
- Exploit GPUs for deep learning predictive ML models
- Programmatically build data visualizations and data wrangling
- Real-time or batch model scoring
- Evaluate model accuracy

Analyze
Cognos Dashboards Embedded







Analyze
IBM Streams

IBM Streams has built-in streaming analytics

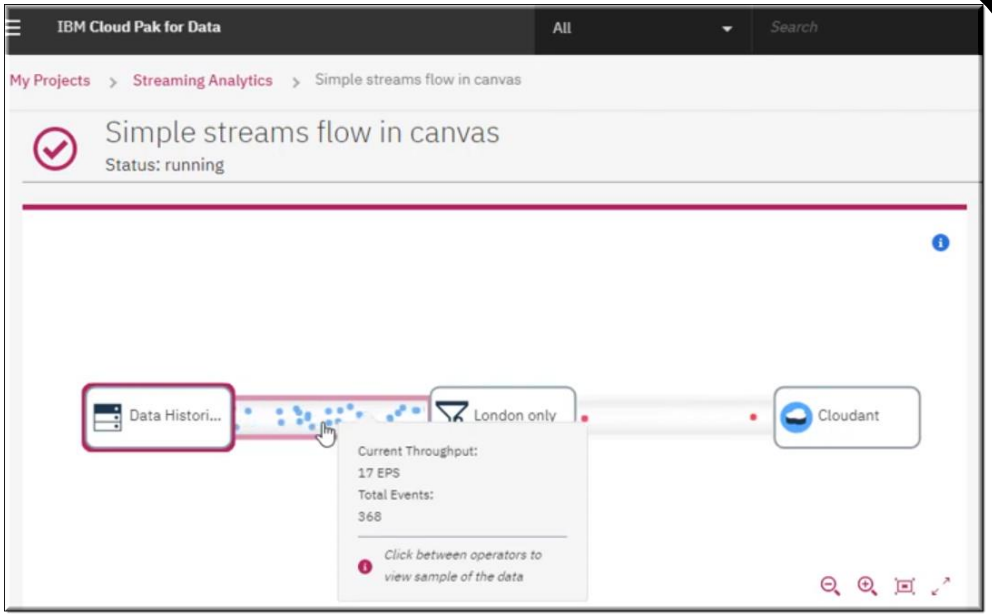
- No business disruption—run, score & update models continuously
- Machine learning, natural language, spatial-temporal, acoustic time series, etc.

Open architecture built for speed

- Millions of events per second for massive amounts of data analytics support
- Ultra-low latency clustered runtime
- Integrate via Kafka, JSON SQL/NoSQL & more

Rapid development

- Wizards, drag/drop development, performance dashboards, debugger
- Python, Java, Scala, PMML, R, C/C++ support
- VS Code and Atom plug-ins
- Export flows to a Python notebook



A streams flow consists of *operators*.

Every node on the streams flow canvas is an operator.

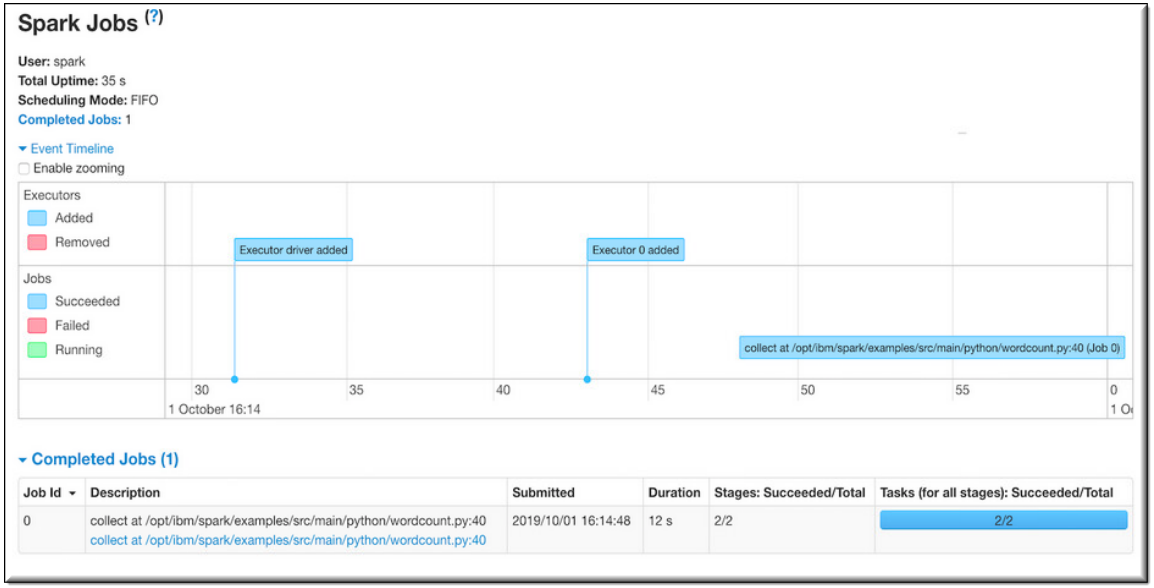
Operator types include: **sources, targets, data processing, alerts and real-time analytics**

Analyze
Analytics Engine - powered by Apache Spark

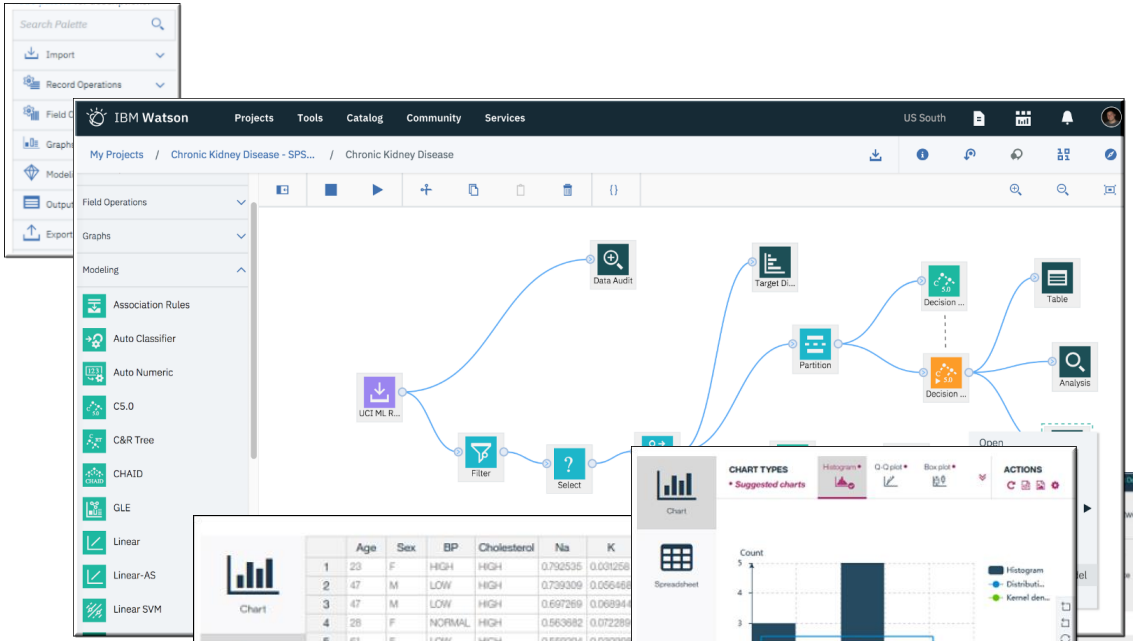
Analytics Engine is a serverless, performant, customizable, and dedicated Spark engine that is available in seconds.

100% open source, Analytics Engine can run a variety of workloads on the CPD cluster:

- **Watson Studio notebooks that call Apache Spark APIs**
- **Spark application that run Spark SQL**
- **Data transformation jobs**
- **Data Science jobs**
- **Machine Learning jobs**



Analyze
Premium Service: SPSS Modeler



SPSS Modeler

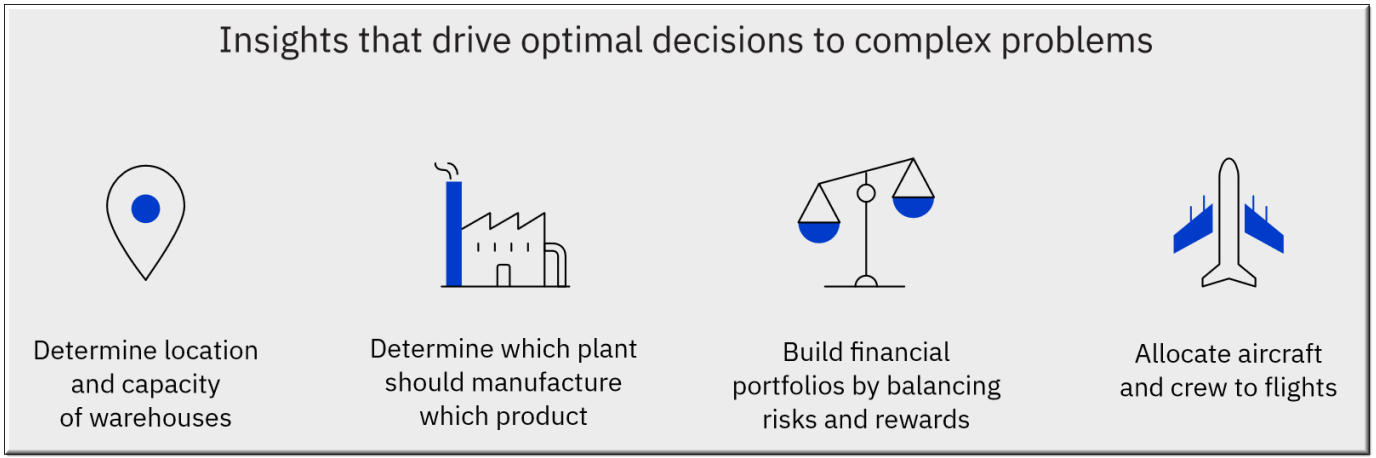
- A leading visual data science and machine-learning and predictive analytics solution
- Helps enterprises accelerate time to value and achieve desired outcomes by speeding up operational tasks for data scientists and business analysts
- Tap into data assets and modern applications, with complete algorithms and models that are ready for immediate use

Analyze
Premium Service: Decision Optimization

Decision Optimization (DO) enables data science teams to capitalize on the power of *prescriptive analytics* and build solutions using a combination of techniques like optimization and machine learning.

