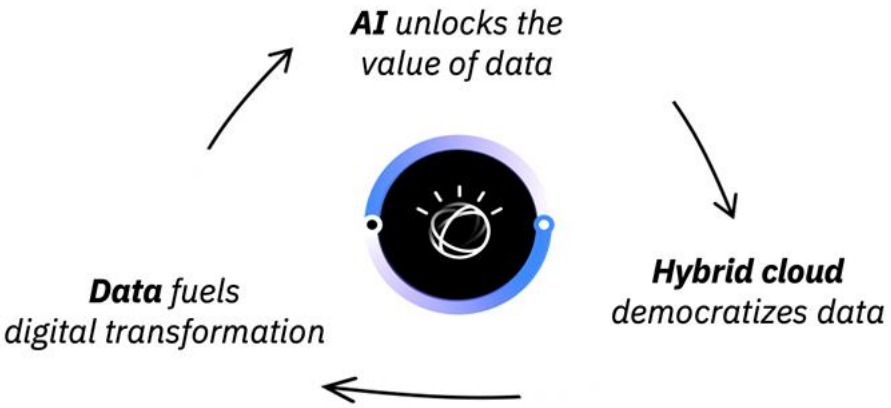
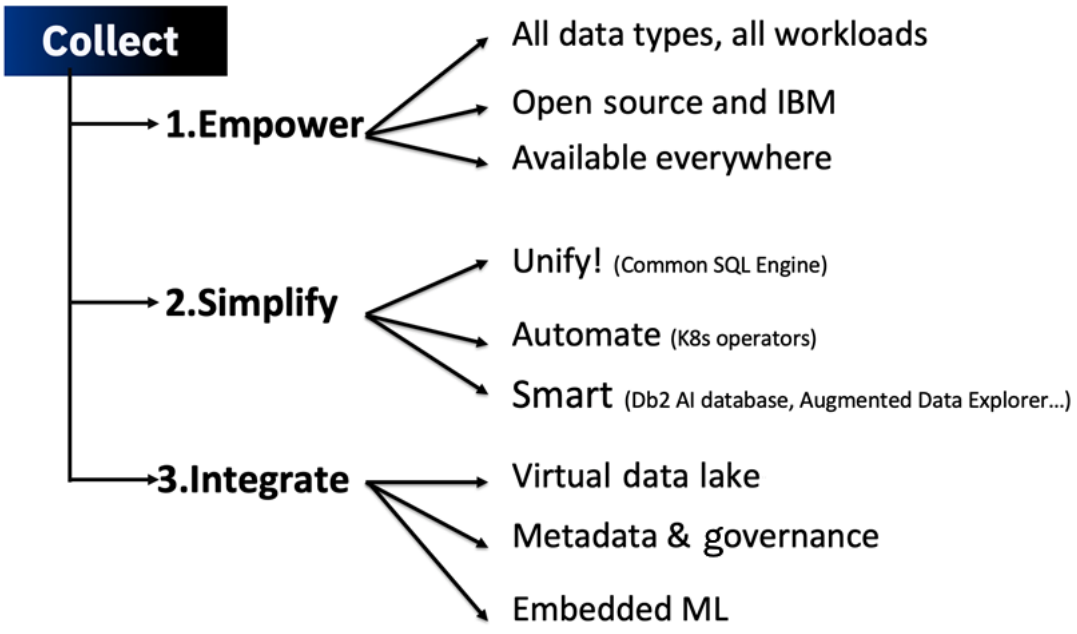
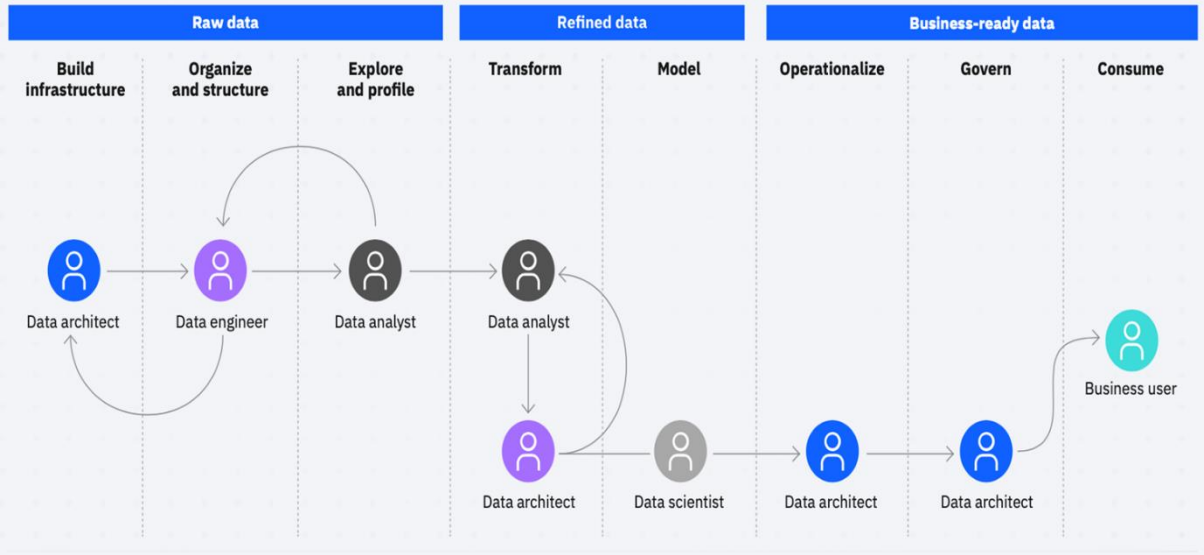
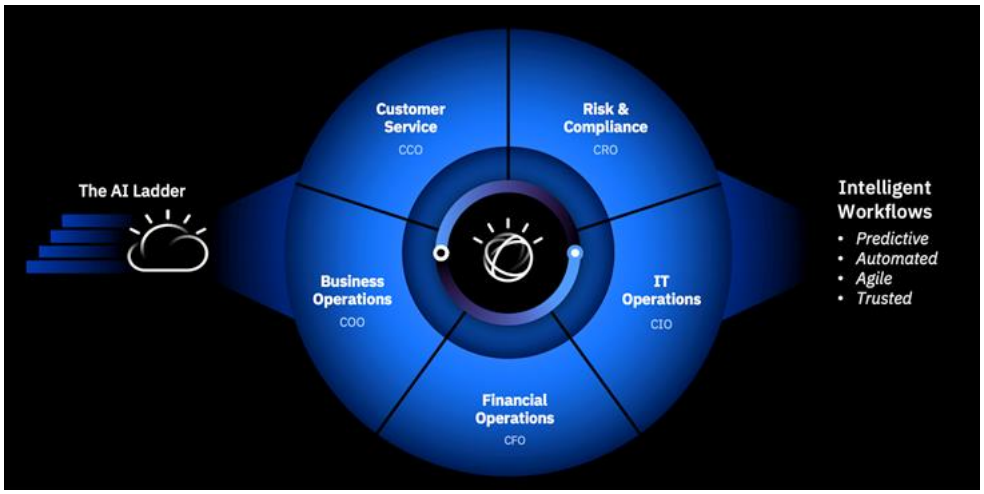
Chapter 1: The AI Ladder: IBM's Prescriptive Approach

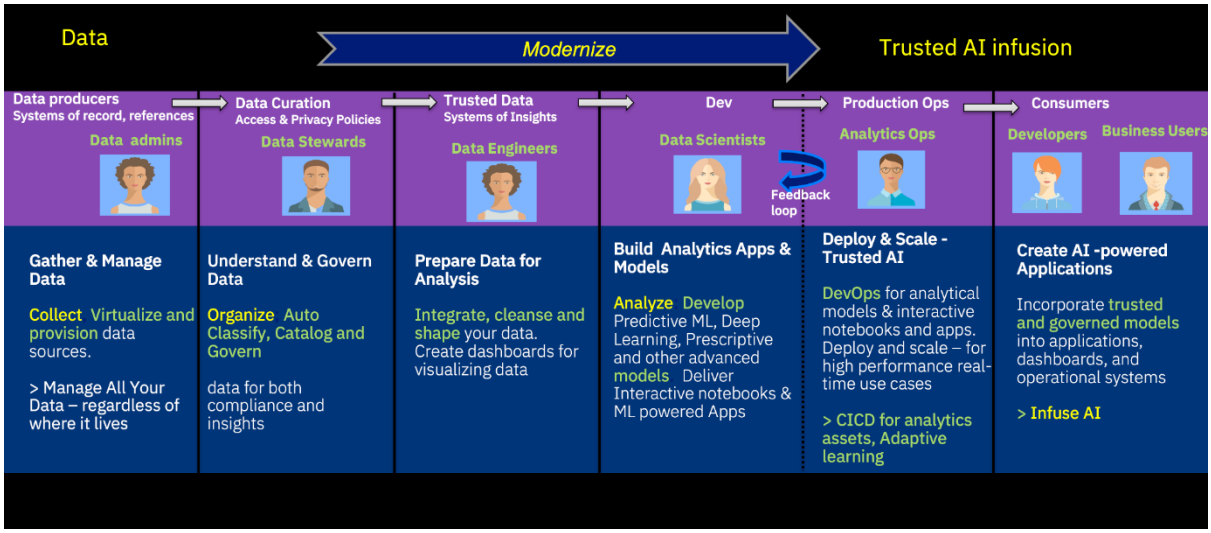


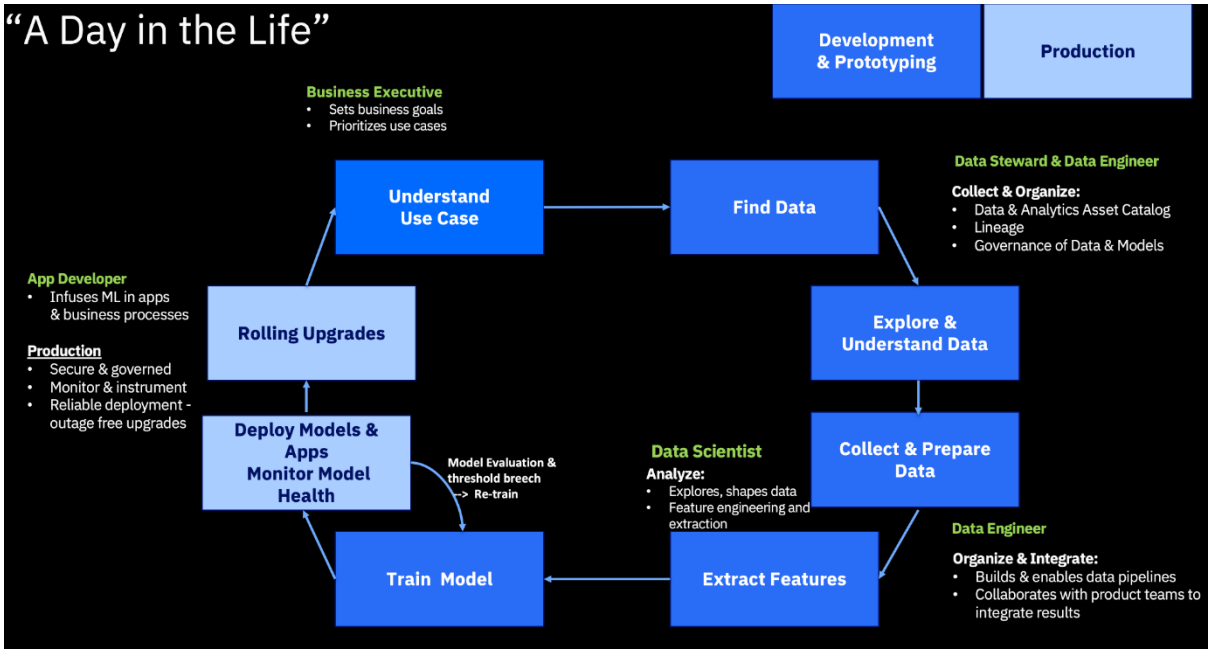



Raw data
Refined data
Business-ready data
Build infrastructure
Organize and structure
Explore and profile
Transform
Model
Operationalize
Govern
Consume
Data architect
Data engineer
Data analyst
Data analyst
Data architect
Data scientist
Data architect
Data architect
Business user

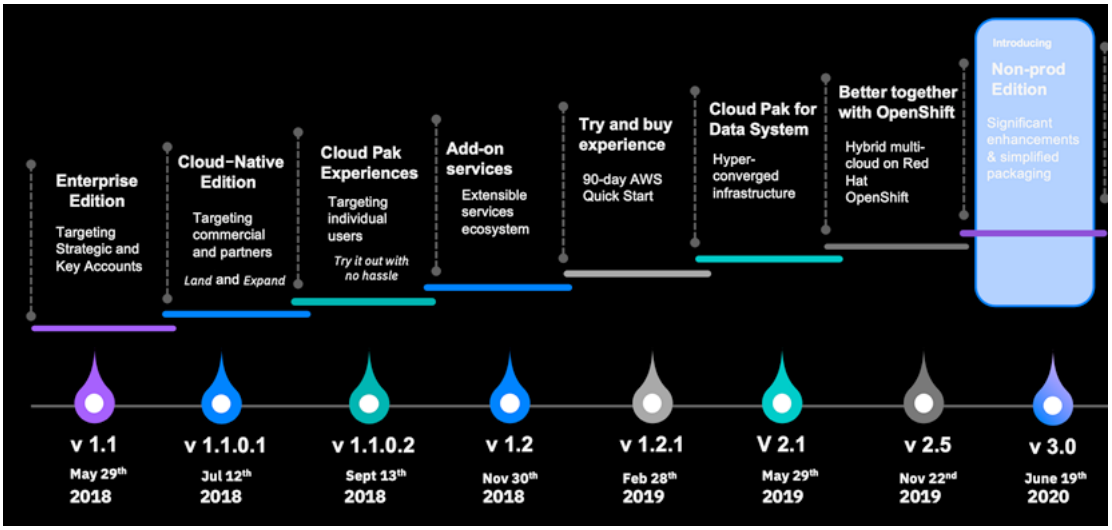
AI life cycle
Business Constraints and Expected Results (Problem Definition)
Data Acquisition
Data Labeling
Data Preparation
Model Building
Verify Business Outcomes Business Input for Model Improvement
AI Life Cycle Lineage
Model Monitoring
Model Deployment
Model Validation

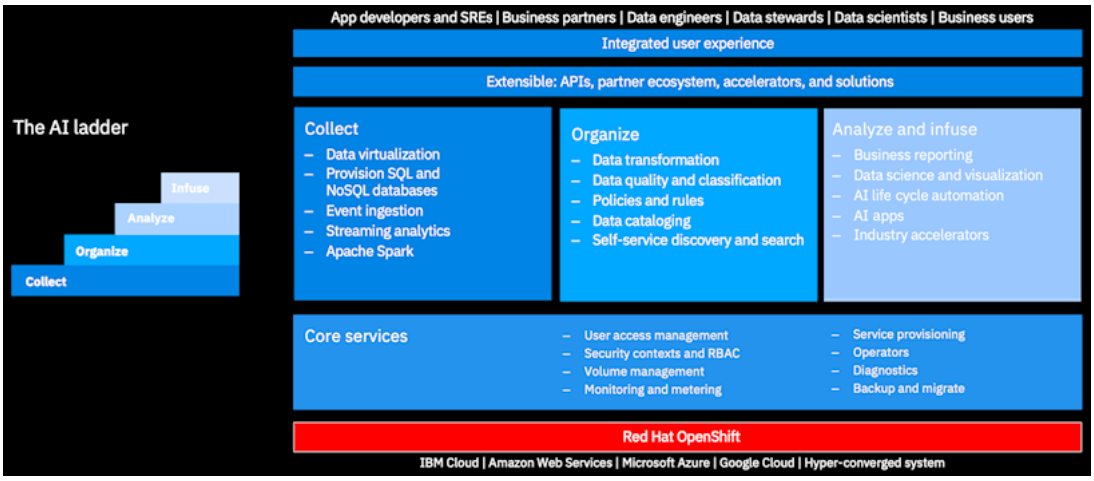

The AI Ladder
Customer Service
CCO
Risk & Compliance
CRO
Business Operations
COO
IT Operations
CIO
Financial Operations
CFO
Intelligent Workflows
Predictive
Automated
Agile
Trusted


Data
Modernize
Trusted AI infusion
Data producers
Systems of record, references
Data admins
Data Curation
Access & Privacy Policies
Data Stewards
Trusted Data
Systems of Insights
Data Engineers
Dev
Data Scientists
Feedback loop
Production Ops
Analytics Ops
Consumers
Developers
Business Users
Gather & Manage Data
Collect Virtualize and provision data sources.
> Manage All Your Data – regardless of where it lives
Understand & Govern Data
Organize Auto Classify, Catalog and Govern
data for both compliance and insights
Prepare Data for Analysis
Integrate, cleanse and shape your data. Create dashboards for visualizing data
Build Analytics Apps & Models
Analyze Develop Predictive ML, Deep Learning, Prescriptive and other advanced models Deliver Interactive notebooks & ML powered Apps
Deploy & Scale - Trusted AI
DevOps for analytical models & interactive notebooks and apps. Deploy and scale – for high performance real-time use cases
> CICD for analytics assets, Adaptive learning
Create AI -powered Applications
Incorporate trusted and governed models into applications, dashboards, and operational systems
> Infuse AI


"A Day in the Life"
Development & Prototyping
Production
Business Executive
• Sets business goals
• Prioritizes use cases
Understand Use Case
Find Data
Data Steward & Data Engineer
Collect & Organize:
• Data & Analytics Asset Catalog
• Lineage
• Governance of Data & Models
App Developer
• Infuses ML in apps & business processes
Production
• Secure & governed
• Monitor & instrument
• Reliable deployment - outage free upgrades
Rolling Upgrades
Explore & Understand Data
Deploy Models & Apps Monitor Model Health
Model Evaluation & threshold breech –> Re-train
Data Scientist
Analyze:
• Explores, shapes data
• Feature engineering and extraction
Collect & Prepare Data
Train Model
Extract Features
Data Engineer
Organize & Integrate:
• Builds & enables data pipelines
• Collaborates with product teams to integrate results

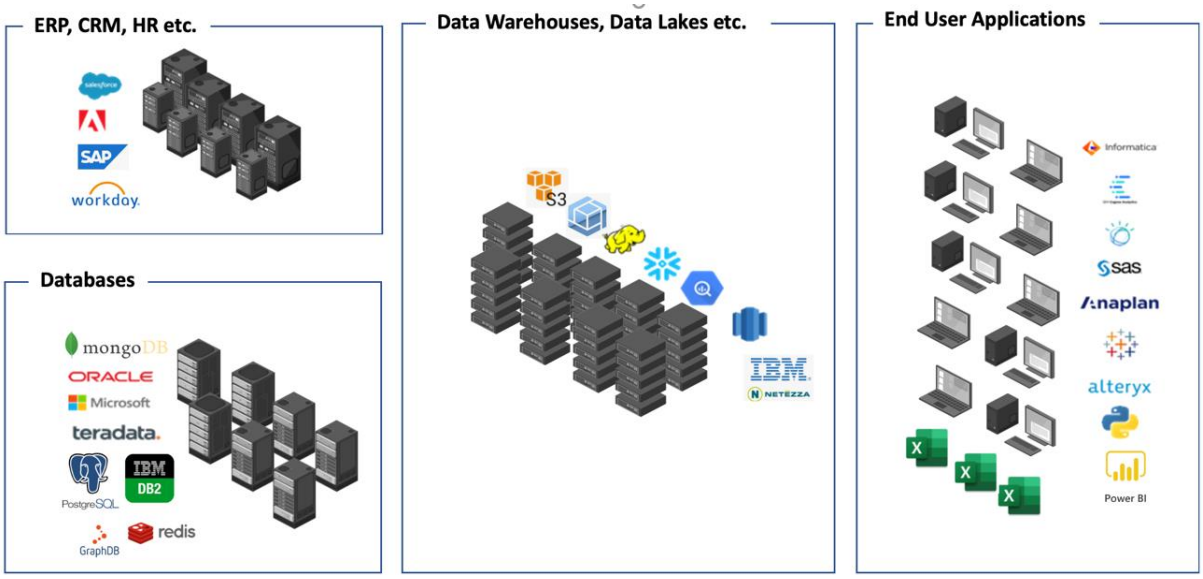
Chapter 2: Cloud Pak for Data: A Brief Introduction

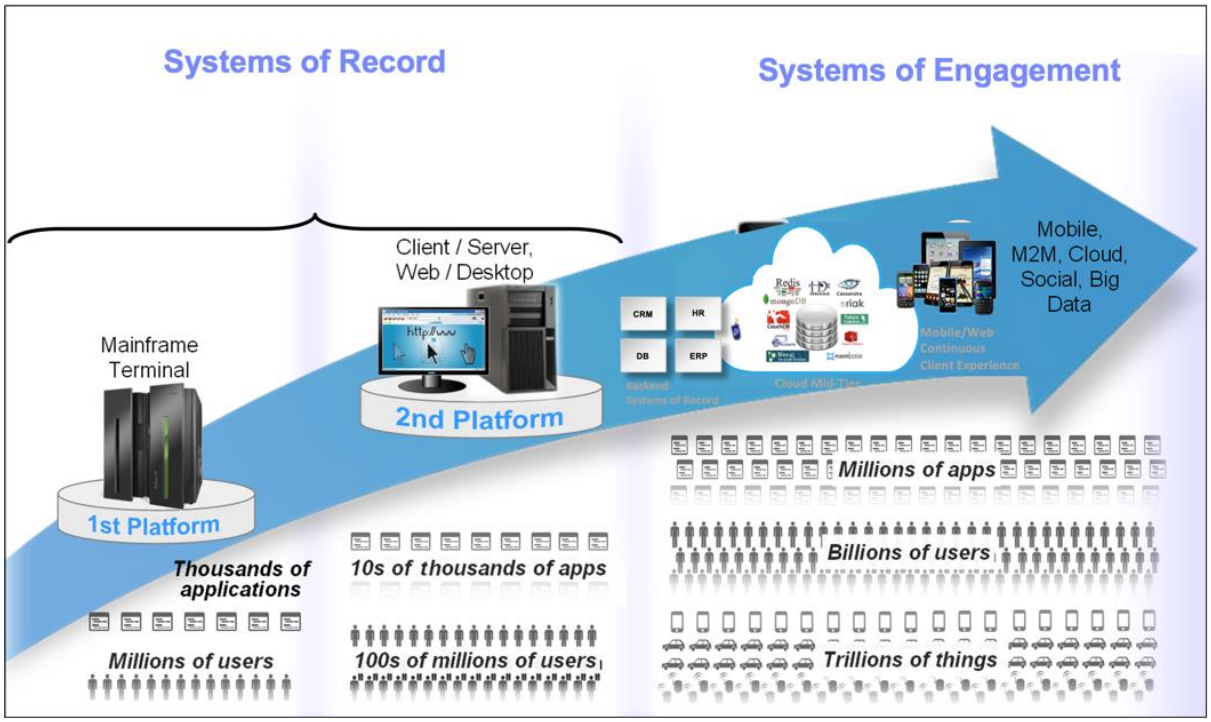


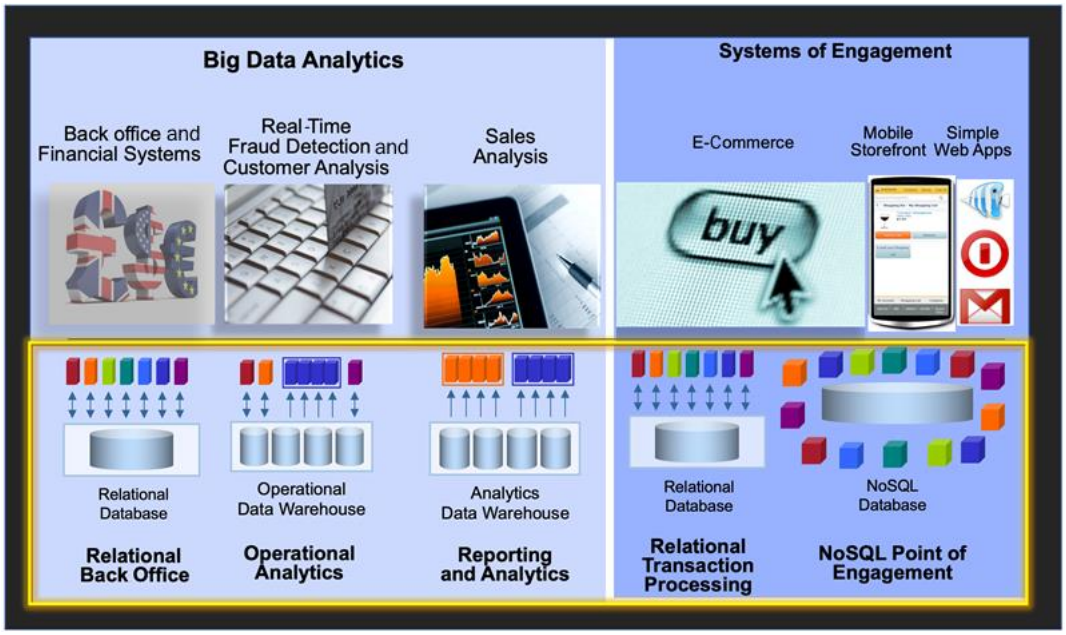


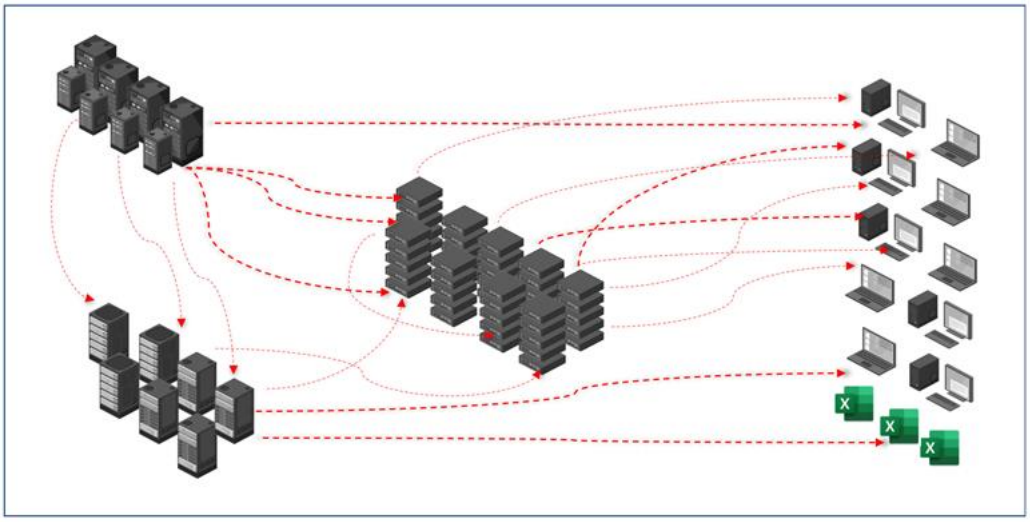
	Kubernetes	OKD	RedHat OpenShift
Platform			
Automated node configuration and cluster updates		X	X
Multi-host container scheduling	X	X	X
Self-Service provisioning	X	X	X
Service discovery	X	X	X
Enterprise operating system			X
Image Registry		X	X
Validated storage plugins		X	X
Validated networking plugins		X	X
Monitoring		X	X
Log aggregation		X	X
Multi-tenancy		X	X
Metering and chargeback			X
Developer Exterience			
Cloud service broker		X	X
Automated image builds		X	X
CI/CD and DevOps workflows		X	X
Validated third-party Kubernetes operators			X
Certified databases			X
Certified middleware			X
Certified ISV Solutions			X
Serverless applications (Knative)		X	X
Operations			
Built-on operational management			X
Zero downtime patching and upgrades			X
Enterprise 24/7 support			X
9-year support life cycle			X
Security response team			X

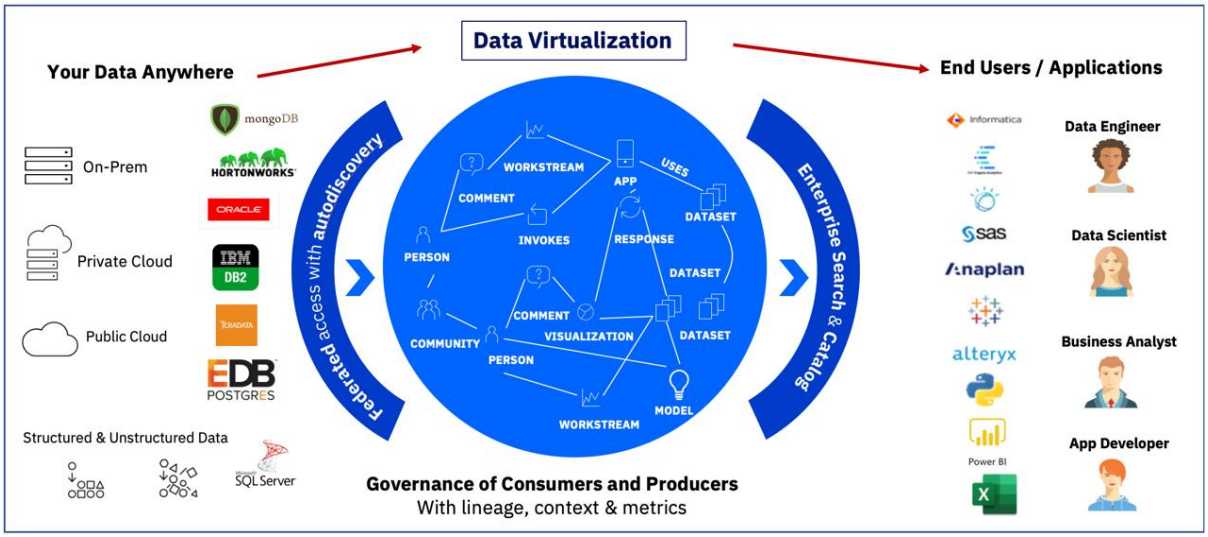
Chapter 3: Collect - Making Data Simple and Accessible

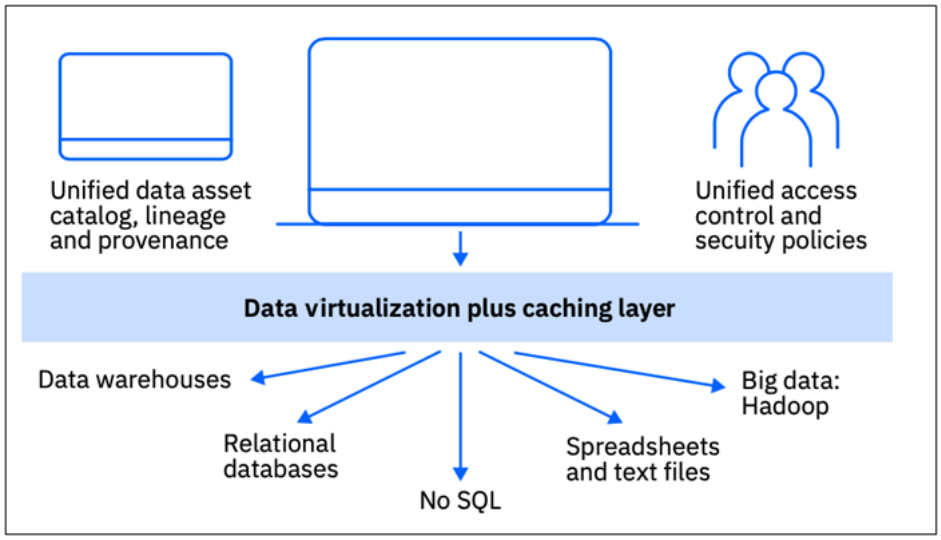


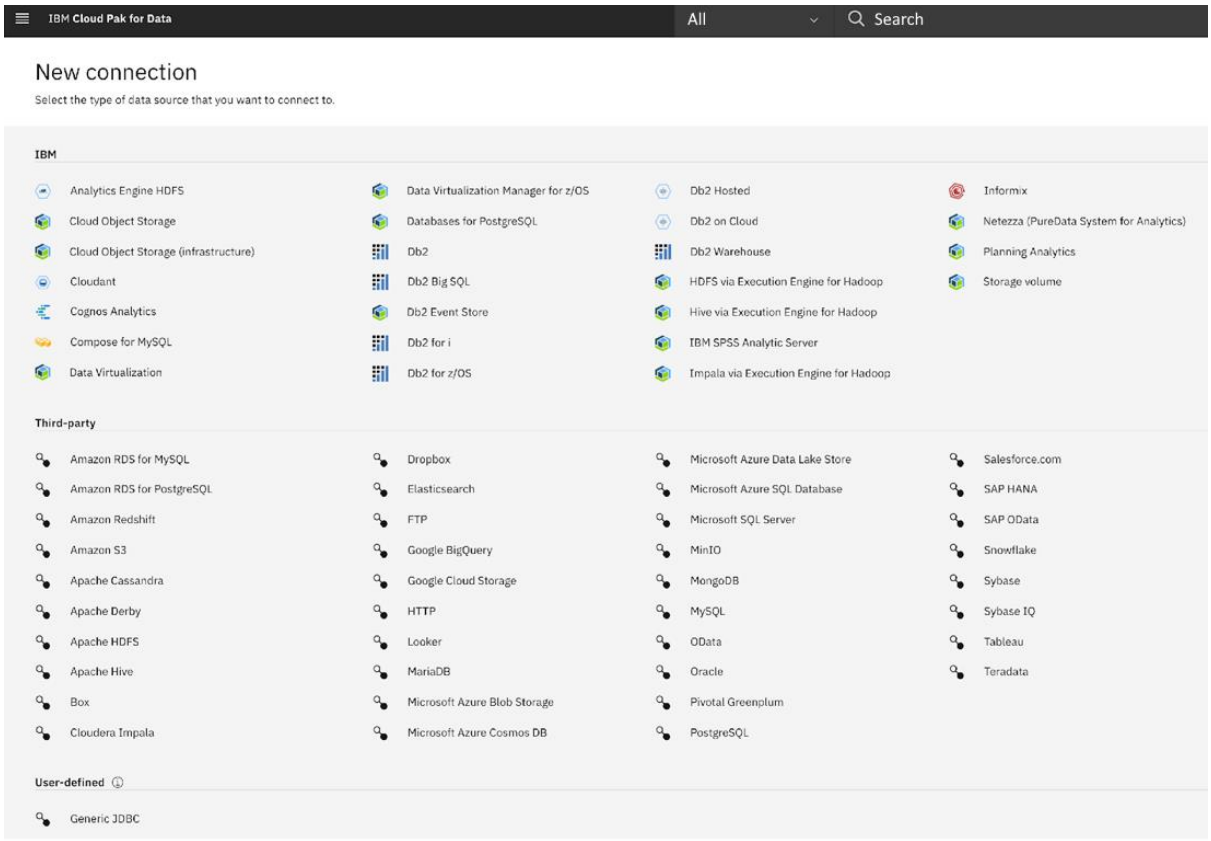



Big Data Analytics
Back office and Financial Systems
Real-Time Fraud Detection and Customer Analysis
Sales Analysis
Systems of Engagement
E-Commerce
Mobile Storefront
Simple Web Apps
buy
Relational Database
Operational Data Warehouse
Analytics Data Warehouse
Relational Database
NoSQL Database
Relational Back Office
Operational Analytics
Reporting and Analytics
Relational Transaction Processing
NoSQL Point of Engagement

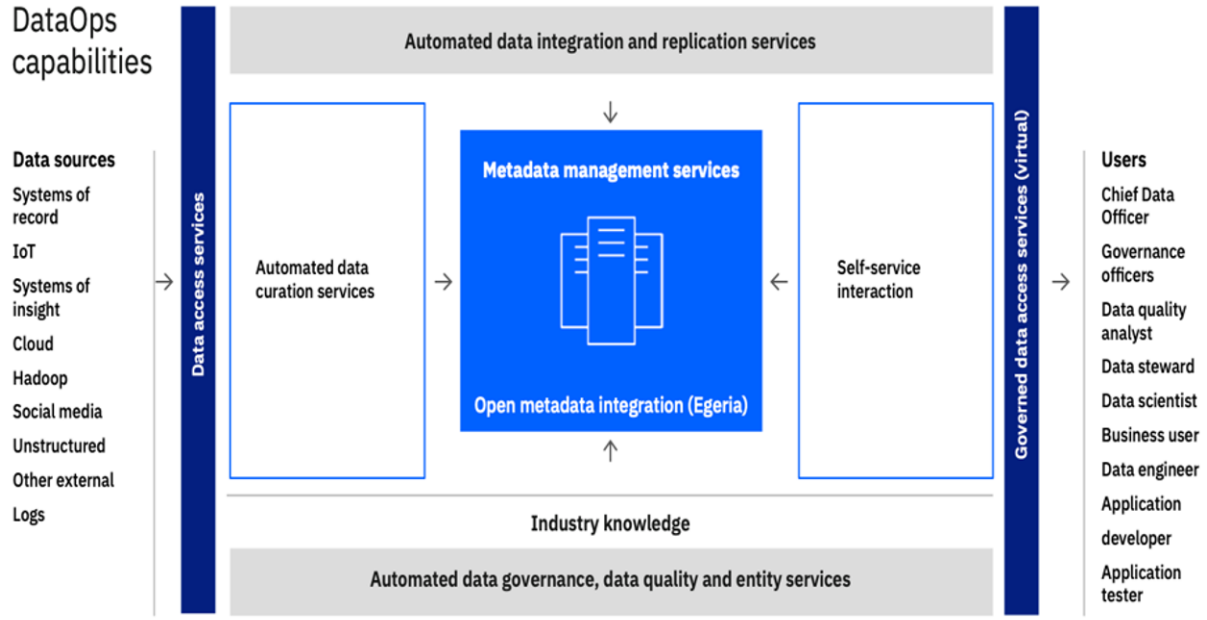



Data Virtualization
Your Data Anywhere
End Users / Applications
On-Prem
Private Cloud
Public Cloud
Structured & Unstructured Data
mongoDB
HORTONWORKS
ORACLE
IBM DB2
TERADATA
EDB POSTGRES
SQL Server
Federated access with autodiscovery
Enterprise Search & Catalog
WORKSTREAM
USES
APP
COMMENT
DATASET
INVOKES
RESPONSE
PERSON
DATASET
COMMENT
VISUALIZATION
DATASET
COMMUNITY
PERSON
MODEL
WORKSTREAM
Governance of Consumers and Producers
With lineage, context & metrics
Informatica
SAS
Anaplan
alteryx
Power BI
Data Engineer
Data Scientist
Business Analyst
App Developer

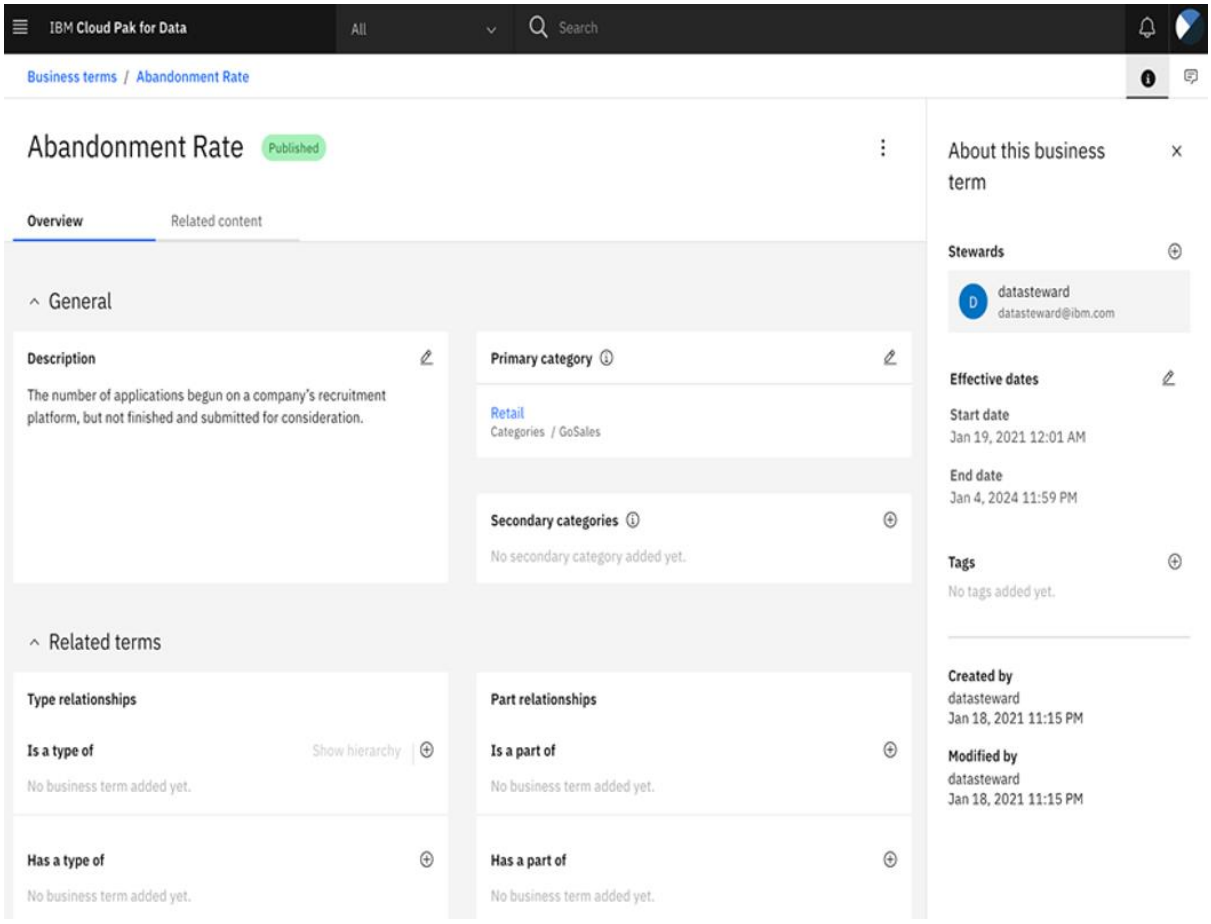

Unified data asset catalog, lineage and provenance
Unified access control and secuity policies
Data virtualization plus caching layer
Data warehouses
Relational databases
No SQL
Spreadsheets and text files
Big data: Hadoop

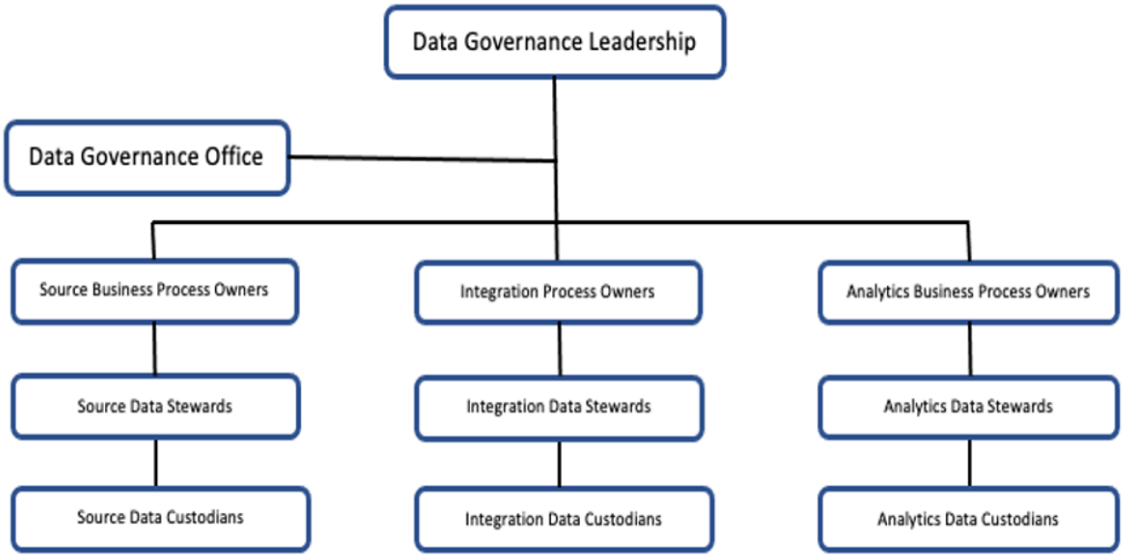

IBM Cloud Pak for Data
All
Search
New connection
Select the type of data source that you want to connect to.
IBM
Analytics Engine HDFS
Cloud Object Storage
Cloud Object Storage (infrastructure)
Cloudant
Cognos Analytics
Compose for MySQL
Data Virtualization
Data Virtualization Manager for z/OS
Databases for PostgreSQL
Db2
Db2 Big SQL
Db2 Event Store
Db2 for i
Db2 for z/OS
Db2 Hosted
Db2 on Cloud
Db2 Warehouse
HDFS via Execution Engine for Hadoop
Hive via Execution Engine for Hadoop
IBM SPSS Analytic Server
Impala via Execution Engine for Hadoop
Informix
Netezza (PureData System for Analytics)
Planning Analytics
Storage volume
Third-party
Amazon RDS for MySQL
Amazon RDS for PostgreSQL
Amazon Redshift
Amazon S3
Apache Cassandra
Apache Derby
Apache HDFS
Apache Hive
Box
Cloudera Impala
Dropbox
Elasticsearch
FTP
Google BigQuery
Google Cloud Storage
HTTP
Looker
MariaDB
Microsoft Azure Blob Storage
Microsoft Azure Cosmos DB
Microsoft Azure Data Lake Store
Microsoft Azure SQL Database
Microsoft SQL Server
MinIO
MongoDB
MySQL
OData
Oracle
Pivotal Greenplum
PostgreSQL
Salesforce.com
SAP HANA
SAP OData
Snowflake
Sybase
Sybase IQ
Tableau
Teradata
User-defined
Generic JDBC

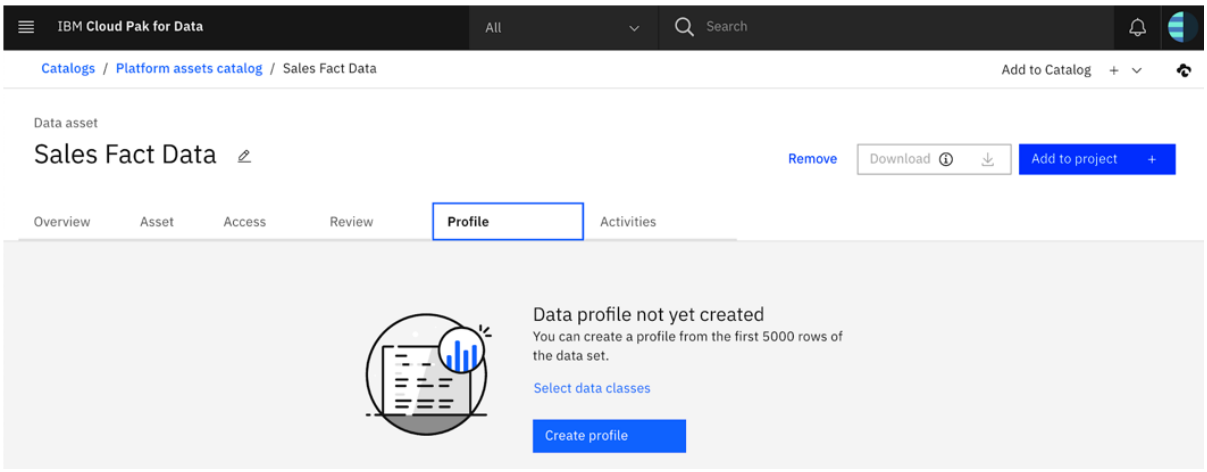
Chapter 4: Organize - Creating a Trusted Analytics Foundation