Integrated with Watson Studio, Decision Optimization can combine optimization techniques with coding and non-coding tools, model management and deployment – as well as other data science capabilities.

Decision Optimization evaluates millions of possibilities – balancing trade-offs and business constraints to find the best possible solution.



Analyze
Watson Machine Learning : WML Accelerator for Deep Learning

The **Experiment Builder** GUI interface is available from a WML project when you install the **WML Accelerator**. It is the simplest method to perform Deep Learning experiments.

The Watson Machine Learning Accelerator has the following components:

- **IBM Spectrum Conductor Deep Learning Impact version 2.1.0:**
  Provides robust, end-to-end workflow support for deep learning application logic for the complete lifecycle management from data ingest and preparation to building, optimizing, training and testing the model.
- **IBM Spectrum Conductor version 2.4.0:**
  A *highly available* and resilient *multitenant distributed* framework, providing deep learning application lifecycle support, centralized management and monitoring, and end-to-end security.
- **Deep Learning Frameworks:**
  TensorFlow, PyTorch, Keras, and Caffe
- **Deep Learning Rest API**

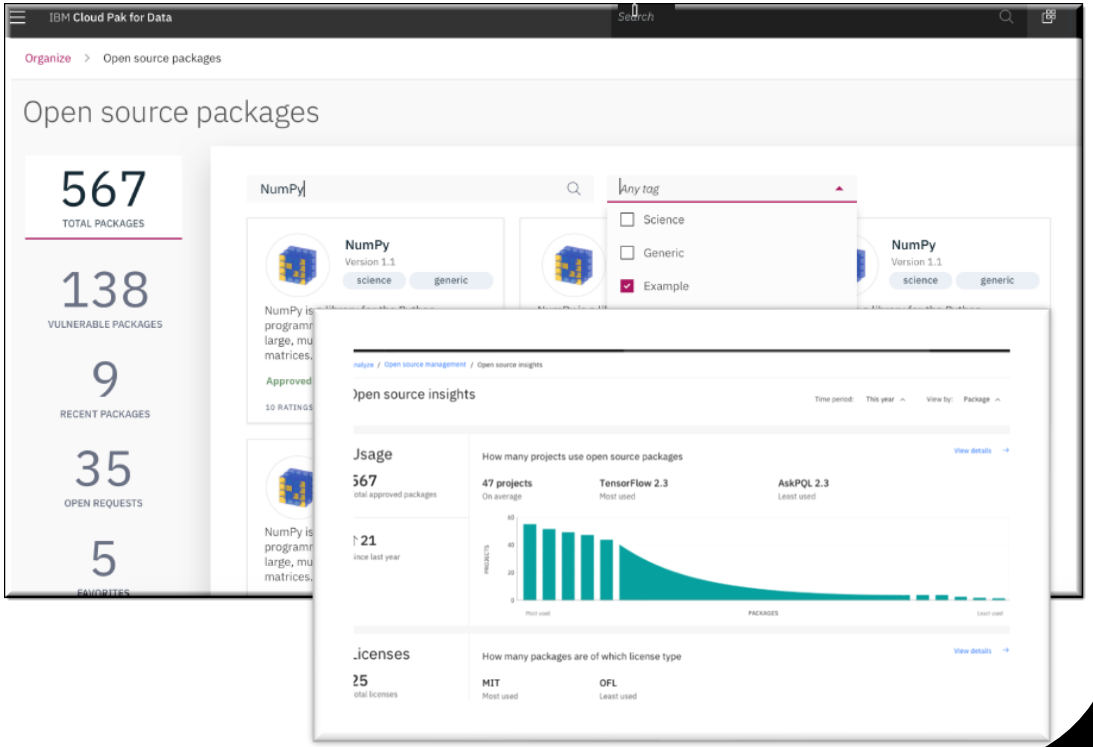
Key Features

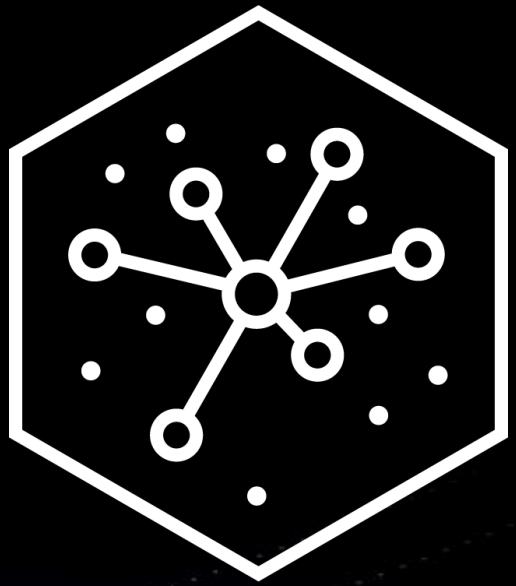
- Supported AI frameworks with performance optimizations for GPU acceleration
- Transparent GPU Topology for Distributed Training
- Auto Hyper Parameter Optimization (HPO)
- Multi-Tenancy for Training and Inference
- Elastic Distributed Training (EDT)
- Resource Utilization, Monitoring, and Reporting
- Elastic Distributed Inference (EDI)
- Secure Deployment Model
- Role Based Access Control / Kerberos Authentication

Analyze
Open Source Software Management

Open Source Software Management is:

- A *centralized inventory* of approved open source packages to ensure code quality & security
- A vehicle for *self-service open source consumption* and workflow requests for developers
- A *correlation of vulnerabilities* across open source portfolio: highlight security risks
- *An infusion of collaboration* by allowing developers to rate and comment on opensource packages
- *A set of dashboards* for CIOs to manage risks and accelerate innovation with OSS





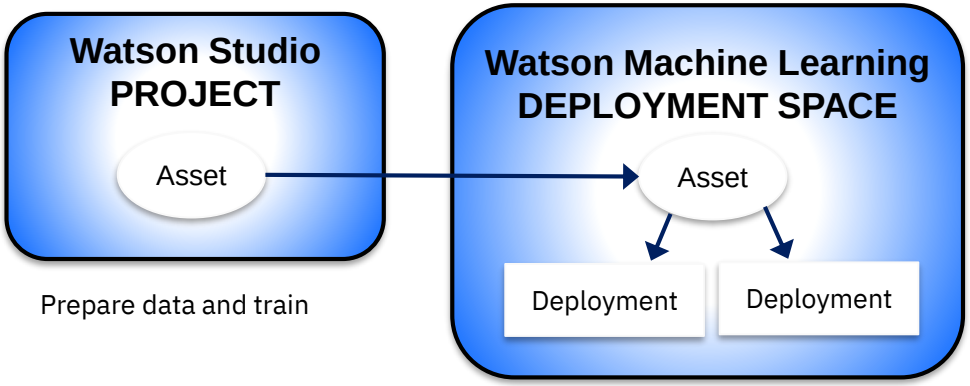
Deploy

Lab 07 – Deploy

Deploy
Watson Machine Learning: Deployment Spaces

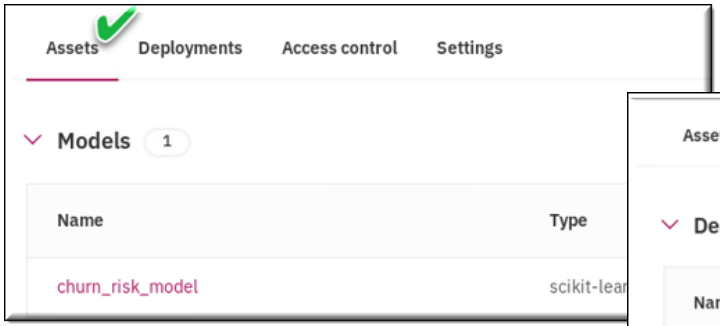
A **Deployment Space** is where you can:

- **Promote and save models**
- **Create the deployments from the models**
- **Find the information you need to score the model and get a prediction back**
- **Embed the deployment in an app so you can interact with it programmatically**



Prepare data and train

Configure, test, deploy and monitor



Deploy
Watson Machine Learning: Deployments

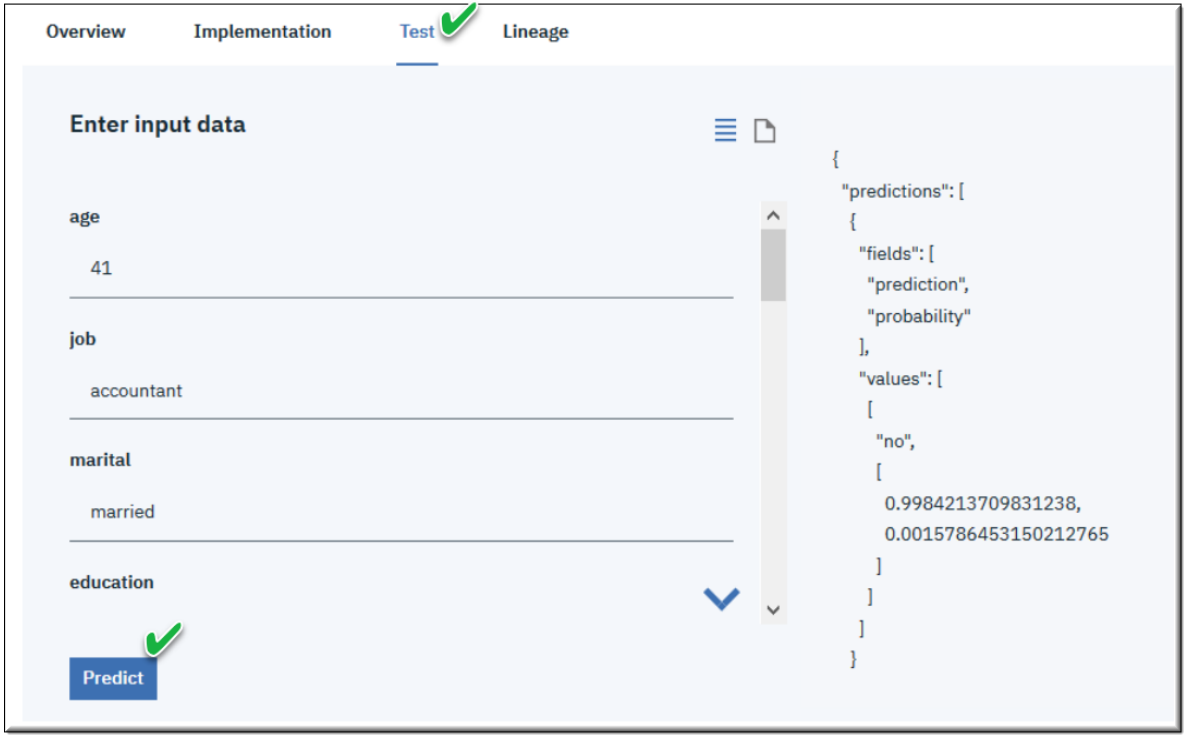
A **Deployment** is the last stage of the model development work. It means you put the model into production so that you can pass data to the model and return a score (or prediction).

After deploying a model, you can access the model *endpoint*, which you will need to make the model available for wider use in applications.

There are three type of WML deployments:

- **Online –** Provides an API endpoint needed to access the deployment programmatically to use in an application. Code snippets are provided in a variety of programming languages that illustrate how to access the deployment.
- **Batch –** Processes input data from a file and writes the output to a file.
- **Virtual** – For models downloaded into WML from different frameworks other than those built on the CPD platform.

Deploy
Watson Machine Learning: Online deployment testing



- You can use the *form* in the *Deployment Space* to easily test an online deployment.
- On the *Test* tab of the Deployment details page, enter test data, and click *Predict* to see the result.

Cloud Pak for Data

We will return for lab review at 4:30 pm. Please work on Lab-6

IBM Analytics Modernization Workshop

Part 3

• Introduction • Business Use Case	• Lab 01 • Lab 02
• Collect: Connect • Organize • Collect: Virtualize	• Lab 03 • Lab 04 • Lab 05
• Analyze • Deploy • Infuse – OpenScale • Infuse – Cognos Analytics • Wrap up	• Lab 06 • Lab 07 • Lab 08 • Lab 09 • Lab 10

IBM Analytics Modernization Workshop

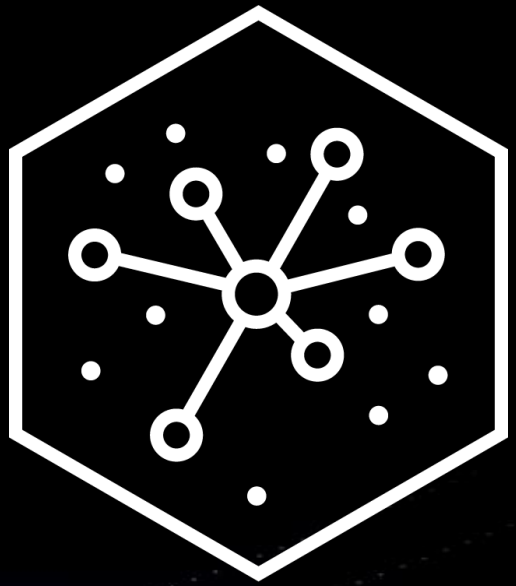
Part 3

- Introduction
- Business Use Case
- Collect: Connect
- Organize
- Collect: Virtualize
- Analyze
- Deploy
- Infuse – OpenScale
- Infuse – Cognos Analytics
- Wrap up

- Lab 01
- Lab 02
- Lab 03
- Lab 04
- Lab 05
- Lab 06
- Lab 07
- Lab 08
- Lab 09
- Lab 10

Infuse

Lab 08 – Infuse: Watson OpenScale



Infuse
Watson OpenScale: Overview

Watson OpenScale:

- Automates and operates AI at scale across its entire lifecycle
- Delivers transparent, explainable outcomes freed from bias and drift
- Provides confidence in AI outcomes and spans the gap between the teams that operate AI and the business units that use these applications
- Monitors models developed in a 3rd party IDE, open source framework and hosted in a 3rd party or private model serve engine

Manage AI at Scale

Watson OpenScale

- Operations Dashboard
- Fairness & Bias Mitigation
- Drift Detection
- Explainability
- Business KPIs
- Payload Logging
- Data Mart

Model build / train frameworks

TensorFlow, scikit learn, Keras, PyTorch, Caffe2, Spark MLlib

Model serving environments

kubernetes, Azure ML, aws

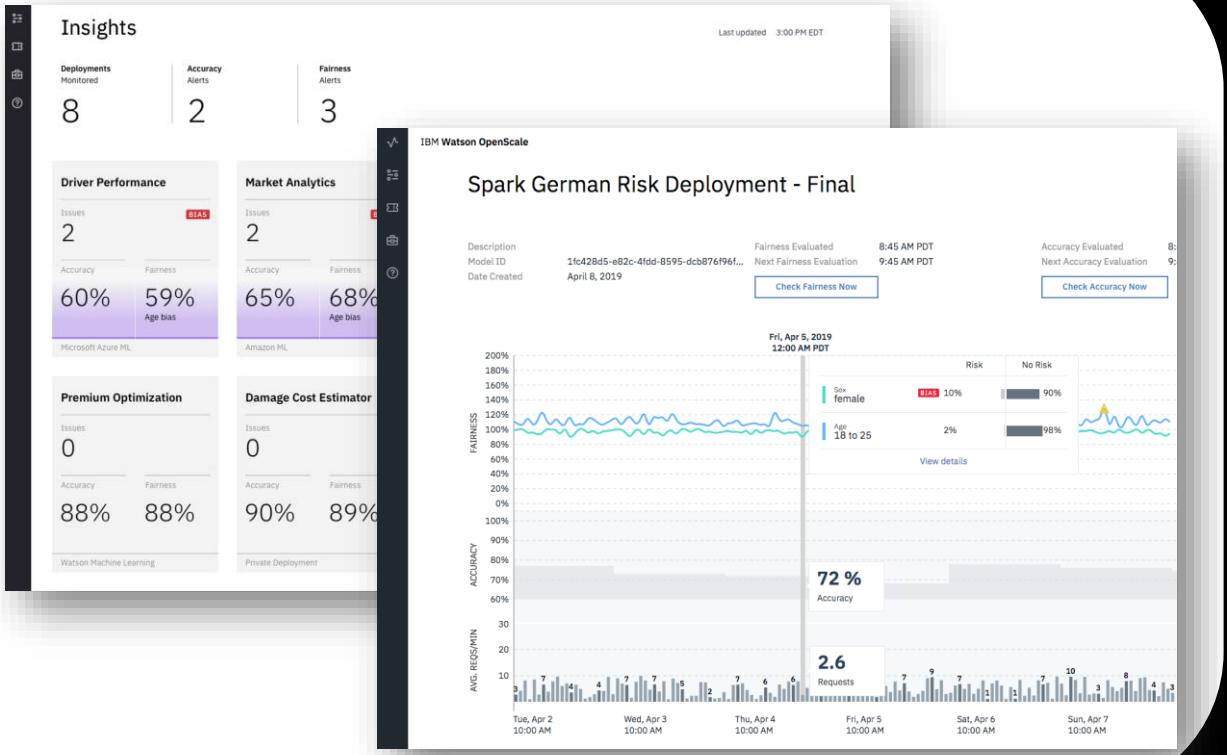
Infuse
Watson OpenScale: Operations dashboard

Description:

Monitor deployed models in a single dashboard that can be filtered by deployment making it easy to manage AI in apps

Value:

- Configure alerts or actions to be triggered when KPIs exceed threshold, ensuring model quality for improve business outcomes
- Measure model accuracy as it pertains to its ability to deliver outcomes more accurate than knowledge workers
- Provides “continuous evolution” for your models