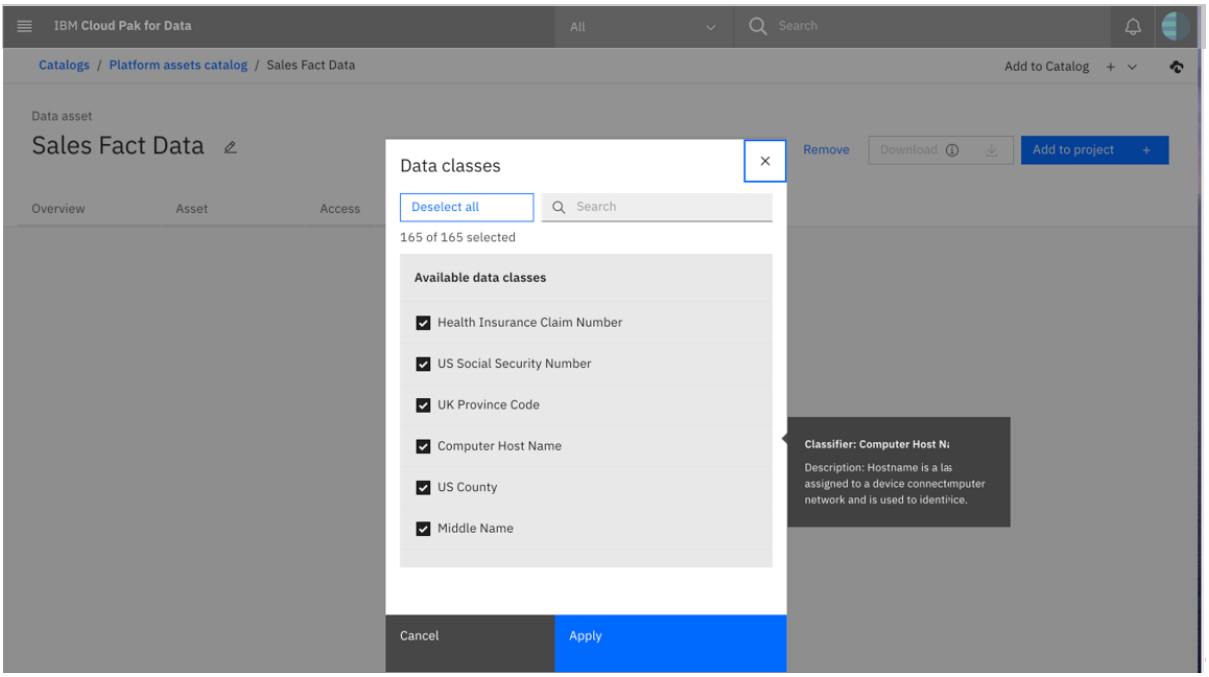


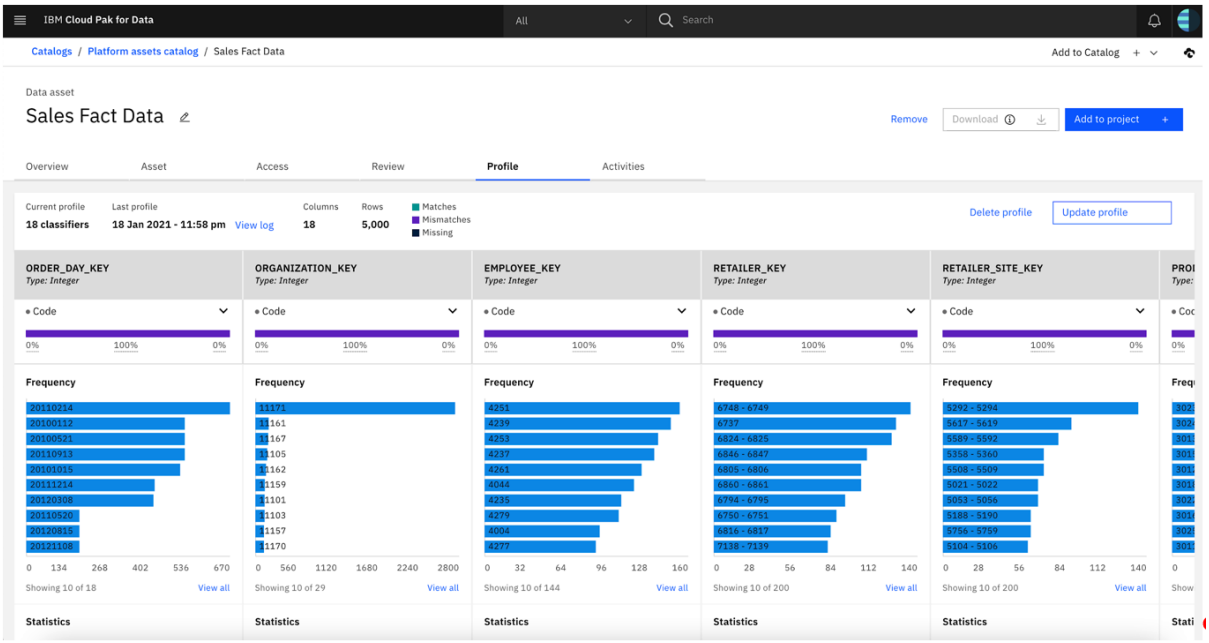
DataOps Toolchain

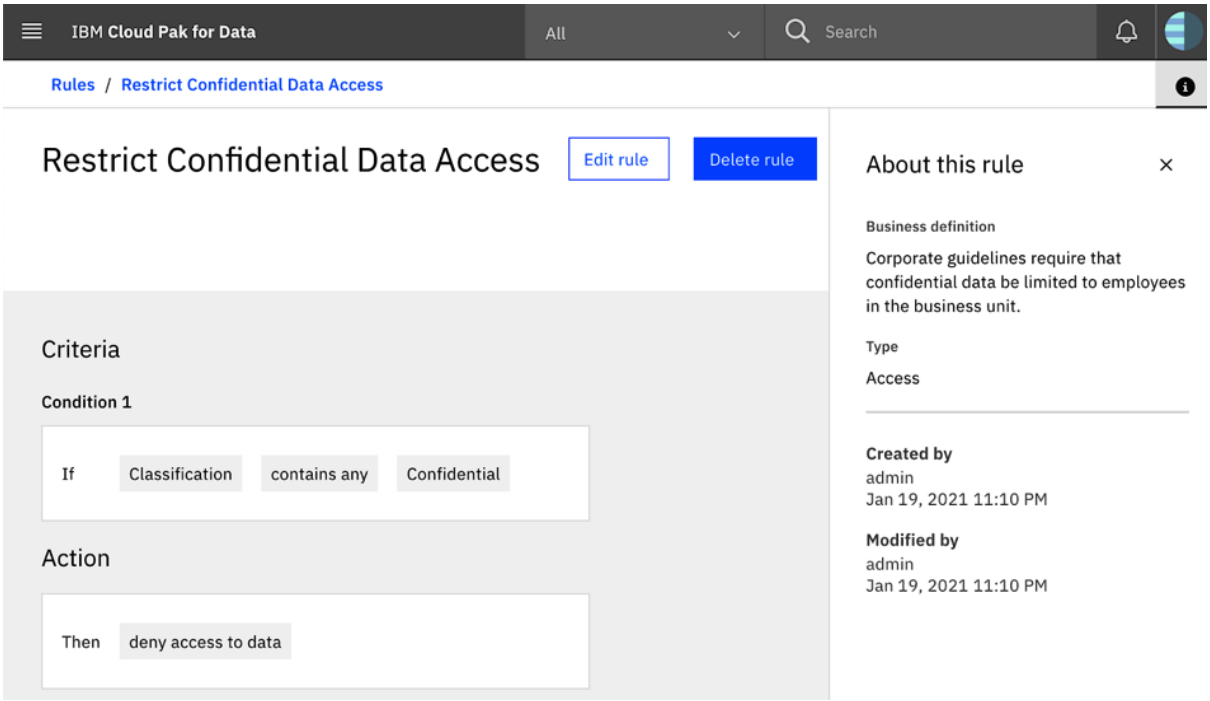


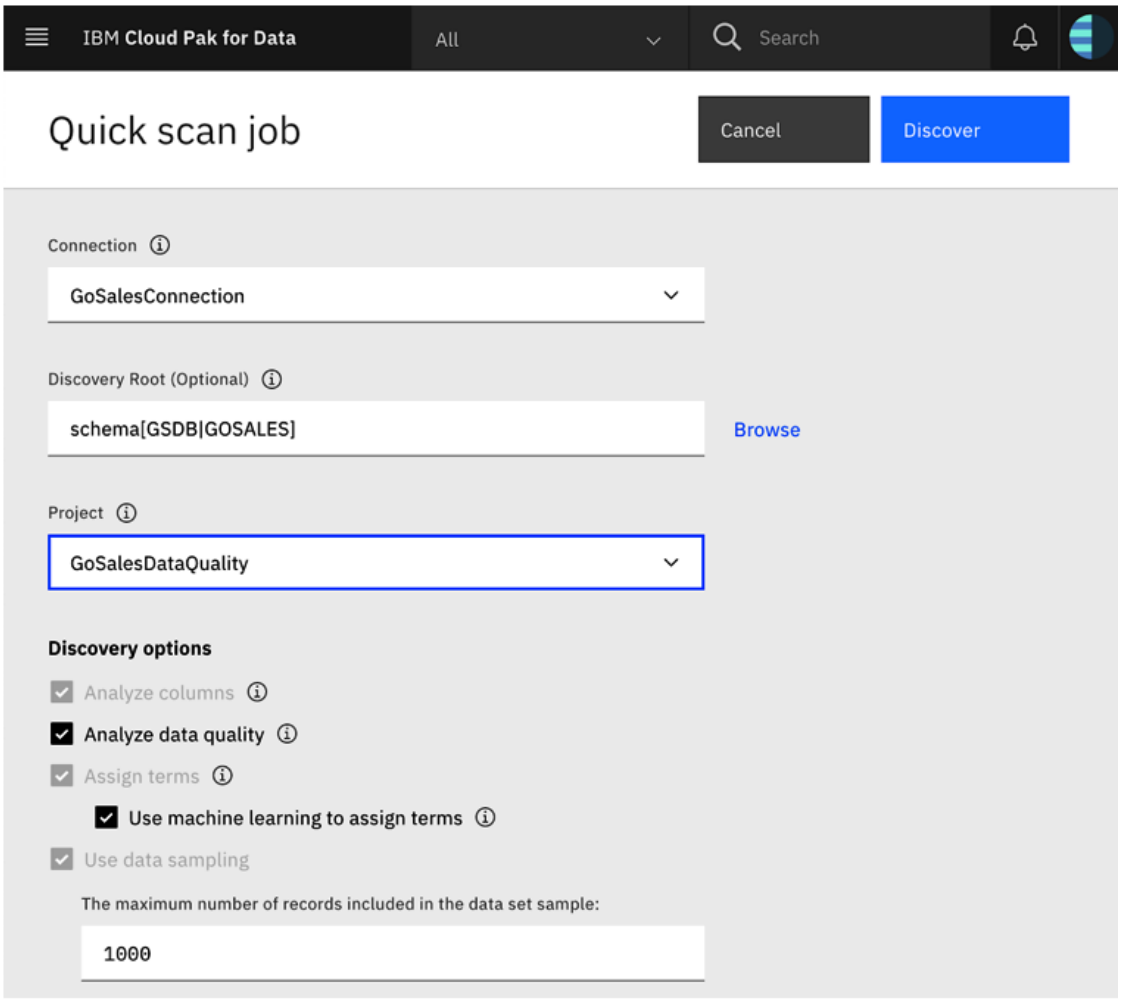

Data Governance Leadership
Data Governance Office
Source Business Process Owners
Integration Process Owners
Analytics Business Process Owners
Source Data Stewards
Integration Data Stewards
Analytics Data Stewards
Source Data Custodians
Integration Data Custodians
Analytics Data Custodians

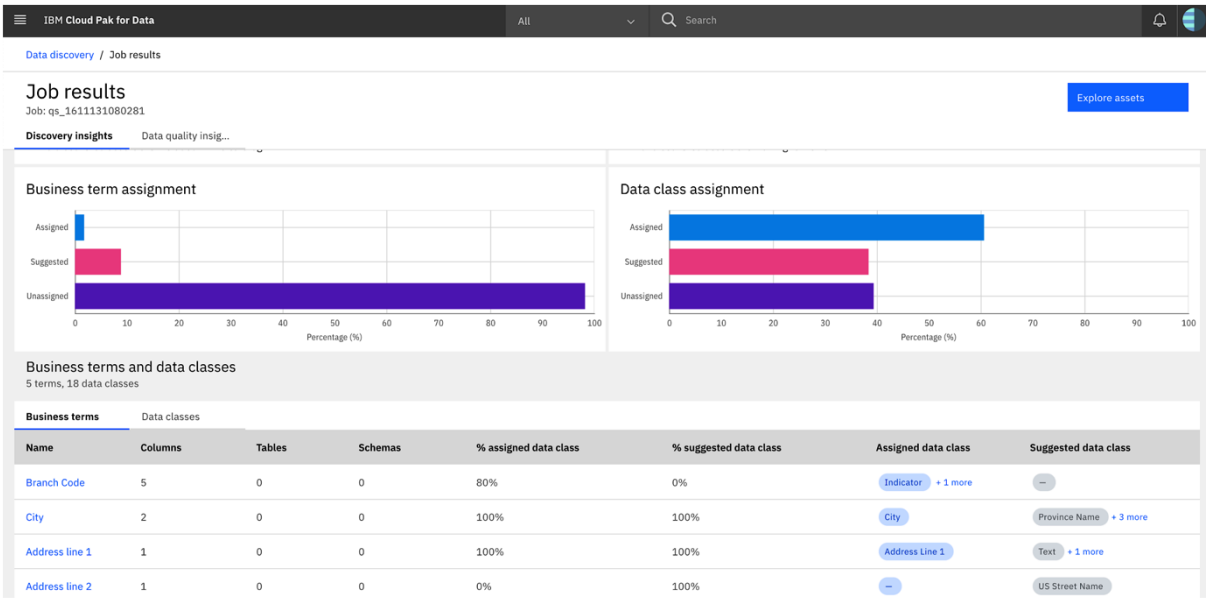

IBM Cloud Pak for Data
All
Search
Catalogs / Platform assets catalog / Sales Fact Data
Add to Catalog
Data asset
Sales Fact Data
Remove
Download
Add to project
Overview
Asset
Access
Review
Profile
Activities
Data profile not yet created
You can create a profile from the first 5000 rows of the data set.
Select data classes
Create profile

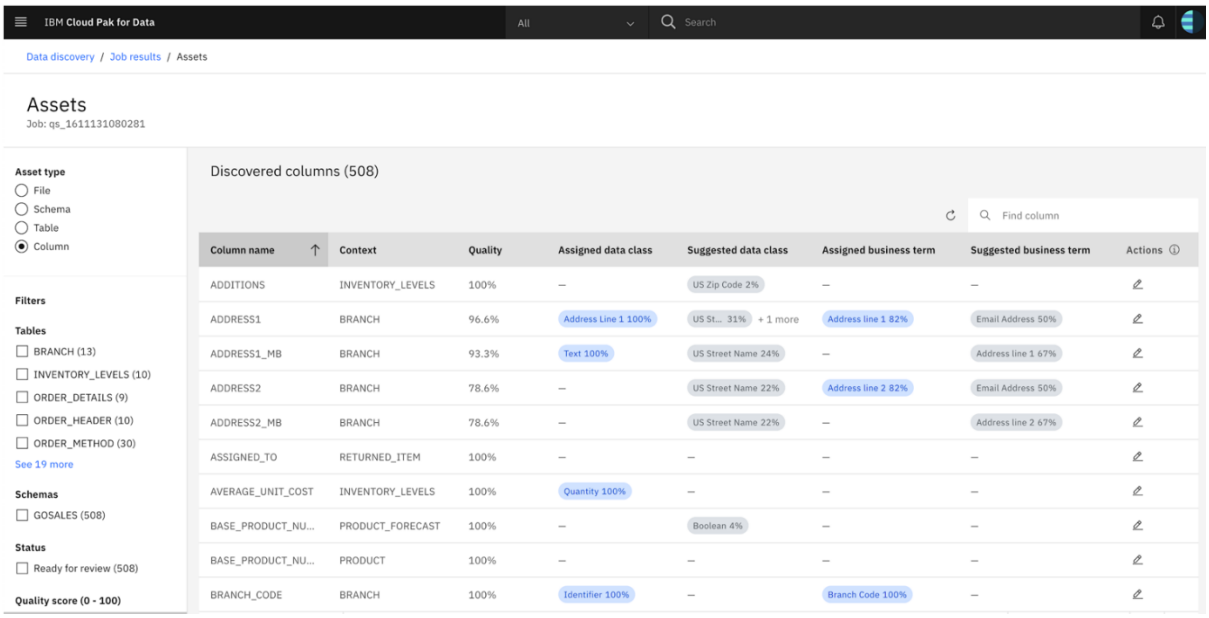

IBM Cloud Pak for Data
All
Search
Catalogs / Platform assets catalog / Sales Fact Data
Add to Catalog
Data asset
Sales Fact Data
Remove
Download
Add to project
Overview
Asset
Access
Data classes
Deselect all
Search
165 of 165 selected
Available data classes
Health Insurance Claim Number
US Social Security Number
UK Province Code
Computer Host Name
US County
Middle Name
Classifier: Computer Host N
Description: Hostname is a la
assigned to a device connect
mputer
network and is used to identi
ice.
Cancel
Apply

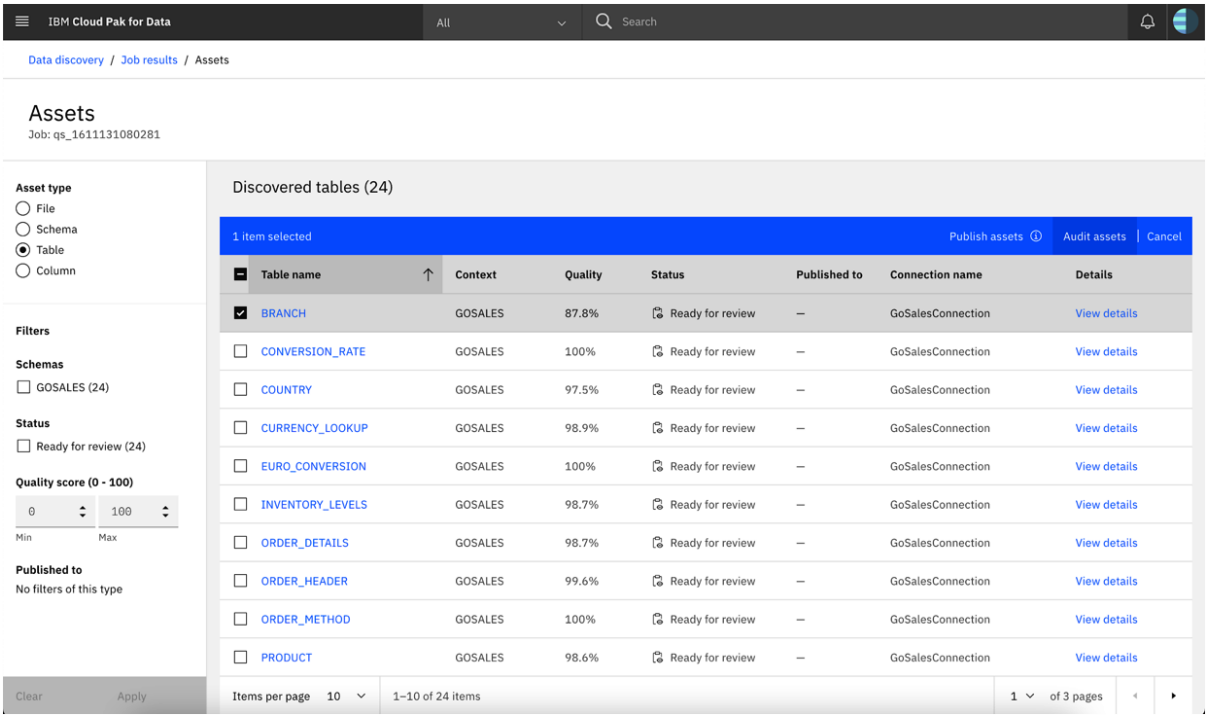

IBM Cloud Pak for Data
All
Search
Catalogs / Platform assets catalog / Sales Fact Data
Add to Catalog
Data asset
Sales Fact Data
Remove
Download
Add to project
Overview
Asset
Access
Review
Profile
Activities
Current profile
18 classifiers
Last profile
18 Jan 2021 - 11:58 pm
View log
Columns
18
Rows
5,000
Matches
Mismatches
Missing
Delete profile
Update profile
ORDER_DAY_KEY
Type: Integer
ORGANIZATION_KEY
Type: Integer
EMPLOYEE_KEY
Type: Integer
RETAILER_KEY
Type: Integer
RETAILER_SITE_KEY
Type: Integer
Code
Frequency
Showing 10 of 18
Showing 10 of 29
Showing 10 of 144
Showing 10 of 200
Showing 10 of 200
View all
Statistics

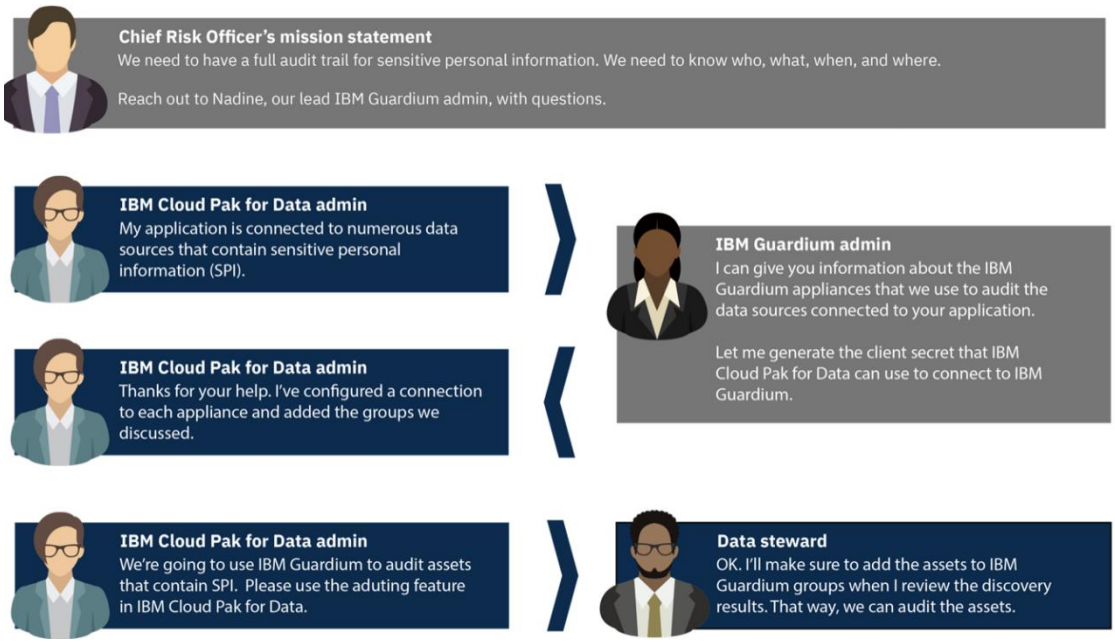

IBM Cloud Pak for Data
All
Search
Rules / Restrict Confidential Data Access
Restrict Confidential Data Access
Edit rule
Delete rule
Criteria
Condition 1
If
Classification
contains any
Confidential
Action
Then
deny access to data
About this rule
Business definition
Corporate guidelines require that confidential data be limited to employees in the business unit.
Type
Access
Created by
admin
Jan 19, 2021 11:10 PM
Modified by
admin
Jan 19, 2021 11:10 PM

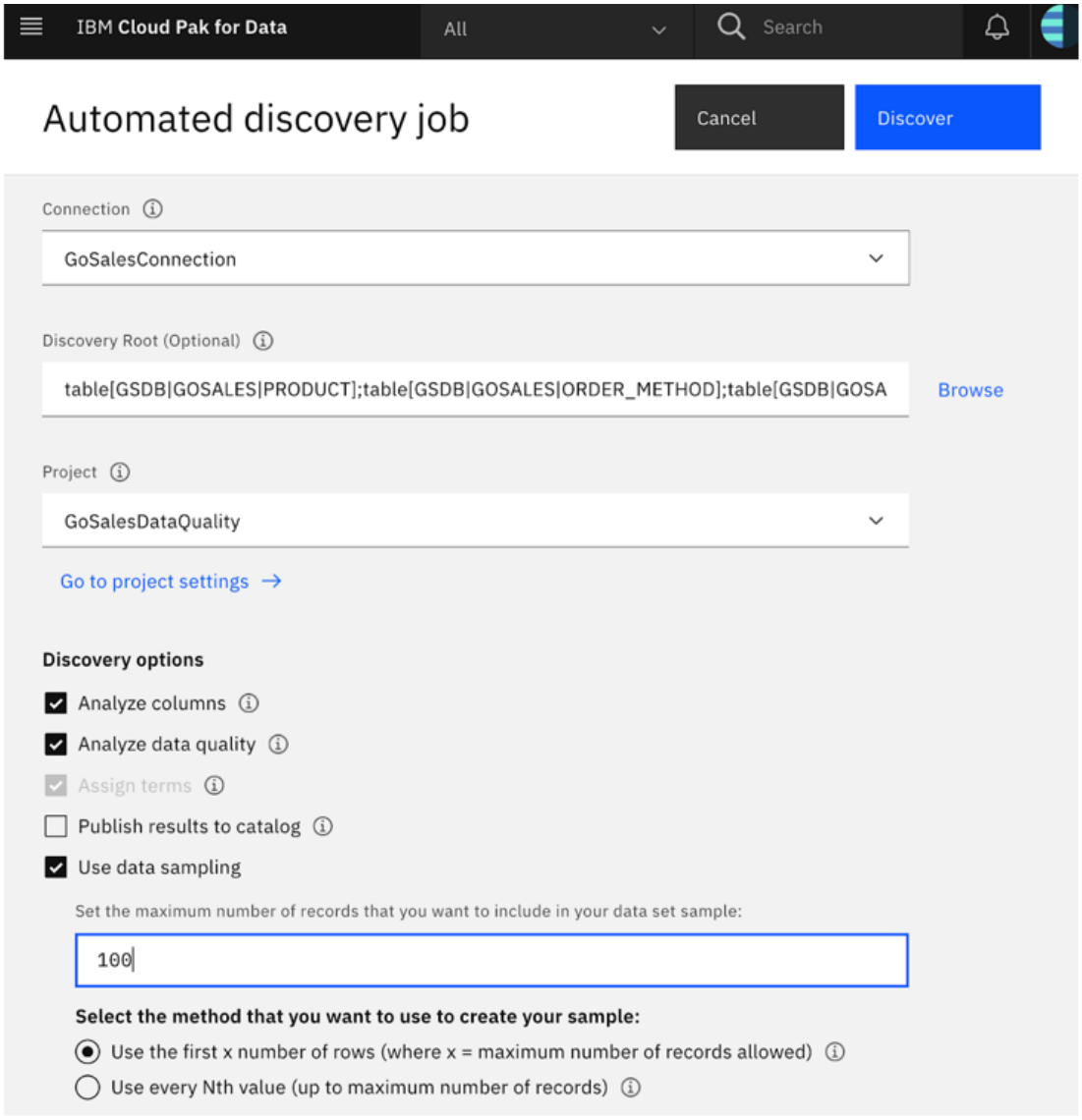

IBM Cloud Pak for Data
All
Search
Quick scan job
Cancel
Discover
Connection
GoSalesConnection
Discovery Root (Optional)
schema[GSDB|GOSALES]
Browse
Project
GoSalesDataQuality
Discovery options
Analyze columns
Analyze data quality
Assign terms
Use machine learning to assign terms
Use data sampling
The maximum number of records included in the data set sample:
1000

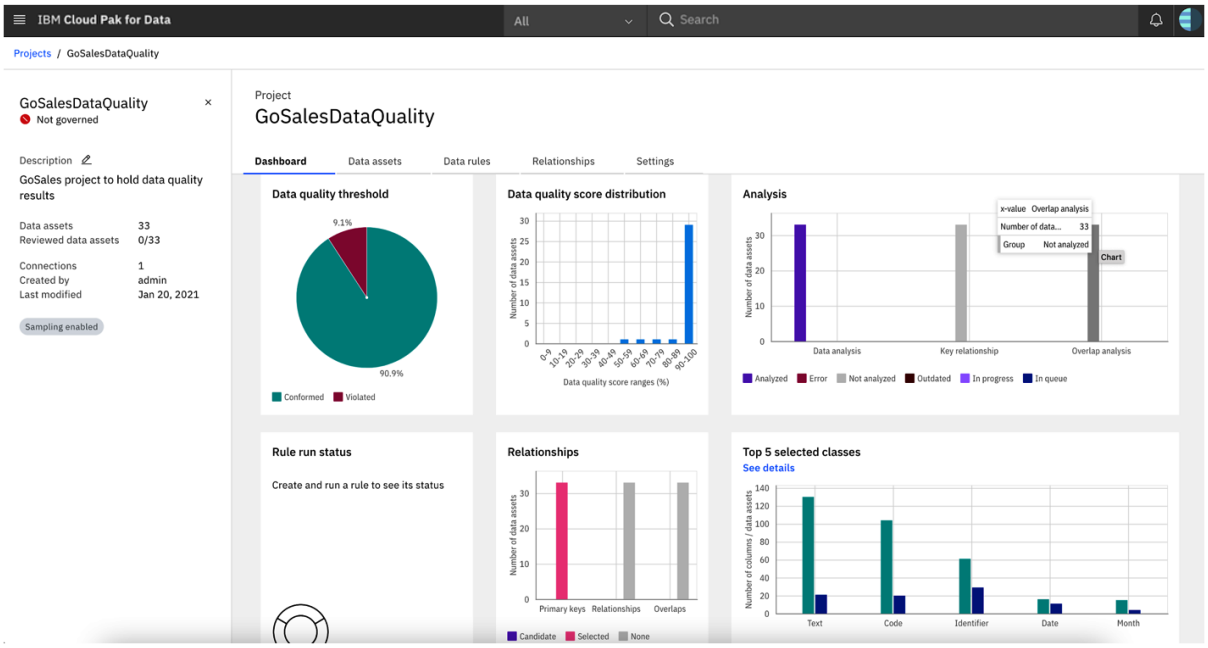

IBM Cloud Pak for Data
All
Search
Data discovery / Job results
Job results
Job: qs_1611131080281
Explore assets
Discovery insights
Data quality insig...
Business term assignment
Assigned
Suggested
Unassigned
0 10 20 30 40 50 60 70 80 90 100
Percentage (%)
Data class assignment
Assigned
Suggested
Unassigned
0 10 20 30 40 50 60 70 80 90 100
Percentage (%)
Business terms and data classes
5 terms, 18 data classes
Business terms
Data classes
Name
Columns
Tables
Schemas
% assigned data class
% suggested data class
Assigned data class
Suggested data class
Branch Code 5 0 0 80% 0% Indicator + 1 more –
City 2 0 0 100% 100% City Province Name + 3 more
Address line 1 1 0 0 100% 100% Address Line 1 Text + 1 more
Address line 2 1 0 0 0% 100% – US Street Name

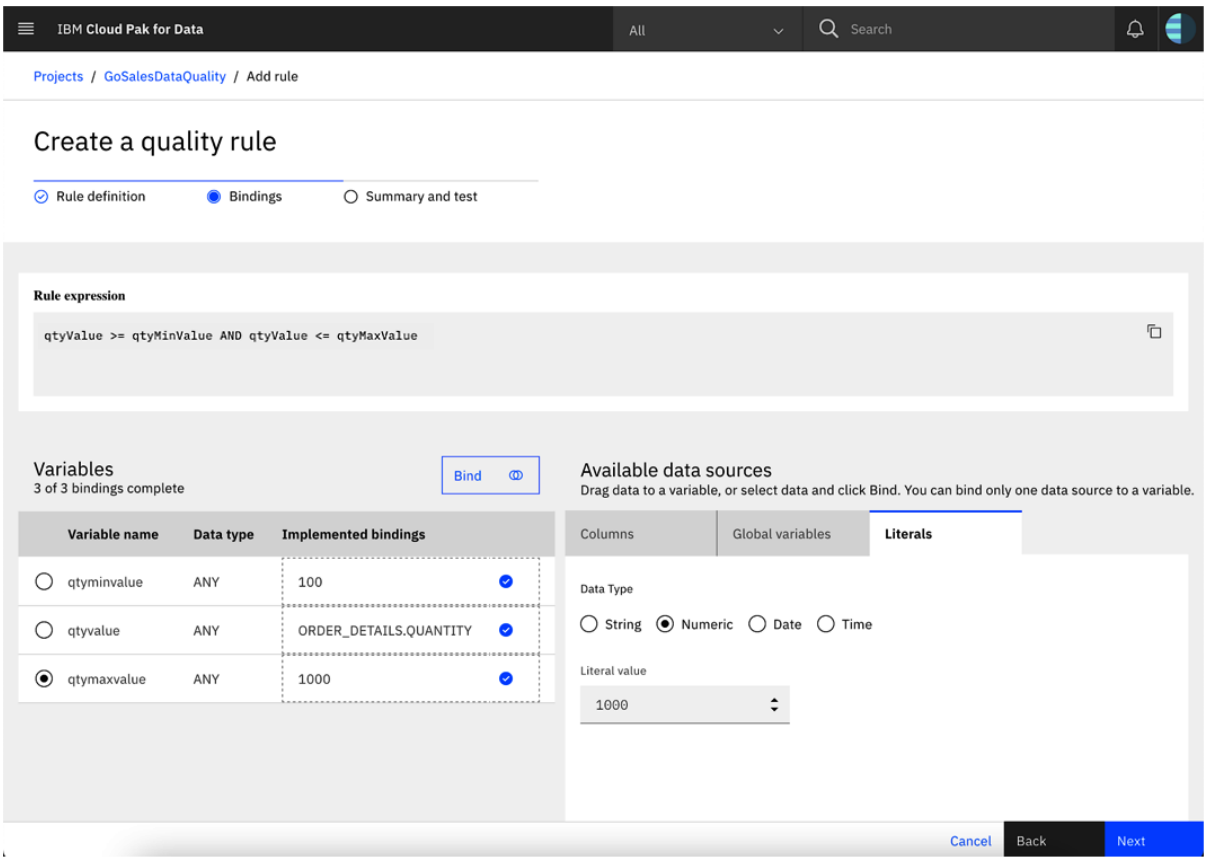

IBM Cloud Pak for Data
All
Search
Data discovery / Job results / Assets
Assets
Job: qs_1611131080281
Asset type
File
Schema
Table
Column
Filters
Tables
BRANCH (13)
INVENTORY_LEVELS (10)
ORDER_DETAILS (9)
ORDER_HEADER (10)
ORDER_METHOD (30)
See 19 more
Schemas
GOSALES (508)
Status
Ready for review (508)
Quality score (0 - 100)
Discovered columns (508)
Find column
Column name
Context
Quality
Assigned data class
Suggested data class
Assigned business term
Suggested business term
Actions
ADDITIONS INVENTORY_LEVELS 100% – US Zip Code 2% – –
ADDRESS1 BRANCH 96.6% Address Line 1 100% US St... 31% + 1 more Address line 1 82% Email Address 50%
ADDRESS1_MB BRANCH 93.3% Text 100% US Street Name 24% – Address line 1 67%
ADDRESS2 BRANCH 78.6% – US Street Name 22% Address line 2 82% Email Address 50%
ADDRESS2_MB BRANCH 78.6% – US Street Name 22% – Address line 2 67%
ASSIGNED_TO RETURNED_ITEM 100% – – – –
AVERAGE_UNIT_COST INVENTORY_LEVELS 100% Quantity 100% – – –
BASE_PRODUCT_NU... PRODUCT_FORECAST 100% – Boolean 4% – –
BASE_PRODUCT_NU... PRODUCT 100% – – – –
BRANCH_CODE BRANCH 100% Identifier 100% – Branch Code 100% –

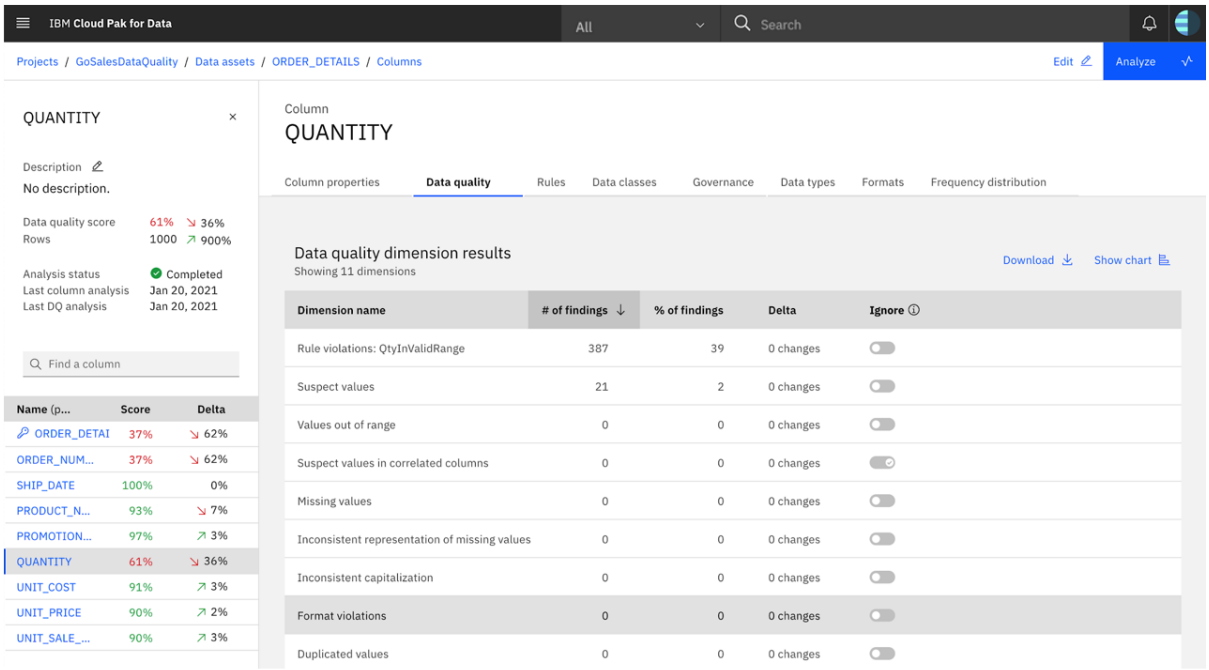

IBM Cloud Pak for Data
All
Search
Data discovery / Job results / Assets
Assets
Job: qs_1611131080281
Asset type
File
Schema
Table
Column
Filters
Schemas
GOSALES (24)
Status
Ready for review (24)
Quality score (0 - 100)
0
100
Min
Max
Published to
No filters of this type
Clear
Apply
Discovered tables (24)
1 item selected
Publish assets
Audit assets
Cancel
Table name
Context
Quality
Status
Published to
Connection name
Details
BRANCH GOSALES 87.8% Ready for review – GoSalesConnection View details
CONVERSION_RATE GOSALES 100% Ready for review – GoSalesConnection View details
COUNTRY GOSALES 97.5% Ready for review – GoSalesConnection View details
CURRENCY_LOOKUP GOSALES 98.9% Ready for review – GoSalesConnection View details
EURO_CONVERSION GOSALES 100% Ready for review – GoSalesConnection View details
INVENTORY_LEVELS GOSALES 98.7% Ready for review – GoSalesConnection View details
ORDER_DETAILS GOSALES 98.7% Ready for review – GoSalesConnection View details
ORDER_HEADER GOSALES 99.6% Ready for review – GoSalesConnection View details
ORDER_METHOD GOSALES 100% Ready for review – GoSalesConnection View details
PRODUCT GOSALES 98.6% Ready for review – GoSalesConnection View details
Items per page 10
1–10 of 24 items
1 of 3 pages

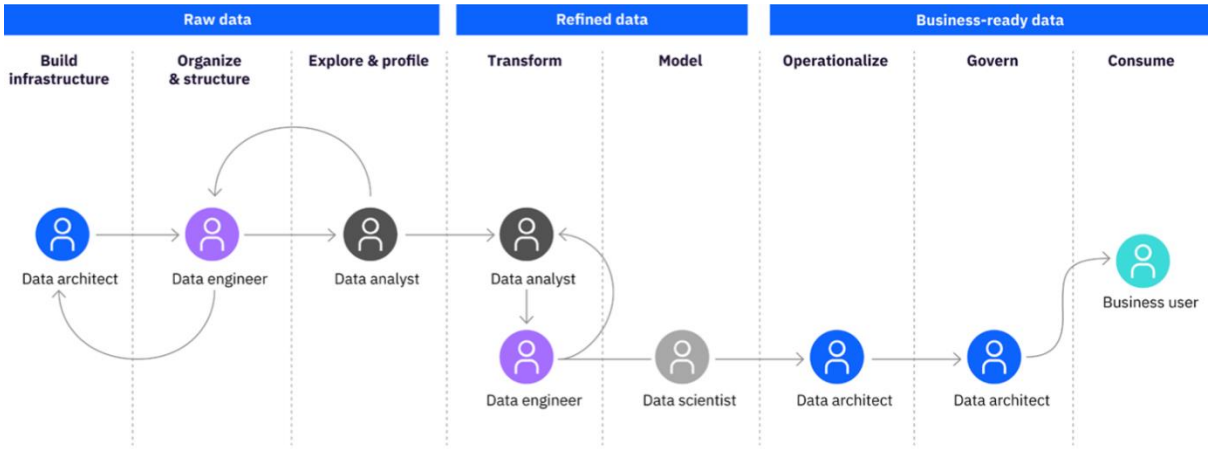

Chief Risk Officer's mission statement
We need to have a full audit trail for sensitive personal information. We need to know who, what, when, and where.
Reach out to Nadine, our lead IBM Guardium admin, with questions.
IBM Cloud Pak for Data admin
My application is connected to numerous data sources that contain sensitive personal information (SPI).
IBM Guardium admin
I can give you information about the IBM Guardium appliances that we use to audit the data sources connected to your application.
Let me generate the client secret that IBM Cloud Pak for Data can use to connect to IBM Guardium.
IBM Cloud Pak for Data admin
Thanks for your help. I've configured a connection to each appliance and added the groups we discussed.
IBM Cloud Pak for Data admin
We're going to use IBM Guardium to audit assets that contain SPI. Please use the aduting feature in IBM Cloud Pak for Data.
Data steward
OK. I'll make sure to add the assets to IBM Guardium groups when I review the discovery results. That way, we can audit the assets.

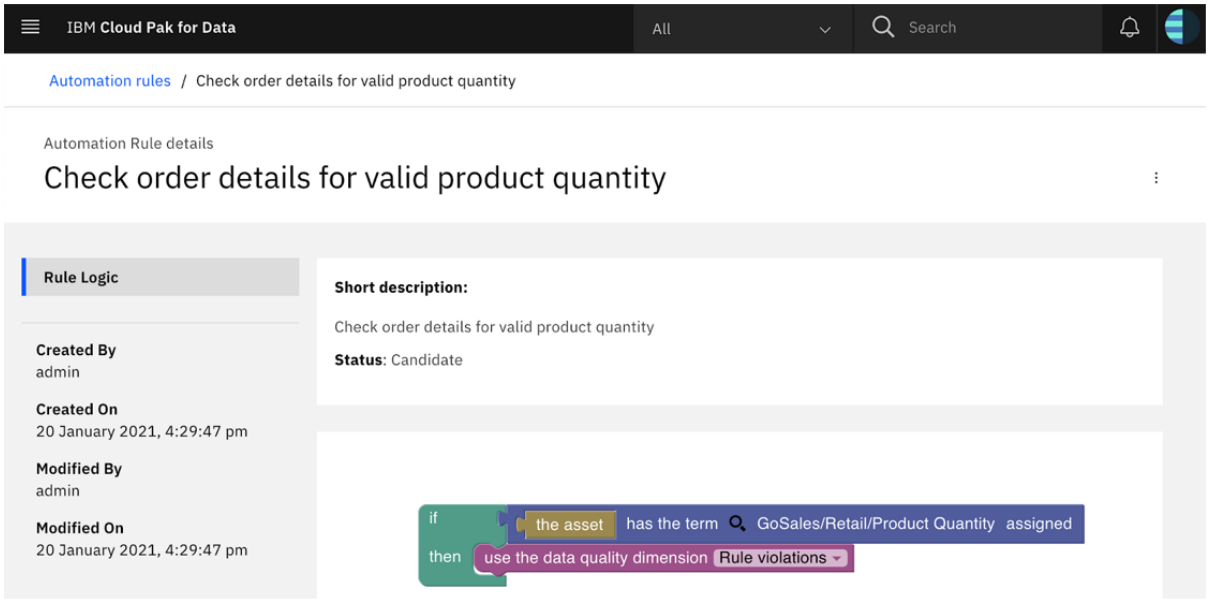

IBM Cloud Pak for Data
All
Search
Automated discovery job
Cancel
Discover
Connection
GoSalesConnection
Discovery Root (Optional)
table[GSDB|GOSALES|PRODUCT];table[GSDB|GOSALES|ORDER_METHOD];table[GSDB|GOSA
Browse
Project
GoSalesDataQuality
Go to project settings
Discovery options
Analyze columns
Analyze data quality
Assign terms
Publish results to catalog
Use data sampling
Set the maximum number of records that you want to include in your data set sample:
100
Select the method that you want to use to create your sample:
Use the first x number of rows (where x = maximum number of records allowed)
Use every Nth value (up to maximum number of records)

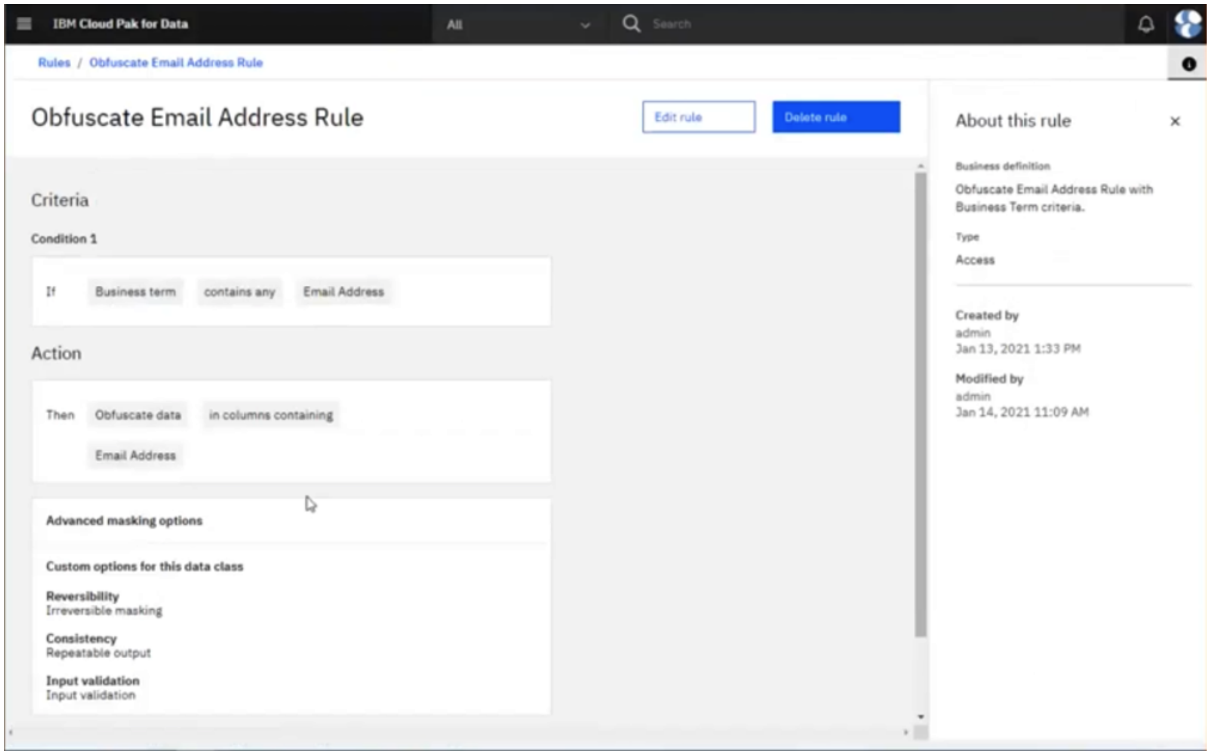

IBM Cloud Pak for Data
All
Search
Projects / GoSalesDataQuality
GoSalesDataQuality
Not governed
Description
GoSales project to hold data quality results
Data assets 33
Reviewed data assets 0/33
Connections 1
Created by admin
Last modified Jan 20, 2021
Sampling enabled
Project
GoSalesDataQuality
Dashboard
Data assets
Data rules
Relationships
Settings
Data quality threshold
9.1%
90.9%
Conformed
Violated
Data quality score distribution
Number of data assets
Data quality score ranges (%)
Analysis
x-value Overlap analysis
Number of data... 33
Group Not analyzed
Chart
Data analysis
Key relationship
Overlap analysis
Analyzed
Error
Not analyzed
Outdated
In progress
In queue
Rule run status
Create and run a rule to see its status
Relationships
Primary keys
Relationships
Overlaps
Candidate
Selected
None
Top 5 selected classes
See details
Number of columns / data assets
Text
Code
Identifier
Date
Month

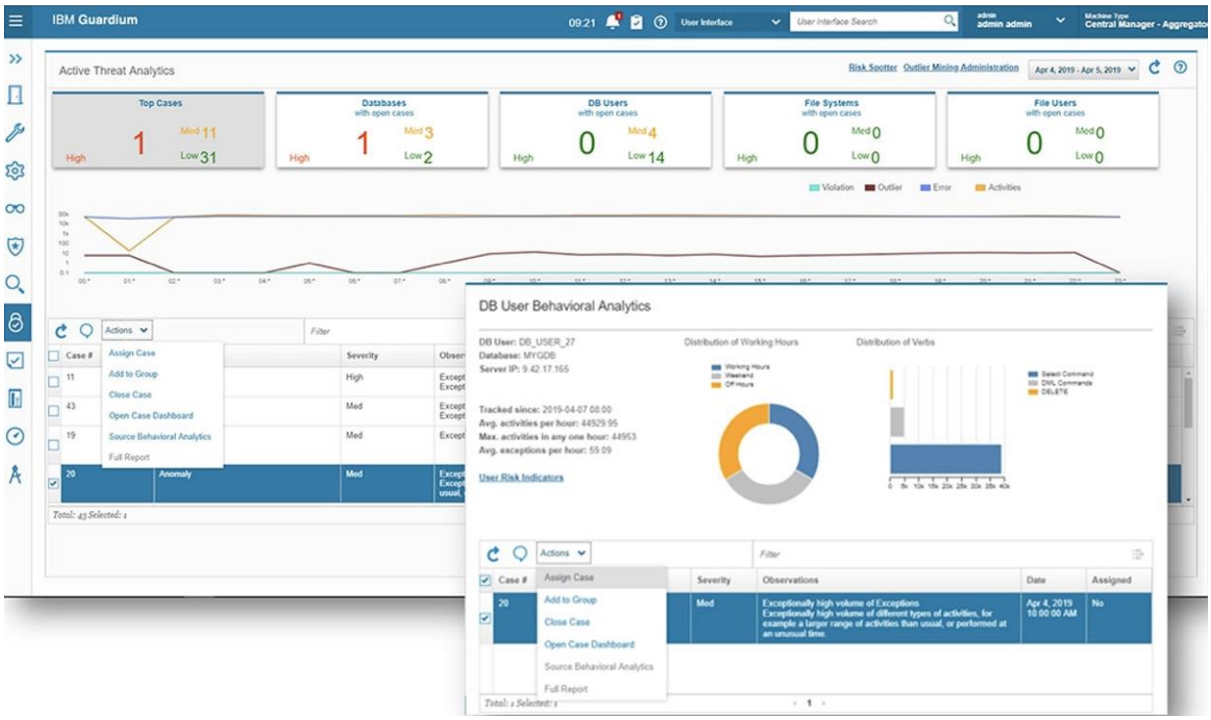

IBM Cloud Pak for Data
All
Search
Projects / GoSalesDataQuality / Add rule
Create a quality rule
Rule definition
Bindings
Summary and test
Rule expression
qtyValue >= qtyMinValue AND qtyValue <= qtyMaxValue
Variables
3 of 3 bindings complete
Bind
Variable name
Data type
Implemented bindings
qtyminvalue ANY 100
qtyvalue ANY ORDER_DETAILS.QUANTITY
qtymaxvalue ANY 1000
Available data sources
Drag data to a variable, or select data and click Bind. You can bind only one data source to a variable.
Columns
Global variables
Literals
Data Type
String
Numeric
Date
Time
Literal value
1000
Cancel
Back
Next

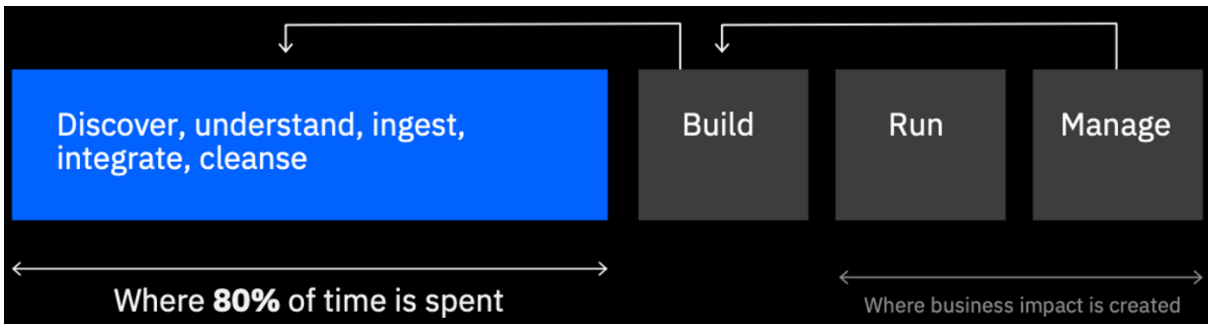

IBM Cloud Pak for Data
All
Search
Projects / GoSalesDataQuality / Data assets / ORDER_DETAILS / Columns
Edit
Analyze
QUANTITY
Description
No description.
Data quality score 61% ↘ 36%
Rows 1000 ↗ 900%
Analysis status Completed
Last column analysis Jan 20, 2021
Last DQ analysis Jan 20, 2021
Find a column
Name (p... Score Delta
ORDER_DETAI 37% ↘ 62%
ORDER_NUM... 37% ↘ 62%
SHIP_DATE 100% 0%
PRODUCT_N... 93% ↘ 7%
PROMOTION... 97% ↗ 3%
QUANTITY 61% ↘ 36%
UNIT_COST 91% ↗ 3%
UNIT_PRICE 90% ↗ 2%
UNIT_SALE_... 90% ↗ 3%
Column
QUANTITY
Column properties
Data quality
Rules
Data classes
Governance
Data types
Formats
Frequency distribution
Data quality dimension results
Showing 11 dimensions
Download
Show chart
Dimension name # of findings % of findings Delta Ignore
Rule violations: QtyInValidRange 387 39 0 changes
Suspect values 21 2 0 changes
Values out of range 0 0 0 changes
Suspect values in correlated columns 0 0 0 changes
Missing values 0 0 0 changes
Inconsistent representation of missing values 0 0 0 changes
Inconsistent capitalization 0 0 0 changes
Format violations 0 0 0 changes
Duplicated values 0 0 0 changes

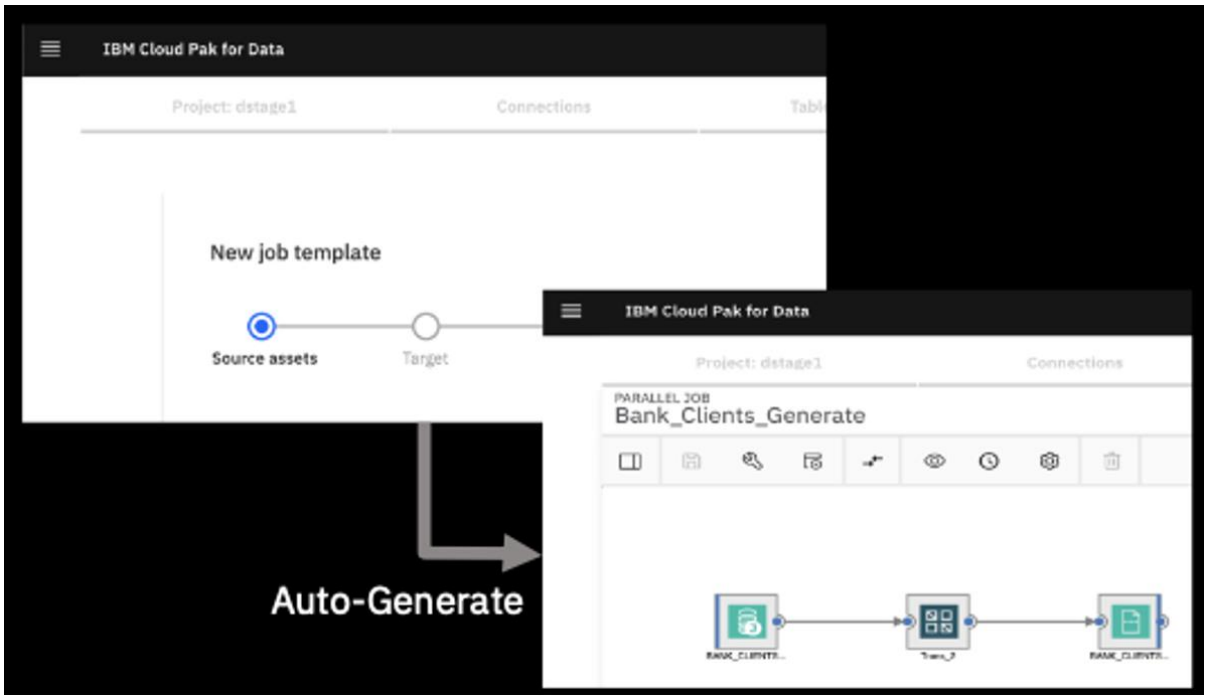

Raw data
Refined data
Business-ready data
Build infrastructure
Organize & structure
Explore & profile
Transform
Model
Operationalize
Govern
Consume
Data architect
Data engineer
Data analyst
Data analyst
Business user
Data engineer
Data scientist
Data architect
Data architect

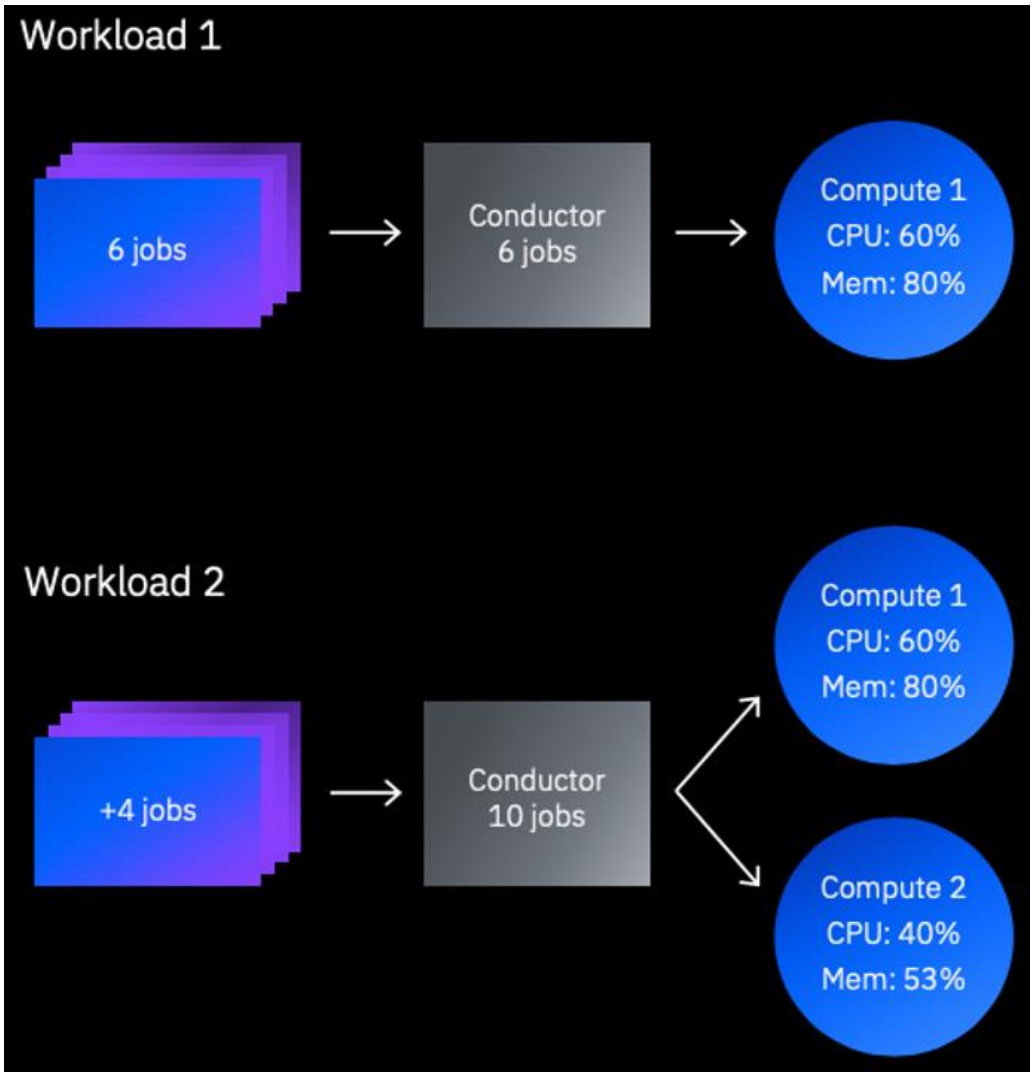

IBM Cloud Pak for Data
All
Search
Automation rules / Check order details for valid product quantity
Automation Rule details
Check order details for valid product quantity
Rule Logic
Created By
admin
Created On
20 January 2021, 4:29:47 pm
Modified By
admin
Modified On
20 January 2021, 4:29:47 pm
Short description:
Check order details for valid product quantity
Status: Candidate
if
the asset
has the term
GoSales/Retail/Product Quantity
assigned
then
use the data quality dimension
Rule violations

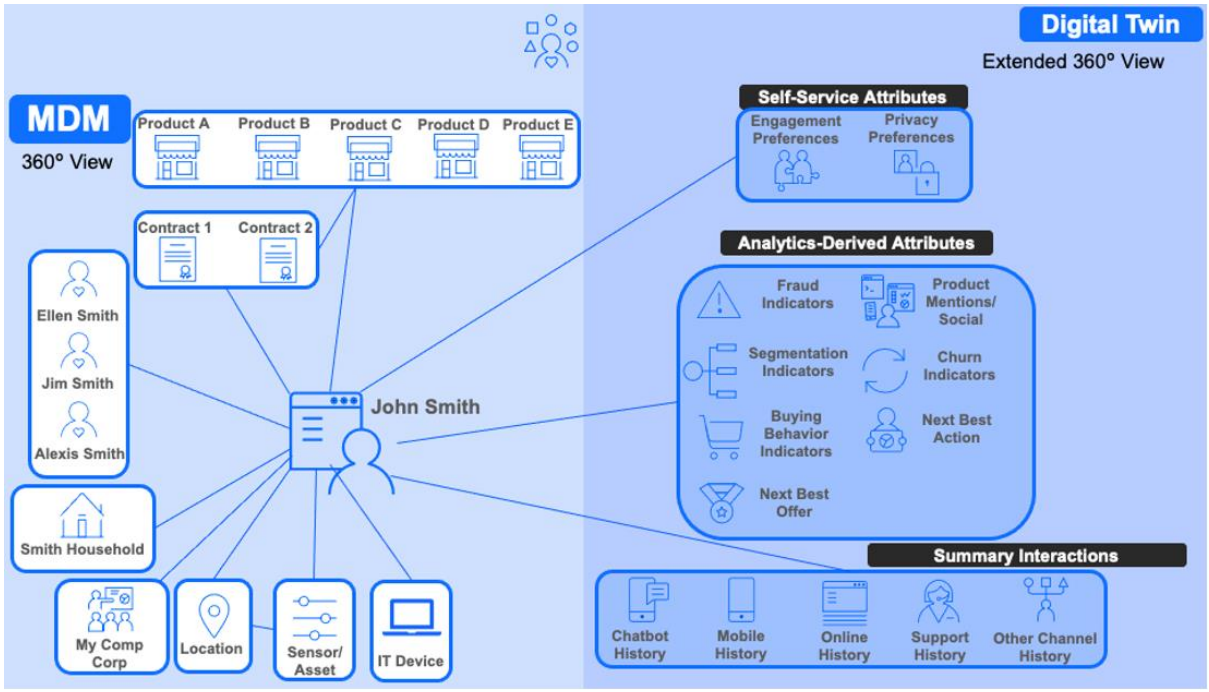

IBM Cloud Pak for Data
Rules / Obfuscate Email Address Rule
Obfuscate Email Address Rule
Edit rule
Delete rule
Criteria
Condition 1
If
Business term
contains any
Email Address
Action
Then
Obfuscate data
in columns containing
Email Address
Advanced masking options
Custom options for this data class
Reversibility
Irreversible masking
Consistency
Repeatable output
Input validation
Input validation
About this rule
Business definition
Obfuscate Email Address Rule with Business Term criteria.
Type
Access
Created by
admin
Jan 13, 2021 1:33 PM
Modified by
admin
Jan 14, 2021 11:09 AM

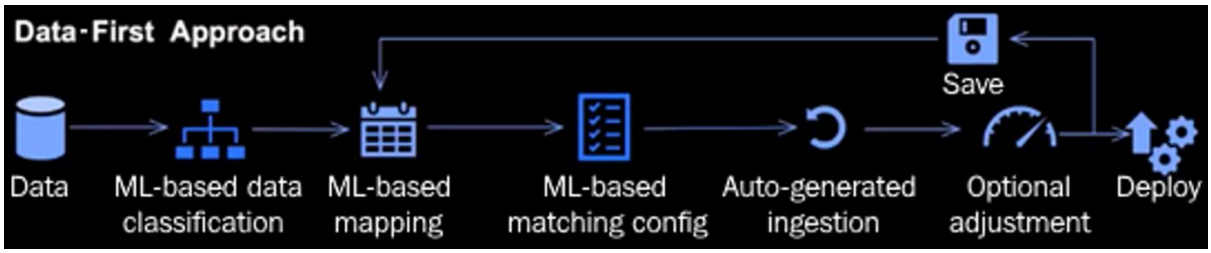

IBM Guardium
Active Threat Analytics
Top Cases
Databases
DB Users
File Systems
File Users
DB User Behavioral Analytics
Assign Case
Add to Group
Close Case
Open Case Dashboard
Source Behavioral Analytics
Full Report

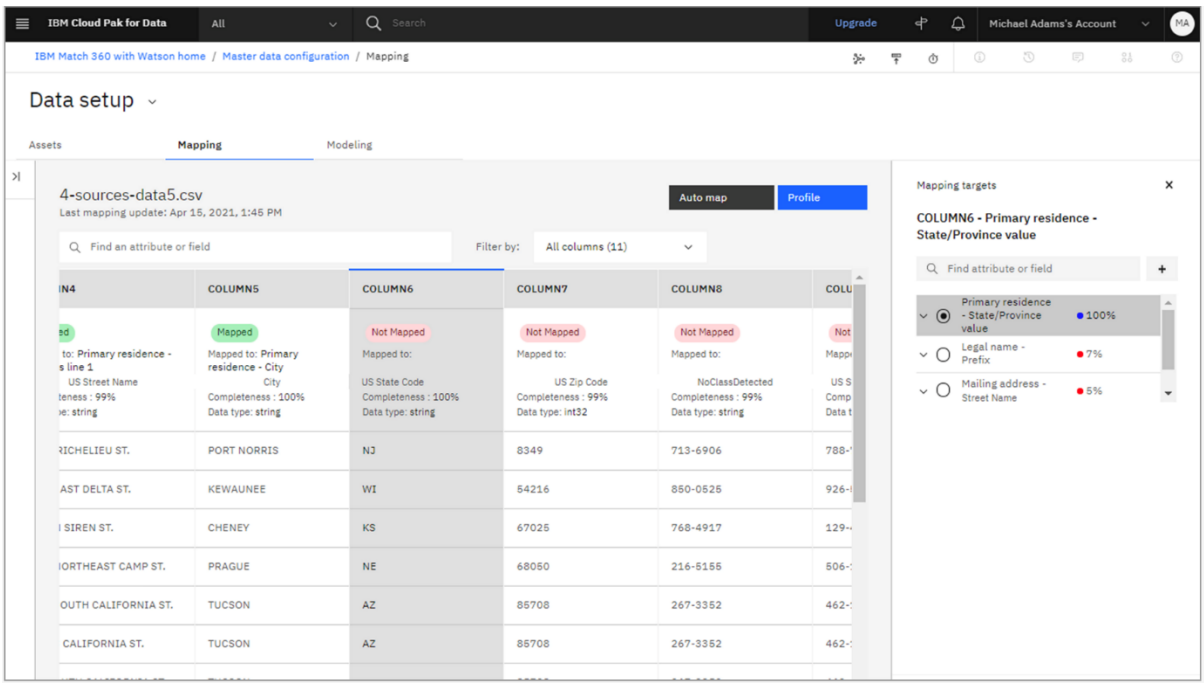

Discover, understand, ingest, integrate, cleanse
Build
Run
Manage
Where 80% of time is spent
Where business impact is created

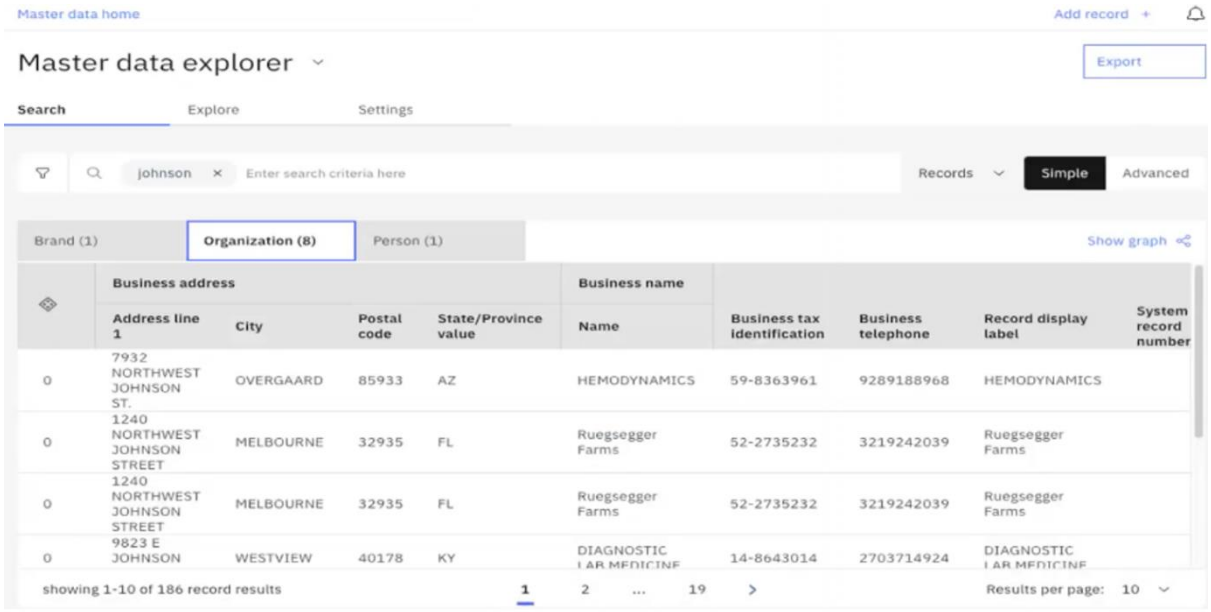

IBM Cloud Pak for Data
Project: dstage1
Connections
New job template
Source assets
Target
IBM Cloud Pak for Data
Project: dstage1
Connections
PARALLEL JOB
Bank_Clients_Generate
Auto-Generate


Workload 1
6 jobs
Conductor
6 jobs
Compute 1
CPU: 60%
Mem: 80%
Workload 2
+4 jobs
Conductor
10 jobs
Compute 1
CPU: 60%
Mem: 80%
Compute 2
CPU: 40%
Mem: 53%

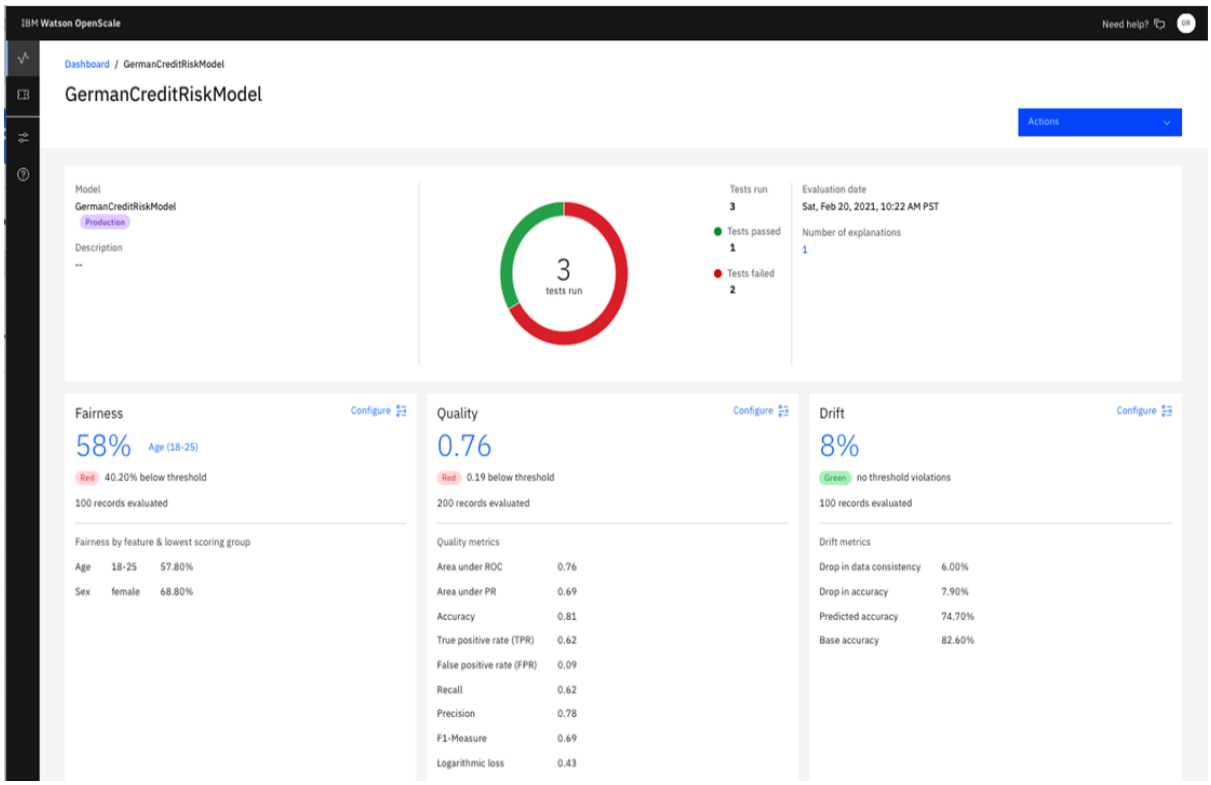

Digital Twin
Extended 360° View
MDM
360° View
Product A
Product B
Product C
Product D
Product E
Contract 1
Contract 2
Ellen Smith
Jim Smith
Alexis Smith
Smith Household
John Smith
My Comp Corp
Location
Sensor/ Asset
IT Device
Self-Service Attributes
Engagement Preferences
Privacy Preferences
Analytics-Derived Attributes
Fraud Indicators
Product Mentions/ Social
Segmentation Indicators
Churn Indicators
Buying Behavior Indicators
Next Best Action
Next Best Offer
Summary Interactions
Chatbot History
Mobile History
Online History
Support History
Other Channel History

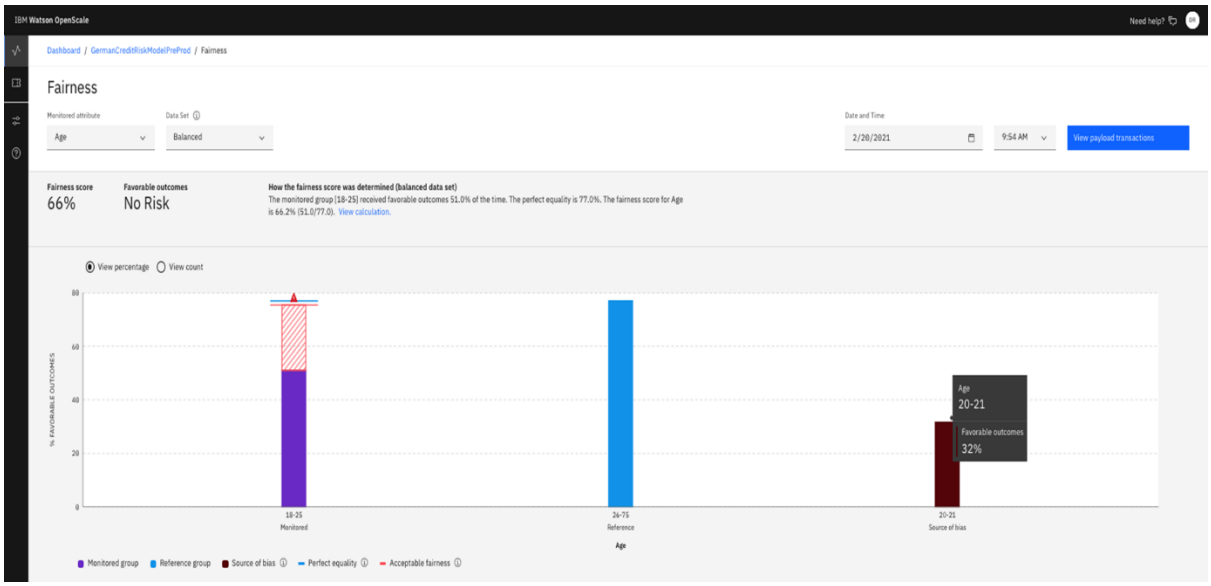

Data-First Approach
Save
Data
ML-based data classification
ML-based mapping
ML-based matching config
Auto-generated ingestion
Optional adjustment
Deploy

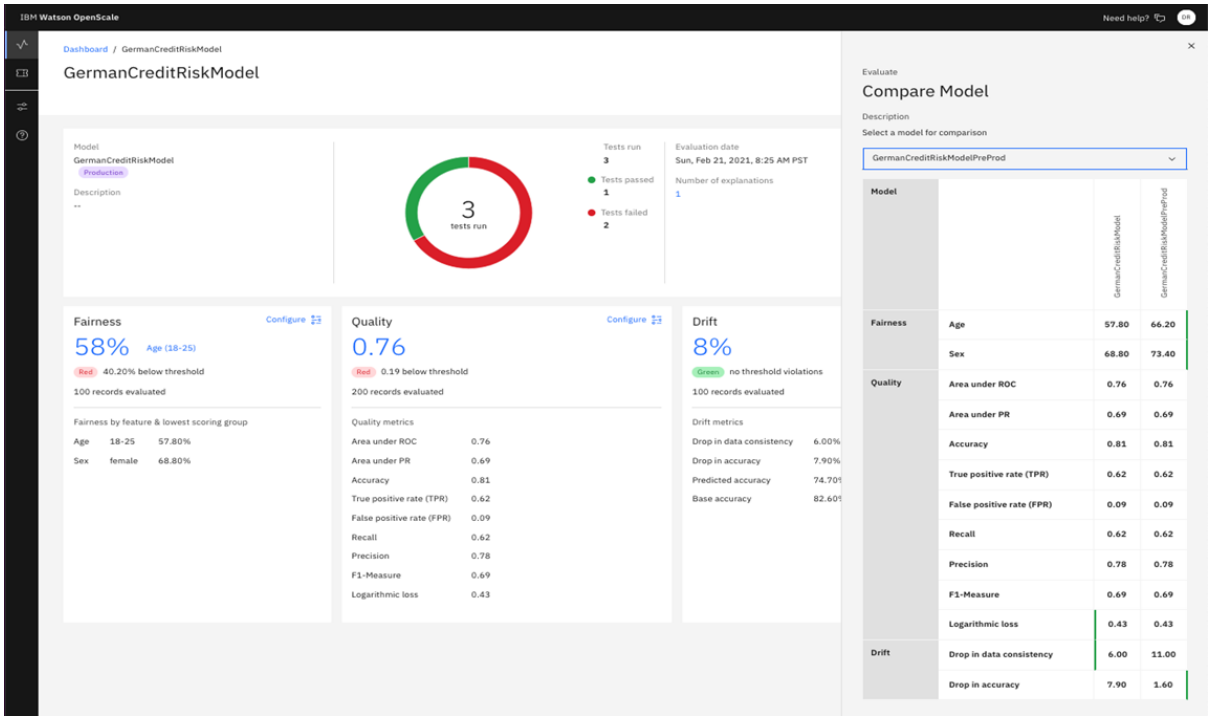

IBM Cloud Pak for Data
Upgrade
Michael Adams's Account
IBM Match 360 with Watson home / Master data configuration / Mapping
Data setup
Assets
Mapping
Modeling
4-sources-data5.csv
Last mapping update: Apr 15, 2021, 1:45 PM
Auto map
Profile
Find an attribute or field
Filter by:
All columns (11)
COLUMN5
COLUMN6
COLUMN7
COLUMN8
Mapped
Not Mapped
PORT NORRIS
NJ
8349
713-6906
KEWAUNEE
WI
54216
850-0525
CHENEY
KS
67025
768-4917
PRAGUE
NE
68050
216-5155
TUCSON
AZ
85708
267-3352
Mapping targets
COLUMN6 - Primary residence - State/Province value
Primary residence - State/Province value
100%
Legal name - Prefix
7%
Mailing address - Street Name
5%

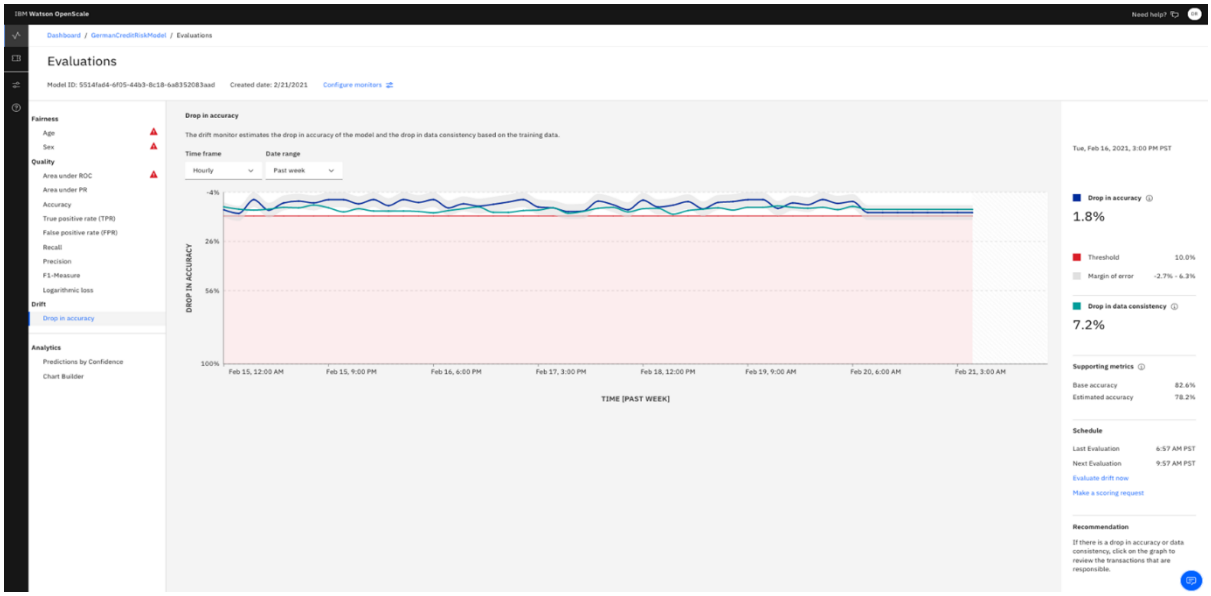

Master data home
Add record +
Master data explorer
Export
Search
Explore
Settings
johnson
Enter search criteria here
Records
Simple
Advanced
Brand (1)
Organization (8)
Person (1)
Show graph
Business address
Business name
Address line 1
City
Postal code
State/Province value
Name
Business tax identification
Business telephone
Record display label
System record number
0
7932 NORTHWEST JOHNSON ST.
OVERGAARD
85933
AZ
HEMODYNAMICS
59-8363961
9289188968
HEMODYNAMICS
0
1240 NORTHWEST JOHNSON STREET
MELBOURNE
32935
FL
Ruegsegger Farms
52-2735232
3219242039
Ruegsegger Farms
0
1240 NORTHWEST JOHNSON STREET
MELBOURNE
32935
FL
Ruegsegger Farms
52-2735232
3219242039
Ruegsegger Farms
0
9823 E JOHNSON
WESTVIEW
40178
KY
DIAGNOSTIC
14-8643014
2703714924
DIAGNOSTIC
showing 1-10 of 186 record results
1
2
...
19
Results per page:
10

Chapter 5: Analyzing: Building, Deploying, and Scaling Models with Trust and Transparency

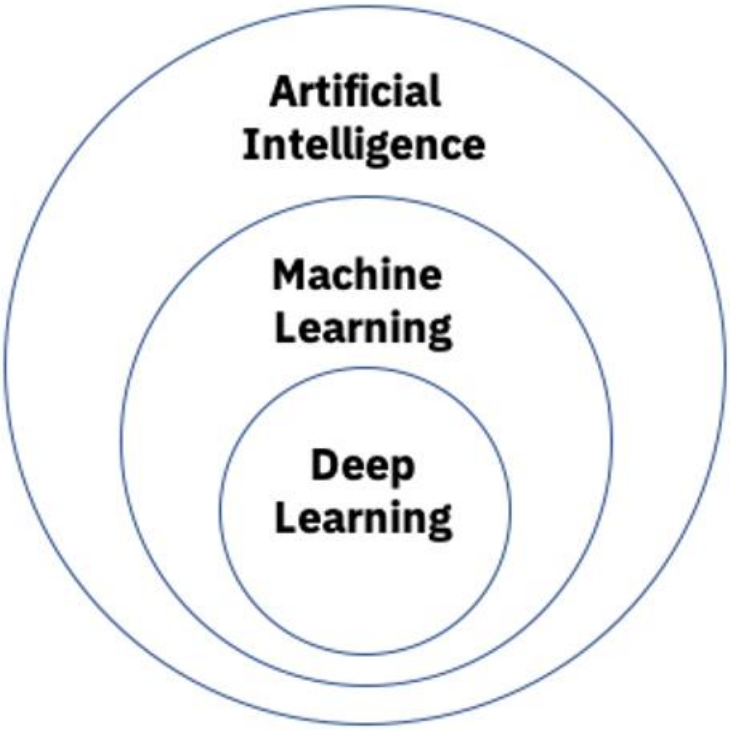


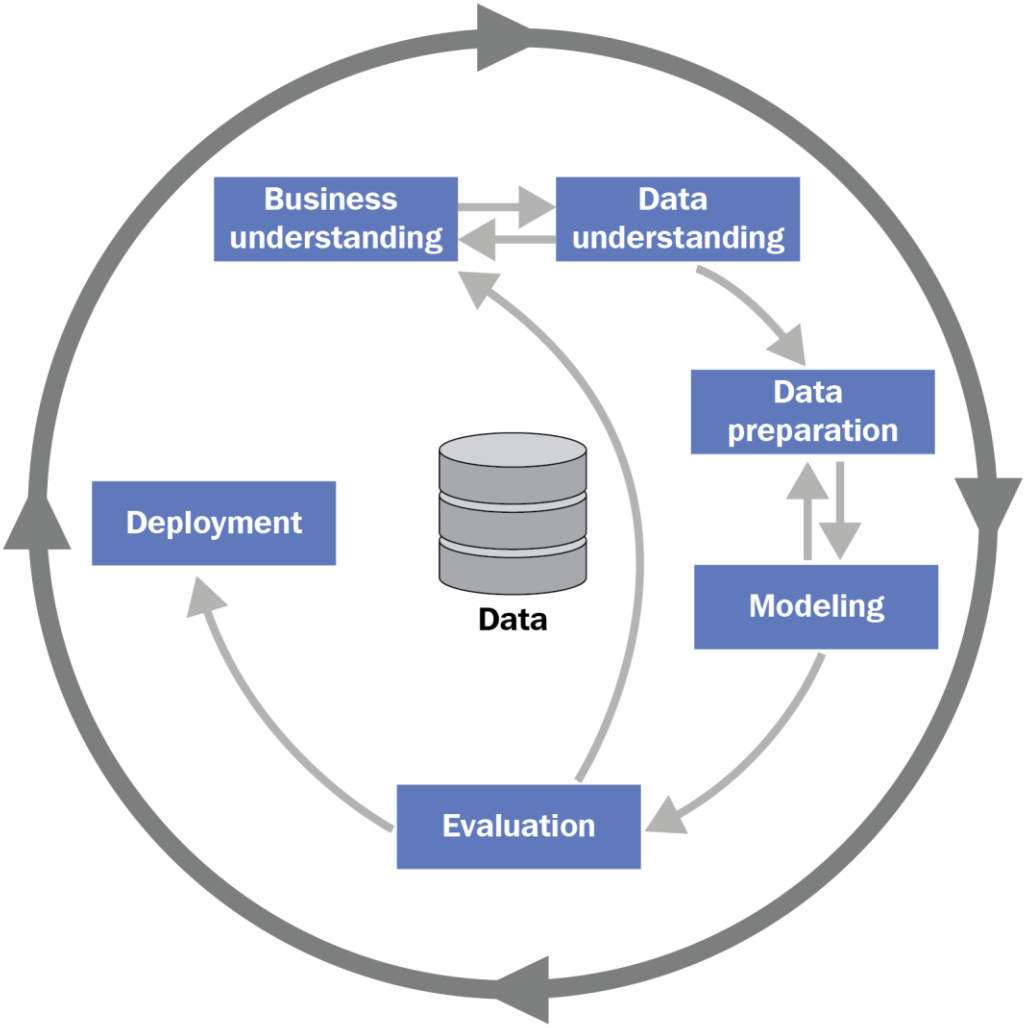


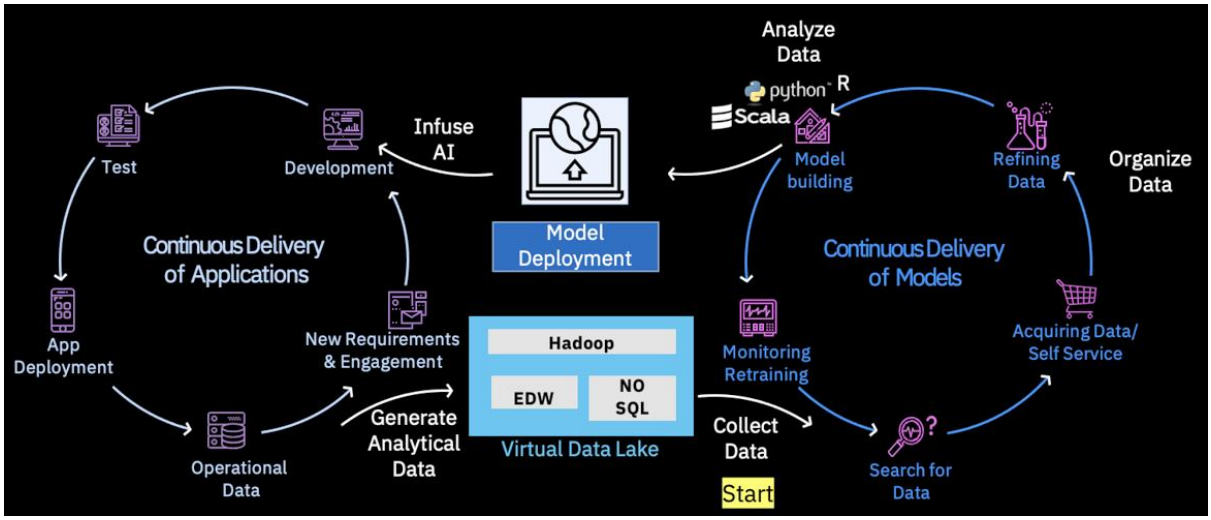

IBM Watson OpenScale
Need help?
Dashboard / GermanCreditRiskModel
GermanCreditRiskModel
Model
GermanCreditRiskModel
Production
Description
--
3
tests run
Tests run
3
Tests passed
1
Tests failed
2
Evaluation date
Sun, Feb 21, 2021, 8:25 AM PST
Number of explanations
1
Fairness
Configure
58%
Age (18-25)
Red 40.20% below threshold
100 records evaluated
Fairness by feature & lowest scoring group
Age 18-25 57.80%
Sex female 68.80%
Quality
Configure
0.76
Red 0.19 below threshold
200 records evaluated
Quality metrics
Area under ROC 0.76
Area under PR 0.69
Accuracy 0.81
True positive rate (TPR) 0.62
False positive rate (FPR) 0.09
Recall 0.62
Precision 0.78
F1-Measure 0.69
Logarithmic loss 0.43
Drift
8%
Green no threshold violations
100 records evaluated
Drift metrics
Drop in data consistency 6.00%
Drop in accuracy 7.90%
Predicted accuracy 74.70%
Base accuracy 82.60%
Evaluate
Compare Model
Description
Select a model for comparison
GermanCreditRiskModelPreProd
Model
GermanCreditRiskModel
GermanCreditRiskModelPreProd
Fairness
Age 57.80 66.20
Sex 68.80 73.40
Quality
Area under ROC 0.76 0.76
Area under PR 0.69 0.69
Accuracy 0.81 0.81
True positive rate (TPR) 0.62 0.62
False positive rate (FPR) 0.09 0.09
Recall 0.62 0.62
Precision 0.78 0.78
F1-Measure 0.69 0.69
Logarithmic loss 0.43 0.43
Drift
Drop in data consistency 6.00 11.00
Drop in accuracy 7.90 1.60

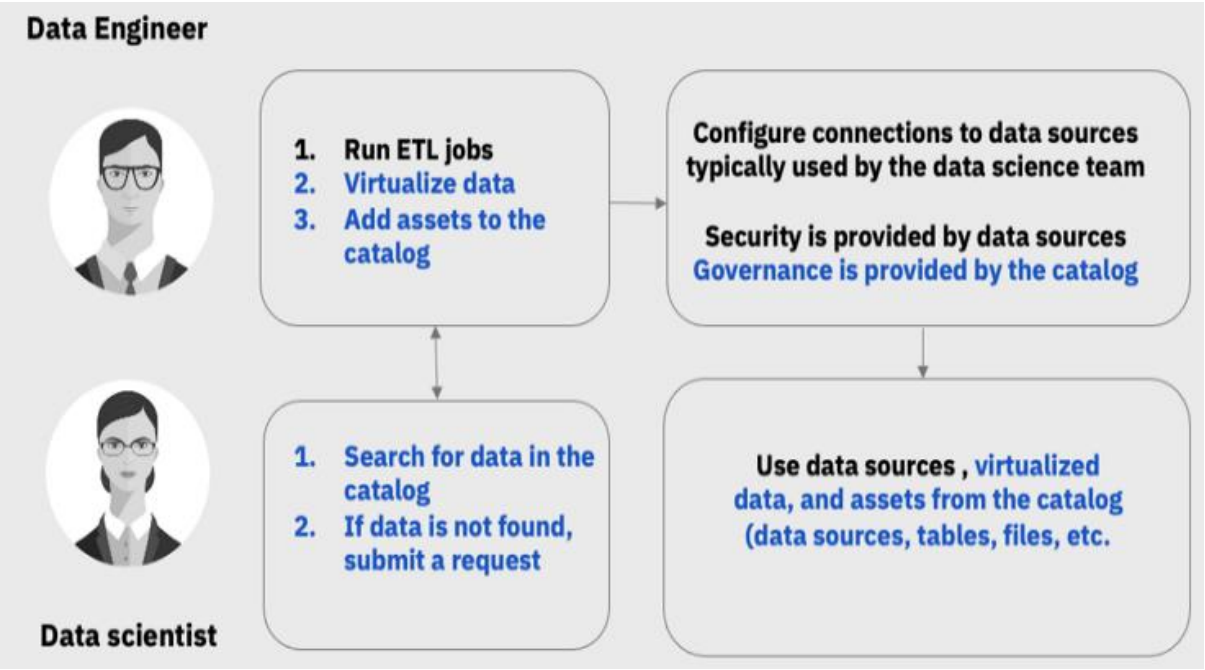

IBM Watson OpenScale
Need help?
Dashboard / GermanCreditRiskModel / Evaluations
Evaluations
Model ID: 5514fad4-6f05-44b3-8c18-6a8352083aad
Created date: 2/21/2021
Configure monitors
Fairness
Age
Sex
Quality
Area under ROC
Area under PR
Accuracy
True positive rate (TPR)
False positive rate (FPR)
Recall
Precision
F1-Measure
Logarithmic loss
Drift
Drop in accuracy
Analytics
Predictions by Confidence
Chart Builder
Drop in accuracy
The drift monitor estimates the drop in accuracy of the model and the drop in data consistency based on the training data.
Time frame
Hourly
Date range
Past week
-4%
26%
56%
100%
DROP IN ACCURACY
Feb 15, 12:00 AM
Feb 15, 9:00 PM
Feb 16, 6:00 PM
Feb 17, 3:00 PM
Feb 18, 12:00 PM
Feb 19, 9:00 AM
Feb 20, 6:00 AM
Feb 21, 3:00 AM
TIME [PAST WEEK]
Tue, Feb 16, 2021, 3:00 PM PST
Drop in accuracy
1.8%
Threshold 10.0%
Margin of error -2.7% - 6.3%
Drop in data consistency
7.2%
Supporting metrics
Base accuracy 82.6%
Estimated accuracy 78.2%
Schedule
Last Evaluation 6:57 AM PST
Next Evaluation 9:57 AM PST
Evaluate drift now
Make a scoring request
Recommendation
If there is a drop in accuracy or data consistency, click on the graph to review the transactions that are responsible.