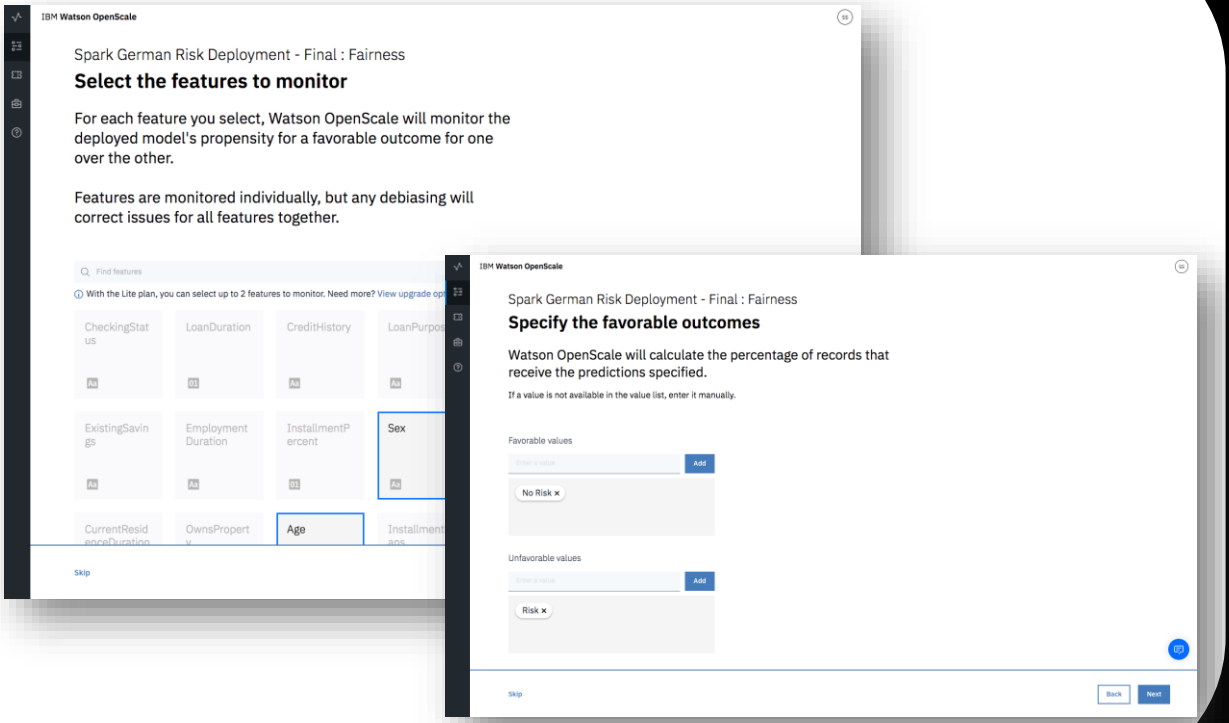
Infuse
Watson OpenScale: Model Fairness

Description:

Production Models need to make fair decisions and *must not be biased* in their recommendations

How it works:

- Outcomes are selected as “favorable or unfavorable”
- “Favored Populations” and “protected populations” are selected where majority and minority groups are found
- A score is calculated based on the probability of favorable outcome for minority vs. Probability of favorable outcome for majority



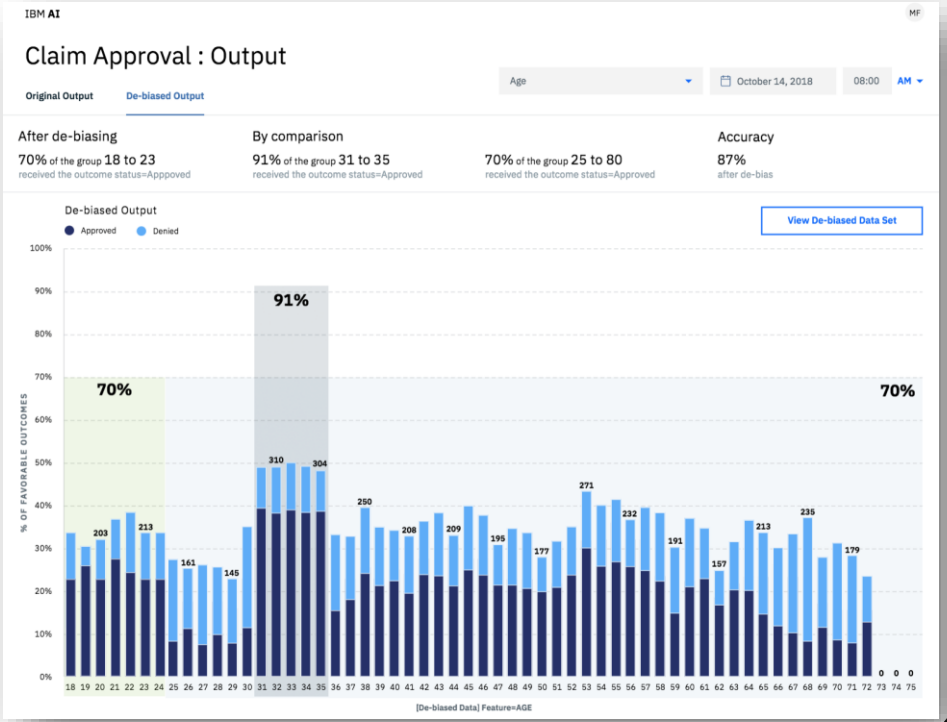
Infuse
Watson OpenScale: Bias Mitigation

Description:

Fairness is enforced with automatic bias mitigation.

How it works:

- Calculated on an *hourly basis* (over a sliding window defined by the user)
- Optimizations identify the *right subset of data to perturb* (rather than perturbing all the data)
- *Perturbed data is sent to the deployed* model to determine effect of perturbations
- An internal bias detection model (logistic regression) is built using perturbed data that *classifies whether new prediction will be biased or not*
- Users receive both the *original prediction* plus the *internal model's classification* of whether the monitored model's prediction is biased or not

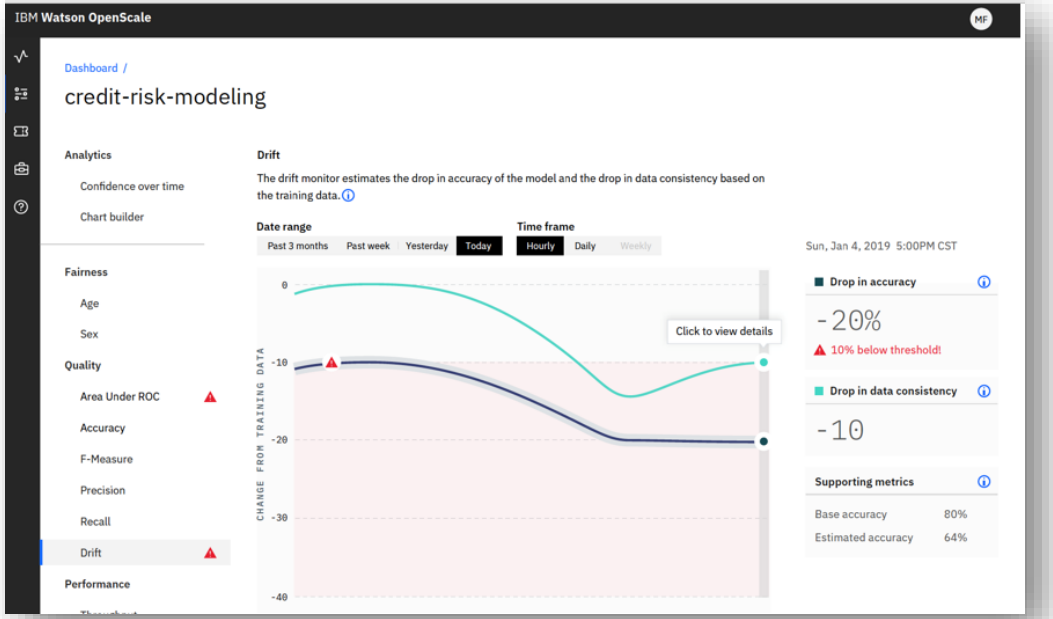


Infuse
Watson OpenScale: Drift detection

Description:

OpenScale monitors for two types of drift:

- **Drop in accuracy**: It estimates the drop in accuracy of the model at runtime. Accuracy could drop if there is an increase in transactions similar to those which the model was unable to evaluate correctly with the training data.
- **Drop in data consistency**: It estimates the drop in consistency of the data at runtime as compared to the characteristics of the data at training time.



OpenScale does drift detection on the entire payload data.

OpenScale measures the drift without requiring labeled data. Accuracy computation using labeled data can be expensive and might not be comprehensive

Infuse
Watson OpenScale: Explainability

Description:

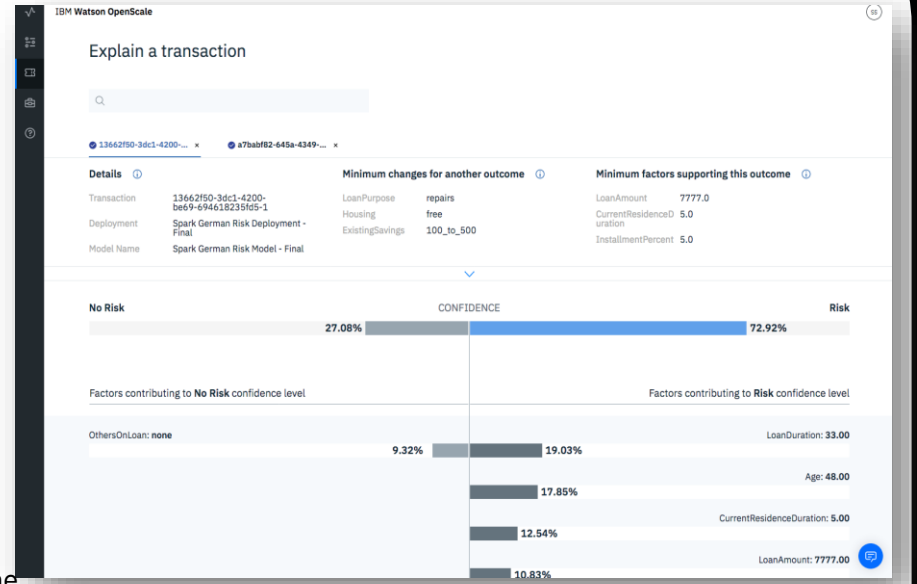
Allows you to understand which feature values of a model that are most influencing a prediction for a specific transaction