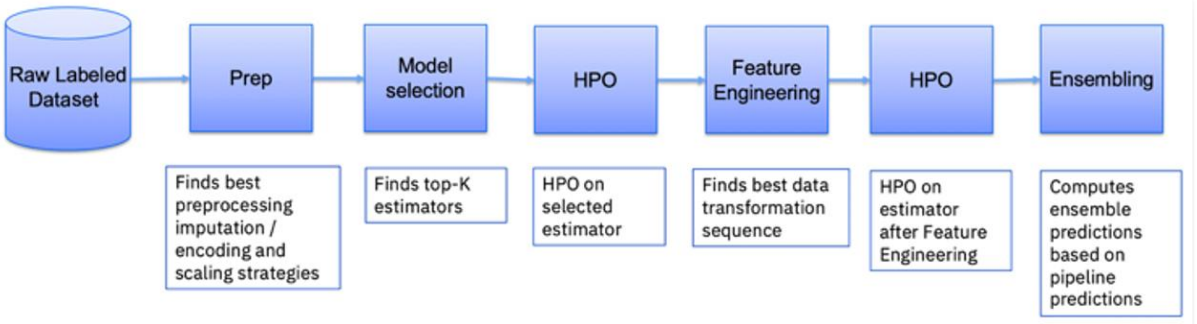
Prescriptive Analytics: What should we do about it?
Planning Analytics: What is our plan?
Descriptive Analytics: What happened?
Diagnostic Analytics: Why did it happen?
Predictive Analytics: What will happen next?

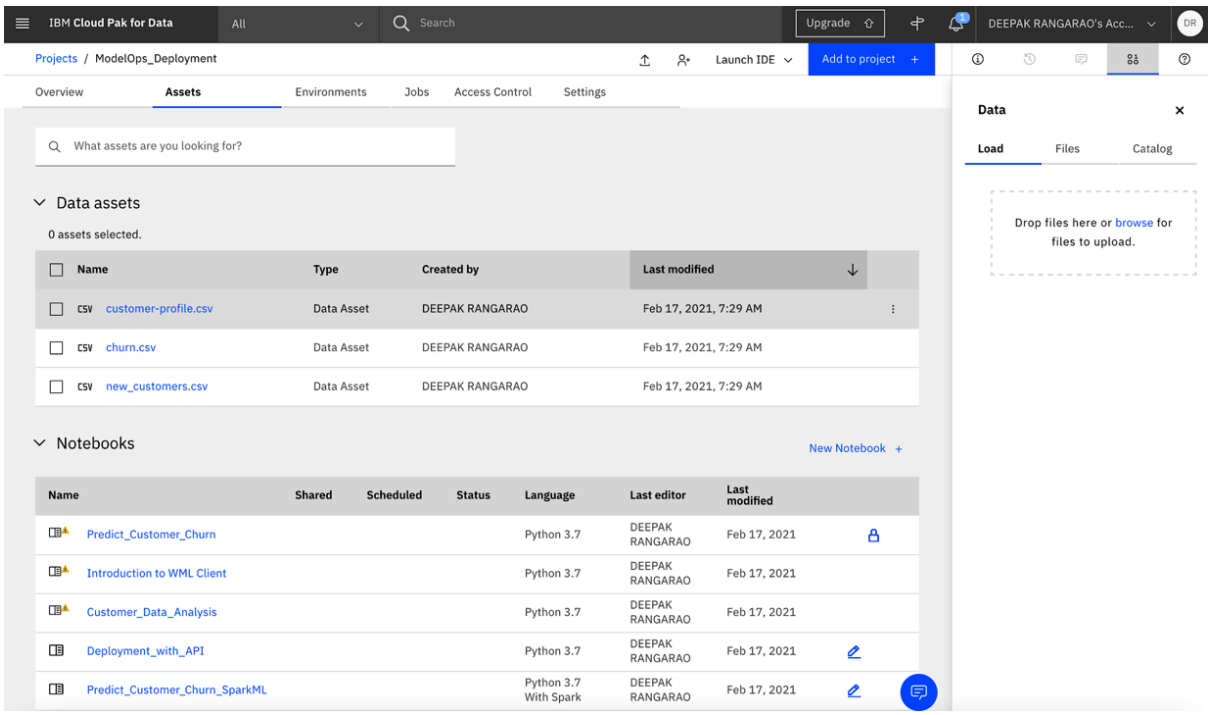

Artificial Intelligence
Machine Learning
Deep Learning

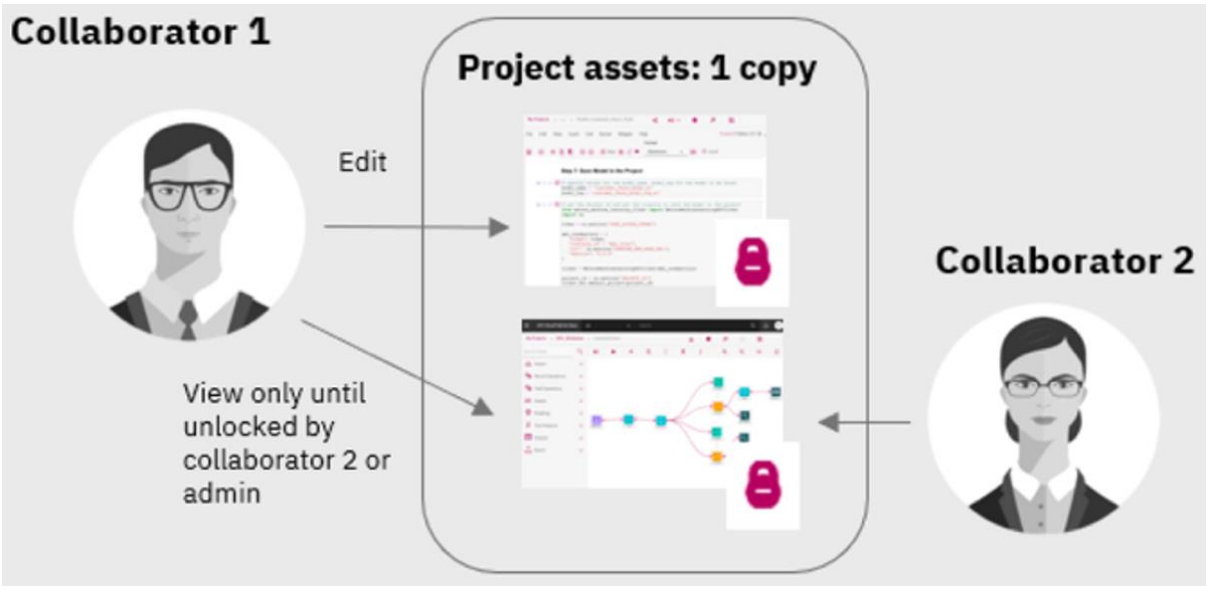

Business understanding
Data understanding
Data preparation
Modeling
Evaluation
Deployment
Data

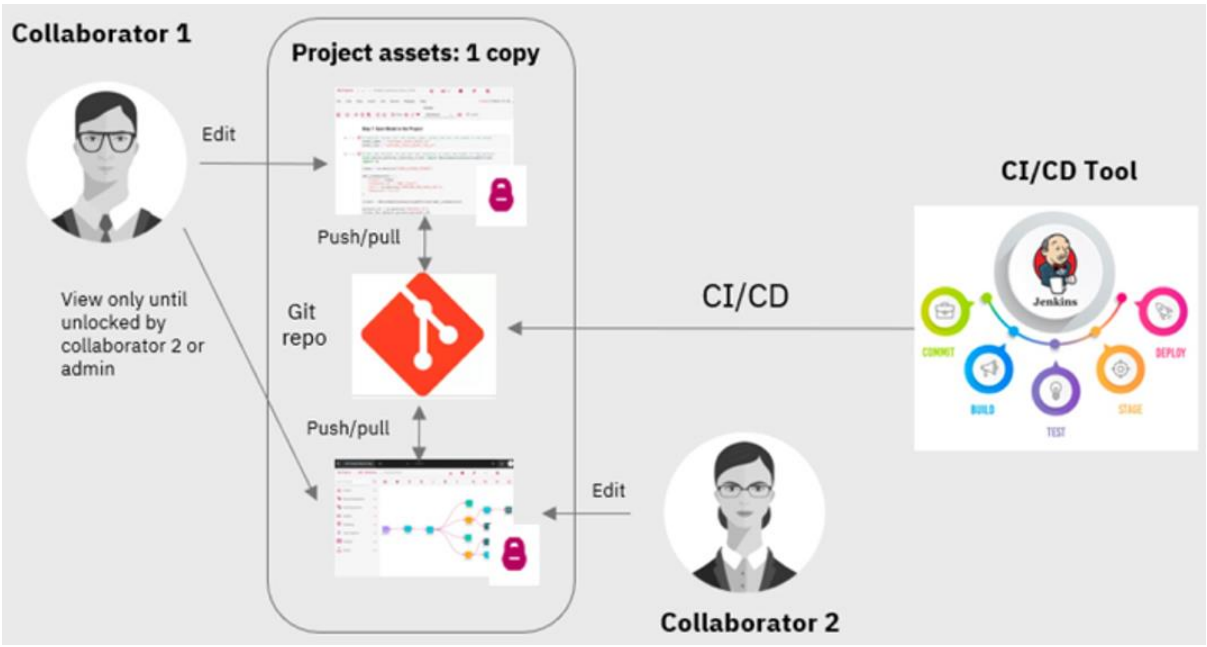

Analyze Data
python
R
Scala
Model building
Refining Data
Organize Data
Test
Development
Infuse AI
Model Deployment
Continuous Delivery of Applications
Continuous Delivery of Models
App Deployment
New Requirements & Engagement
Hadoop
EDW
NO SQL
Virtual Data Lake
Monitoring Retraining
Acquiring Data/ Self Service
Operational Data
Generate Analytical Data
Collect Data
Start
Search for Data

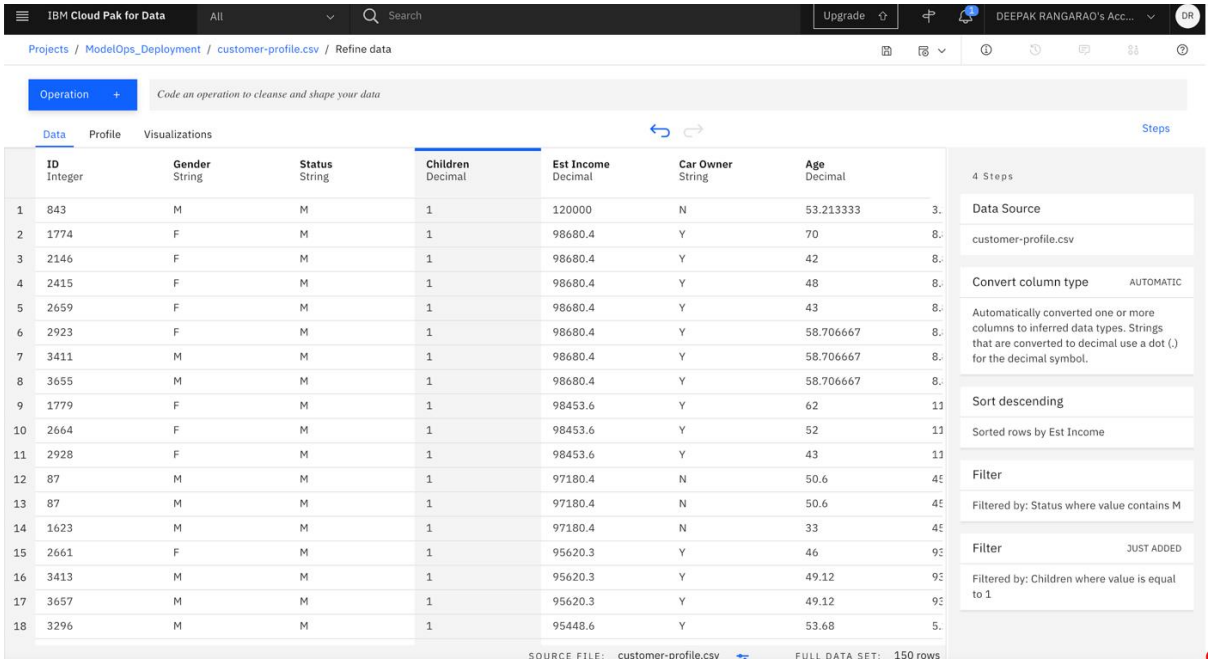

Data Engineer
1. Run ETL jobs
2. Virtualize data
3. Add assets to the catalog
Configure connections to data sources typically used by the data science team
Security is provided by data sources
Governance is provided by the catalog
1. Search for data in the catalog
2. If data is not found, submit a request
Use data sources , virtualized data, and assets from the catalog (data sources, tables, files, etc.
Data scientist

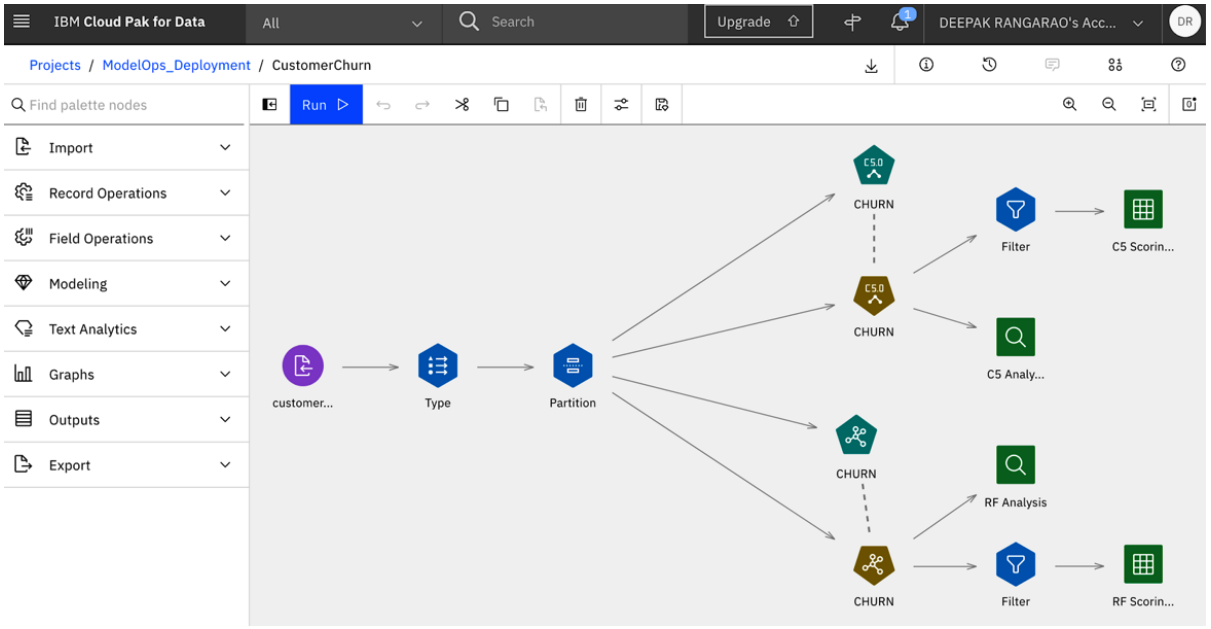

Raw Labeled Dataset
Prep
Model selection
HPO
Feature Engineering
HPO
Ensembling
Finds best preprocessing imputation / encoding and scaling strategies
Finds top-K estimators
HPO on selected estimator
Finds best data transformation sequence
HPO on estimator after Feature Engineering
Computes ensemble predictions based on pipeline predictions

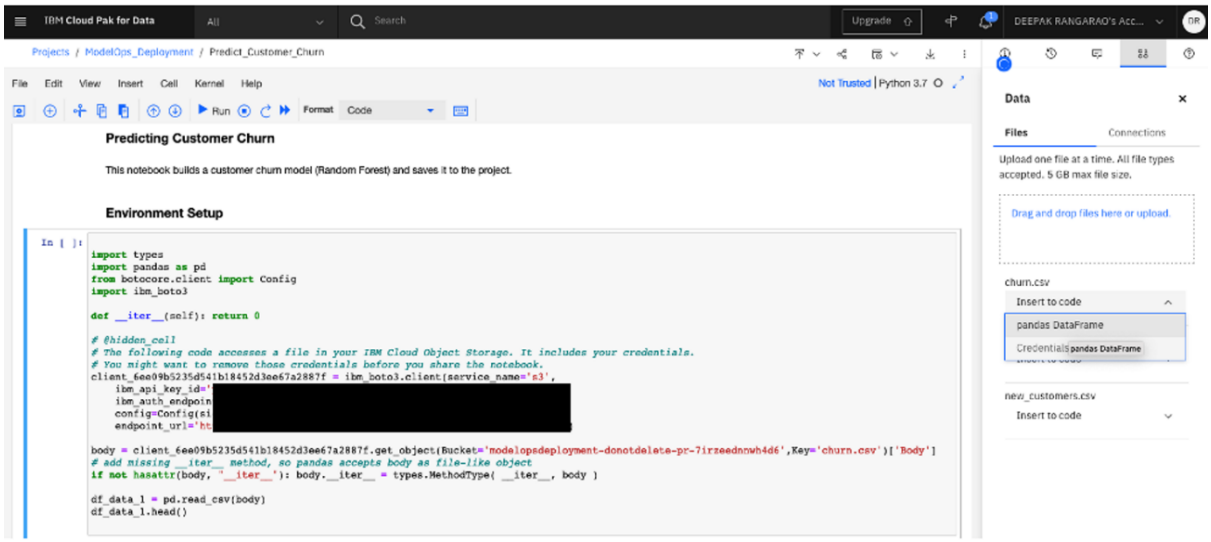

IBM Cloud Pak for Data
All
Search
Upgrade
DEEPAK RANGARAO's Acc...
Projects / ModelOps_Deployment
Launch IDE
Add to project
Overview
Assets
Environments
Jobs
Access Control
Settings
What assets are you looking for?
Data assets
0 assets selected.
Name
Type
Created by
Last modified
customer-profile.csv
Data Asset
DEEPAK RANGARAO
Feb 17, 2021, 7:29 AM
churn.csv
Data Asset
DEEPAK RANGARAO
Feb 17, 2021, 7:29 AM
new_customers.csv
Data Asset
DEEPAK RANGARAO
Feb 17, 2021, 7:29 AM
Notebooks
New Notebook
Name
Shared
Scheduled
Status
Language
Last editor
Last modified
Predict_Customer_Churn
Python 3.7
DEEPAK RANGARAO
Feb 17, 2021
Introduction to WML Client
Python 3.7
DEEPAK RANGARAO
Feb 17, 2021
Customer_Data_Analysis
Python 3.7
DEEPAK RANGARAO
Feb 17, 2021
Deployment_with_API
Python 3.7
DEEPAK RANGARAO
Feb 17, 2021
Predict_Customer_Churn_SparkML
Python 3.7 With Spark
DEEPAK RANGARAO
Feb 17, 2021
Data
Load
Files
Catalog
Drop files here or browse for files to upload.

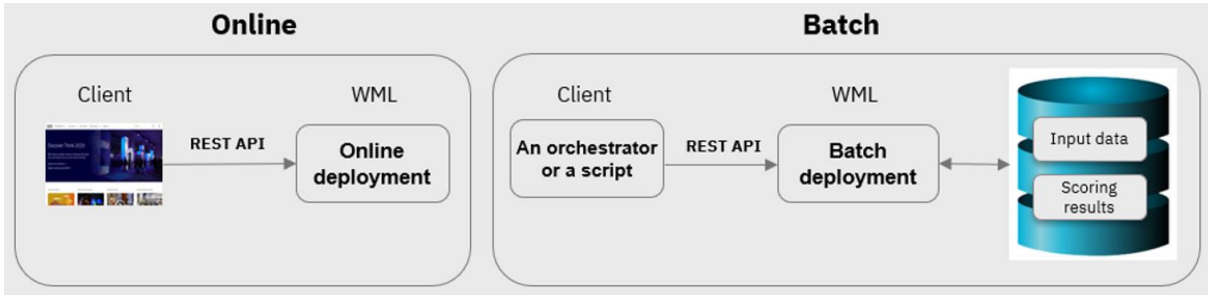

Collaborator 1
Project assets: 1 copy
Edit
Collaborator 2
View only until unlocked by collaborator 2 or admin

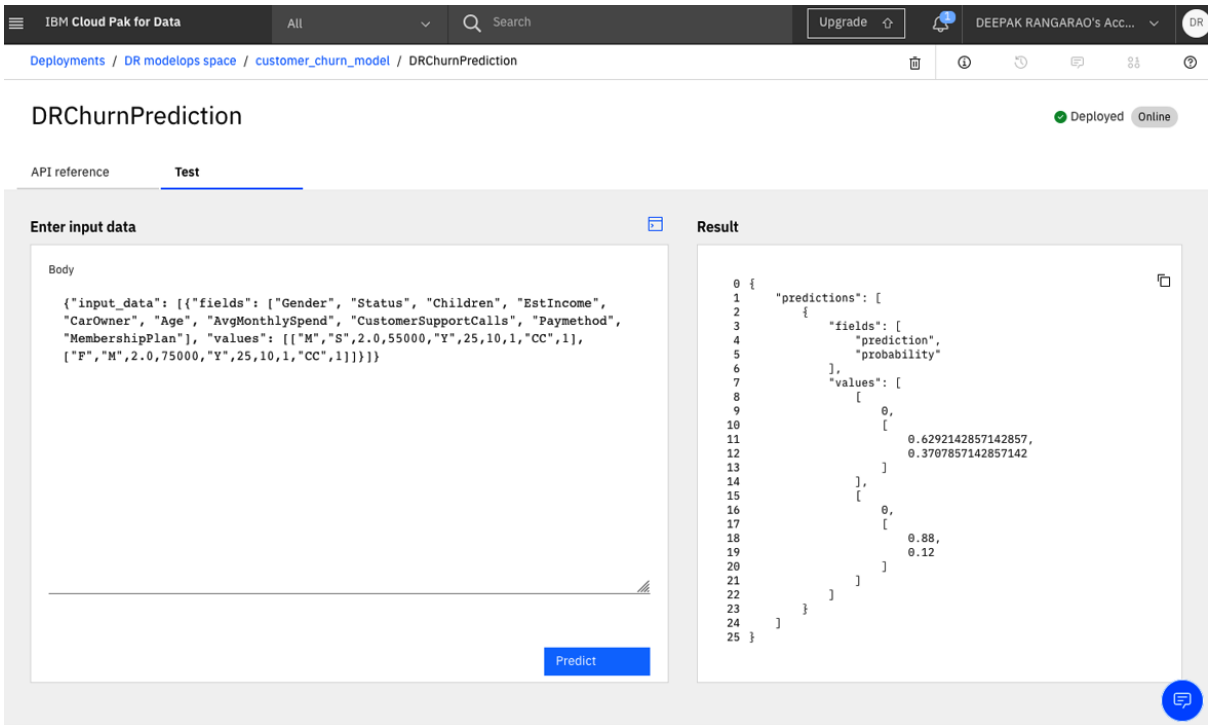

Collaborator 1
Project assets: 1 copy
Edit
Push/pull
Git repo
Push/pull
View only until unlocked by collaborator 2 or admin
CI/CD Tool
CI/CD
Jenkins
Edit
Collaborator 2

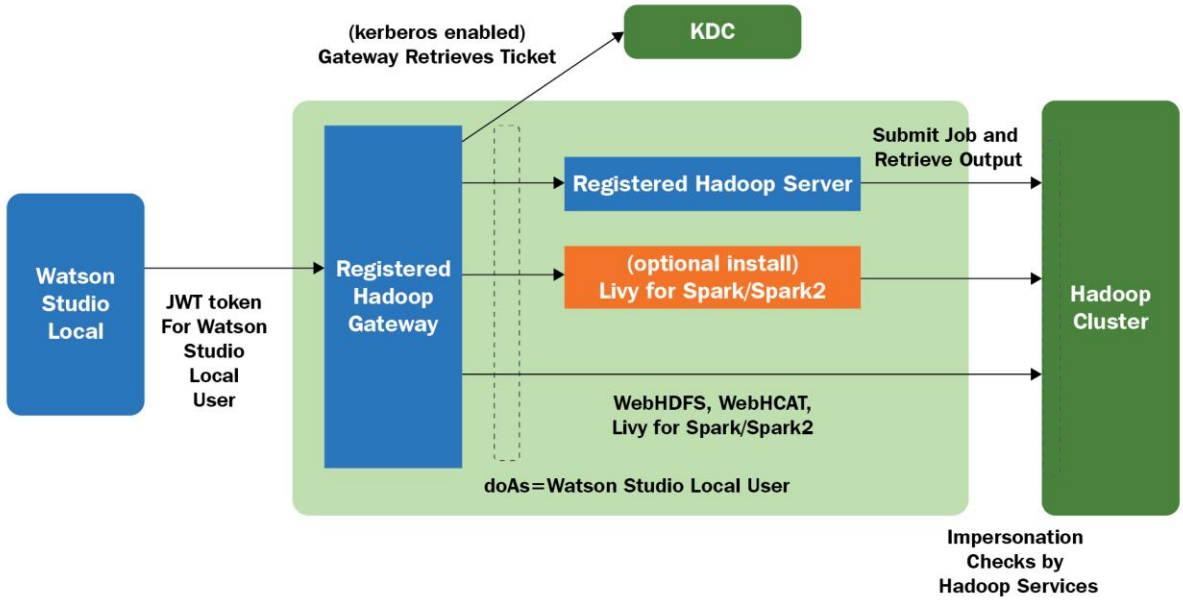

IBM Cloud Pak for Data
All
Search
Upgrade
DEEPAK RANGARAO's Acc...
DR
Projects / ModelOps_Deployment / customer-profile.csv / Refine data
Operation +
Code an operation to cleanse and shape your data
Data
Profile
Visualizations
Steps
ID Integer
Gender String
Status String
Children Decimal
Est Income Decimal
Car Owner String
Age Decimal
1 843 M M 1 120000 N 53.213333
2 1774 F M 1 98680.4 Y 70
3 2146 F M 1 98680.4 Y 42
4 2415 F M 1 98680.4 Y 48
5 2659 F M 1 98680.4 Y 43
6 2923 F M 1 98680.4 Y 58.706667
7 3411 M M 1 98680.4 Y 58.706667
8 3655 M M 1 98680.4 Y 58.706667
9 1779 F M 1 98453.6 Y 62
10 2664 F M 1 98453.6 Y 52
11 2928 F M 1 98453.6 Y 43
12 87 M M 1 97180.4 N 50.6
13 87 M M 1 97180.4 N 50.6
14 1623 M M 1 97180.4 N 33
15 2661 F M 1 95620.3 Y 46
16 3413 M M 1 95620.3 Y 49.12
17 3657 M M 1 95620.3 Y 49.12
18 3296 M M 1 95448.6 Y 53.68
SOURCE FILE: customer-profile.csv
FULL DATA SET: 150 rows
4 Steps
Data Source
customer-profile.csv
Convert column type AUTOMATIC
Automatically converted one or more columns to inferred data types. Strings that are converted to decimal use a dot (.) for the decimal symbol.
Sort descending
Sorted rows by Est Income
Filter
Filtered by: Status where value contains M
Filter JUST ADDED
Filtered by: Children where value is equal to 1

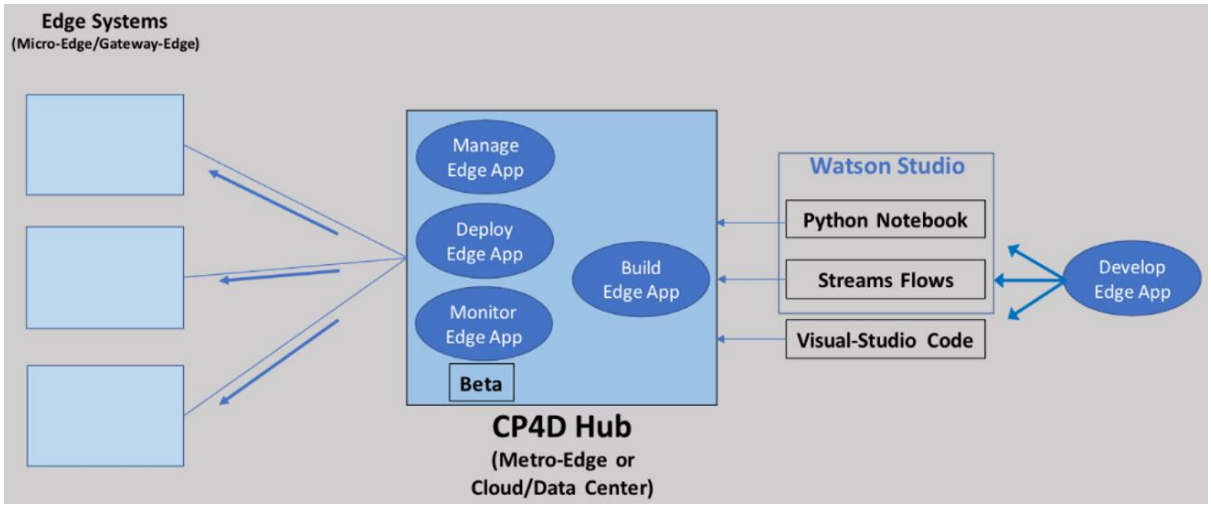

IBM Cloud Pak for Data
All
Search
Upgrade
DEEPAK RANGARAO's Acc...
DR
Projects / ModelOps_Deployment / CustomerChurn
Find palette nodes
Run
Import
Record Operations
Field Operations
Modeling
Text Analytics
Graphs
Outputs
Export
customer...
Type
Partition
C5.0
CHURN
C5.0
CHURN
CHURN
CHURN
Filter
C5 Scorin...
C5 Analy...
RF Analysis
Filter
RF Scorin...


IBM Cloud Pak for Data
Projects / ModelOps_Deployment / Predict_Customer_Churn
File Edit View Insert Cell Kernel Help
Not Trusted | Python 3.7
Predicting Customer Churn
This notebook builds a customer churn model (Random Forest) and saves it to the project.
Environment Setup
In [ ]:
import types
import pandas as pd
from botocore.client import Config
import ibm_boto3
def __iter__(self): return 0
# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
body = client_6ee09b5235d541b18452d3ee67a2887f.get_object(Bucket='modelopsdeployment-donotdelete-pr-7irzeednnwh4d6',Key='churn.csv')['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType(__iter__, body)
df_data_1 = pd.read_csv(body)
df_data_1.head()
Data
Files
Connections
Upload one file at a time. All file types accepted. 5 GB max file size.
Drag and drop files here or upload.
churn.csv
Insert to code
pandas DataFrame
Credentials
new_customers.csv
Insert to code

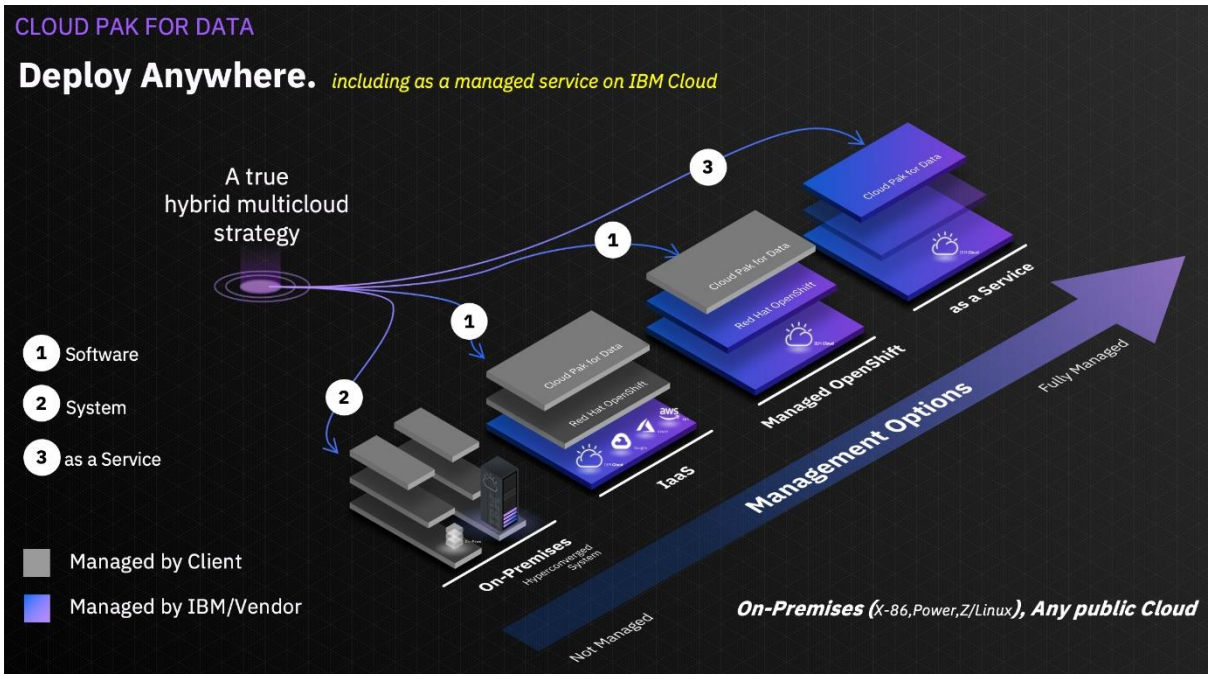

Online
Client
WML
REST API
Online deployment
Batch
Client
WML
An orchestrator or a script
REST API
Batch deployment
Input data
Scoring results

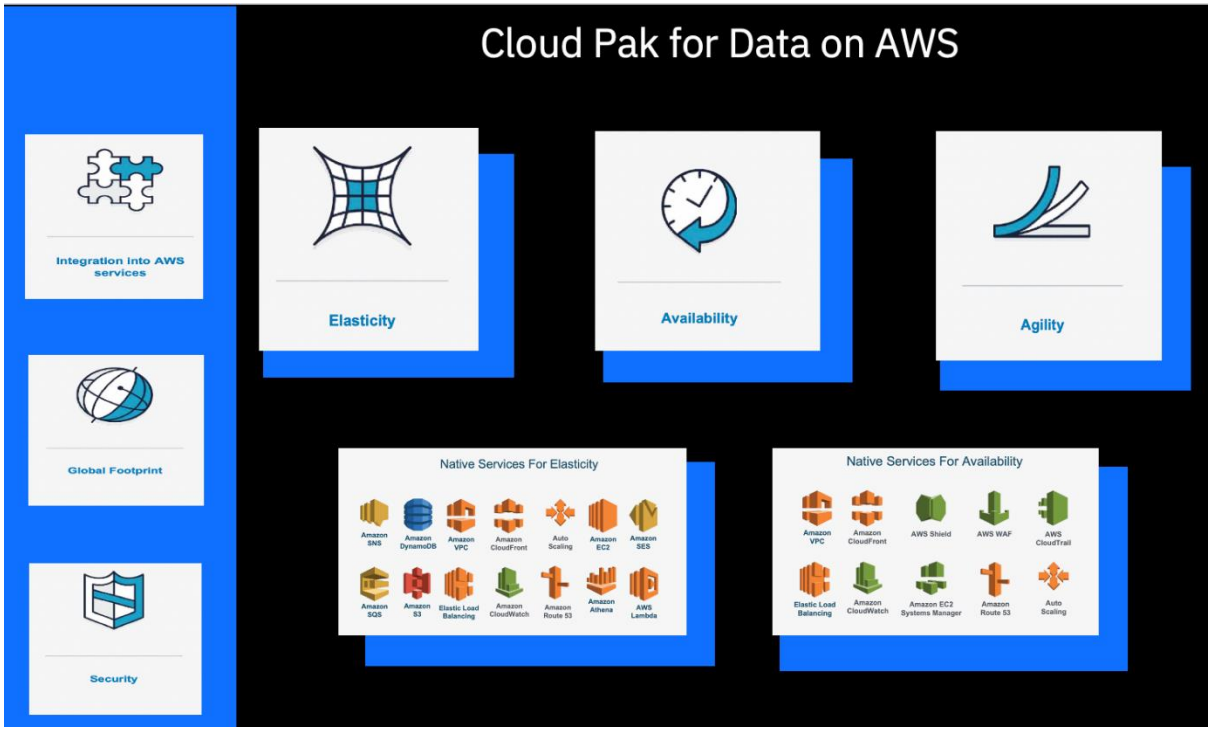

IBM Cloud Pak for Data
Deployments / DR modelops space / customer_churn_model / DRChurnPrediction
DRChurnPrediction
Deployed
Online
API reference
Test
Enter input data
Body
{"input_data": [{"fields": ["Gender", "Status", "Children", "EstIncome", "CarOwner", "Age", "AvgMonthlySpend", "CustomerSupportCalls", "Paymethod", "MembershipPlan"], "values": [["M","S",2.0,55000,"Y",25,10,1,"CC",1], ["F","M",2.0,75000,"Y",25,10,1,"CC",1]]}]}
Predict
Result
{
"predictions": [
{
"fields": [
"prediction",
"probability"
],
"values": [
[
0,
[
0.6292142857142857,
0.3707857142857142
]
],
[
0,
[
0.88,
0.12
]
]
]
}
]
}

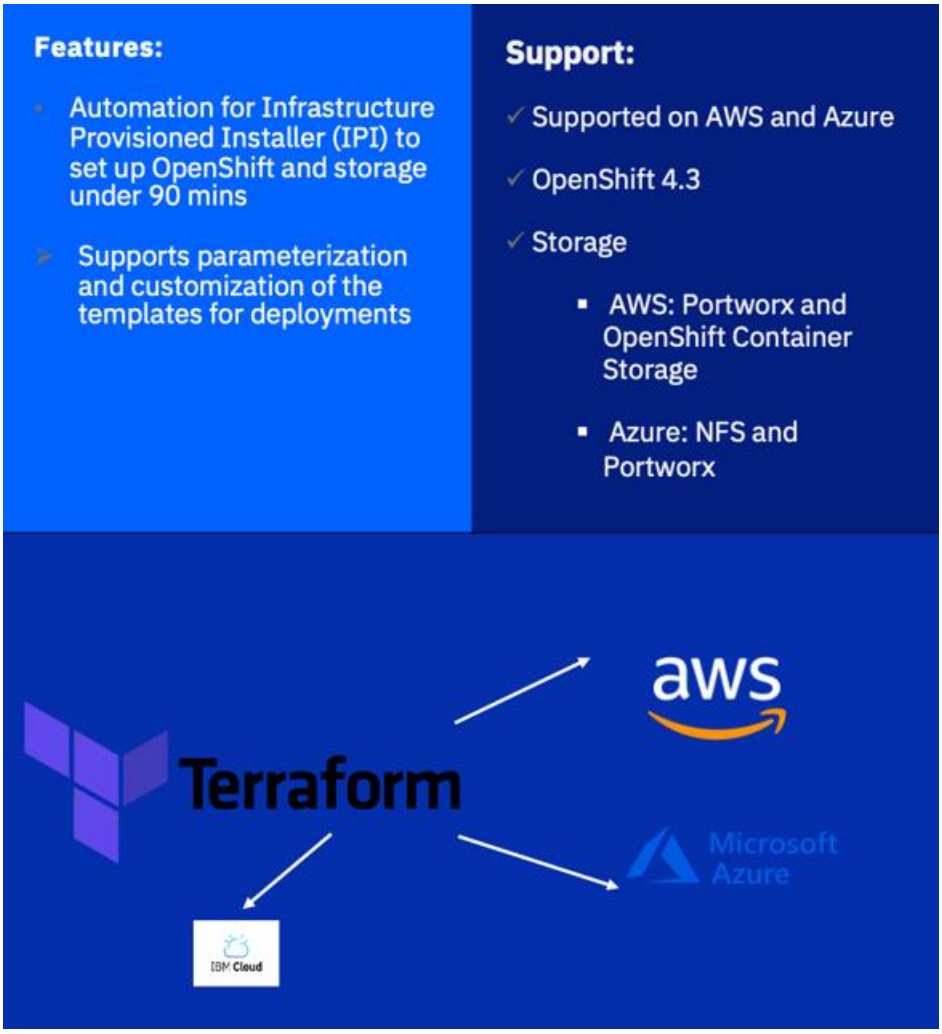

(kerberos enabled)
Gateway Retrieves Ticket
KDC
Watson Studio Local
JWT token For Watson Studio Local User
Registered Hadoop Gateway
Registered Hadoop Server
(optional install) Livy for Spark/Spark2
Submit Job and Retrieve Output
Hadoop Cluster
WebHDFS, WebHCAT, Livy for Spark/Spark2
doAs=Watson Studio Local User
Impersonation Checks by Hadoop Services