Example:

A loan is not approved by a model prediction - explainability will tell you why

How it works:

- Perturbation analysis on thousands of variations
- Risk model is created for two variations:
  - **LIME (local) Explanation:** set of features which played a positive or negative role in the prediction - also identifies the feature weights which helps to identify the most or least important features
  - **Contrastive Explanation:** Explains the behavior of the model in the vicinity of the data point whose explanation is being generated – assumption: the most common value is the least interesting from an explanation point of view

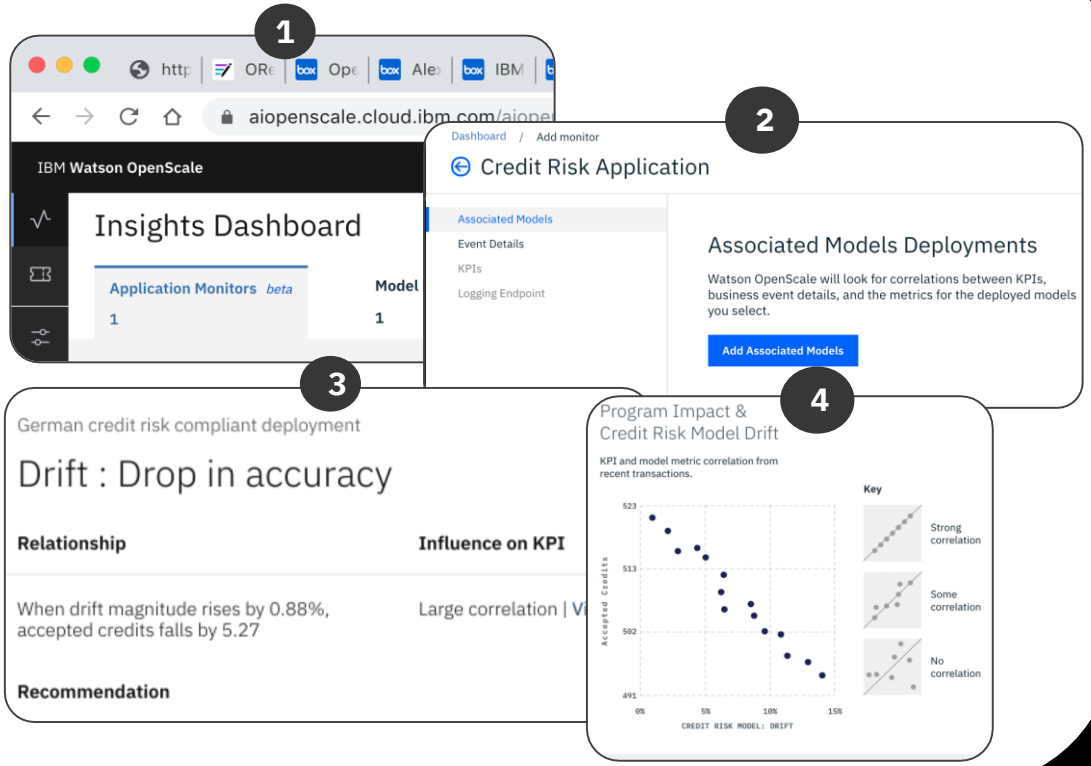


Infuse
Watson OpenScale: Business (Application) KPIs

Description:

The OpenScale UI dashboard contains an Application centric dashboard in addition to the Model centric dashboard:

- Bring in business event data into OpenScale and use it to compute Key Performance Indicators (KPIs)
- Monitor correlation between model monitors and KPIs - get alerts and recommendation from OpenScale
- Visualize the correlation through a plot



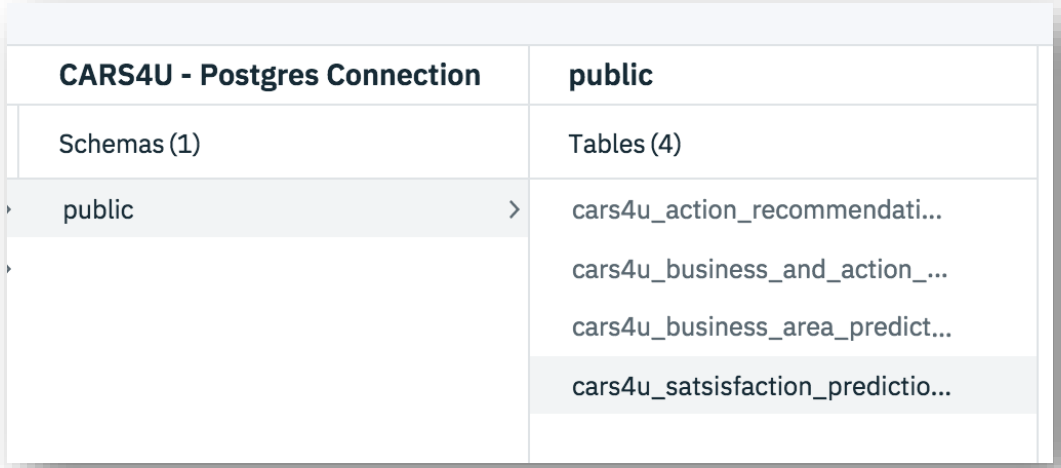
Infuse
Watson OpenScale: Payload logging

Description:

Payload logs capture (in Postgres) the request sent to the model or python function along with statistics about its health which when combined with feedback data provides insights into AI and application behavior

Value:

- Enable logging of payloads for *traceability* of business outcomes to AI recommendations
- Payload *data powers visualizations* for the OpenScale dashboard, making it easy to monitor health of deployed models
- Payload *fuels the Open Datamart* for custom reporting and business KPI integration

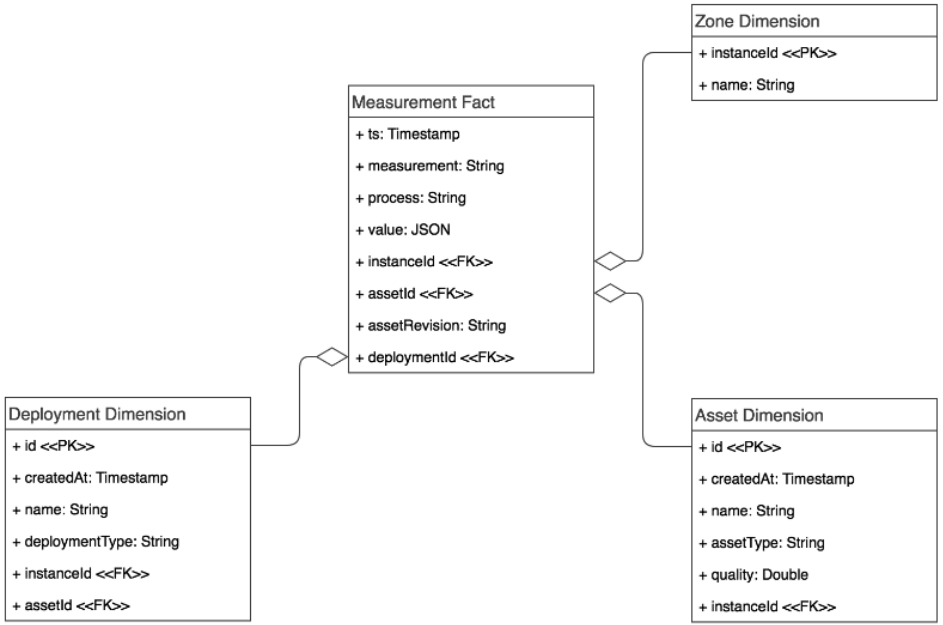


Infuse
Watson OpenScale: Data mart

Description:

The Data Mart is a data system to support the following use cases:

- Enable *OpenScale UI dashboards* for operations staff
- *Enable 3rd party reporting tools* so operations staff can customize their own dashboards using the underlying OpenScale data with their own tools
- *Enable data engineers to integrate OpenScale data* with existing enterprise data marts, warehouses, and data lakes by exporting data into those systems
- *Provide data scientists with the actual runtime payload, scoring and feedback data* that can be utilized in continuous learning



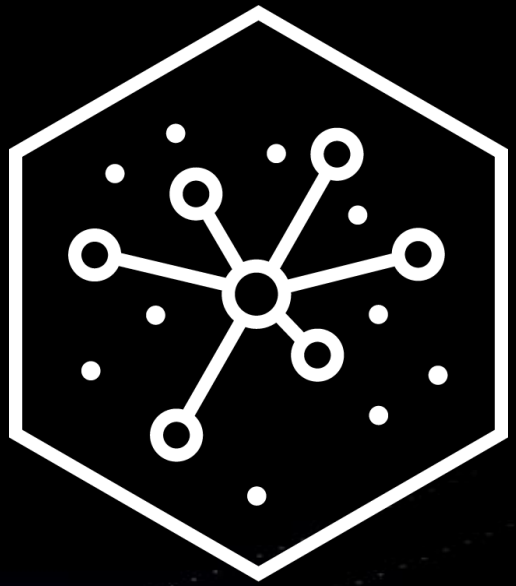
Cloud Pak for Data

We will return for lab review at
5:45 pm. Please work on Lab-8

(Note Lab 7 should be done after issuing the OpenScale set up).

Infuse

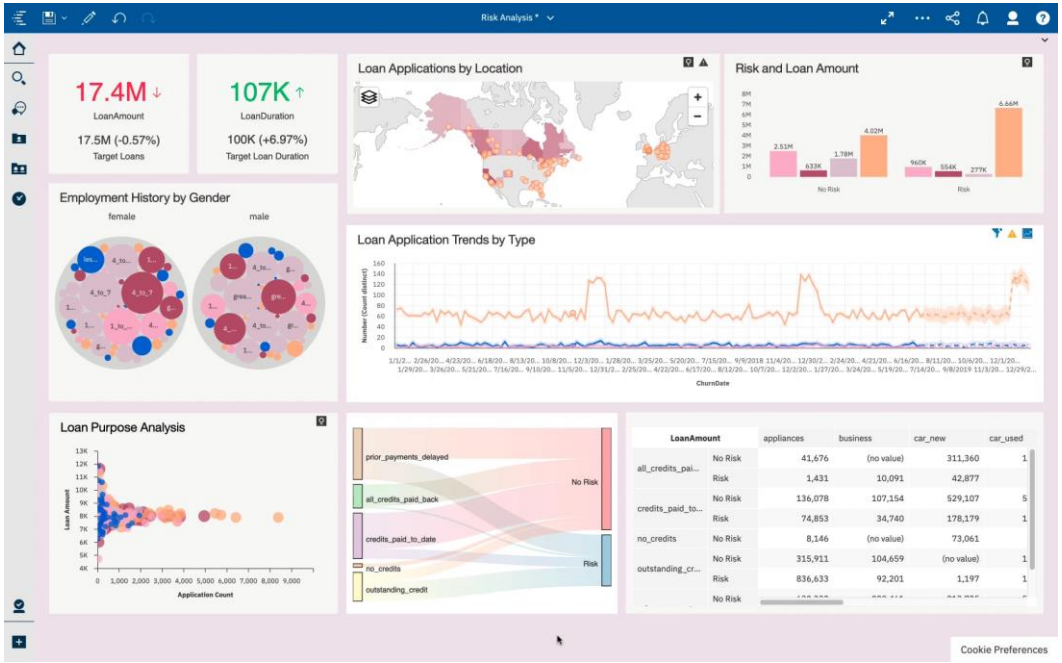
Lab 09 – Infuse: Cognos Analytics



Infuse
Premium Service: Cognos Analytics

Cognos Analytics is self-service analytics, infused with AI and machine learning.

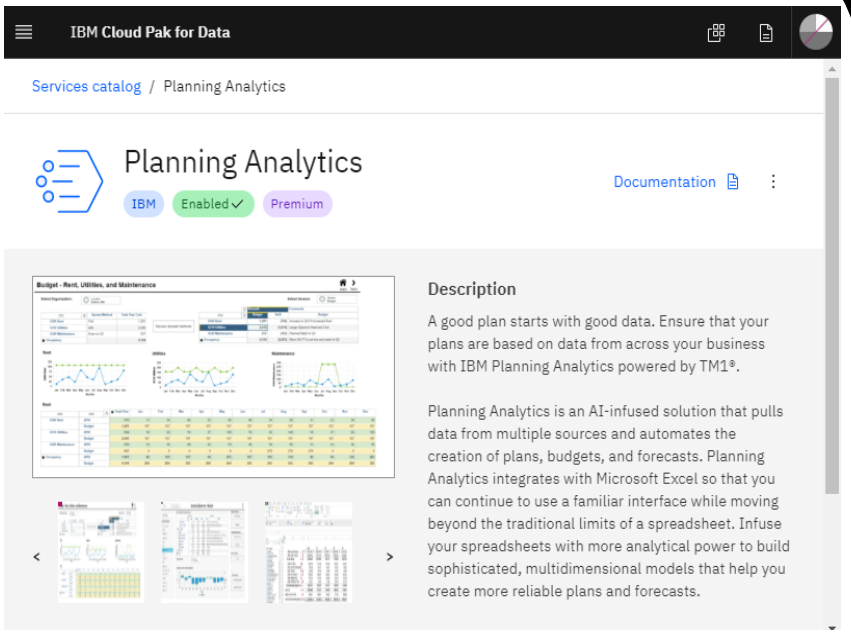
- Enables you to create stunning visualizations to share your findings through *dashboards* and *reports*
- These can be embedded (infused) into your applications
- The Cognos Analytics service makes it easier for you to extract meaning from your data with features such as:
  - **Automated data preparation**
  - **Automated modeling**
  - **Automated creation of visualizations and dashboards**
  - **Data exploration**



Infuse
Premium Service: Planning Analytics

Use Planning Analytics to:
- Automatically create plans, budgets and forecasts
- Steer business performance by bridging operations and finance for any department allowing you to adapt to changing business conditions
- See impact before executing – explore what-if scenarios and assess impact to determine the best course of action
- Make changes in real-time – pivot plans, budgets, and forecasts quickly to meet changing demands and priorities

Differentiators:
- Adjust financial plans in real time across departments
- Protect your investment in Microsoft Excel while transcending limitations of spreadsheets
- Uncover deep insights through AI-infused planning, without the need for help from a data scientist



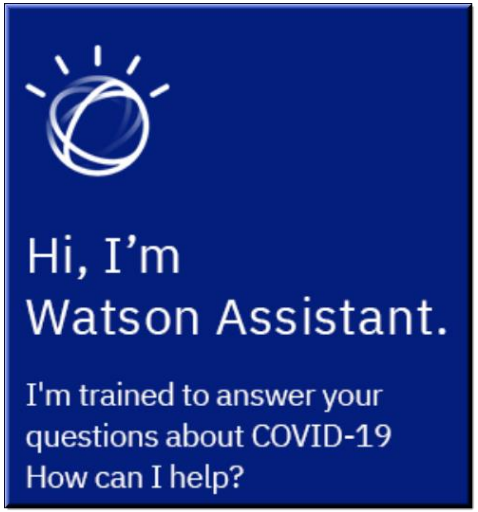
Infuse
Premium Service: Watson Assistant

Watson Assistant enables you to build conversational interfaces (chat bots) into any application, device, or channel

- Intuitive tooling for dialog building
- Intent recommendations from chat logs for continuous improvements
- Digression handling when user changes topics mod-conversation
- Out-of-the-box integration with search capabilities when coupled with Watson Discovery
- Common misspellings are automatically handled by Watson Assistant.
- Contextual entities: Support for additional languages not yet in Cloud Pak
- Fuzzy matching: Support for additional languages not yet in Cloud Pak

Benefits:

- Get up and running quickly
- Easily improve customer experience
- Reduce costs and speed of resolution
- Increase value by integrating apps and channels
- Ensure security, resiliency and data privacy - anywhere



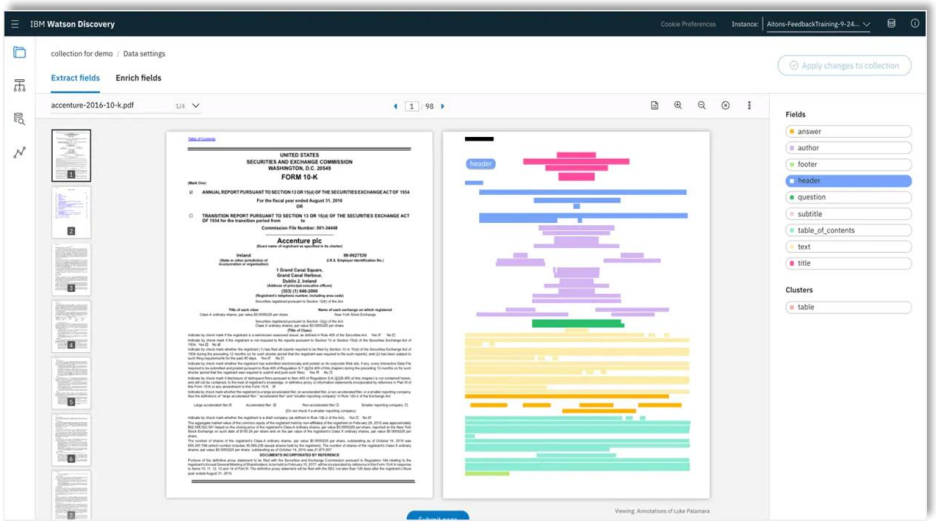
Infuse
Premium Service: Watson Discovery

Watson Discovery Surface answers and rich insights from your enterprise data.

- Award-winning enterprise search and AI search technology that breaks open data silos and retrieves specific answers to your questions while analyzing trends and relationships buried in enterprise data.
- Applies the latest breakthroughs in machine learning, including natural language processing capabilities, and is easily trained on the language of your domain.