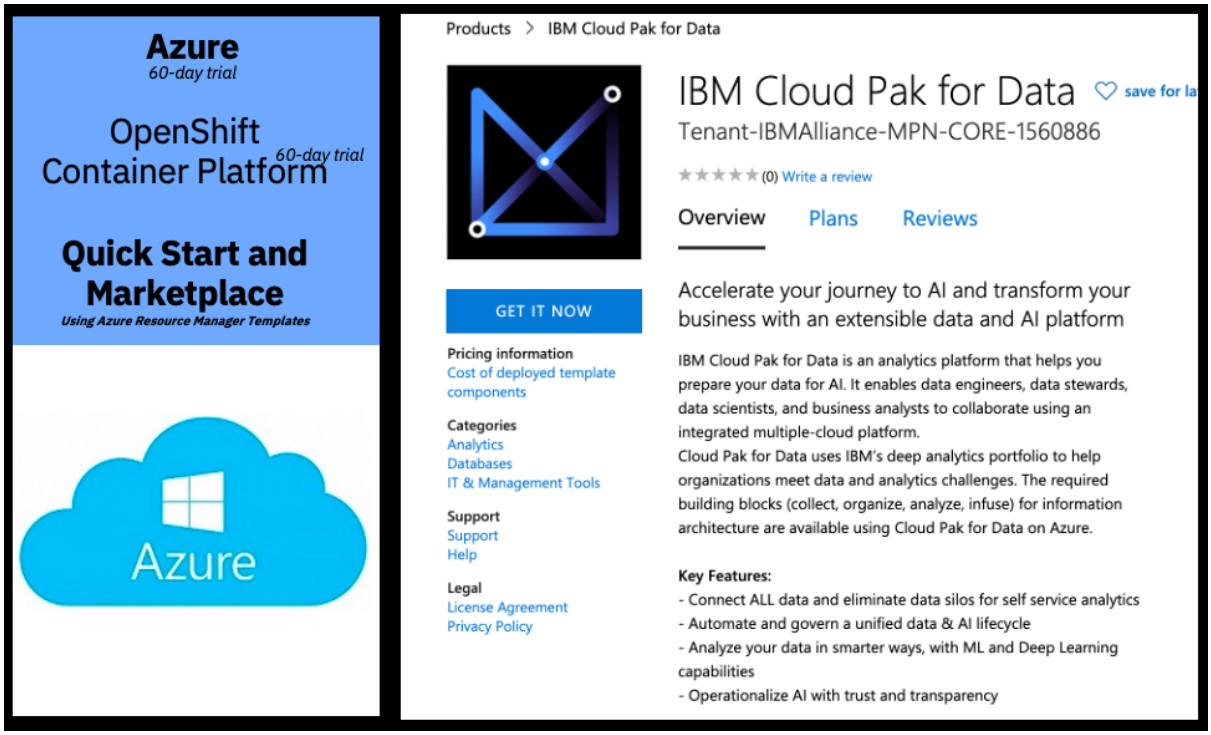

Edge Systems
(Micro-Edge/Gateway-Edge)
Manage Edge App
Deploy Edge App
Monitor Edge App
Beta
Build Edge App
CP4D Hub
(Metro-Edge or Cloud/Data Center)
Watson Studio
Python Notebook
Streams Flows
Visual-Studio Code
Develop Edge App

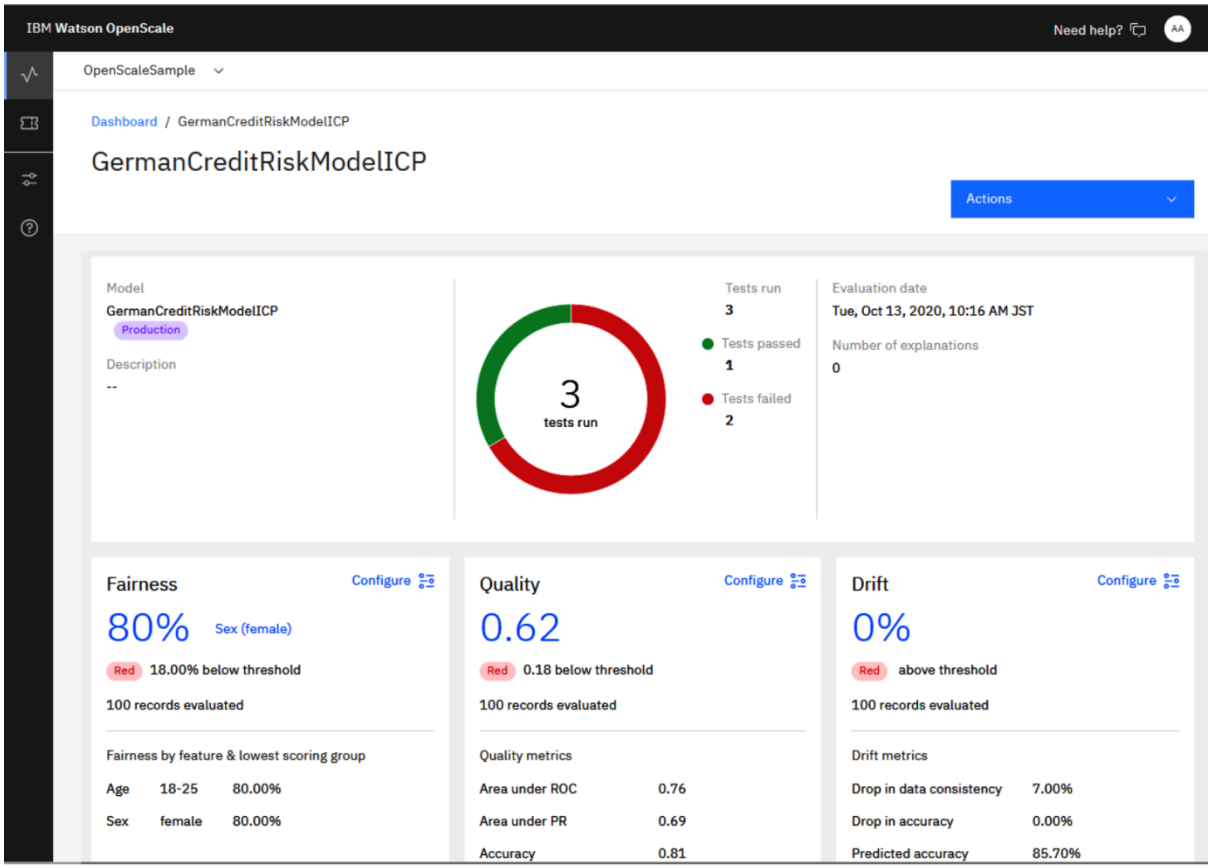
Chapter 6: Multi-Cloud Strategy and Cloud Satellite

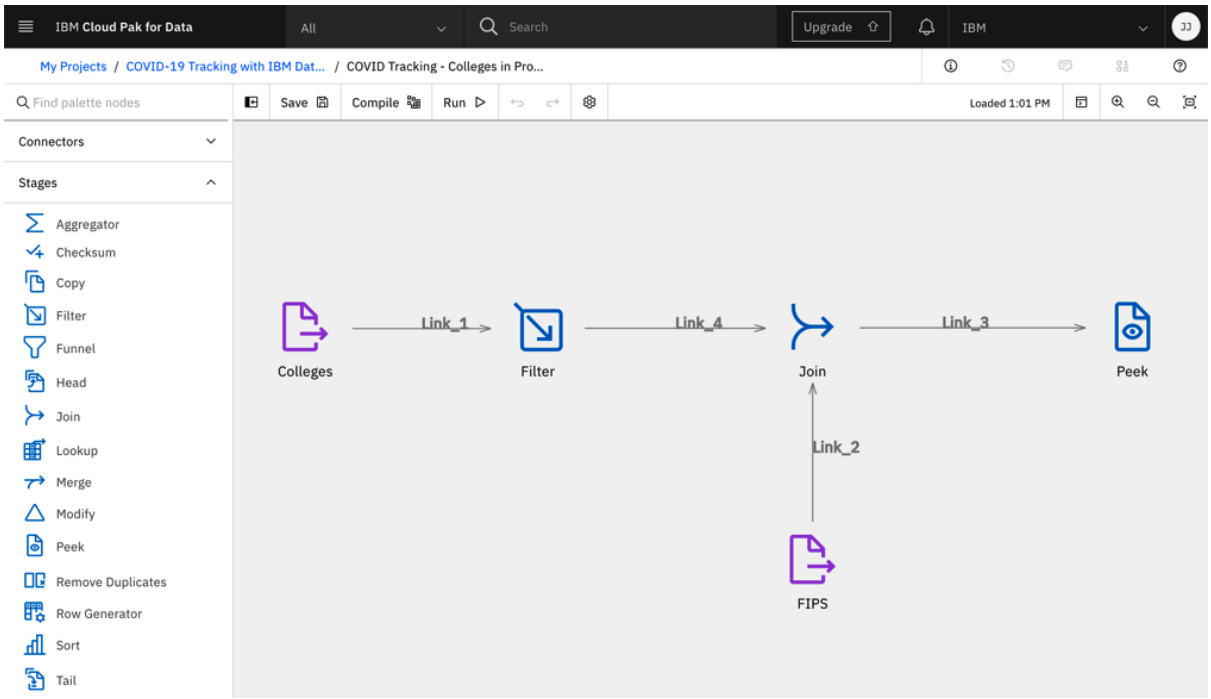


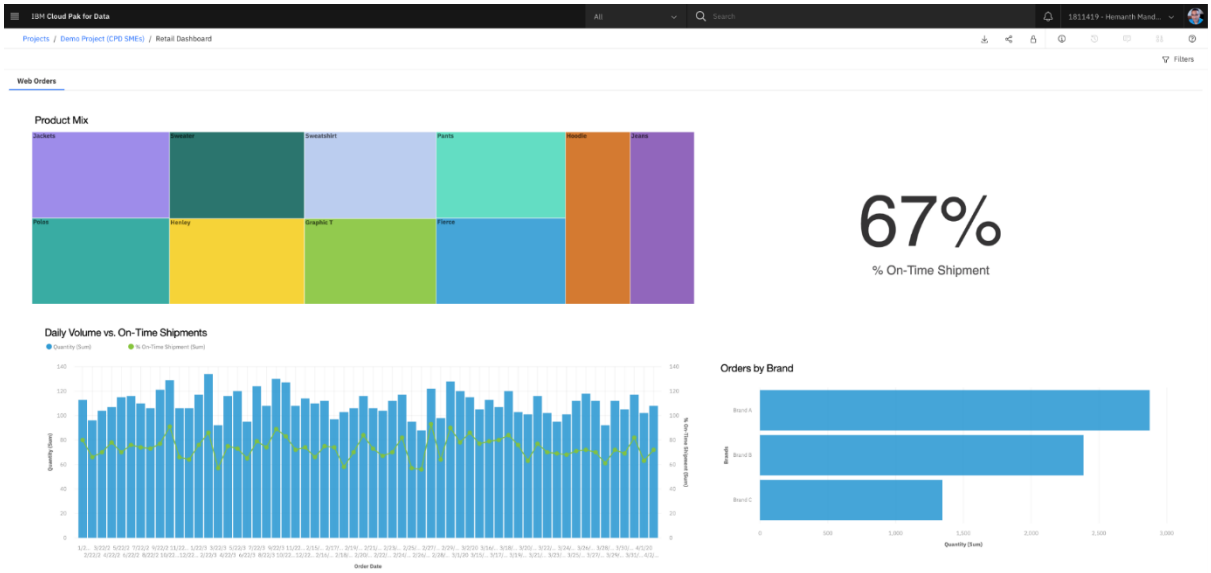


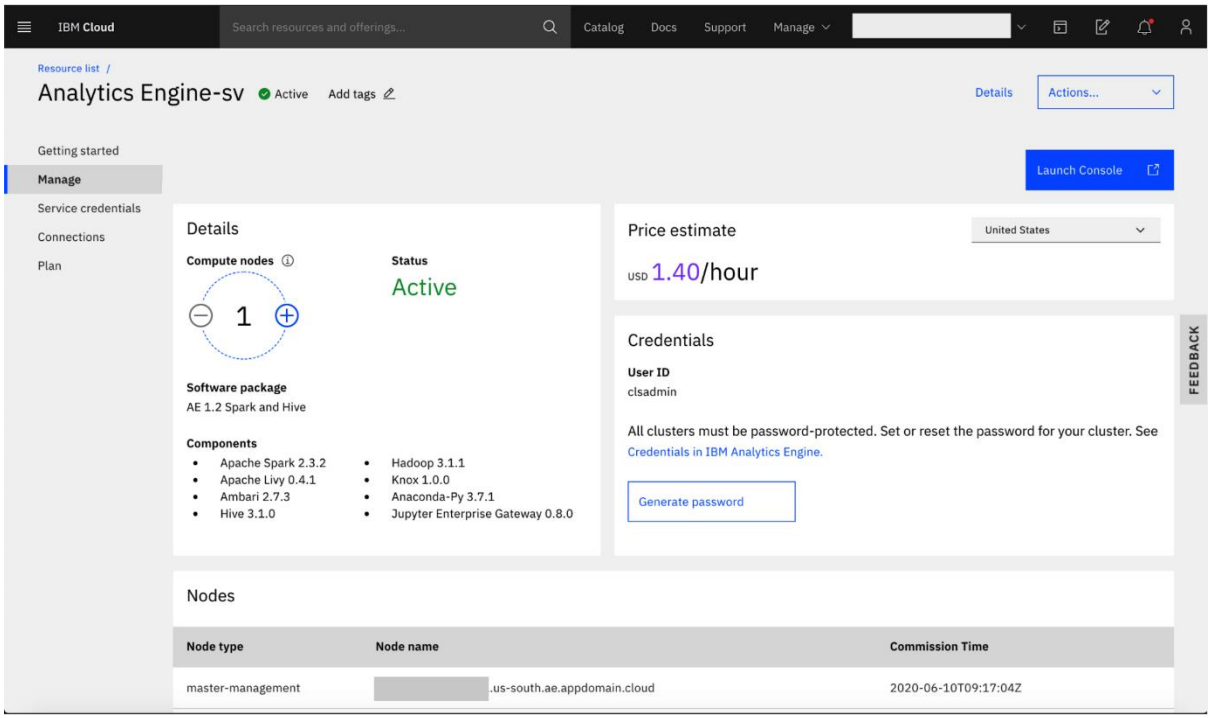

Features:
Automation for Infrastructure Provisioned Installer (IPI) to set up OpenShift and storage under 90 mins
Supports parameterization and customization of the templates for deployments
Support:
Supported on AWS and Azure
OpenShift 4.3
Storage
AWS: Portworx and OpenShift Container Storage
Azure: NFS and Portworx
Terraform
aws
Microsoft Azure
IBM Cloud

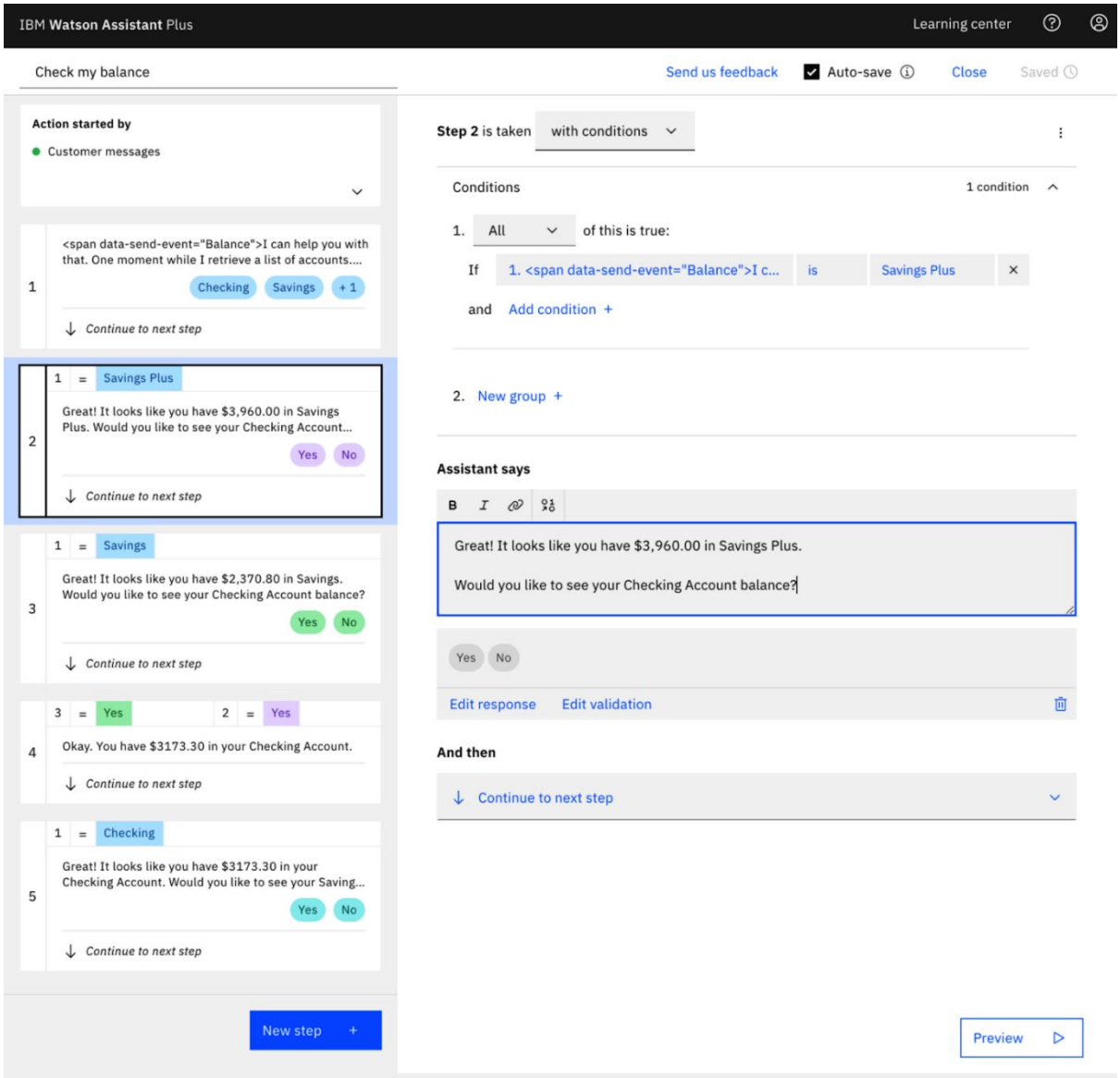

Azure
60-day trial
OpenShift Container Platform
60-day trial
Quick Start and Marketplace
Using Azure Resource Manager Templates
Azure
Products > IBM Cloud Pak for Data
IBM Cloud Pak for Data
save for la
Tenant-IBMAlliance-MPN-CORE-1560886
(0) Write a review
Overview
Plans
Reviews
GET IT NOW
Pricing information
Cost of deployed template components
Categories
Analytics
Databases
IT & Management Tools
Support
Support
Help
Legal
License Agreement
Privacy Policy
Accelerate your journey to AI and transform your business with an extensible data and AI platform
IBM Cloud Pak for Data is an analytics platform that helps you prepare your data for AI. It enables data engineers, data stewards, data scientists, and business analysts to collaborate using an integrated multiple-cloud platform.
Cloud Pak for Data uses IBM's deep analytics portfolio to help organizations meet data and analytics challenges. The required building blocks (collect, organize, analyze, infuse) for information architecture are available using Cloud Pak for Data on Azure.
Key Features:
- Connect ALL data and eliminate data silos for self service analytics
- Automate and govern a unified data & AI lifecycle
- Analyze your data in smarter ways, with ML and Deep Learning capabilities
- Operationalize AI with trust and transparency


IBM Watson OpenScale
Need help?
AA
OpenScaleSample
Dashboard / GermanCreditRiskModelICP
GermanCreditRiskModelICP
Actions
Model
GermanCreditRiskModelICP
Production
Description
--
3
tests run
Tests run
3
Tests passed
1
Tests failed
2
Evaluation date
Tue, Oct 13, 2020, 10:16 AM JST
Number of explanations
0
Fairness
Configure
80%
Sex (female)
Red
18.00% below threshold
100 records evaluated
Fairness by feature & lowest scoring group
Age 18-25 80.00%
Sex female 80.00%
Quality
Configure
0.62
Red
0.18 below threshold
100 records evaluated
Quality metrics
Area under ROC 0.76
Area under PR 0.69
Accuracy 0.81
Drift
Configure
0%
Red
above threshold
100 records evaluated
Drift metrics
Drop in data consistency 7.00%
Drop in accuracy 0.00%
Predicted accuracy 85.70%

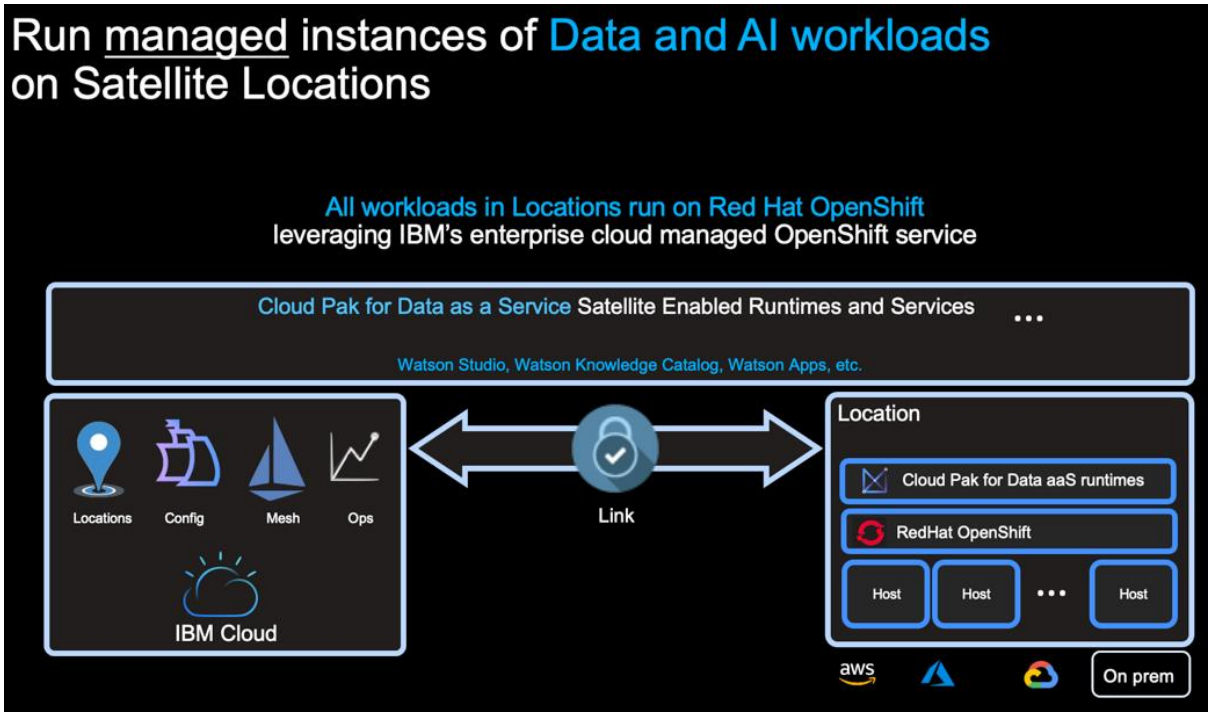

IBM Cloud Pak for Data
All
Search
Upgrade
IBM
JJ
My Projects / COVID-19 Tracking with IBM Dat... / COVID Tracking - Colleges in Pro...
Find palette nodes
Save
Compile
Run
Loaded 1:01 PM
Connectors
Stages
Aggregator
Checksum
Copy
Filter
Funnel
Head
Join
Lookup
Merge
Modify
Peek
Remove Duplicates
Row Generator
Sort
Tail
Colleges
Link_1
Filter
Link_4
Join
Link_3
Peek
Link_2
FIPS


IBM Cloud Pak for Data
Projects / Demo Project (CPD SMEs) / Retail Dashboard
Web Orders
Product Mix
Jackets
Sweater
Sweatshirt
Pants
Hoodie
Jeans
Polos
Henley
Graphic T
Fleece
67%
% On-Time Shipment
Daily Volume vs. On-Time Shipments
Orders by Brand


IBM Cloud
Search resources and offerings...
Catalog
Docs
Support
Manage
Resource list /
Analytics Engine-sv
Active
Add tags
Details
Actions...
Getting started
Manage
Service credentials
Connections
Plan
Launch Console
Details
Compute nodes
1
Status
Active
Software package
AE 1.2 Spark and Hive
Components
Apache Spark 2.3.2
Apache Livy 0.4.1
Ambari 2.7.3
Hive 3.1.0
Hadoop 3.1.1
Knox 1.0.0
Anaconda-Py 3.7.1
Jupyter Enterprise Gateway 0.8.0
Price estimate
United States
USD 1.40/hour
Credentials
User ID
clsadmin
All clusters must be password-protected. Set or reset the password for your cluster. See Credentials in IBM Analytics Engine.
Generate password
FEEDBACK
Nodes
Node type
Node name
Commission Time
master-management
.us-south.ae.appdomain.cloud
2020-06-10T09:17:04Z


IBM Watson Assistant Plus
Learning center
Check my balance
Send us feedback
Auto-save
Close
Saved
Action started by
Customer messages
1
I can help you with that. One moment while I retrieve a list of accounts....
Checking
Savings
+ 1
Continue to next step
2
1 = Savings Plus
Great! It looks like you have $3,960.00 in Savings Plus. Would you like to see your Checking Account...
Yes
No
Continue to next step
3
1 = Savings
Great! It looks like you have $2,370.80 in Savings. Would you like to see your Checking Account balance?
Yes
No
Continue to next step
4
3 = Yes
2 = Yes
Okay. You have $3173.30 in your Checking Account.
Continue to next step
5
1 = Checking
Great! It looks like you have $3173.30 in your Checking Account. Would you like to see your Saving...
Yes
No
Continue to next step
New step
Step 2 is taken with conditions
Conditions
1 condition
1. All of this is true:
If 1. I c... is Savings Plus
and Add condition +
2. New group +
Assistant says
Great! It looks like you have $3,960.00 in Savings Plus.
Would you like to see your Checking Account balance?
Yes
No
Edit response
Edit validation
And then
Continue to next step
Preview

Run managed instances of Data and AI workloads on Satellite Locations



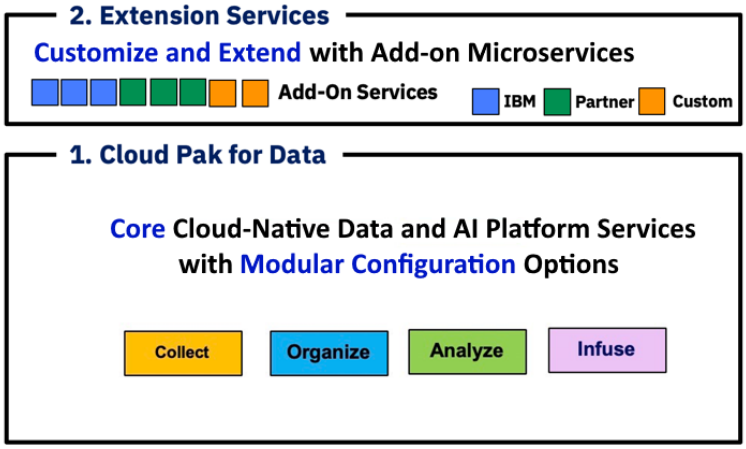
Data Fabric : Manage your Data and AI across hybrid and multi-cloud stacks

1. **Data Orchestration & Virtualization**
   - Provision & manage databases across multiple clouds
   - Access data irrespective of where it resides (Data Virtualization)
   - Clone databases, create sample dev-test instances with masking across clouds
2. **Data Governance (Distributed & Synchronized)**
   - Auto-catalog all enterprise data & analytics assets, define policies & rules
   - Distributed data governance & meta data synchronization through OMRS
   - Location sensitive data governance : track data lineage, define & enforce policies etc.
3. **Model Management & Deployment**
   - Develop, train , deploy and manage ML & AI models for multi-cloud
   - Ensure trust and governance of models across enterprise
4. **Data Movement (Backup/DR/Operational)**
   - Ad-hoc data movement/copies across different clouds
   - Backup/disaster recovery of data: lift and shift
5. **Data Privacy & Security**
   - Policy enforcement and data masking to ensure data privacy across Cloud
   - Data access and security across clouds

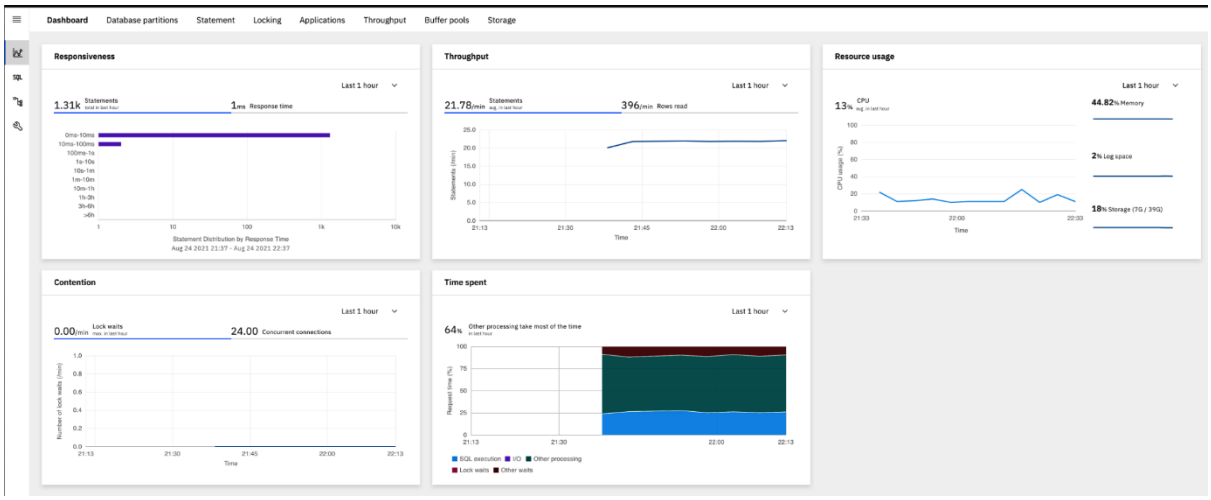
Chapter 7: IBM and Partner Extension Services

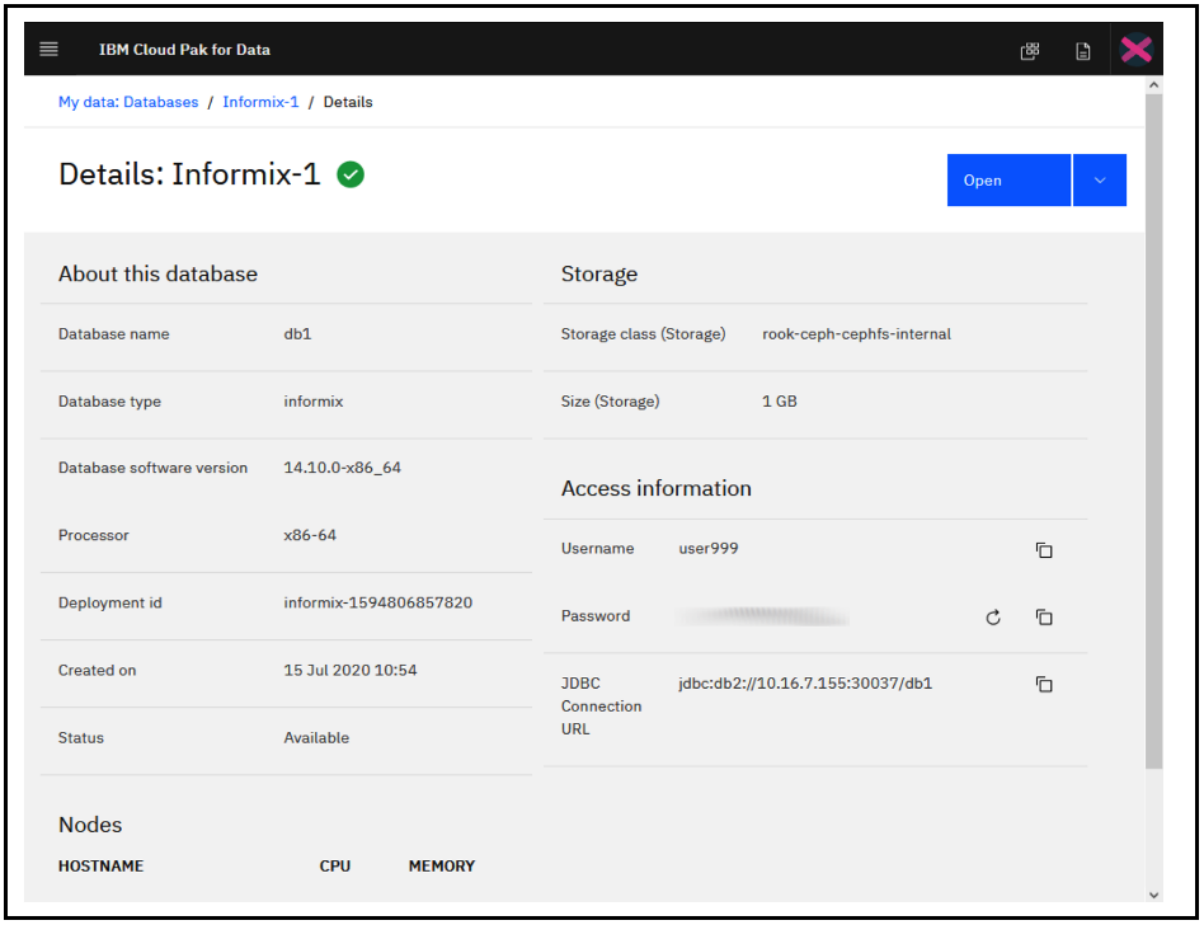
Analogy : iPhone Apps

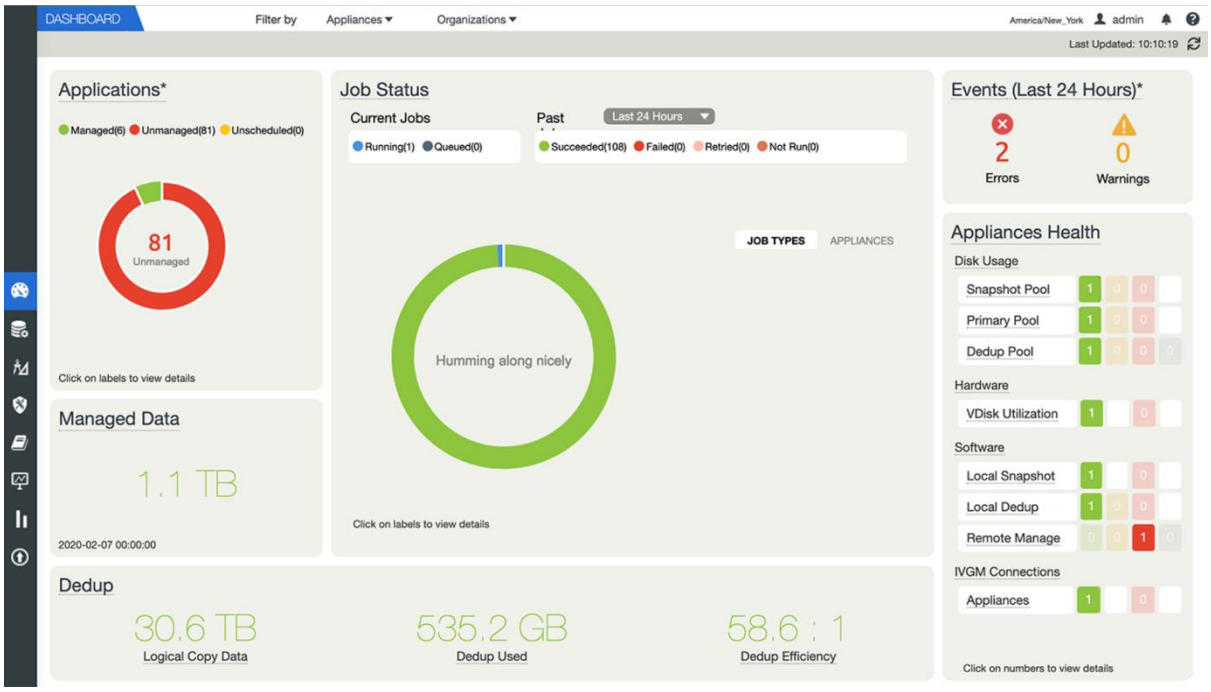
Apps by Apple, Third-Party Apps Sold through the Marketplace and Custom

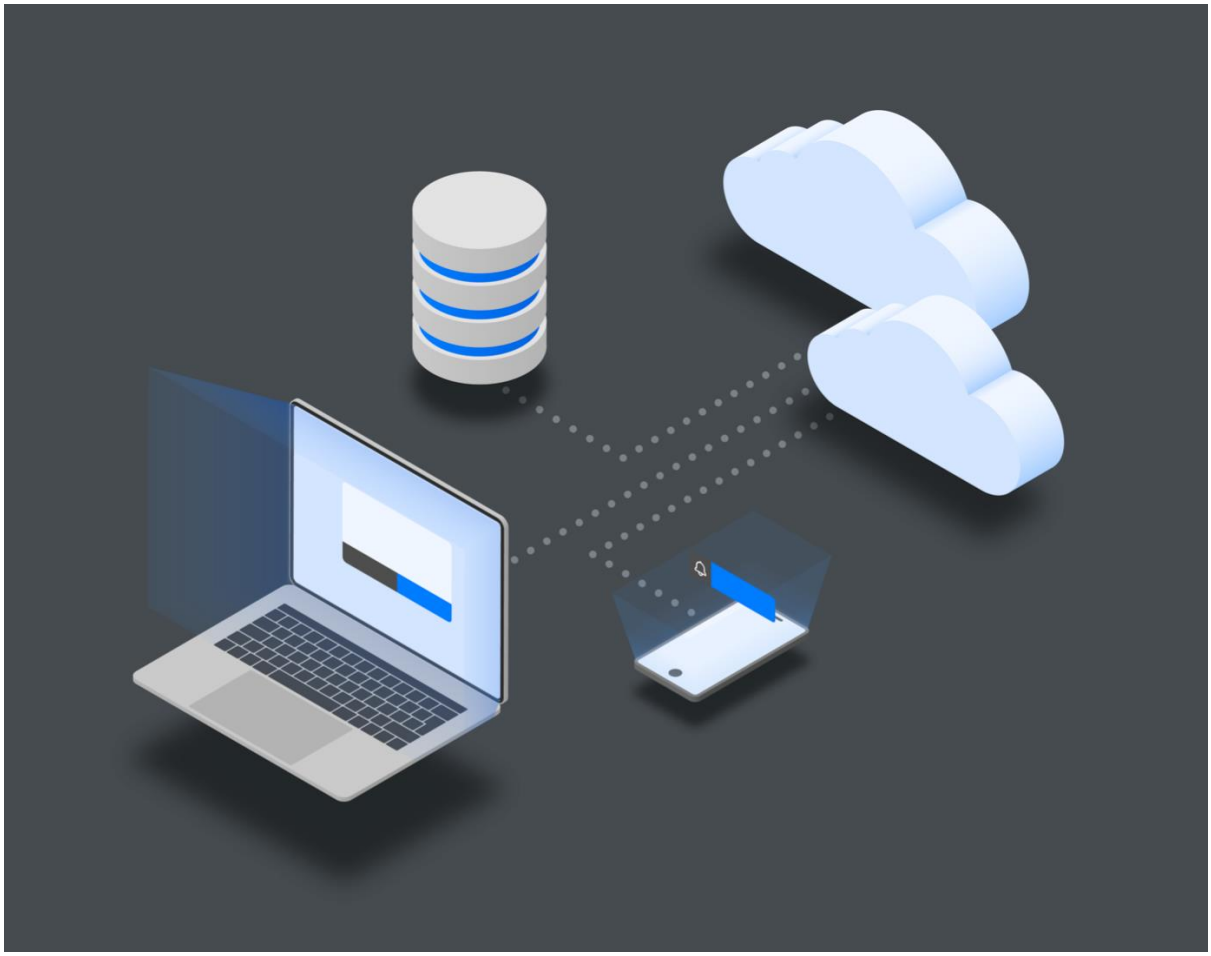


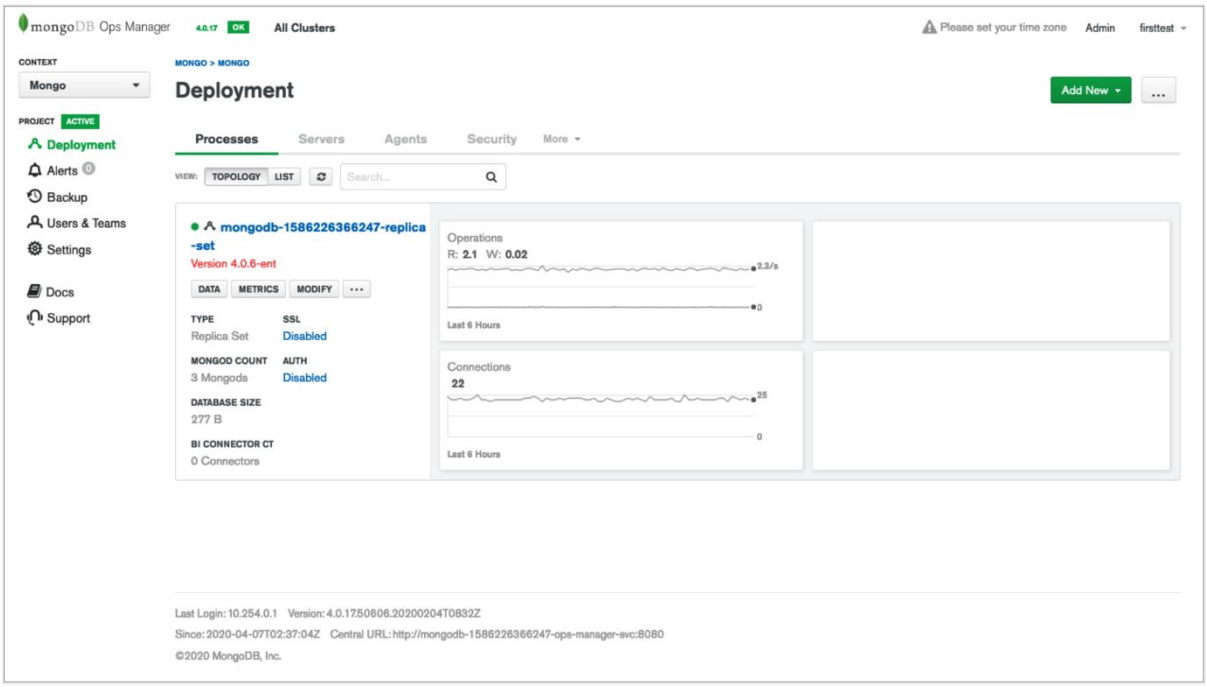
Multi-Cloud

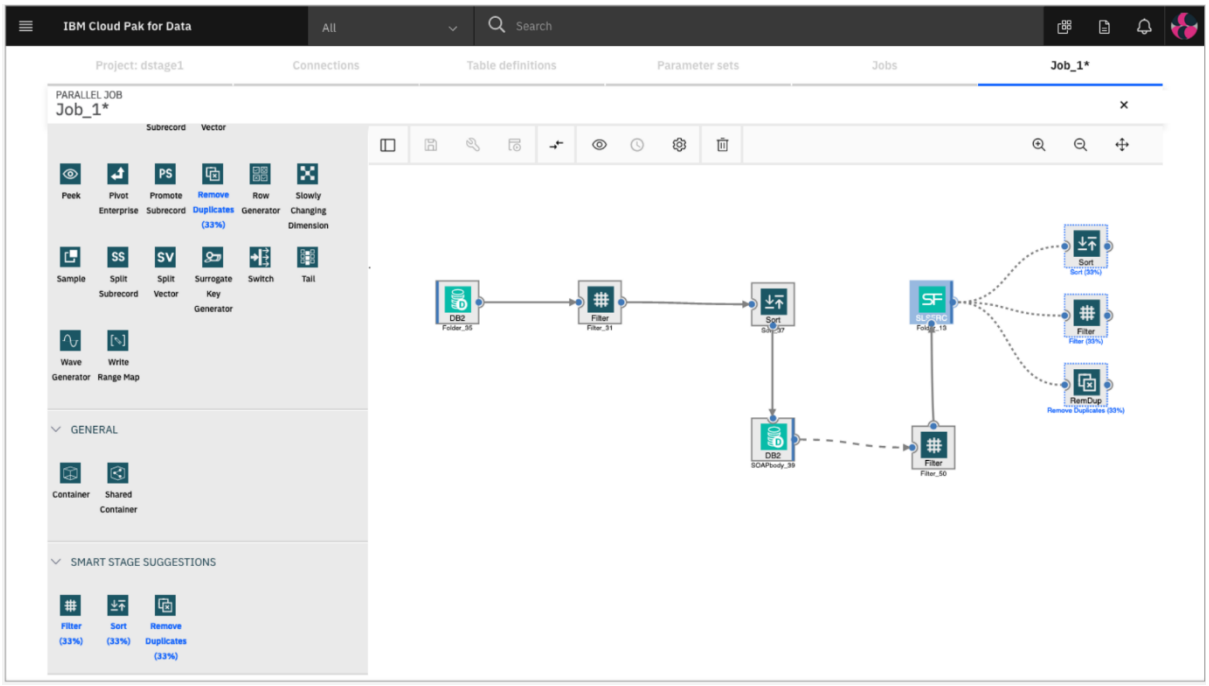


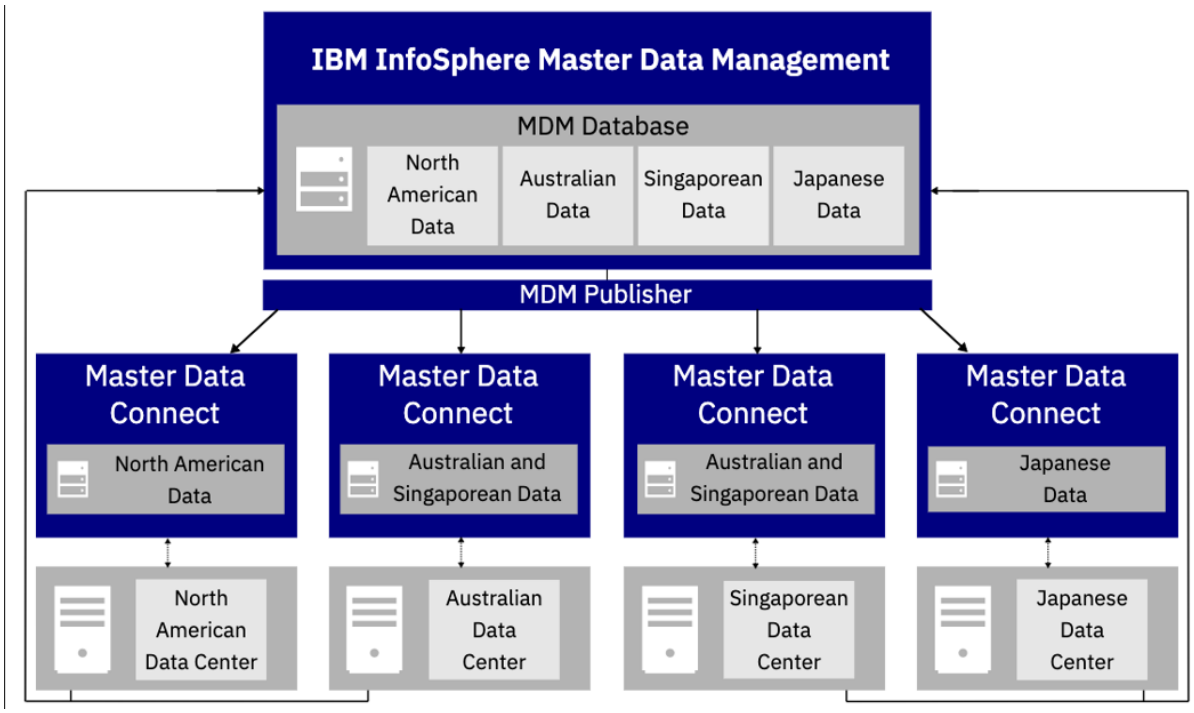

IBM Cloud Pak for Data
My data: Databases / Informix-1 / Details
Details: Informix-1
Open
About this database
Database name db1
Database type informix
Database software version 14.10.0-x86_64
Processor x86-64
Deployment id informix-1594806857820
Created on 15 Jul 2020 10:54
Status Available
Storage
Storage class (Storage) rook-ceph-cephfs-internal
Size (Storage) 1 GB
Access information
Username user999
Password
JDBC Connection URL jdbc:db2://10.16.7.155:30037/db1
Nodes
HOSTNAME CPU MEMORY