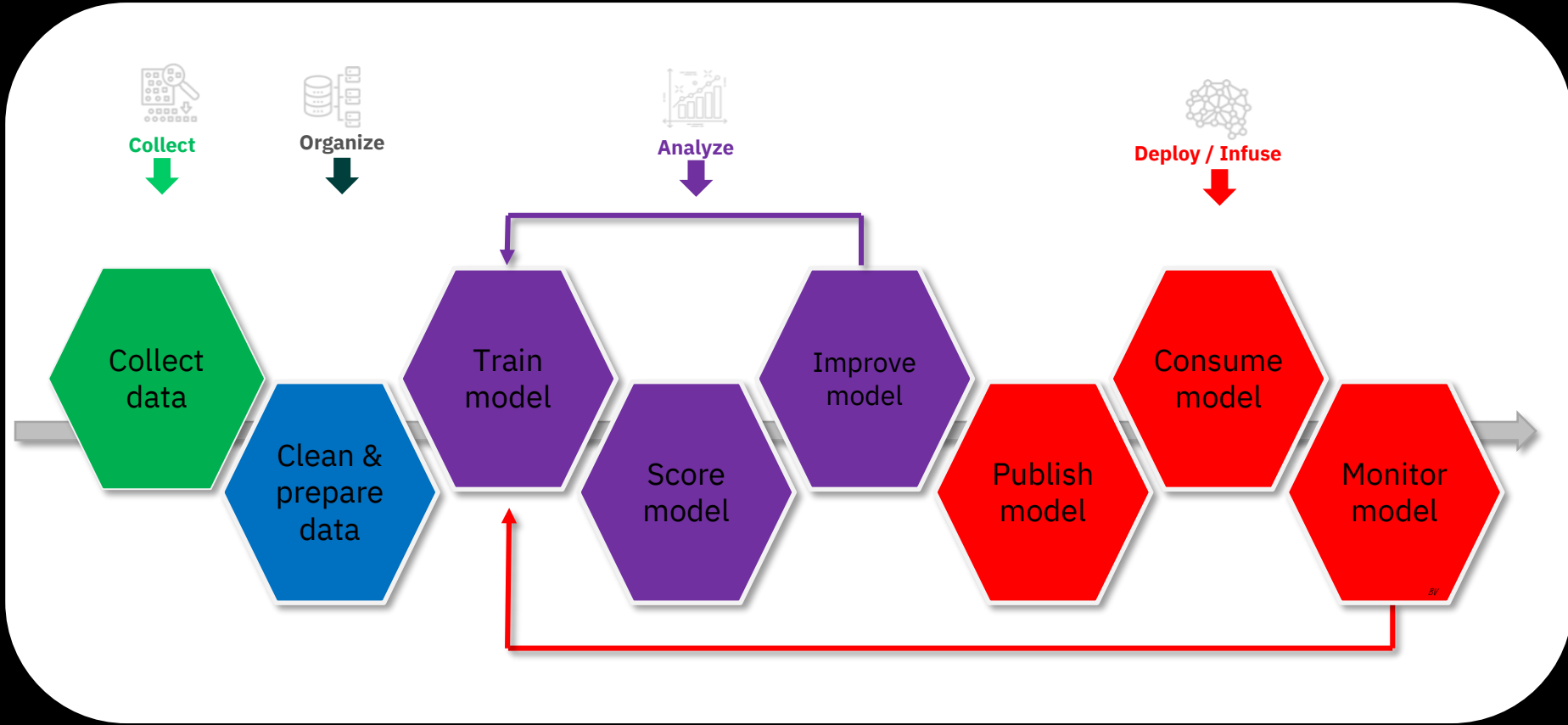
Benefits:

- Traditional enterprise search engines don't provide exact answers because they can't understand the nuances of phrases and acronyms in your industry and accurately search through your complex documents in a timely manner.
- Delivers specific answers to your queries while also serving up the entire document and supporting links, allowing your employees and customers to make informed decisions with confidence.



Machine Learning Model Lifecycle

CPD simplifies the entire process



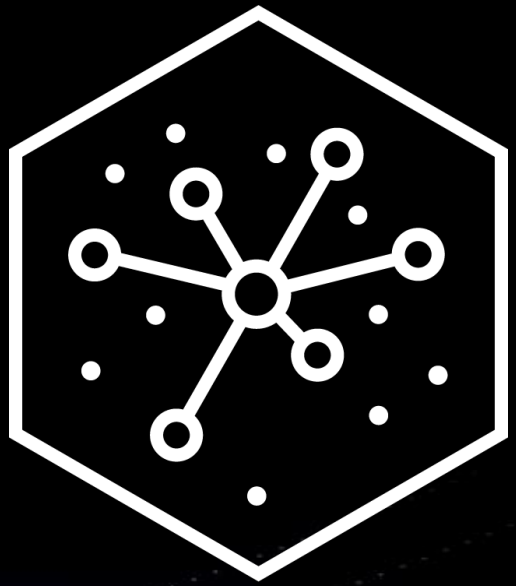
IBM Analytics Modernization Workshop

Part 3

• Introduction • Business Use Case	• Lab 01 • Lab 02
• Collect: Connect • Organize • Collect: Virtualize	• Lab 03 • Lab 04 • Lab 05
• Analyze • Deploy • Infuse – OpenScale • Infuse – Cognos Analytics • Wrap up	• Lab 06 • Lab 07 • Lab 08 • Lab 09 • Lab 10

Wrap up

Lab 10 – Wrap up



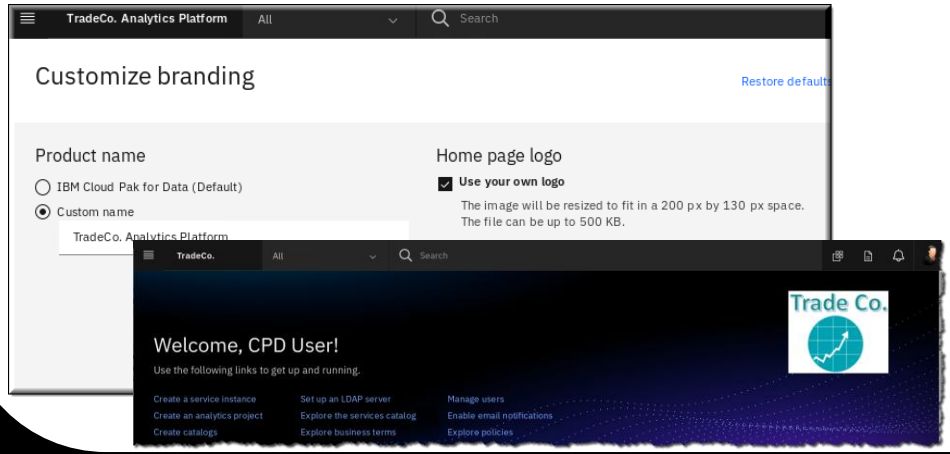
Cloud Pak for Data
Unified Experience

Customized Logo and Branding

Example: Customize the web client interface per tenant.

Customizable components in the Home Page:

- **Product name**
- **Home page logo**



Group 1 language support

All services in Cloud Pak for Data are translated for the following languages:

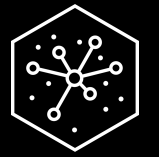
- **Simplified Chinese**
- **Traditional Chinese**
- **Japanese**
- **French**
- **German**
- **Italian**
- **Spanish**
- **Brazilian Portuguese**

Carbon 10

Modernized, modular and flexible open source design framework

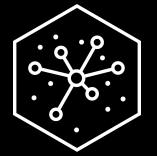
- Consistent look and feel
- Reusable components

Cloud Pak for Data
Version 3.0.1 notable updates



Update	Comment
Support for POWER systems	Build an AI foundation for speed and scale with the premier, built-in GPU acceleration platform for faster time to AI.
OpenShift Container Storage	Software-defined Storage automated for quicker and efficient hybrid multi-cloud deployments and optimized for Red Hat OpenShift Platform.
Red Hat OpenShift 4.x support	Or continue to run on RHOCP 3.11.
IAM Integration	IBM Cloud Platform Common Services (CPCS) IAM for all Cloud Paks provides authentication support via the OpenID Connect (OIDC) specification for SSO capability.
Fine-grained permissions	4 New fine-grained permissions: • Configure authentication • Configure platform • Manage users • Monitor platform
Operations Management	New utilities for upgrade, backup and restore, import and export

Cloud Pak for Data
Security Considerations



Security Features
- ✓ Security Architecture and Design
- ✓ Access Control, Authentication and Authorization (e.g. Integrates with leading LDAPs)
- ✓ Data Protection
- ✓ Security Logging

Security Engineering
- ✓ Development trained in Secure Coding Practices
- ✓ Secure Engineering Development Practices: threat modeling, risk assessment, static and dynamic code analysis, penetration testing, container scanning, etc.

Security Operations
- ✓ Audit Log consolidation and analysis
- ✓ User access management
- ✓ Security Incident Management

Governance & Compliance
- ✓ Compliance Controls defined by Outside Agencies
- ✓ System Security Plans for maintaining compliance security postures

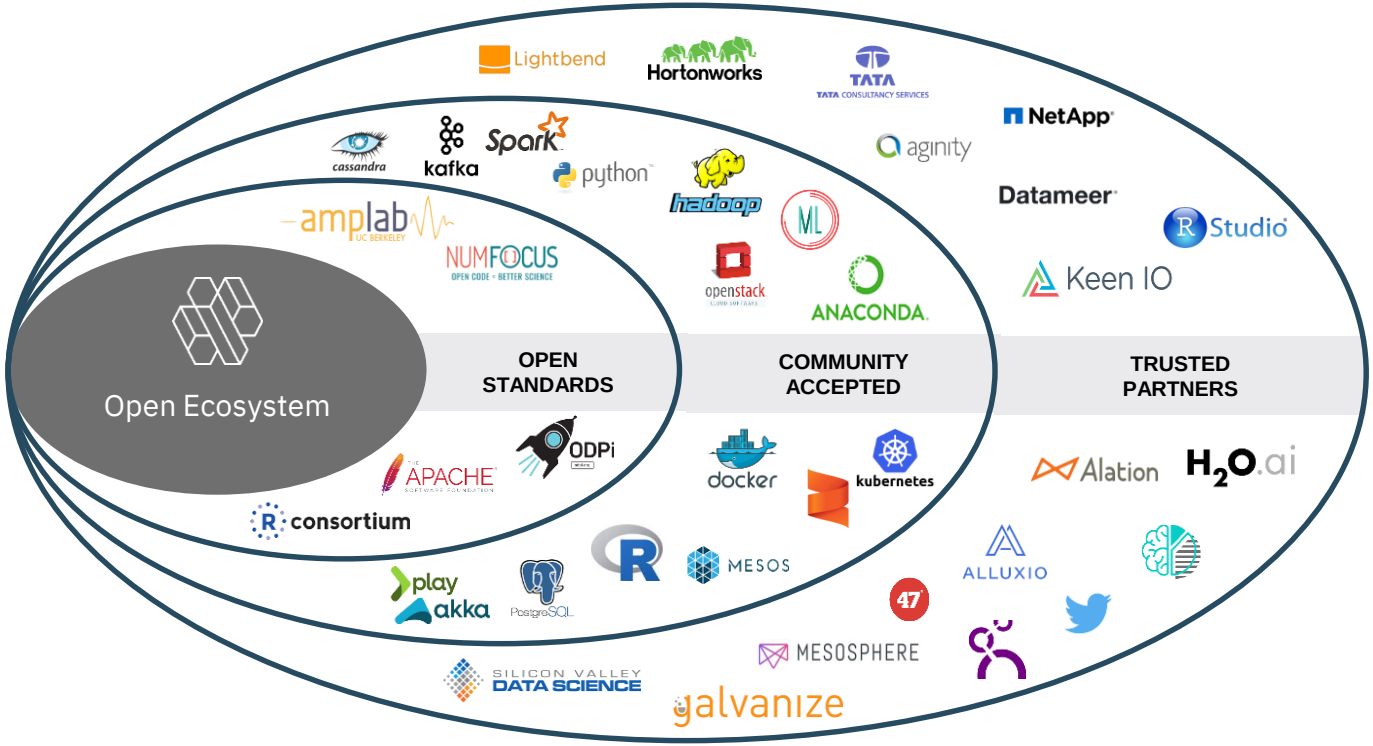
Compliance Best Practices:

FISMA High *"ready"* with System Security Plans, spanning 350 controls:
- Risk Assessment
- Certification, Accreditation and Security Assessments
- System Services and Acquisition
- Security Planning
- Configuration Management
- System and Communications Protection
- Personnel Security
- Awareness and Training
- Physical and Environmental Protection
- Media Protection
- Contingency Planning
- System and Information Integrity
- Incident Response
- Identification and Authentication
- Access Control, Accountability and Audit

GDPR *"readiness"* considerations

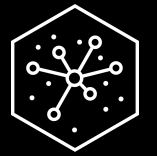
CPD built on an open Ecosystem

Where IBM leads, partners and co-creates



IBM's approach to Open technology: https://developer.ibm.com/articles/cl-open-architecture-update/

CPD Packaging
Edition Overview



Cloud Pak for Data Editions

Make your data ready for AI – Cloud Agility, Lightning Fast & AI-ready

Standard Edition

- Beginning with a minimum of 24 VPCs
- Expands in increments of 1 VPC
- Up to a *maximum* of 64 VPCs
- You cannot use separately priced premium services with this edition
- 50% Enterprise Edition list price

Enterprise Edition

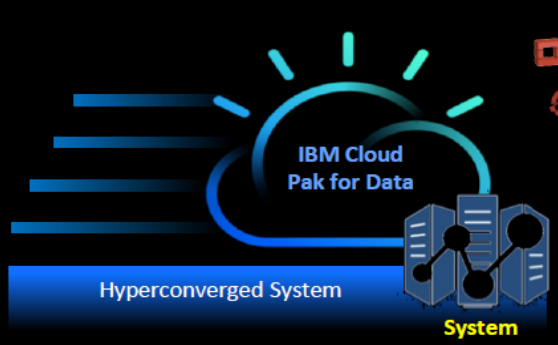
- Deploy in any form: on premises, on public cloud or CPD System
- Beginning w/ recommended minimum of 48 VPCs
- Select 24 VPC configurations supported
- Expands in increments of 1 VPC
- No maximum

Non-Prod Edition

- Can only be used for non-production scenarios
- No restriction in terms capabilities or VPC quantities
- Needs to be in separate namespace from production licenses
- Need parallel and appropriately sized Standard or Enterprise Edition licenses for production use
- 50% of Enterprise Edition list price

IBM Cloud Pak for Data System

True plug-and-play enterprise data & AI in hours right out of the box



- Brings the elasticity & scalability of public clouds securely behind the firewall
- Connect all your data for self-service analytics
- Operationalize AI with trust & transparency
- Deploy dynamic cloud native data workloads

An **all-in-one** data & AI system with all the necessary systems and software components pre-integrated

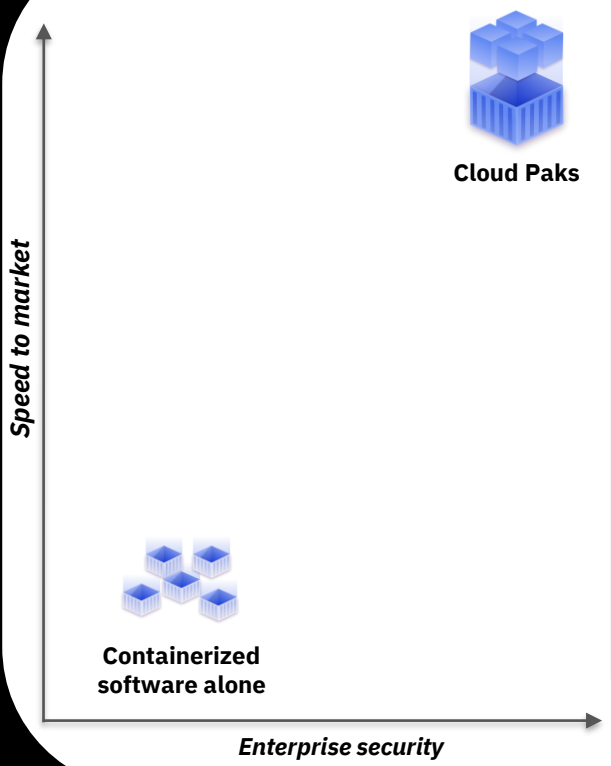
Deploy a complete private cloud in under 4 hours, with no assembly required

Dynamically scale compute, storage and networking resources with plug and play of new hardware nodes

Simplify management and optimization with a unified and intuitive dashboard

Cloud Pak for Data vs. Containers Alone

IBM Certified and production ready



	Containers Alone Client creates containers or receives software as standalone container(s)	IBM Cloud Paks Complete solutions certified for enterprise use cases
Runs anywhere	Yes	Yes
Vulnerability scanned	Yes	Yes
Red Hat container certification	Depends on product	Yes
Complete solution w/ container platform	No	Yes
Flexible & modular	No	Yes
IBM certified/orchestrated for production (Built for Kubernetes by experts; certified against 200+ criteria)	No	Yes
Multicloud validation	No	Yes
Integrated deployment experience	No	Yes
Full stack support by IBM (Base OS, software, and container platform)	No	Yes
License metering integration	No	Yes
Scalable and resilient	No	Yes
Encrypted secrets / limited privileges	Do it yourself	Yes
Management and operations	Build your own	Yes
Lifecycle Management	Manage it yourself	Yes

Cloud Pak for Data
IBM Kubernetes Certified

http://ibm.biz/cp-certify

Production Grade	Security	Quality Assurance	Lifecycle Management
• Multi-cloud • Storage • Networking • Resiliency • Scalability • Self Healing • Recoverable	• Vulnerability Management • Limited Security Privilege • Secure Access, Keys, Certs • Network & Data Protection • Secrets and SPI • Security and Privacy by Design	• Comprehensive Testing • Architectures • OCP Versions • Currency • Airgap • Upgrades	• Patch • Upgradeability • Rollback/Restore • Backup/Recovery • OCP Version Alignment • Kubernetes Currency

Consistency and Standards

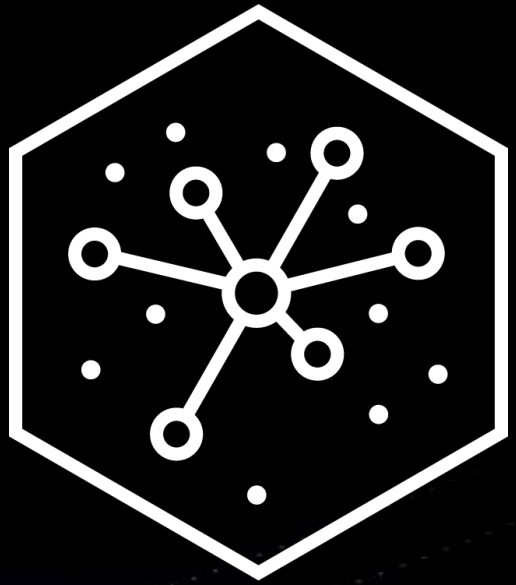
- Consistent Packaging / Publishing
- Supporting Operators and Helm
- Consistent Entitlement management
- Common management of OSS elements
- UBI and Red Hat Certified
- Consistent use of OCP and IBM Services
- ~200 Code Standards enforced
- Governed Best Practice / Anti Practices

Red Hat OpenShift | OPERATOR FRAMEWORK | redhat CERTIFIED

- Managed Image CVEs
- Packaging
- Publishing
- Trusted Source
- E2E Support

Thank you for your time!

Begin your journey now on the IBM Platform built for AI...



We appreciate your feedback.

Presentation by Burt Vialpando and Kent Rubin - IBM

Copyright and trademarks

© Copyright IBM Corporation 2020

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
July 2020

IBM, the IBM logo, ibm.com, API Connect, Db2, Elastic Storage, FlashCore, POWER, Spectrum Scale, UrbanCode, WebSphere and IBM Z are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
VMware, the VMware logo, VMware Cloud Foundation, VMware Cloud Foundation Service, VMware vCenter Server, and VMware vSphere are registered trademarks or trademarks of VMware, Inc. Or its subsidiaries in the United States and/or other jurisdictions.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.