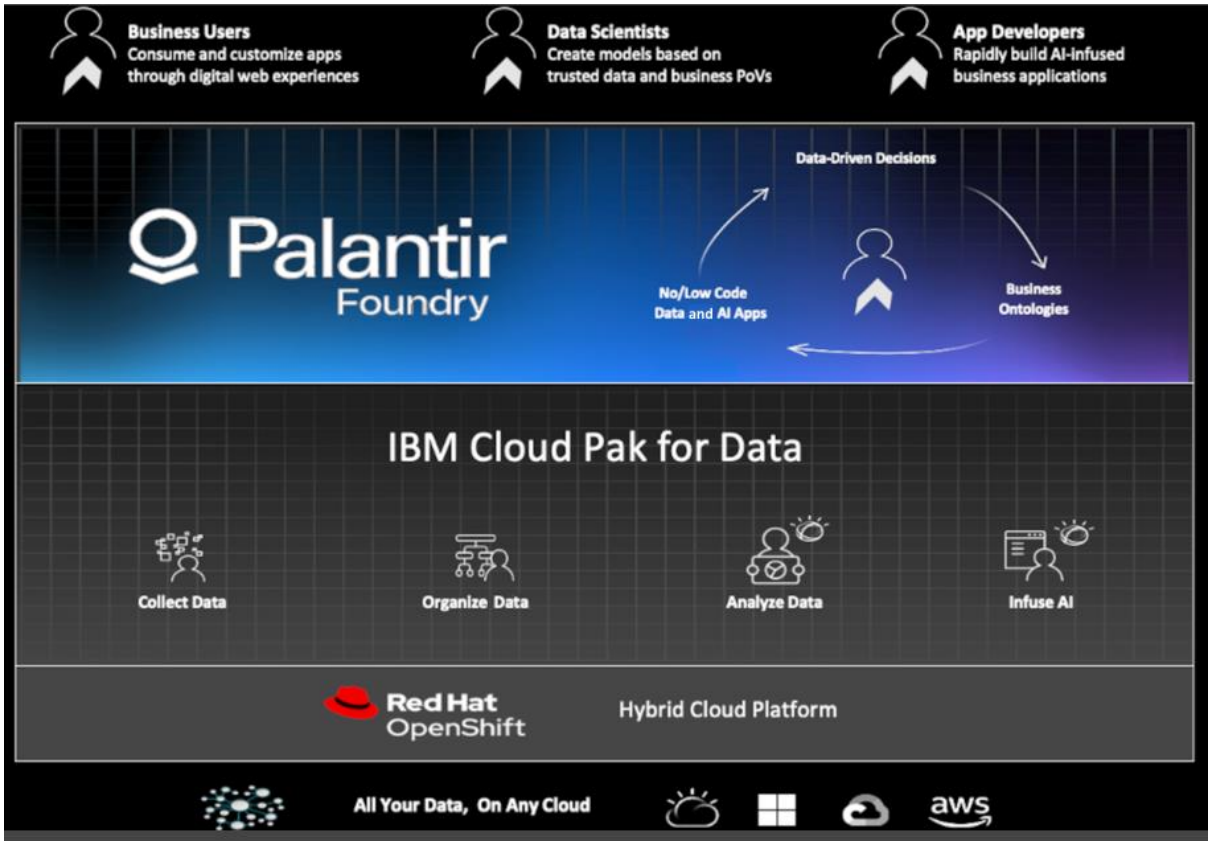

DASHBOARD
Filter by Appliances Organizations
America/New_York admin
Last Updated: 10:10:19
Applications*
Managed(6) Unmanaged(81) Unscheduled(0)
81
Unmanaged
Click on labels to view details
Managed Data
1.1 TB
2020-02-07 00:00:00
Job Status
Current Jobs
Running(1) Queued(0)
Past
Last 24 Hours
Succeeded(108) Failed(0) Retried(0) Not Run(0)
JOB TYPES APPLIANCES
Humming along nicely
Click on labels to view details
Events (Last 24 Hours)*
2
Errors
0
Warnings
Appliances Health
Disk Usage
Snapshot Pool
Primary Pool
Dedup Pool
Hardware
VDisk Utilization
Software
Local Snapshot
Local Dedup
Remote Manage
IVGM Connections
Appliances
Click on numbers to view details
Dedup
30.6 TB
Logical Copy Data
535.2 GB
Dedup Used
58.6 : 1
Dedup Efficiency

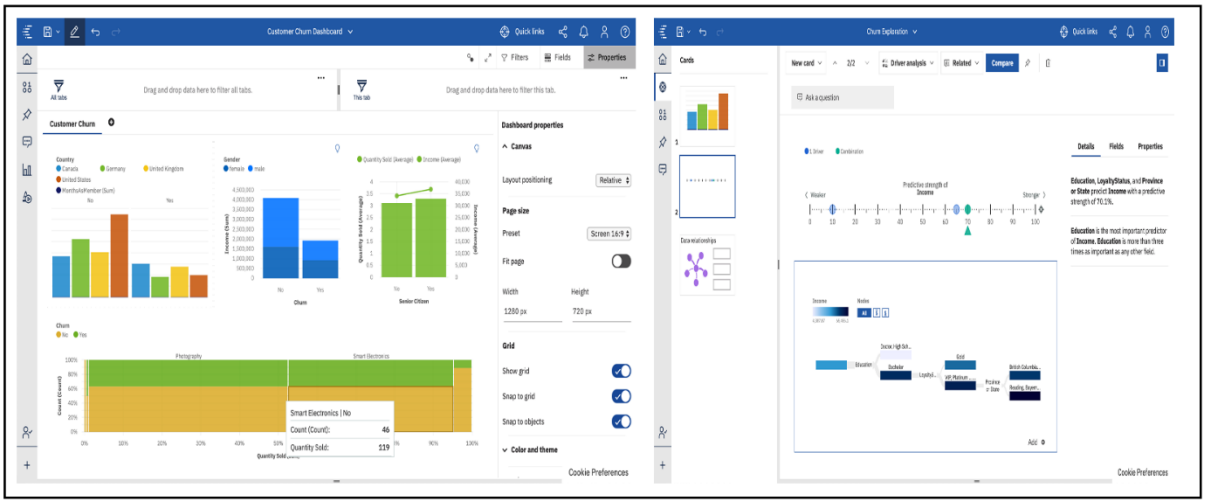


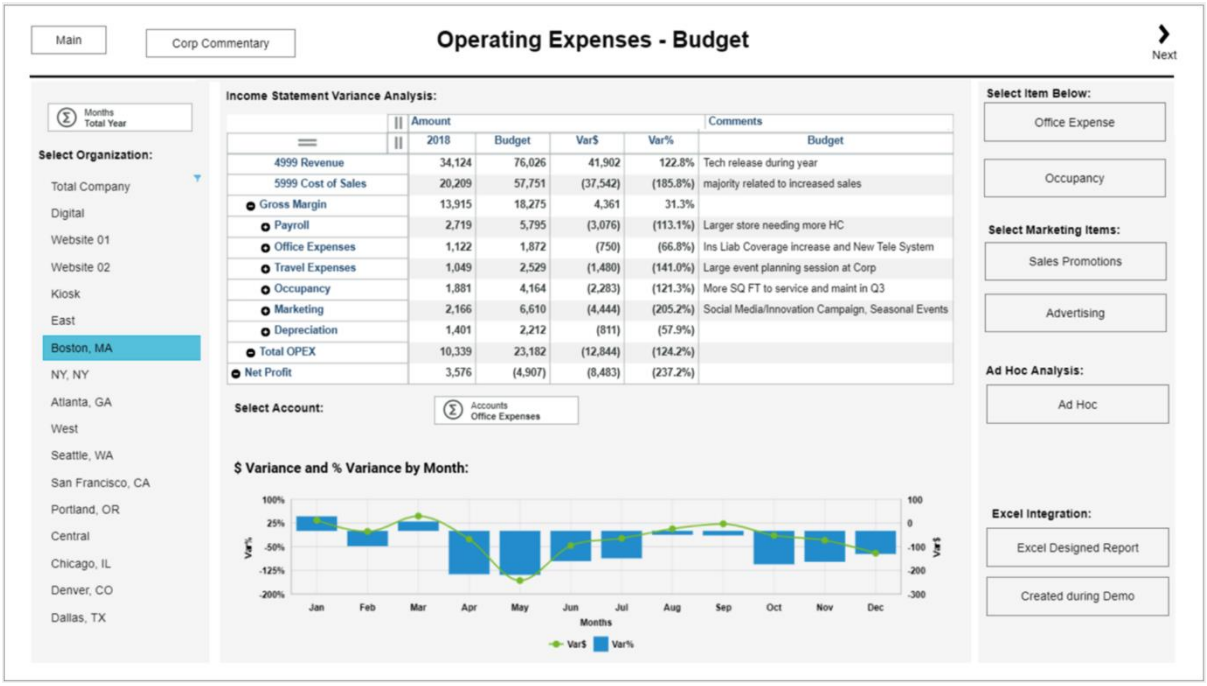

mongoDB Ops Manager
4.0.17
OK
All Clusters
Please set your time zone
Admin
firsttest
CONTEXT
Mongo
PROJECT
ACTIVE
Deployment
Alerts
Backup
Users & Teams
Settings
Docs
Support
MONGO > MONGO
Deployment
Add New
Processes
Servers
Agents
Security
More
VIEW:
TOPOLOGY
LIST
Search...
mongodb-1586226366247-replica-set
Version 4.0.6-ent
DATA
METRICS
MODIFY
TYPE
Replica Set
SSL
Disabled
MONGOD COUNT
3 Mongods
AUTH
Disabled
DATABASE SIZE
277 B
BI CONNECTOR CT
0 Connectors
Operations
R: 2.1 W: 0.02
2.3/s
0
Last 6 Hours
Connections
22
25
0
Last 6 Hours
Last Login: 10.254.0.1 Version: 4.0.17.50606.20200204T0832Z
Since: 2020-04-07T02:37:04Z Central URL: http://mongodb-1586226366247-ops-manager-svc:8080
©2020 MongoDB, Inc.

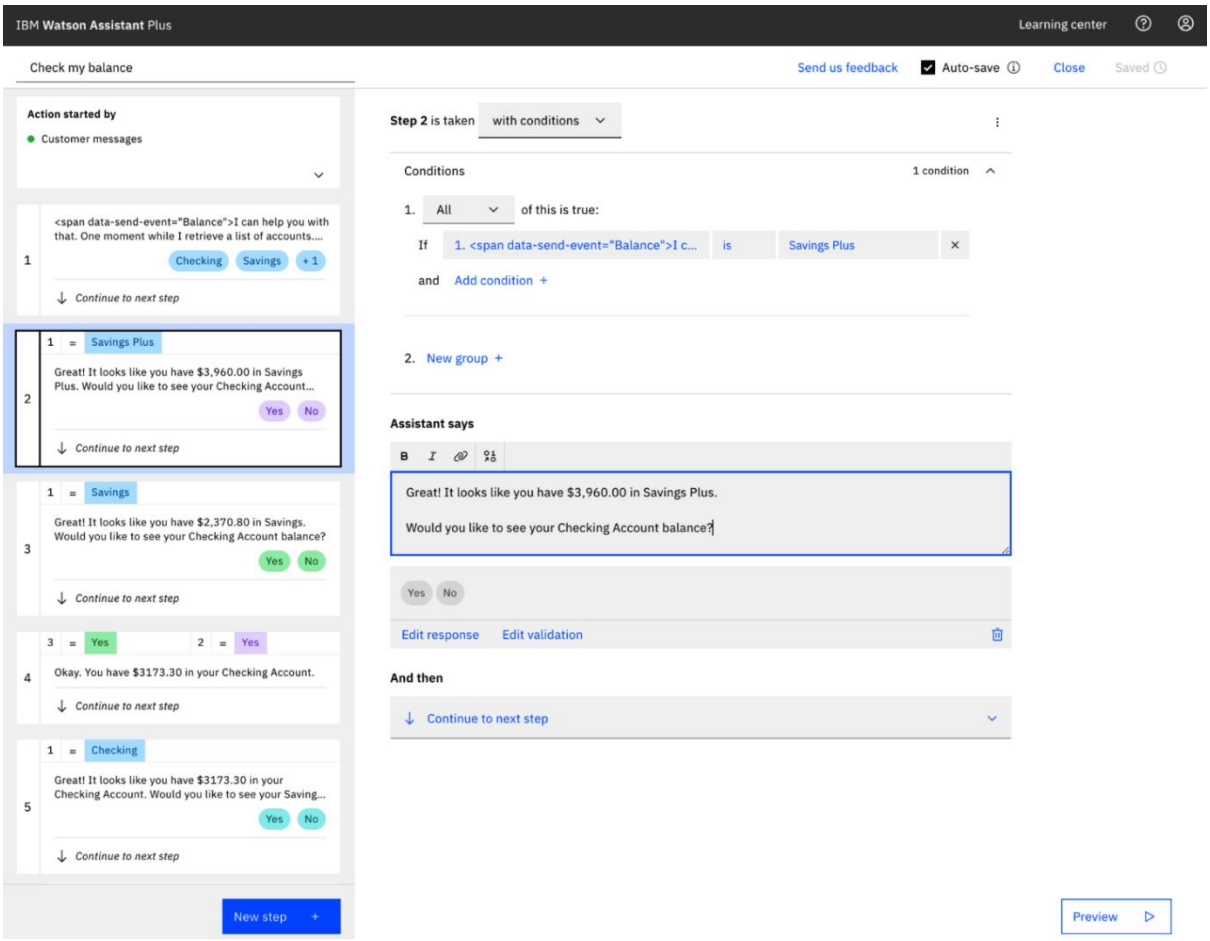

IBM Cloud Pak for Data
All
Search
Project: dstage1
Connections
Table definitions
Parameter sets
Jobs
Job_1*
PARALLEL JOB
Job_1*
Peek
Pivot Enterprise
Promote Subrecord
Remove Duplicates (33%)
Row Generator
Slowly Changing Dimension
Sample
Split Subrecord
Split Vector
Surrogate Key Generator
Switch
Tail
Wave Generator
Write Range Map
GENERAL
Container
Shared Container
SMART STAGE SUGGESTIONS
Filter (33%)
Sort (33%)
Remove Duplicates (33%)

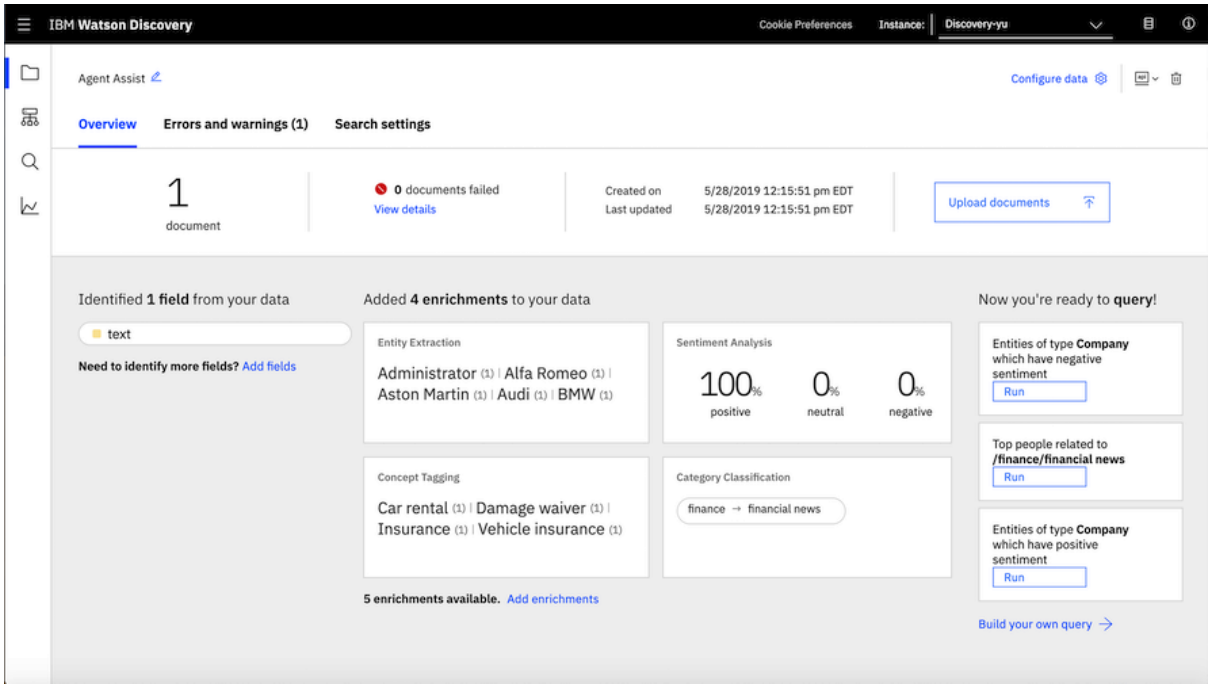

IBM InfoSphere Master Data Management
MDM Database
North American Data
Australian Data
Singaporean Data
Japanese Data
MDM Publisher
Master Data Connect
North American Data
Master Data Connect
Australian and Singaporean Data
Master Data Connect
Australian and Singaporean Data
Master Data Connect
Japanese Data
North American Data Center
Australian Data Center
Singaporean Data Center
Japanese Data Center


Business Users
Consume and customize apps through digital web experiences
Data Scientists
Create models based on trusted data and business PoVs
App Developers
Rapidly build AI-infused business applications
Palantir
Foundry
Data-Driven Decisions
No/Low Code
Data and AI Apps
Business
Ontologies
IBM Cloud Pak for Data
Collect Data
Organize Data
Analyze Data
Infuse AI
Red Hat
OpenShift
Hybrid Cloud Platform
All Your Data, On Any Cloud
aws

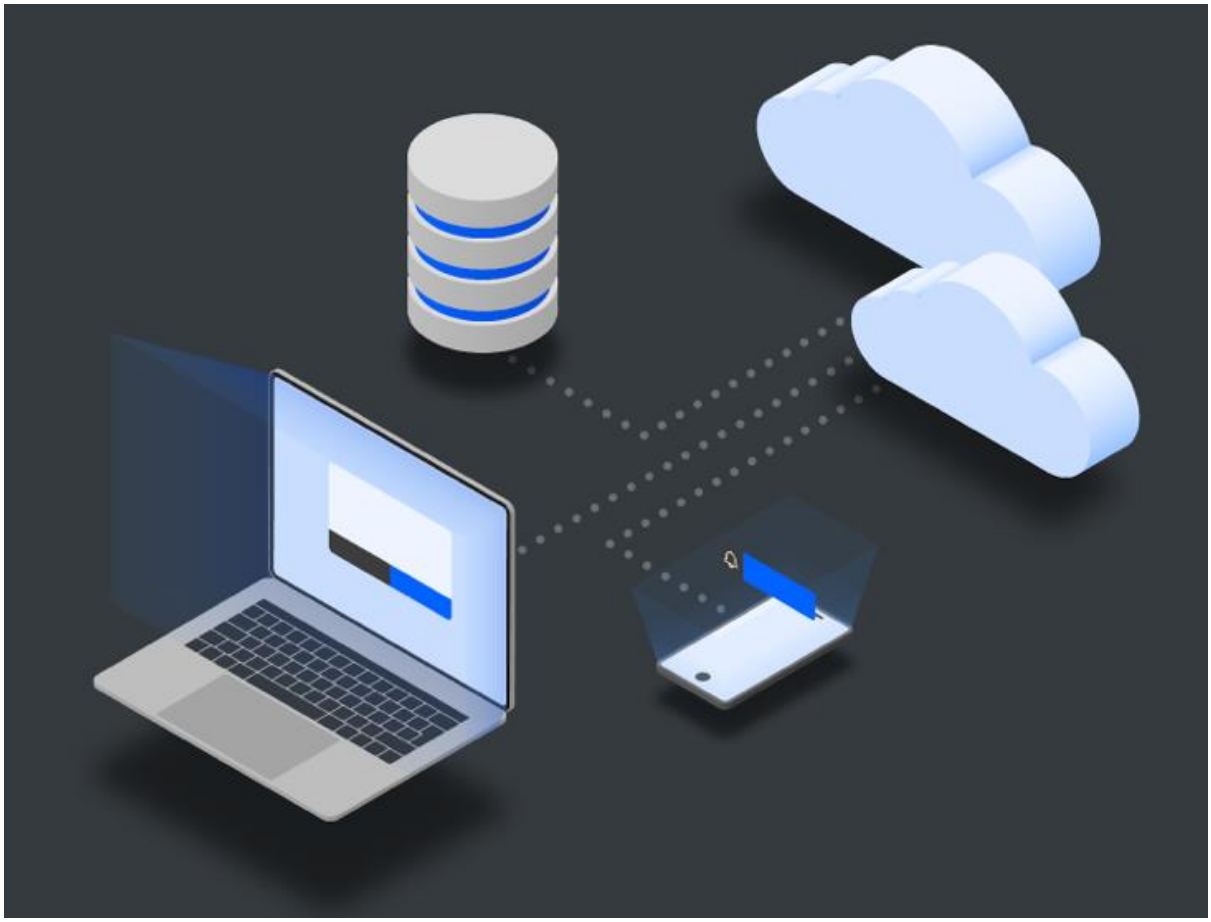

Customer Churn Dashboard
Quick links
Filters
Fields
Properties
Drag and drop data here to filter all tabs.
Drag and drop data here to filter this tab.
Customer Churn
Dashboard properties
Canvas
Layout positioning
Relative
Page size
Preset
Screen 16:9
Fit page
Width
1280 px
Height
720 px
Grid
Show grid
Snap to grid
Snap to objects
Color and theme
Cookie Preferences
Churn Explanation
Cards
New card
Driver analysis
Related
Compare
Ask a question
Details
Fields
Properties
Add
Cookie Preferences

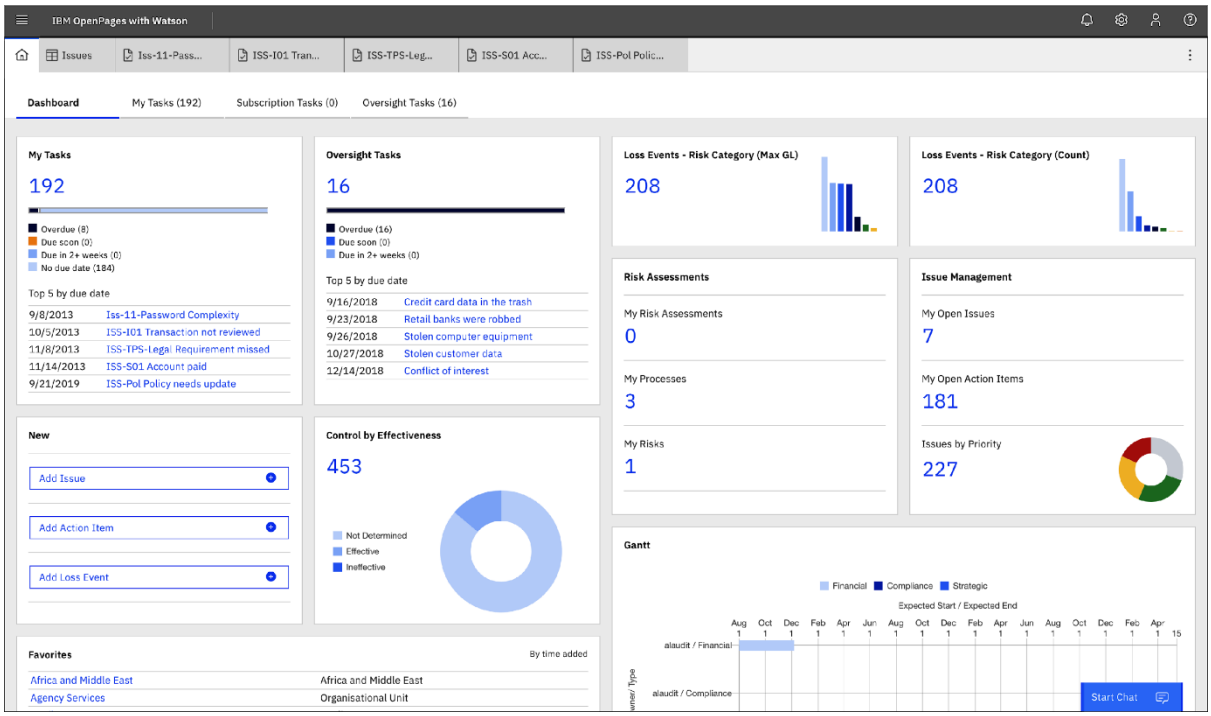

Main
Corp Commentary
Operating Expenses - Budget
Next
Months
Total Year
Select Organization:
Total Company
Digital
Website 01
Website 02
Kiosk
East
Boston, MA
NY, NY
Atlanta, GA
West
Seattle, WA
San Francisco, CA
Portland, OR
Central
Chicago, IL
Denver, CO
Dallas, TX
Income Statement Variance Analysis:
Amount
Comments
2018
Budget
Var$
Var%
Budget
4999 Revenue 34,124 76,026 41,902 122.8% Tech release during year
5999 Cost of Sales 20,209 57,751 (37,542) (185.8%) majority related to increased sales
Gross Margin 13,915 18,275 4,361 31.3%
Payroll 2,719 5,795 (3,076) (113.1%) Larger store needing more HC
Office Expenses 1,122 1,872 (750) (66.8%) Ins Liab Coverage increase and New Tele System
Travel Expenses 1,049 2,529 (1,480) (141.0%) Large event planning session at Corp
Occupancy 1,881 4,164 (2,283) (121.3%) More SQ FT to service and maint in Q3
Marketing 2,166 6,610 (4,444) (205.2%) Social Media/Innovation Campaign, Seasonal Events
Depreciation 1,401 2,212 (811) (57.9%)
Total OPEX 10,339 23,182 (12,844) (124.2%)
Net Profit 3,576 (4,907) (8,483) (237.2%)
Select Account:
Accounts
Office Expenses
$ Variance and % Variance by Month:
100%
25%
-50%
-125%
-200%
Var$
100
0
-100
-200
-300
Var%
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Months
Var$ Var%
Select Item Below:
Office Expense
Occupancy
Select Marketing Items:
Sales Promotions
Advertising
Ad Hoc Analysis:
Ad Hoc
Excel Integration:
Excel Designed Report
Created during Demo

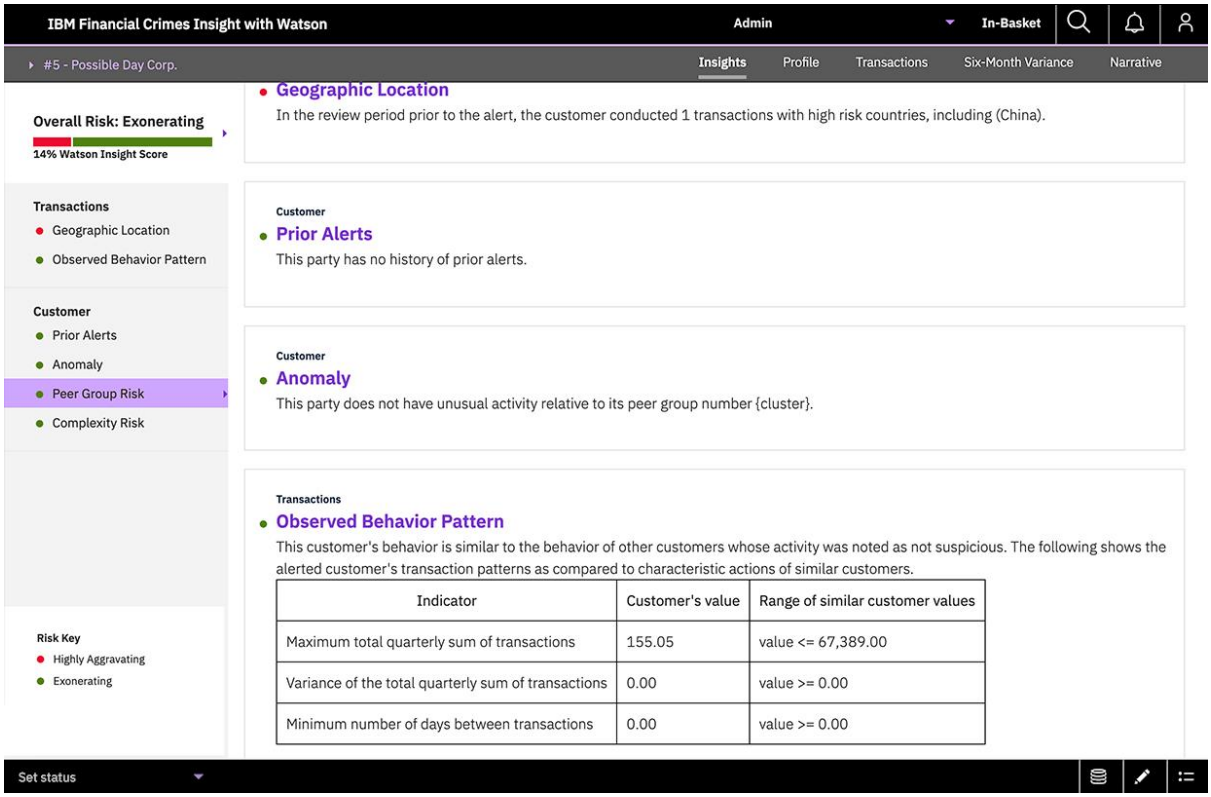

IBM Watson Assistant Plus
Learning center
Check my balance
Send us feedback
Auto-save
Close
Saved
Action started by
Customer messages
1
I can help you with that. One moment while I retrieve a list of accounts....
Checking
Savings
+ 1
Continue to next step
2
1 = Savings Plus
Great! It looks like you have $3,960.00 in Savings Plus. Would you like to see your Checking Account...
Yes
No
Continue to next step
3
1 = Savings
Great! It looks like you have $2,370.80 in Savings. Would you like to see your Checking Account balance?
Yes
No
Continue to next step
4
3 = Yes
2 = Yes
Okay. You have $3173.30 in your Checking Account.
Continue to next step
5
1 = Checking
Great! It looks like you have $3173.30 in your Checking Account. Would you like to see your Saving...
Yes
No
Continue to next step
New step
Step 2 is taken
with conditions
Conditions
1 condition
1. All of this is true:
If
1. I c...
is
Savings Plus
and
Add condition +
2. New group +
Assistant says
Great! It looks like you have $3,960.00 in Savings Plus.
Would you like to see your Checking Account balance?
Yes
No
Edit response
Edit validation
And then
Continue to next step
Preview

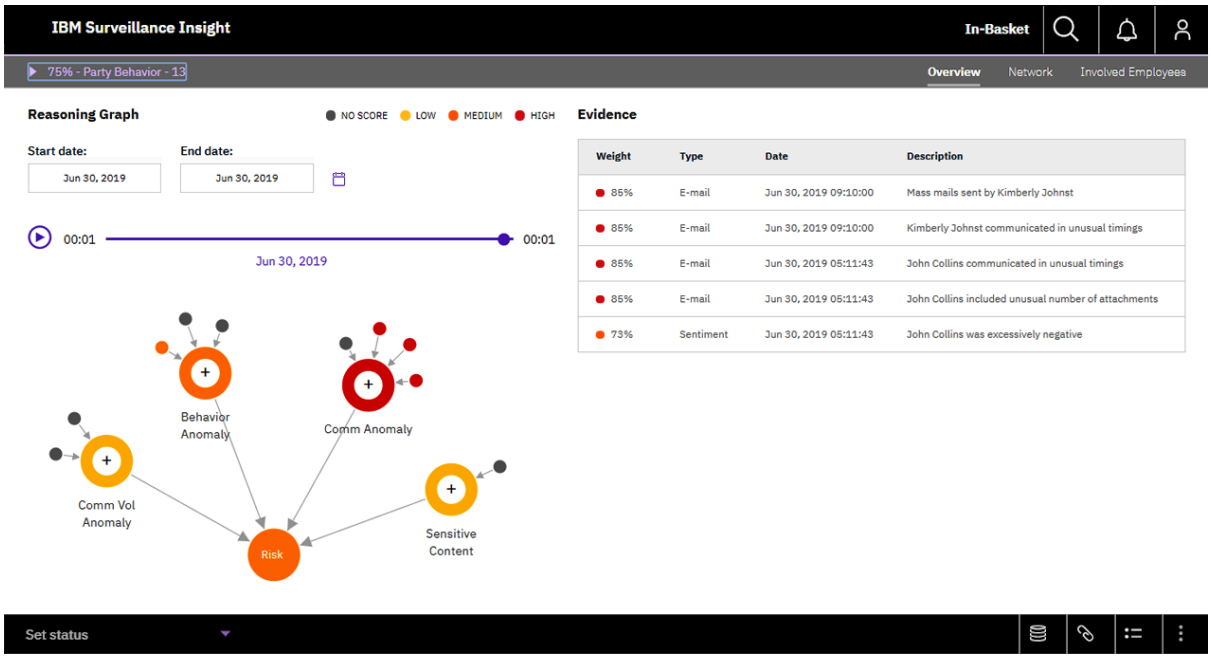

IBM Watson Discovery
Cookie Preferences
Instance:
Discovery-yu
Agent Assist
Configure data
Overview
Errors and warnings (1)
Search settings
1
document
0 documents failed
View details
Created on
5/28/2019 12:15:51 pm EDT
Last updated
5/28/2019 12:15:51 pm EDT
Upload documents
Identified 1 field from your data
text
Need to identify more fields? Add fields
Added 4 enrichments to your data
Entity Extraction
Administrator (1) | Alfa Romeo (1) |
Aston Martin (1) | Audi (1) | BMW (1)
Sentiment Analysis
100% positive
0% neutral
0% negative
Concept Tagging
Car rental (1) | Damage waiver (1) |
Insurance (1) | Vehicle insurance (1)
Category Classification
finance → financial news
5 enrichments available. Add enrichments
Now you're ready to query!
Entities of type Company which have negative sentiment
Run
Top people related to /finance/financial news
Run
Entities of type Company which have positive sentiment
Run
Build your own query

IBM Watson Knowledge Studio
reggie.raye@ibm.com
Back to Workspaces
Assets
Documents
Entity Types
Relation Types
Dictionaries
Rule-based Model
Rules
Versions
Machine Learning Model
Pre-annotation
Annotation Tasks
Performance
Versions
Settings
Help
Back to Task
Open document list
Mention
Relation
Coreference
View Details
Replace
Concordance
Attribute View
View Guidelines
In Progress
Save
race.txt
1 Freddy Fraye, captain of the Harvard University squad, hit the tarmac with 10K to go in the Boston Beanpot Road Race.
2 Though his Cannondale's derailleur was damaged, he managed to regain the wheel of the Westpoint riders who'd been his breakaway companions.
3 Penn State, which had come with a stacked team, took up the chase in the peloton.
4 The gap to the break fell to 3 seconds with 100m to go, but the Penn boys ran out of road, allowing Freddy to cross the line first.
Alphab...
18pt
1
Entity
Mention
Type
Subtype
Role
bicycle
crash
distance
race
rider
team
win

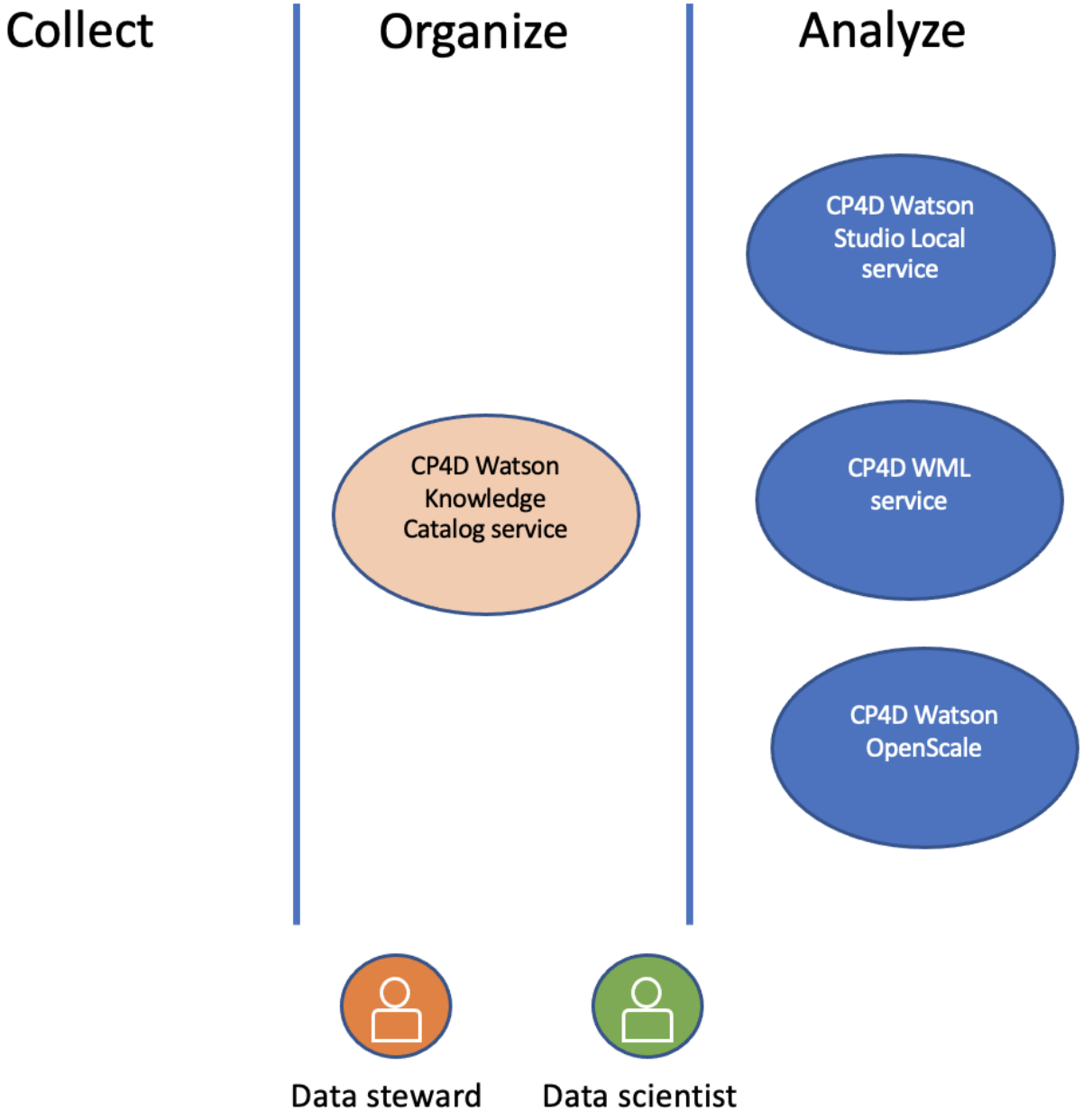


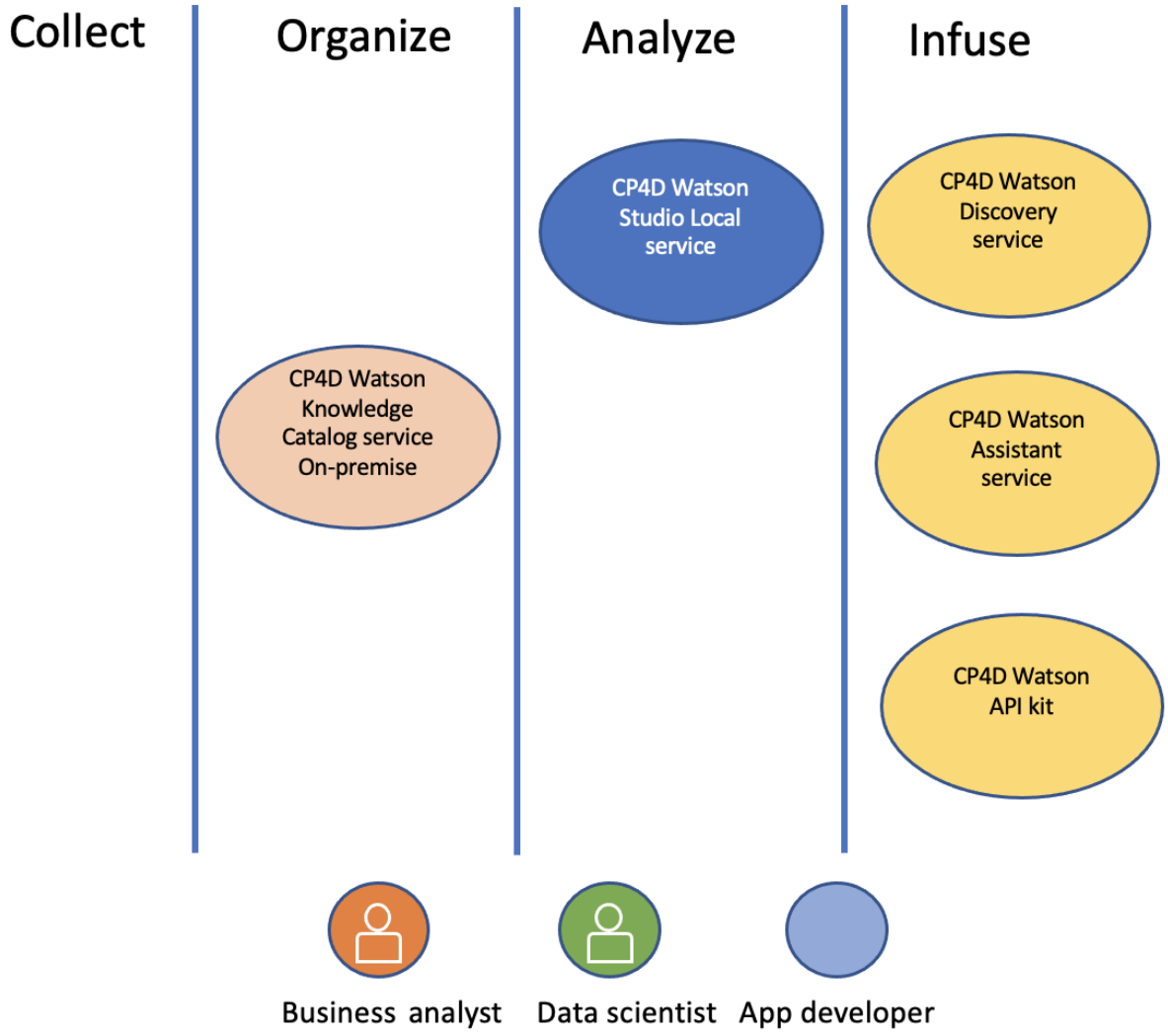

IBM OpenPages with Watson
Issues
Iss-11-Pass...
ISS-I01 Tran...
ISS-TPS-Leg...
ISS-S01 Acc...
ISS-Pol Polic...
Dashboard
My Tasks (192)
Subscription Tasks (0)
Oversight Tasks (16)
My Tasks
192
Overdue (8)
Due soon (0)
Due in 2+ weeks (0)
No due date (184)
Top 5 by due date
9/8/2013 Iss-11-Password Complexity
10/5/2013 ISS-I01 Transaction not reviewed
11/8/2013 ISS-TPS-Legal Requirement missed
11/14/2013 ISS-S01 Account paid
9/21/2019 ISS-Pol Policy needs update
Oversight Tasks
16
Overdue (16)
Due soon (0)
Due in 2+ weeks (0)
Top 5 by due date
9/16/2018 Credit card data in the trash
9/23/2018 Retail banks were robbed
9/26/2018 Stolen computer equipment
10/27/2018 Stolen customer data
12/14/2018 Conflict of interest
Loss Events - Risk Category (Max GL)
208
Loss Events - Risk Category (Count)
208
Risk Assessments
My Risk Assessments
0
My Processes
3
My Risks
1
Issue Management
My Open Issues
7
My Open Action Items
181
Issues by Priority
227
New
Add Issue
Add Action Item
Add Loss Event
Control by Effectiveness
453
Not Determined
Effective
Ineffective
Gantt
Financial
Compliance
Strategic
Expected Start / Expected End
alaudit / Financial
alaudit / Compliance
Start Chat
Favorites
By time added
Africa and Middle East
Africa and Middle East
Agency Services
Organisational Unit

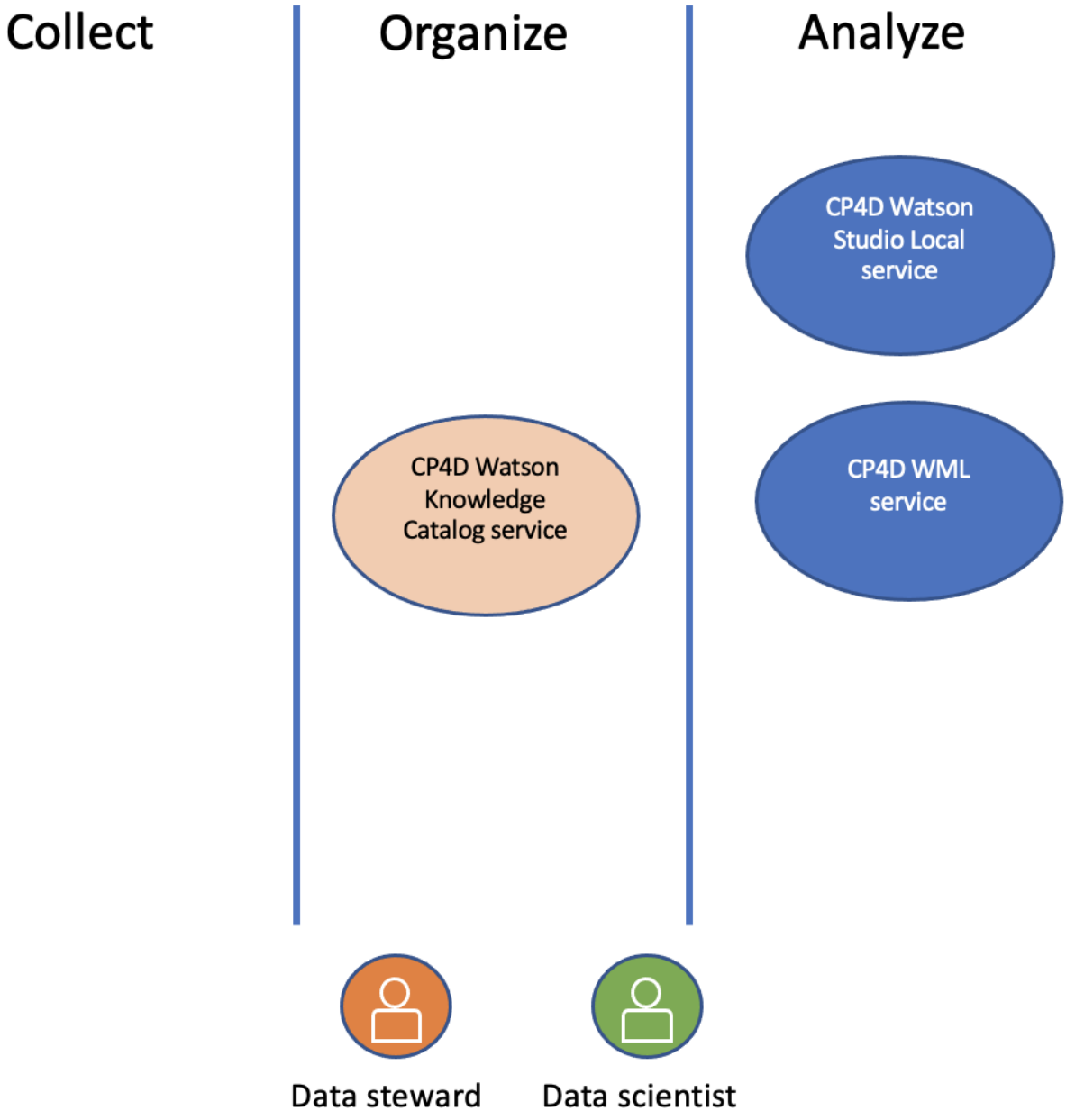

IBM Financial Crimes Insight with Watson
Admin
In-Basket
#5 - Possible Day Corp.
Insights
Profile
Transactions
Six-Month Variance
Narrative
Overall Risk: Exonerating
14% Watson Insight Score
Transactions
Geographic Location
Observed Behavior Pattern
Customer
Prior Alerts
Anomaly
Peer Group Risk
Complexity Risk
Risk Key
Highly Aggravating
Exonerating
Geographic Location
In the review period prior to the alert, the customer conducted 1 transactions with high risk countries, including (China).
Customer
Prior Alerts
This party has no history of prior alerts.
Customer
Anomaly
This party does not have unusual activity relative to its peer group number {cluster}.
Transactions
Observed Behavior Pattern
This customer's behavior is similar to the behavior of other customers whose activity was noted as not suspicious. The following shows the alerted customer's transaction patterns as compared to characteristic actions of similar customers.
Indicator
Customer's value
Range of similar customer values
Maximum total quarterly sum of transactions
155.05
value <= 67,389.00
Variance of the total quarterly sum of transactions
0.00
value >= 0.00
Minimum number of days between transactions
0.00
value >= 0.00
Set status

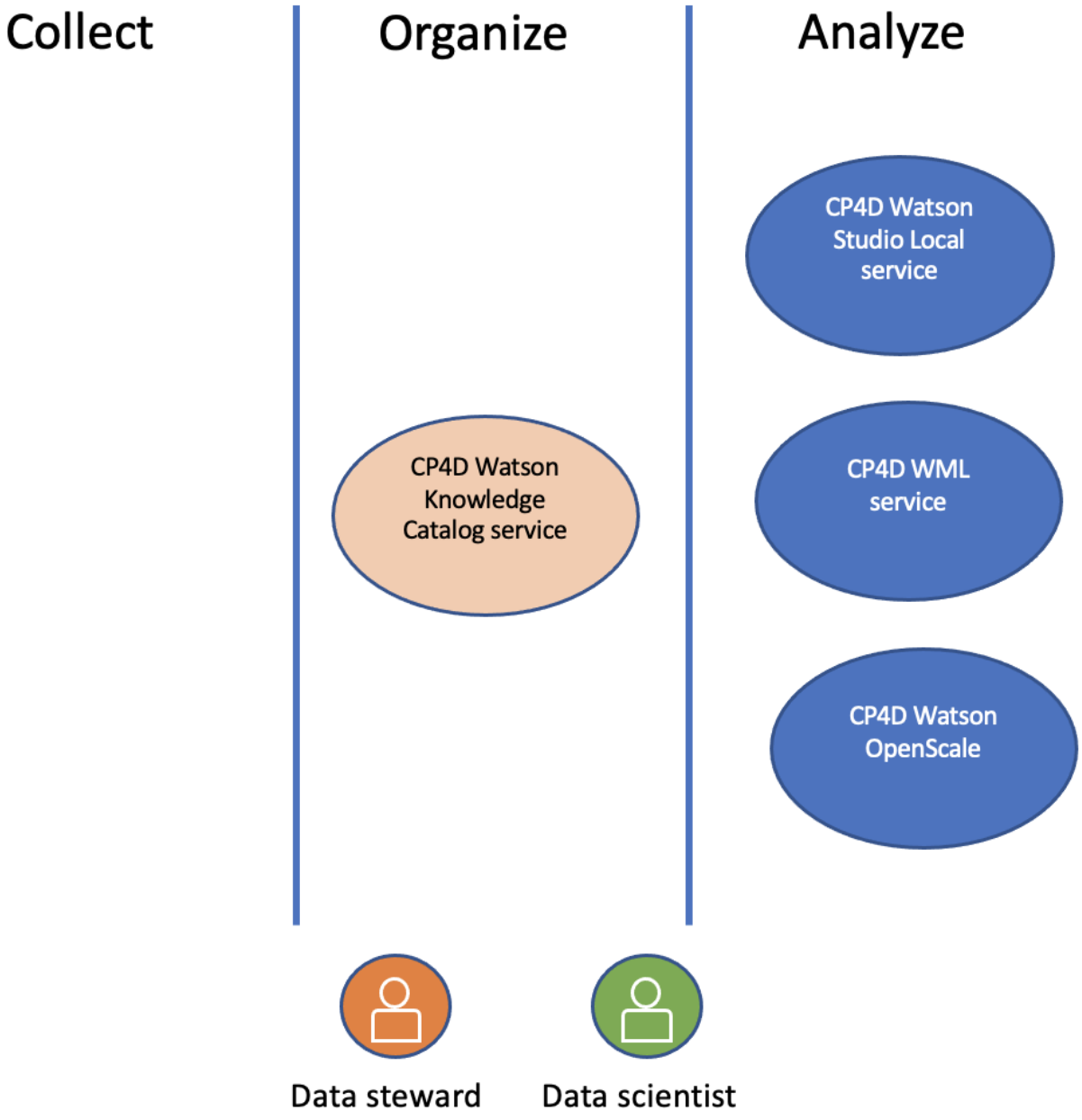

IBM Surveillance Insight
In-Basket
75% - Party Behavior - 13
Overview
Network
Involved Employees
Reasoning Graph
NO SCORE
LOW
MEDIUM
HIGH
Start date:
End date:
Jun 30, 2019
Jun 30, 2019
00:01
00:01
Jun 30, 2019
Behavior Anomaly
Comm Anomaly
Comm Vol Anomaly
Sensitive Content
Risk
Evidence
Weight
Type
Date
Description
85%
E-mail
Jun 30, 2019 09:10:00
Mass mails sent by Kimberly Johnst
85%
E-mail
Jun 30, 2019 09:10:00
Kimberly Johnst communicated in unusual timings
85%
E-mail
Jun 30, 2019 05:11:43
John Collins communicated in unusual timings
85%
E-mail
Jun 30, 2019 05:11:43
John Collins included unusual number of attachments
73%
Sentiment
Jun 30, 2019 05:11:43
John Collins was excessively negative
Set status

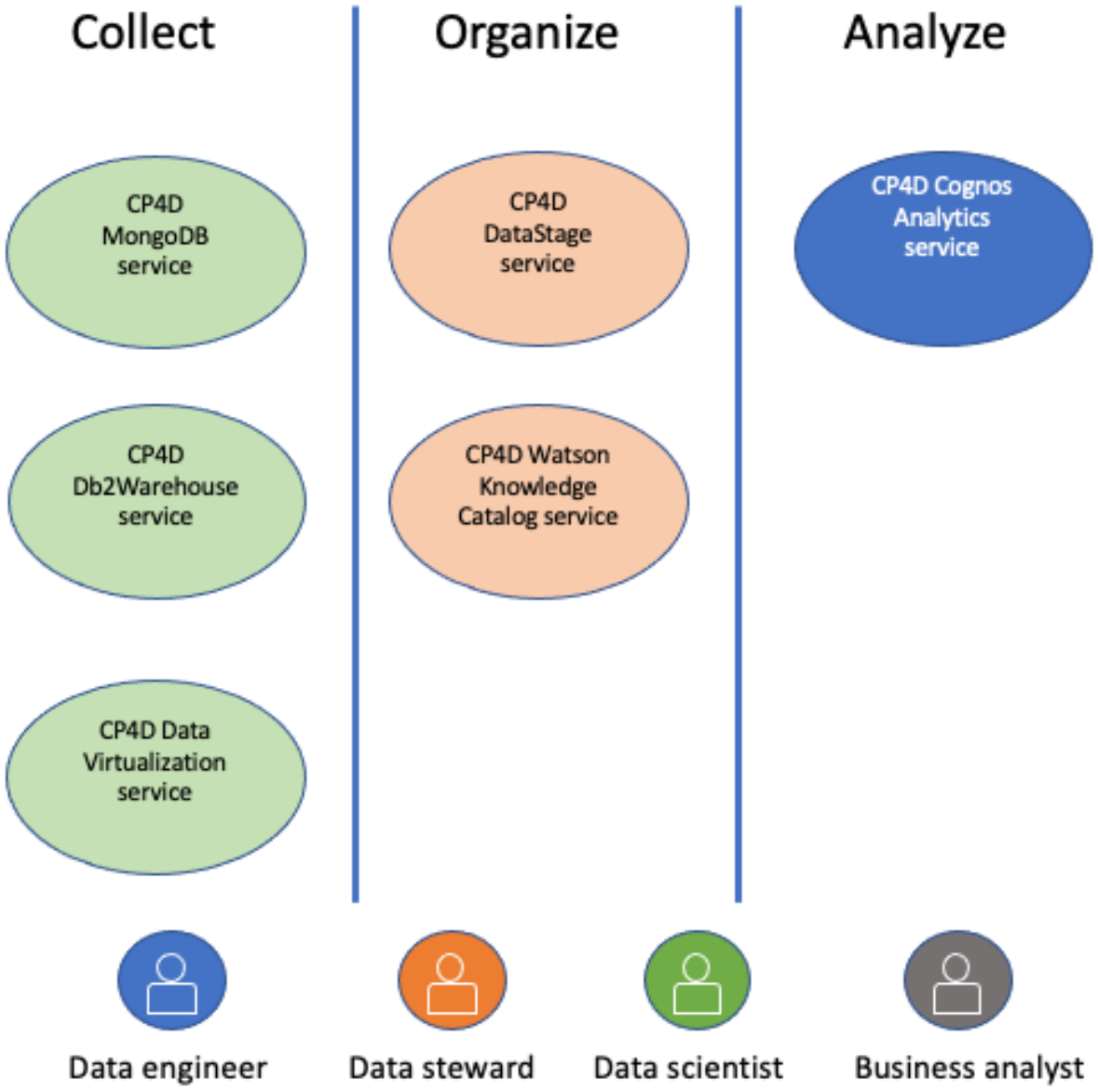
Chapter 8: Customer Use Cases

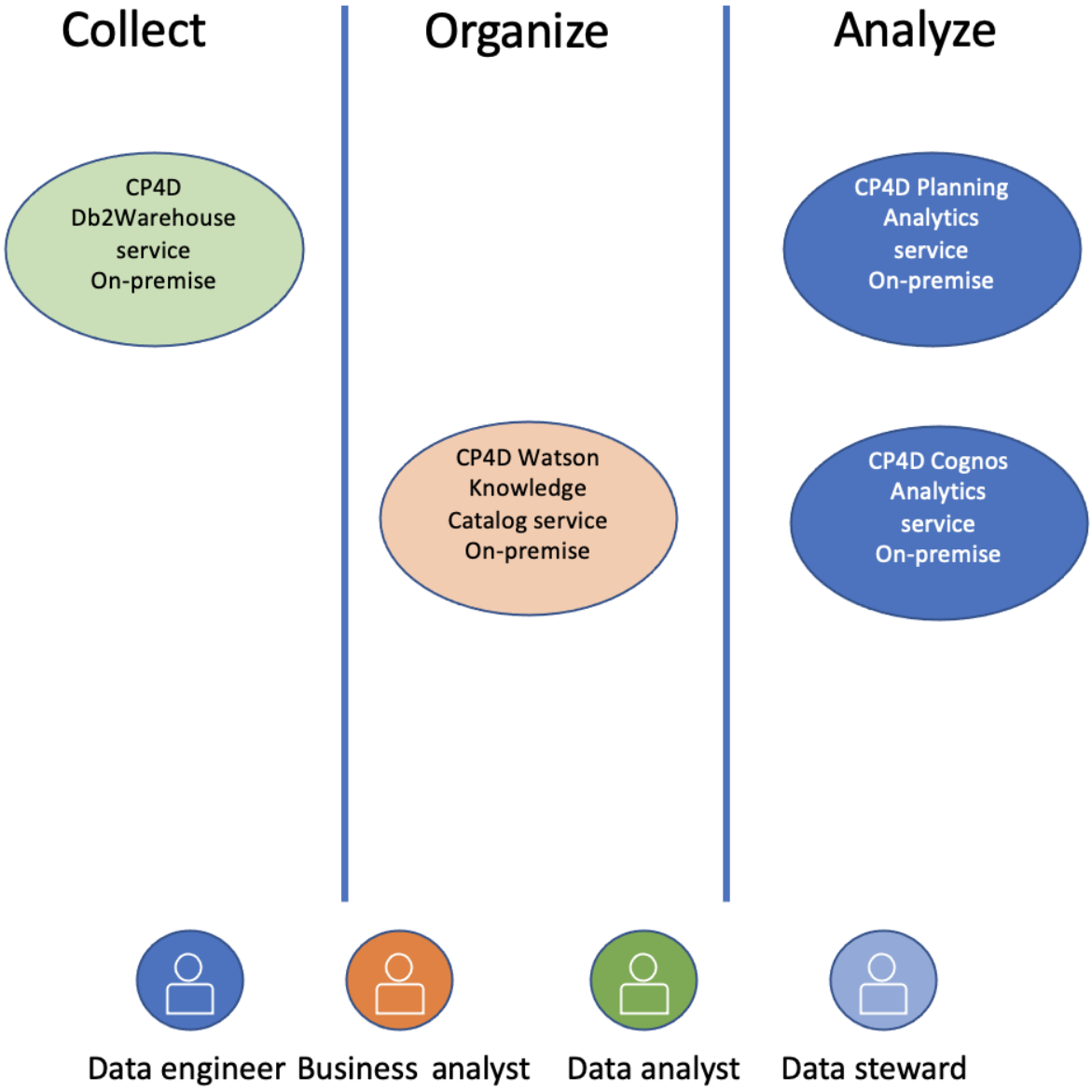



Collect
Organize
Analyze
Infuse
CP4D Watson Studio Local service
CP4D Watson Discovery service
CP4D Watson Knowledge Catalog service On-premise
CP4D Watson Assistant service
CP4D Watson API kit
Business analyst
Data scientist
App developer

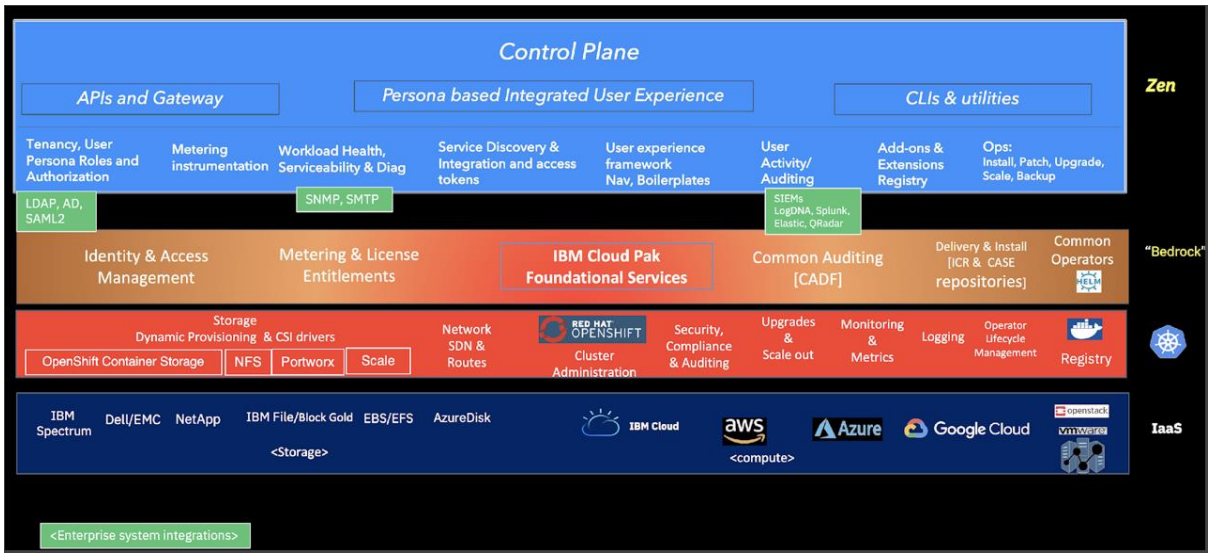

Collect
Organize
Analyze
CP4D Watson Studio Local service
CP4D Watson Knowledge Catalog service
CP4D WML service
Data steward
Data scientist

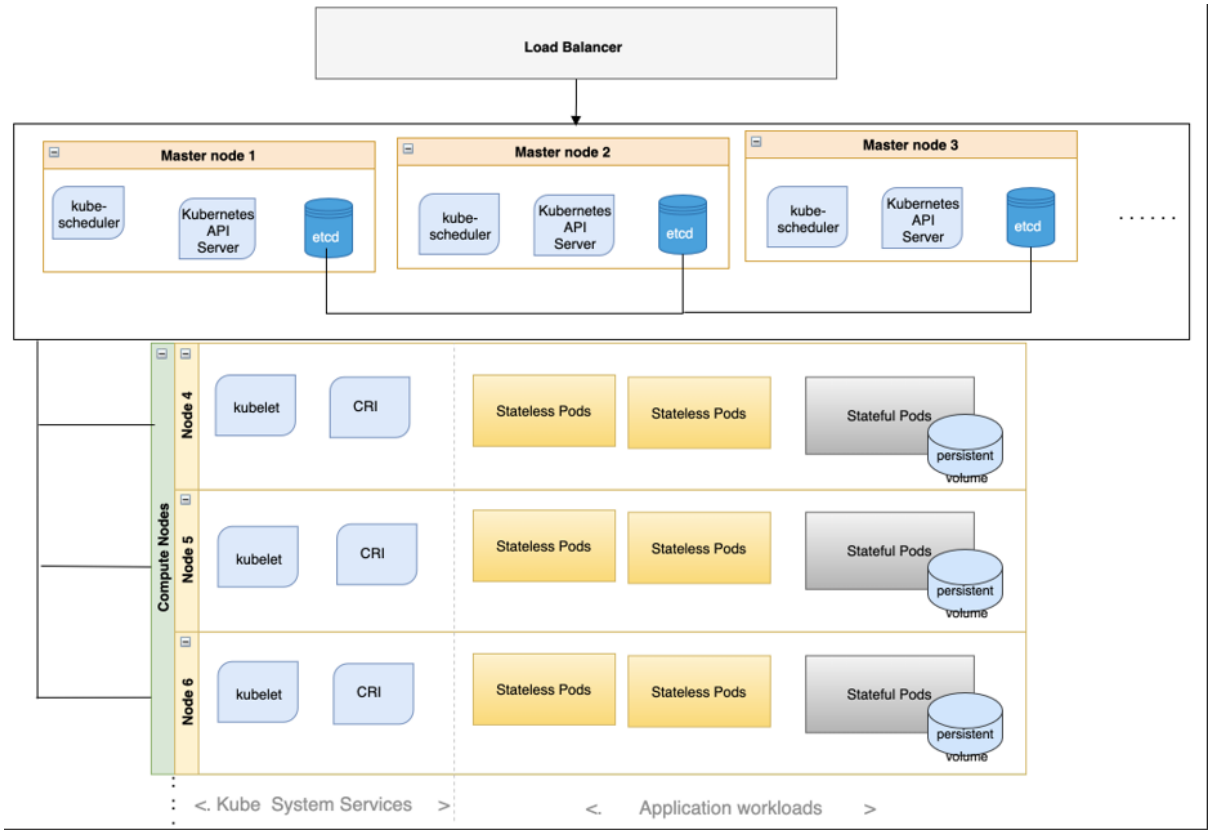

Collect
Organize
Analyze
CP4D Watson Studio Local service
CP4D WML service
CP4D Watson OpenScale
CP4D Watson Knowledge Catalog service
Data steward
Data scientist

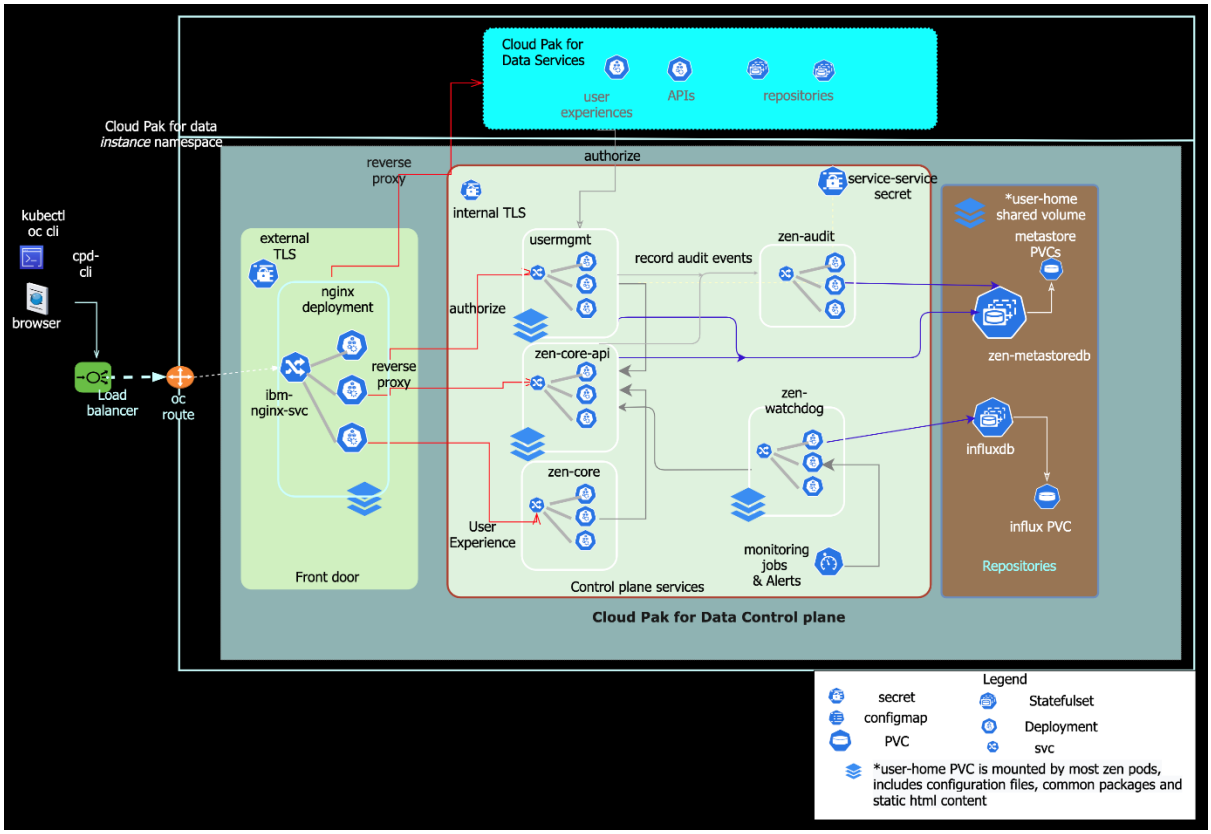

Collect
Organize
Analyze
CP4D MongoDB service
CP4D Db2Warehouse service
CP4D Data Virtualization service
CP4D DataStage service
CP4D Watson Knowledge Catalog service
CP4D Cognos Analytics service
Data engineer
Data steward
Data scientist
Business analyst

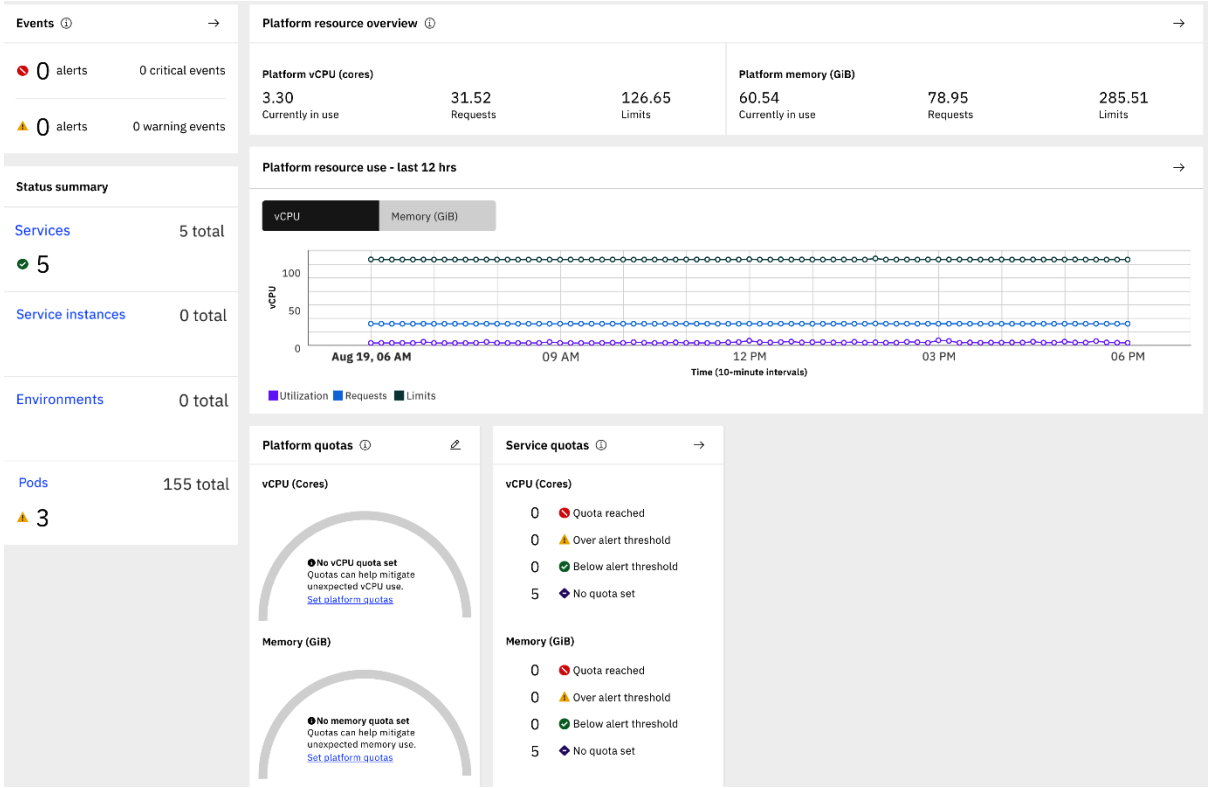

Collect
Organize
Analyze
CP4D Db2Warehouse service On-premise
CP4D Watson Knowledge Catalog service On-premise
CP4D Planning Analytics service On-premise
CP4D Cognos Analytics service On-premise
Data engineer
Business analyst
Data analyst
Data steward

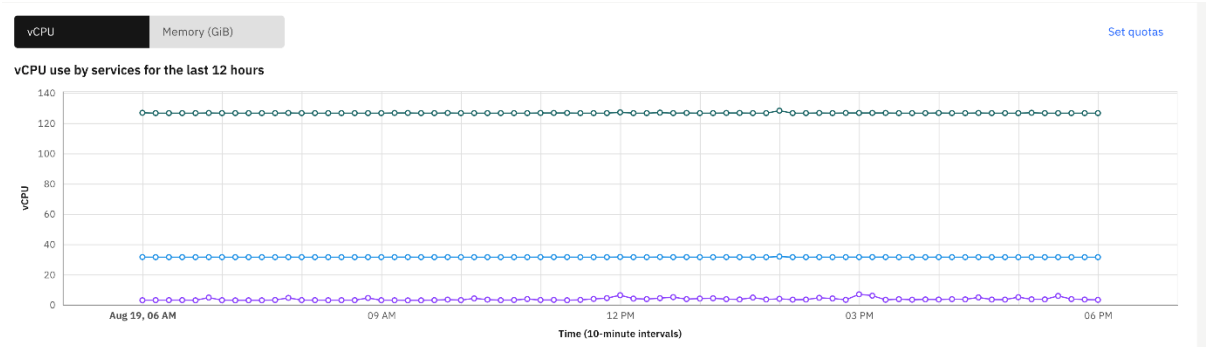
Chapter 9: Technical Overview, Management, and Administration






Cloud Pak for Data Services
user experiences
APIs
repositories
Cloud Pak for data instance namespace
reverse proxy
authorize
service-service secret
internal TLS
*user-home shared volume
metastore PVCs
kubectl oc cli
cpd-cli
browser
external TLS
usermgmt
record audit events
zen-audit
nginx deployment
authorize
zen-metastoredb
zen-core-api
reverse proxy
ibm-nginx-svc
zen-watchdog
Load balancer
oc route
influxdb
zen-core
influx PVC
User Experience
monitoring jobs & Alerts
Repositories
Front door
Control plane services
Cloud Pak for Data Control plane
Legend
secret
configmap
PVC
Statefulset
Deployment
svc
*user-home PVC is mounted by most zen pods, includes configuration files, common packages and static html content


Events
0 alerts
0 critical events
0 alerts
0 warning events
Status summary
Services
5 total
5
Service instances
0 total
Environments
0 total
Pods
155 total
3
Platform resource overview
Platform vCPU (cores)
3.30
Currently in use
31.52
Requests
126.65
Limits
Platform memory (GiB)
60.54
Currently in use
78.95
Requests
285.51
Limits
Platform resource use - last 12 hrs
vCPU
Memory (GiB)
100
50
0
vCPU
Aug 19, 06 AM
09 AM
12 PM
03 PM
06 PM
Time (10-minute intervals)
Utilization
Requests
Limits
Platform quotas
vCPU (Cores)
No vCPU quota set
Quotas can help mitigate unexpected vCPU use.
Set platform quotas
Memory (GiB)
No memory quota set
Quotas can help mitigate unexpected memory use.
Set platform quotas
Service quotas
vCPU (Cores)
0 Quota reached
0 Over alert threshold
0 Below alert threshold
5 No quota set
Memory (GiB)
0 Quota reached
0 Over alert threshold
0 Below alert threshold
5 No quota set



Detailed breakdown by service for: 8/19/2021 6:06 PM (Latest)

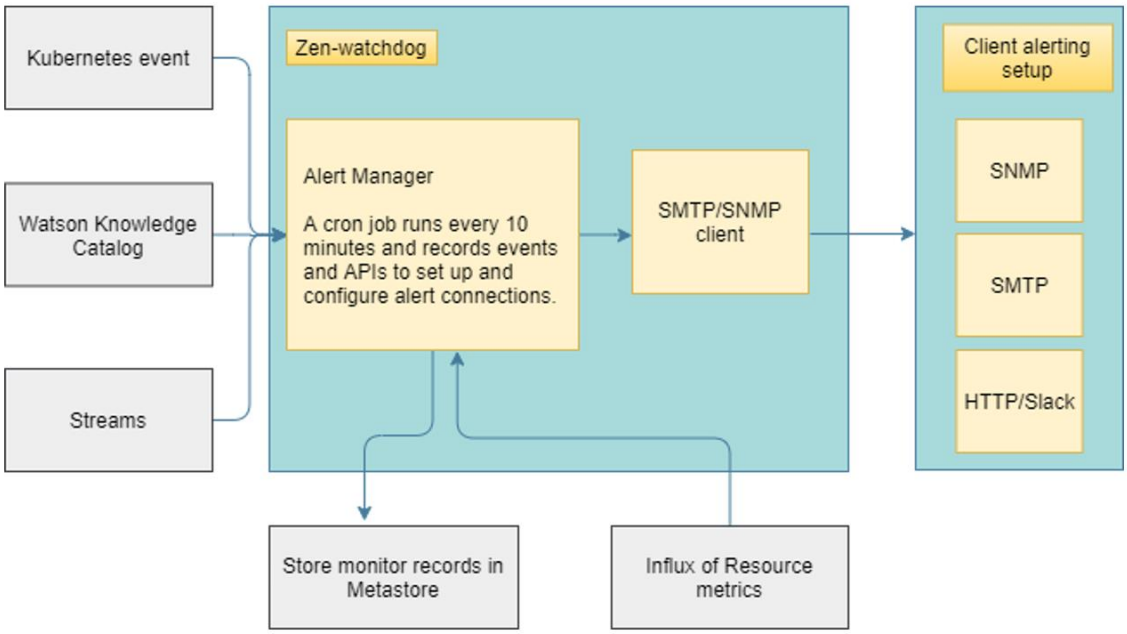
Find services | Filter by: Status | vCPU quota status | Memory quota status | Clear all

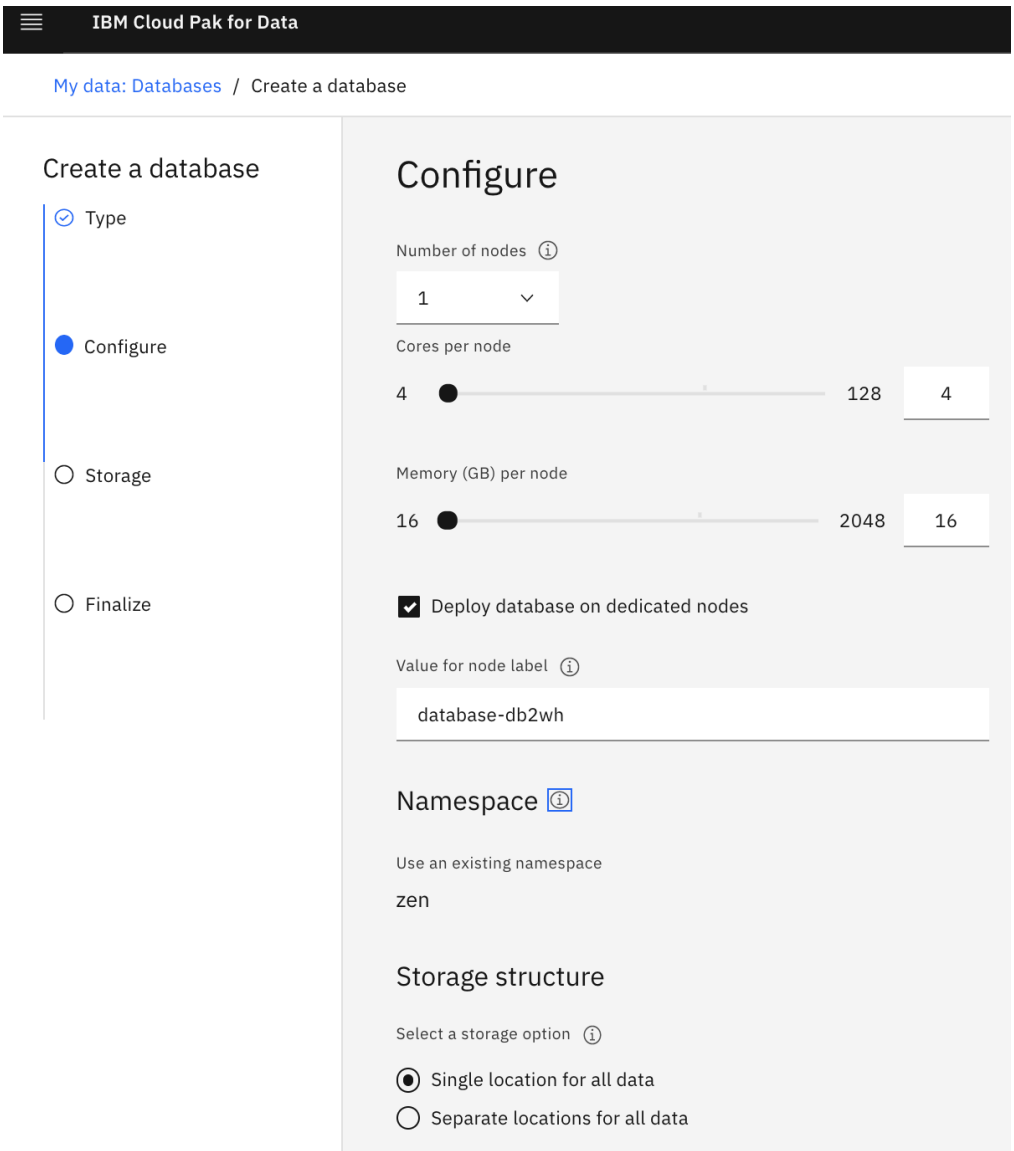
Service name	Status	Service instances	Environments	Jobs	vCPU status	Memory status
Total						
Common Core Services	Normal	0	0	0	No quota set	No quota set
Common Database Services	Normal	0	0	0	No quota set	No quota set
Data Refinery	Normal	0	0	0	No quota set	No quota set
Watson Knowledge Catalog	Normal	0	0	0	No quota set	No quota set
IBM Cloud Pak for Data Control Plane	Normal	0	0	0	No quota set	No quota set

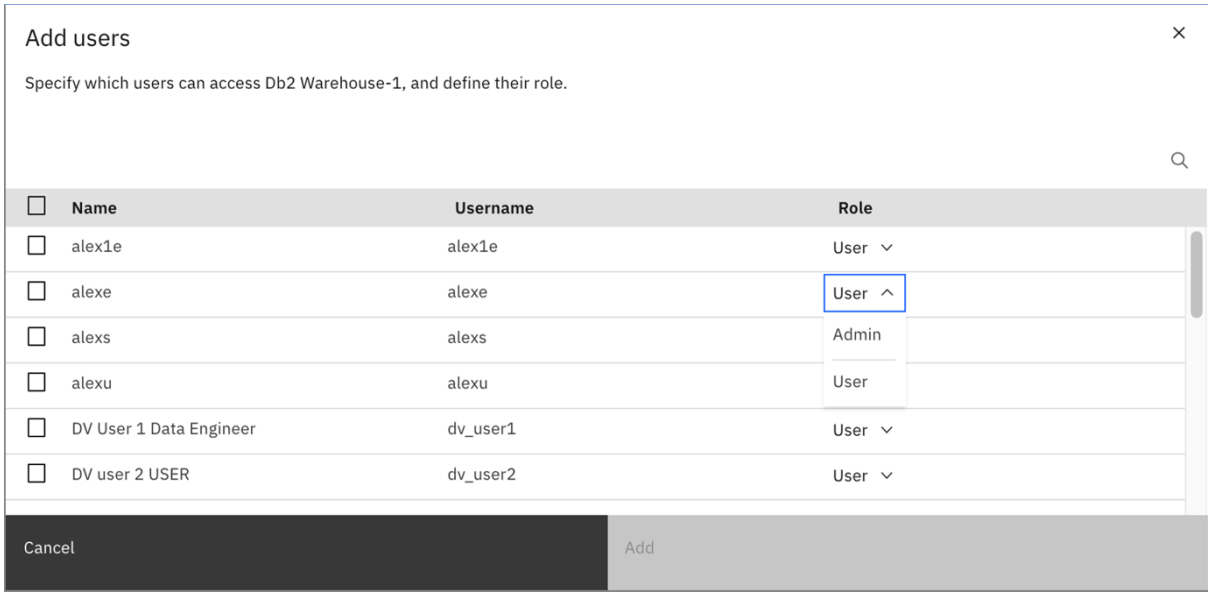
Find pods | Filter by: Pod status | Service | Namespace | Type | Clear all | Refresh

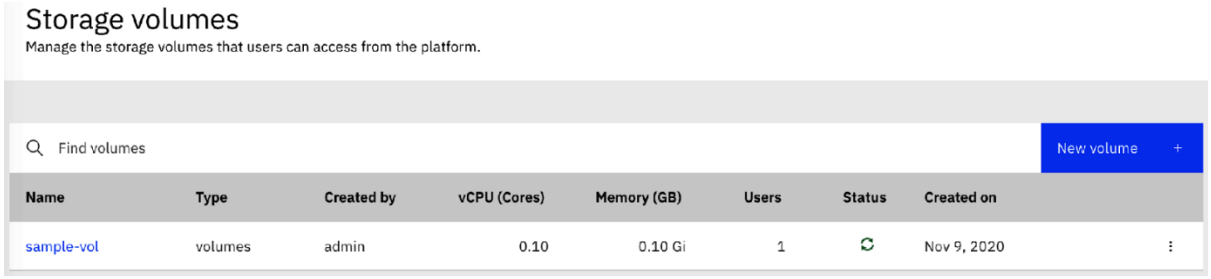
Pod name	Service	Current vCPU use	vCPU requests	Current memory use (GiB)	Memory requests (GiB)
Total		3.79	31.52	60.44	78.95
shop4info-mappers-service-0	Watson Knowledg...	0.01	0.20	0.29	0.50
shop4info-rest-0	Watson Knowledg...	0.00	0.20	0.48	1.00
shop4info-scheduler-0	Watson Knowledg...	0.00	0.10	0.33	1.00
shop4info-type-registry-servic...	Watson Knowledg...	0.00	0.10	0.32	0.50
solr-0	Watson Knowledg...	0.01	0.15	0.84	1.00
spaces-7556bf4568-5jz7w	Common Core Ser...	0.01	0.50	0.16	0.50
spawner-api-54bfc796d5-hx6j2	Common Core Ser...	0.00	0.01	0.03	0.06
usermgmt-5b79c55dd9-5j49b	IBM Cloud Pak for...	0.01	0.20	0.09	0.25
usermgmt-5b79c55dd9-kr7j6	IBM Cloud Pak for...	0.00	0.20	0.09	0.25
wdp-activities-85	Watson Knowledg...	0.06	0.10	0.51	0.78
wdp-connect-con	Common Core Ser...	0.02	0.15	1.40	0.63
wdp-connect-con	Common Core Ser...	0.02	0.20	1.40	0.63
wdp-connect-flig	Common Core Ser...	0.01	0.20	1.50	0.63
wdp-couchdb-0	Common Core Ser...	0.11	1.00	1.80	0.50
wdp-couchdb-1	Common Core Ser...	0.11	1.00	1.90	0.50
wdp-couchdb-2	Common Core Ser...	0.11	1.00	1.90	0.50

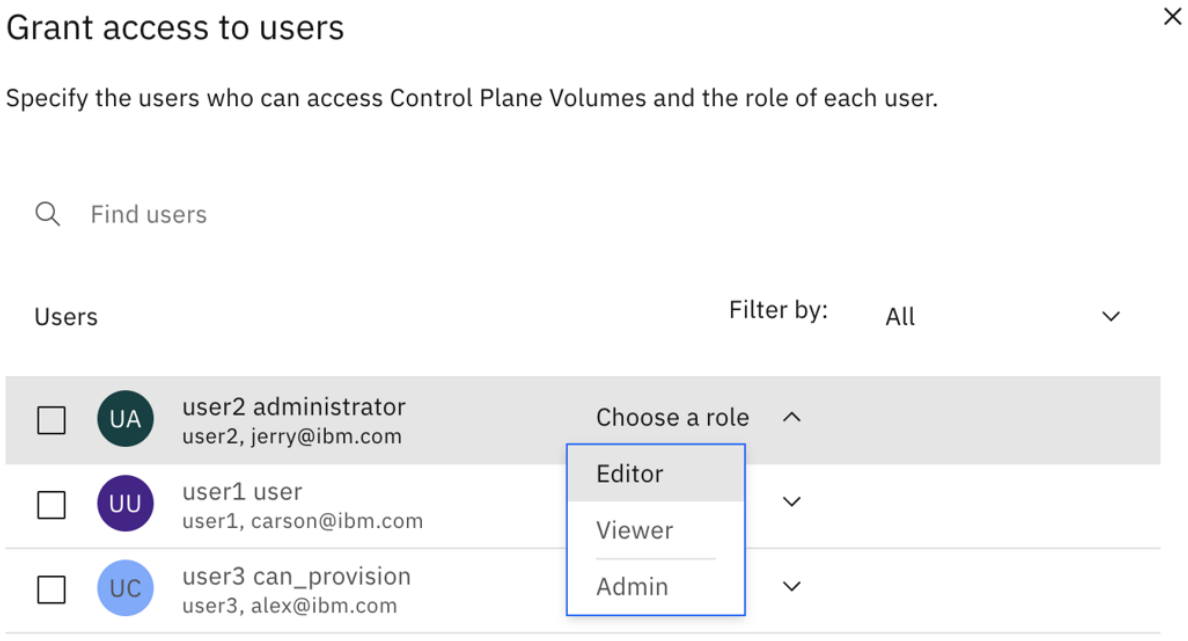
Details
View log
Restart pod

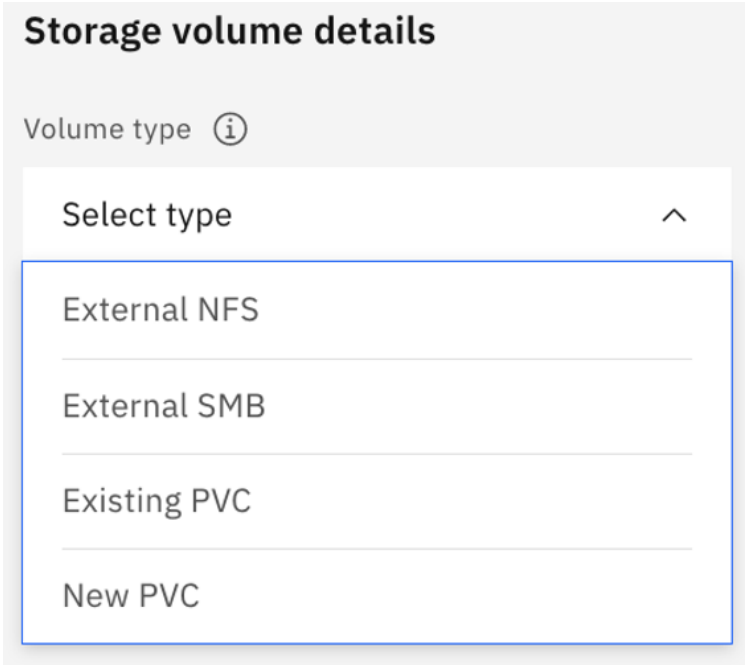

Kubernetes event
Watson Knowledge Catalog
Streams
Zen-watchdog
Alert Manager
A cron job runs every 10 minutes and records events and APIs to set up and configure alert connections.
SMTP/SNMP client
Client alerting setup
SNMP
SMTP
HTTP/Slack
Store monitor records in Metastore
Influx of Resource metrics

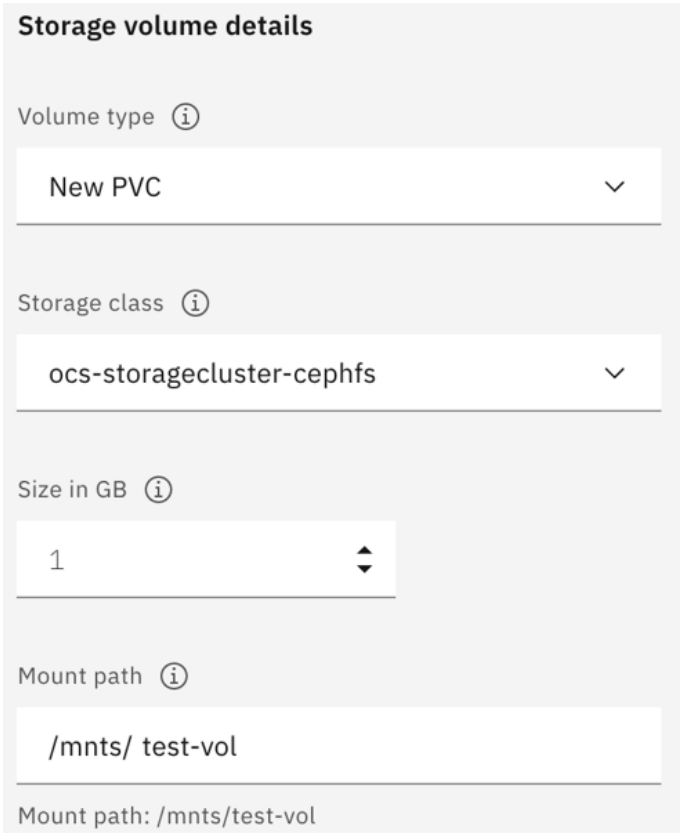

IBM Cloud Pak for Data
My data: Databases / Create a database
Create a database
Type
Configure
Storage
Finalize
Configure
Number of nodes
1
Cores per node
4
128
4
Memory (GB) per node
16
2048
16
Deploy database on dedicated nodes
Value for node label
database-db2wh
Namespace
Use an existing namespace
zen
Storage structure
Select a storage option
Single location for all data
Separate locations for all data


Add users
Specify which users can access Db2 Warehouse-1, and define their role.
Name
Username
Role
alex1e
alex1e
User
alexe
alexe
User
alexs
alexs
Admin
alexu
alexu
User
DV User 1 Data Engineer
dv_user1
User
DV user 2 USER
dv_user2
User
Cancel
Add


Storage volumes
Manage the storage volumes that users can access from the platform.
Find volumes
New volume
Name
Type
Created by
vCPU (Cores)
Memory (GB)
Users
Status
Created on
sample-vol
volumes
admin
0.10
0.10 Gi
1
Nov 9, 2020


Grant access to users
Specify the users who can access Control Plane Volumes and the role of each user.
Find users
Users
Filter by:
All
UA
user2 administrator
user2, jerry@ibm.com
Choose a role
Editor
Viewer
Admin
UU
user1 user
user1, carson@ibm.com
UC
user3 can_provision
user3, alex@ibm.com


Storage volume details
Volume type
Select type
External NFS
External SMB
Existing PVC
New PVC


Storage volume details
Volume type
New PVC
Storage class
ocs-storagecluster-cephfs
Size in GB
1
Mount path
/mnts/ test-vol
Mount path: /mnts/test-vol

Storage volumes

Manage the storage volumes that users can access from the platform.

You can enable access to a volume on an external NFS server or to a persistent volume claim (PVC).

Storage volume overview

Name

sample-vo1

Description (optional)

Description for the volume

Storage volume details

Volume type

External NFS

NFS server

zen-dev-bart-master.fyre.ibm.com

Exported path

/sample-data

Mount path

/mnts/ data

Mount path: /mnts/data

Watson Knowledge Catalog quotas

Monitor platform resource use against your target use

Current requests and limits

vCPU requests	18.51
vCPU limits	57.35
Memory requests (GiB)	70.90
Memory limits (GiB)	154.27

What do you want to set the quotas on?

Requests Limits

Quotas

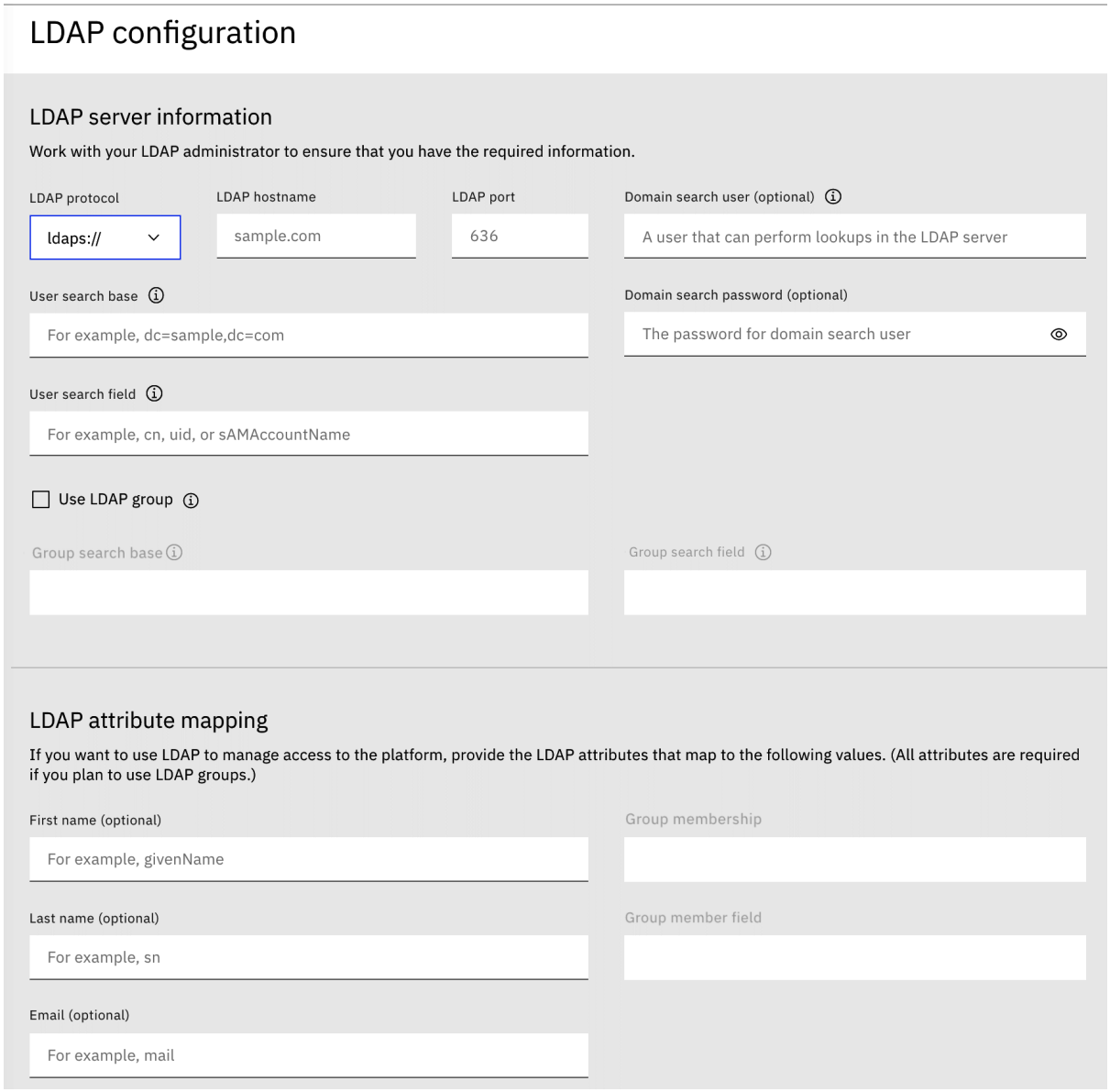
vCPU quota: 24

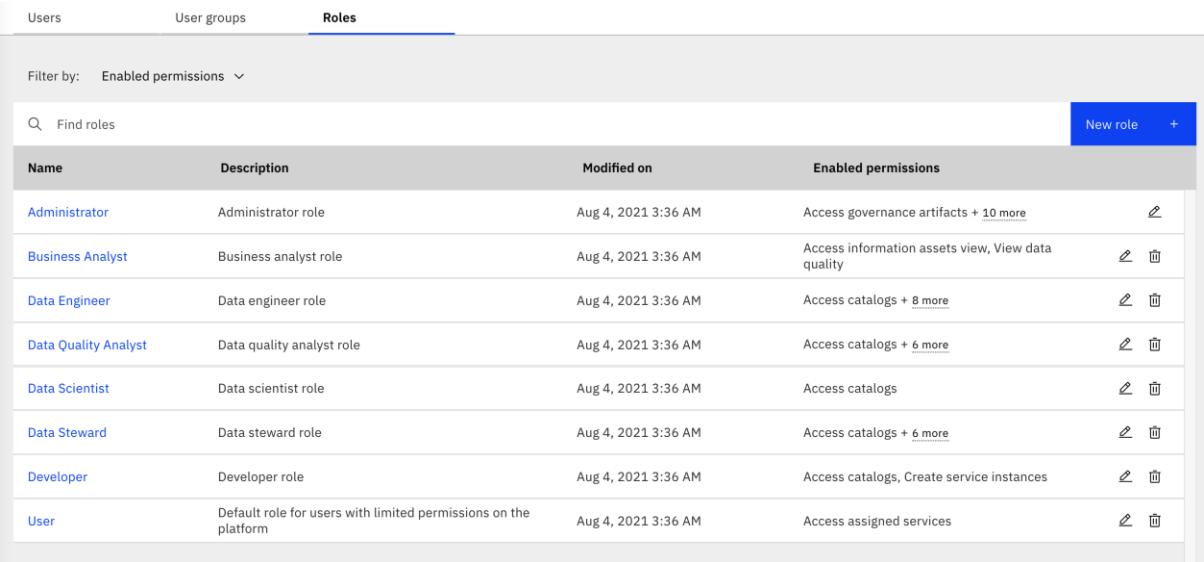
vCPU alert threshold: 90 % of quota

Memory quota (GiB): 100

Memory alert threshold: 95 % of quota

Chapter 10: Security and Compliance

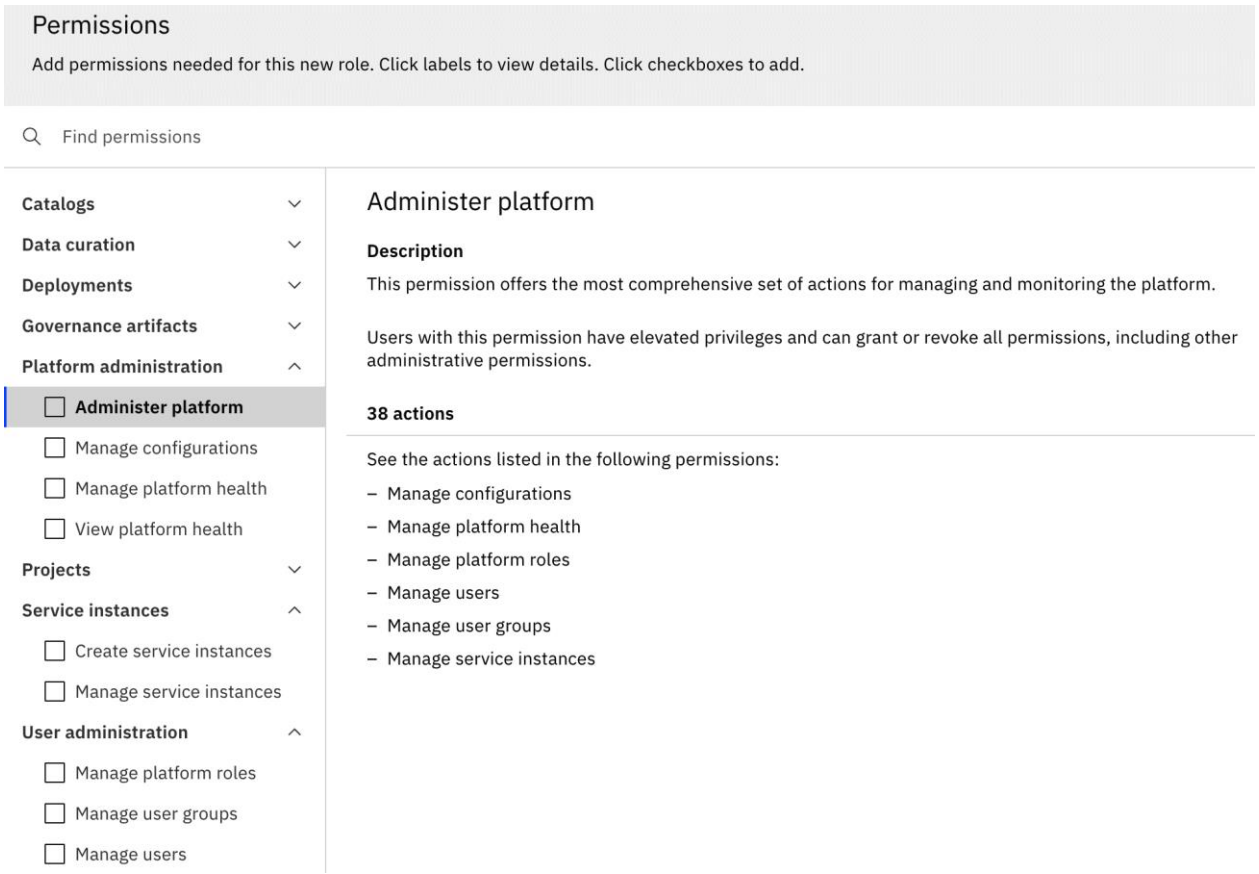


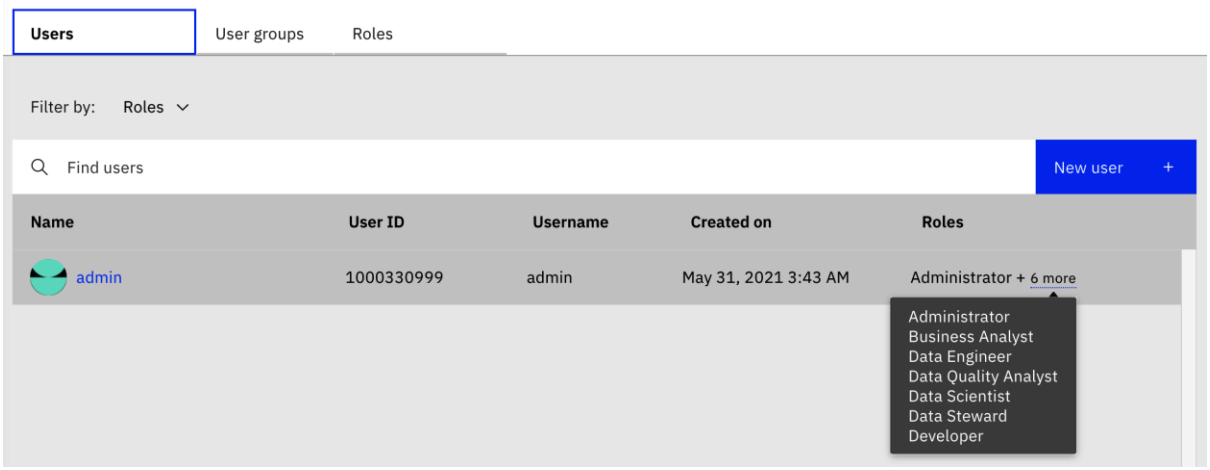


Permission	Category	Services that contribute the permission
Access governance artifacts	Governance artifacts	Watson™ Knowledge Catalog
Administer platform	Platform administration	Cloud Pak for Data control plane
Create deployment spaces	Deployments	– Watson Knowledge Catalog – Watson Studio
Create projects	Projects	– Watson Knowledge Catalog – Watson Studio
Create service instances	Service instances	Cloud Pak for Data control plane
Manage asset discovery	Data curation	Watson Knowledge Catalog
Manage catalogs	Catalogs	– Watson Knowledge Catalog – Watson Studio
Manage data protection rules	Governance artifacts	Watson Knowledge Catalog
Manage data quality	Data curation	Watson Knowledge Catalog
Manage data transformation	Data integration	DataStage®
Manage governance categories	Governance artifacts	Watson Knowledge Catalog
Manage information assets	Catalogs	Watson Knowledge Catalog
Manage metadata	Data curation	Watson Knowledge Catalog
Manage workflows	Workflows	Watson Knowledge Catalog

Permission	Category	Services that contribute the permission
Access catalogs	Catalogs	– Watson Knowledge Catalog – Watson Studio
Create deployment spaces	Deployments	– Watson Knowledge Catalog – Watson Studio
Create projects	Projects	– Watson Knowledge Catalog – Watson Studio

Permission	Category	Services that contribute the permission
Create deployment spaces	Deployments	– Watson Knowledge Catalog – Watson Studio
Create projects	Projects	– Watson Knowledge Catalog – Watson Studio





Add user

Authorize user access to the platform.

- Profile information
- Platform access
- Roles
- Summary

Roles

Assign roles to this user or create a new role.

administrator

Administrator

Administrator

Description

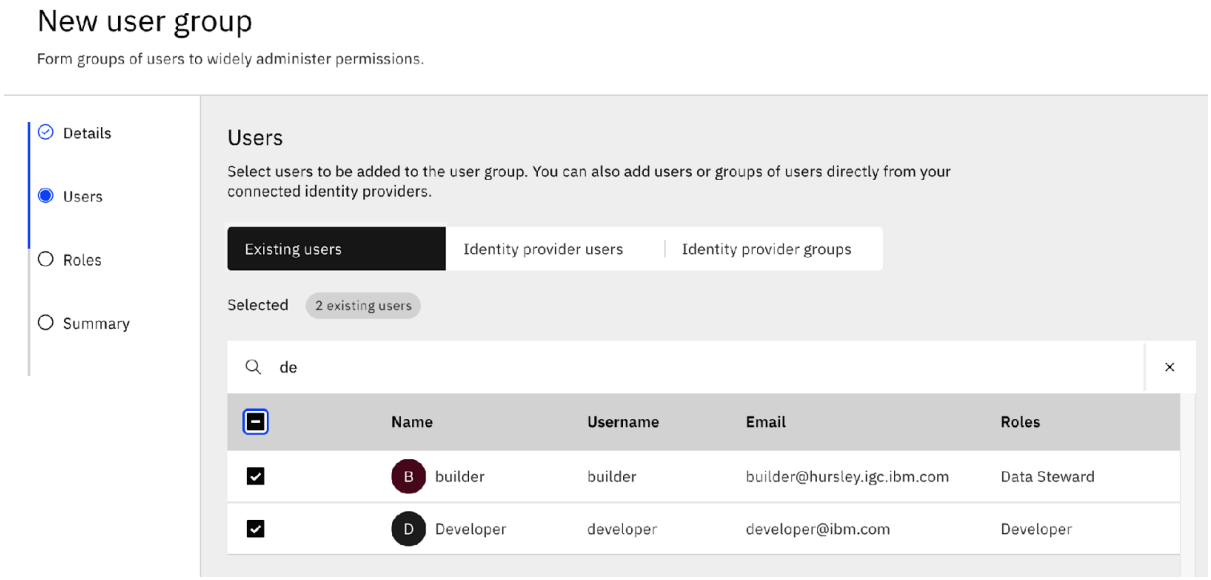
Administrator role

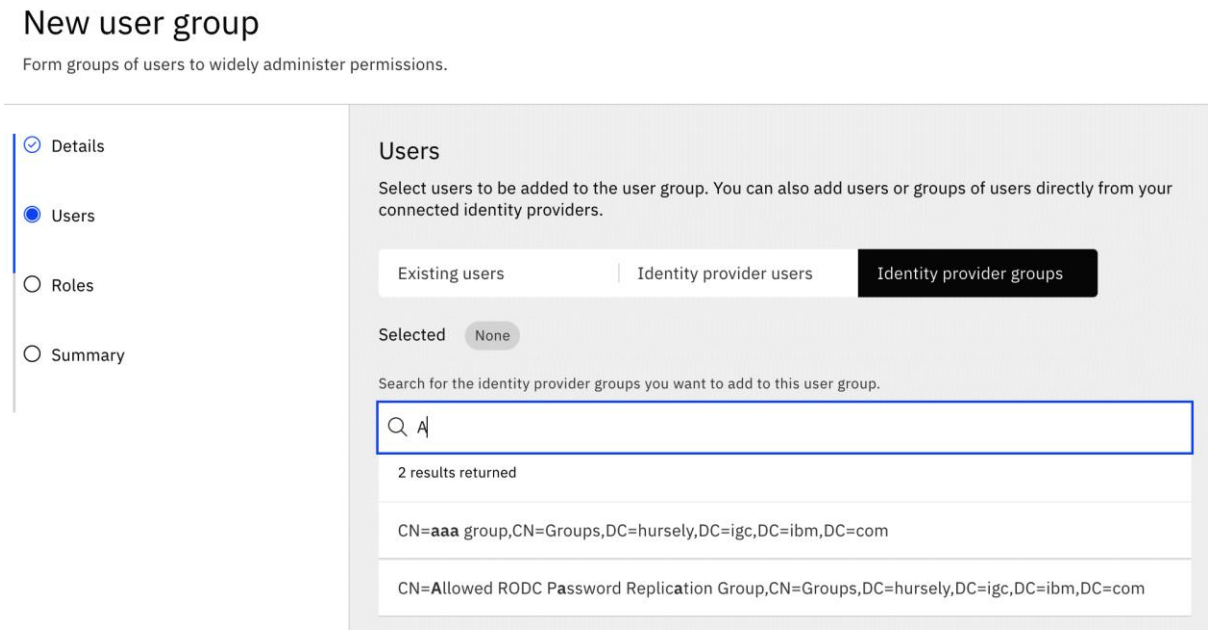
Modified on

Jul 20, 2021 4:34 AM

24 permissions, 110 actions Expand all

- Access advanced governance
- Access advanced mapping
- Access catalogs
- Access data quality
- Access governance artifacts
- Administer platform


New user group
Form groups of users to widely administer permissions.
Details
Users
Roles
Summary
Users
Select users to be added to the user group. You can also add users or groups of users directly from your connected identity providers.
Existing users
Identity provider users
Identity provider groups
Selected
2 existing users
de
Name
Username
Email
Roles
builder
builder
builder@hursley.igc.ibm.com
Data Steward
Developer
developer
developer@ibm.com
Developer


New user group
Form groups of users to widely administer permissions.
Details
Users
Roles
Summary
Users
Select users to be added to the user group. You can also add users or groups of users directly from your connected identity providers.
Existing users
Identity provider users
Identity provider groups
Selected
None
Search for the identity provider groups you want to add to this user group.
A
2 results returned
CN=aaa group,CN=Groups,DC=hursely,DC=igc,DC=ibm,DC=com
CN=Allowed RODC Password Replication Group,CN=Groups,DC=hursely,DC=igc,DC=ibm,DC=com

New user group

Cancel | Back | Next

Details | Users | Roles | Summary

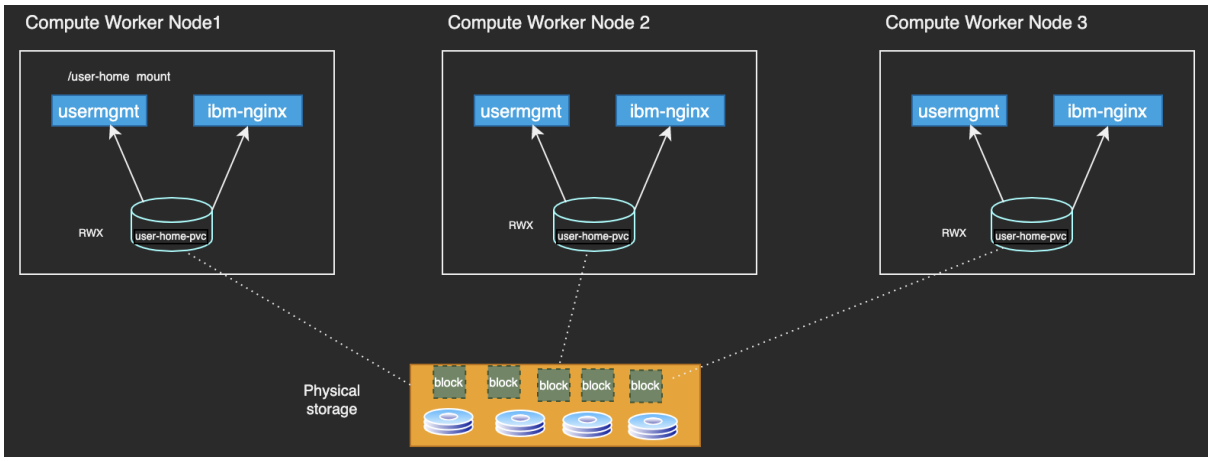
Roles

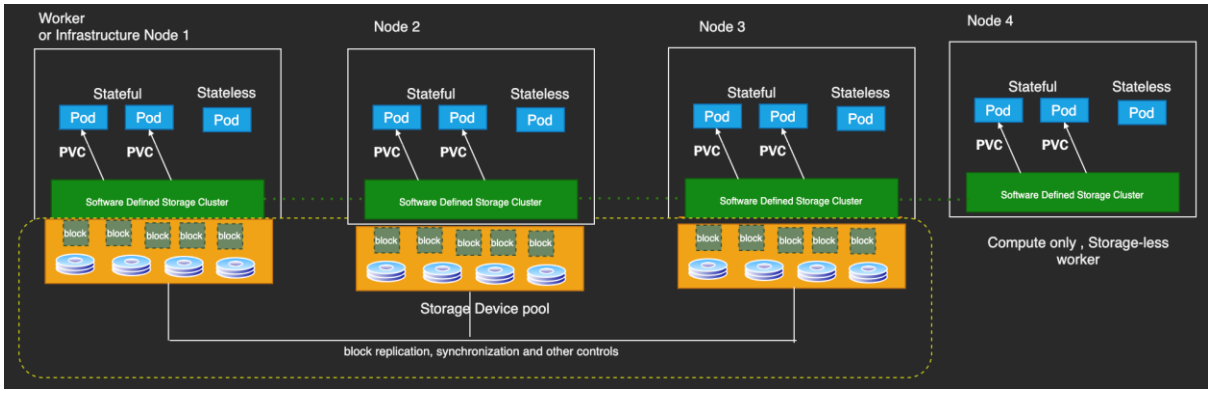
Select the roles to assign to this group. If you can't find the appropriate role in the list, you can create a new role.

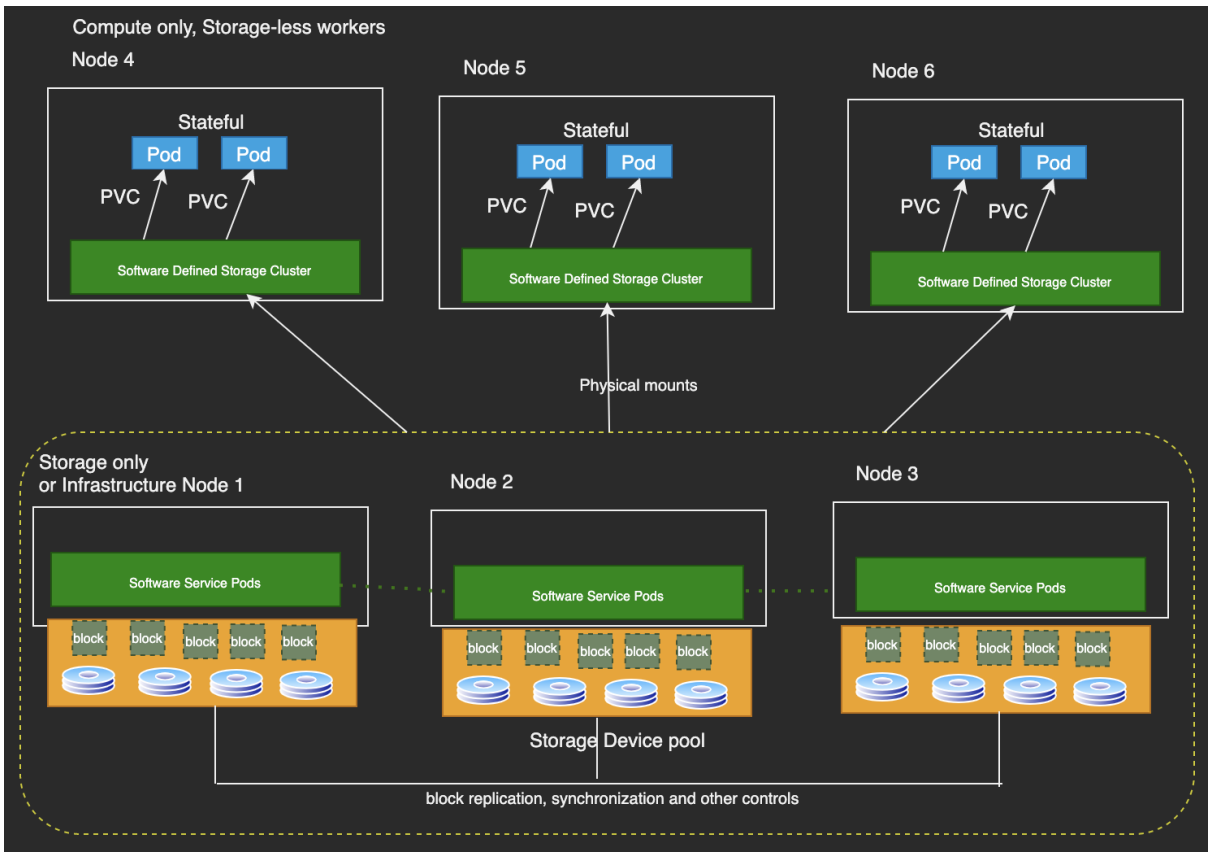
Find roles

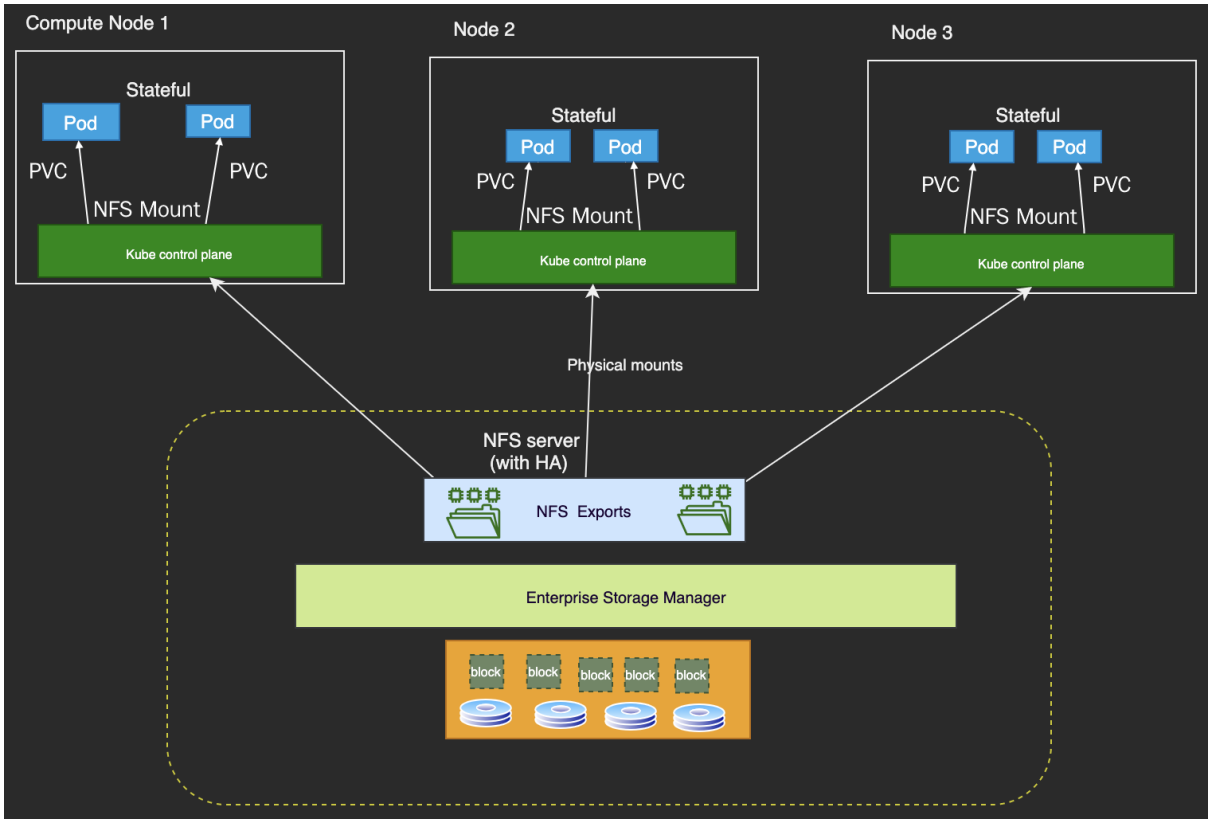
	Name	Enabled permissions
☐	Administrator	Access governance artifacts + 10 more
☐	Business Analyst	Access information assets view, View data quality
☐	Data Engineer	Access catalogs + 8 more
☐	Data Quality Analyst	Access catalogs + 6 more
☐	Data Scientist	Access catalogs
☑	Data Steward	Access catalogs + 6 more
☐	Developer	Access catalogs, Create service instances
☐	User	Access assigned services

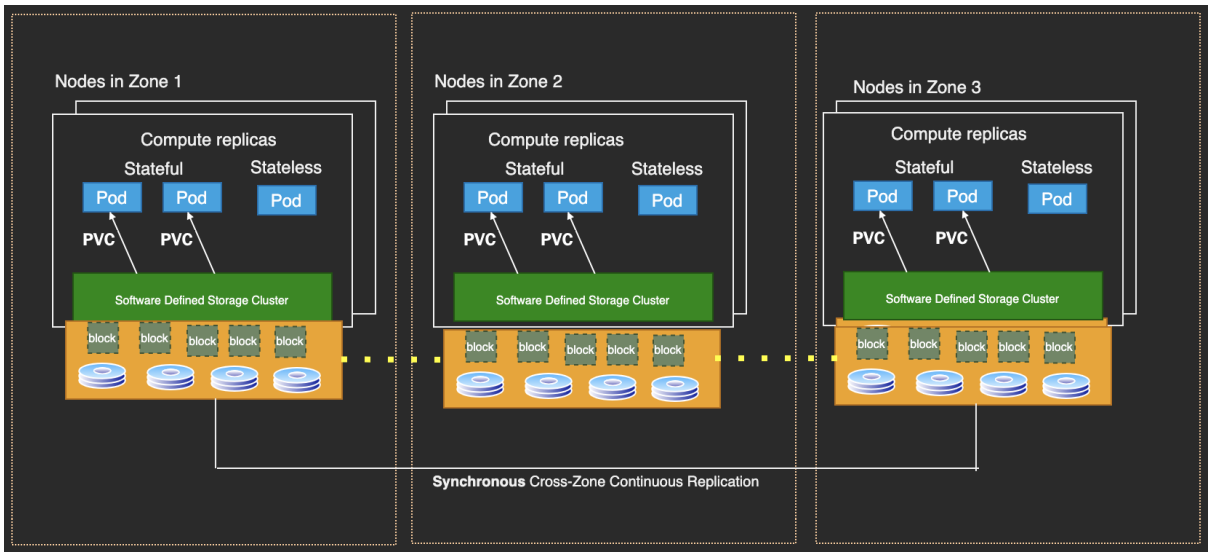
Chapter 11: Storage



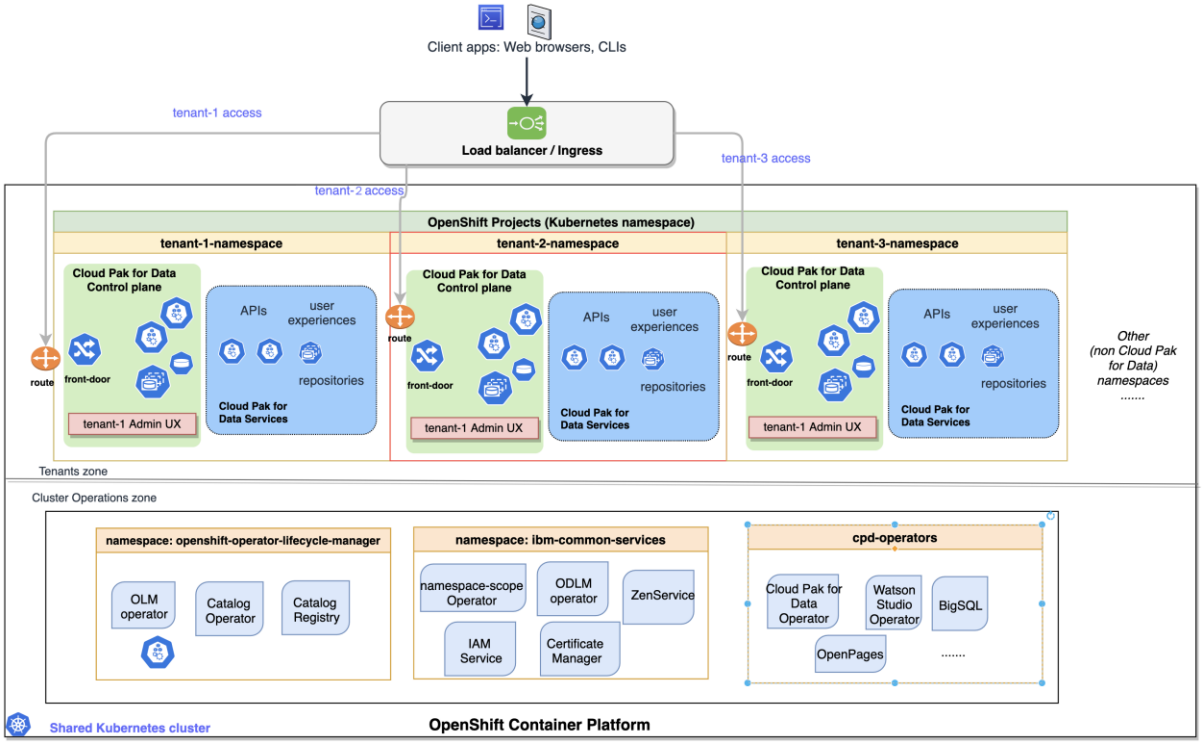


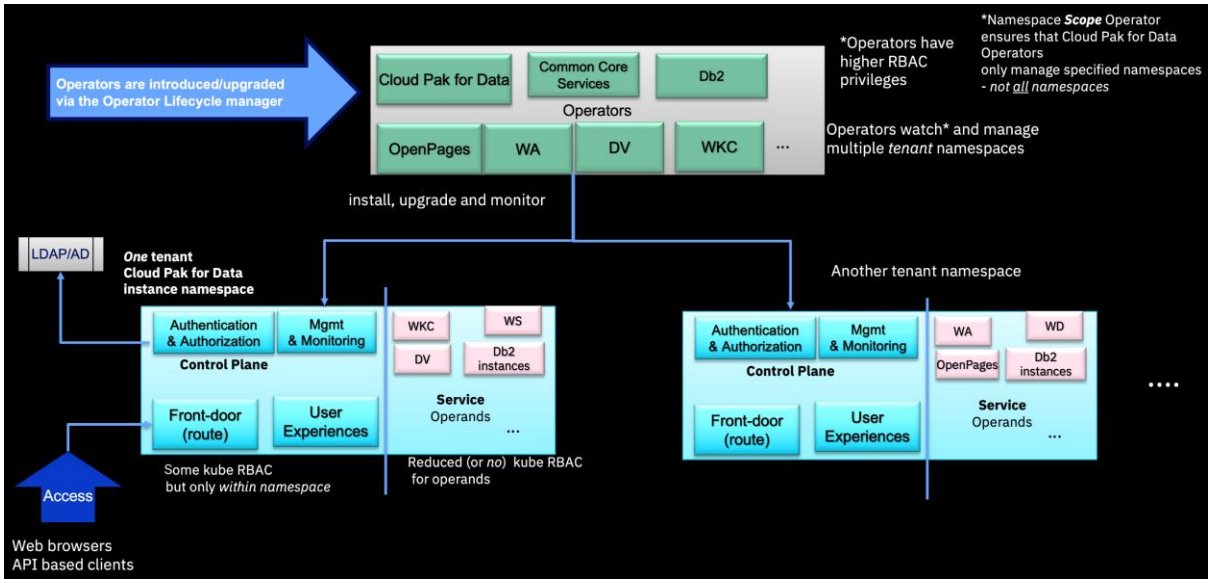


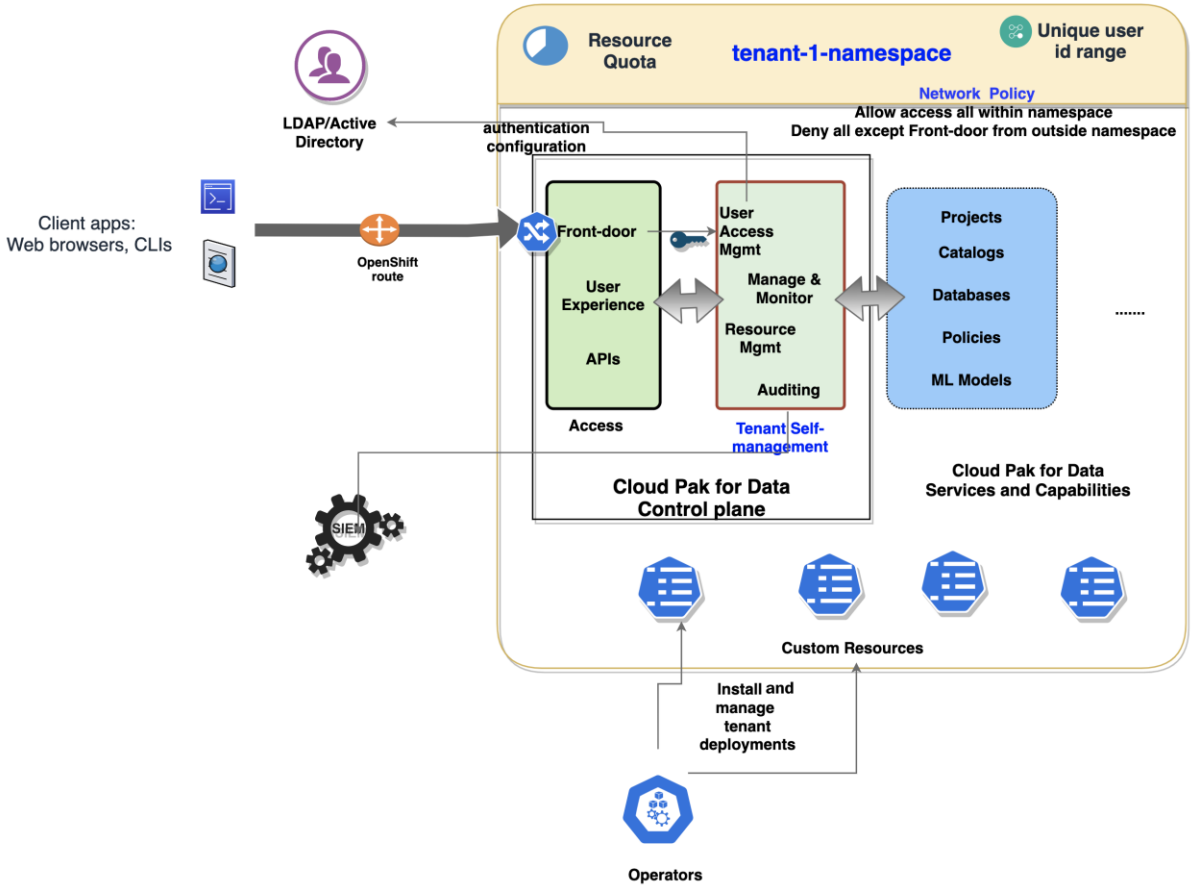

Compute Node 1
Node 2
Node 3
Stateful
Pod
Pod
PVC
PVC
NFS Mount
Kube control plane
Physical mounts
NFS server
(with HA)
NFS Exports
Enterprise Storage Manager
block


Nodes in Zone 1
Nodes in Zone 2
Nodes in Zone 3
Compute replicas
Stateful
Stateless
Pod
PVC
Software Defined Storage Cluster
block
Synchronous Cross-Zone Continuous Replication

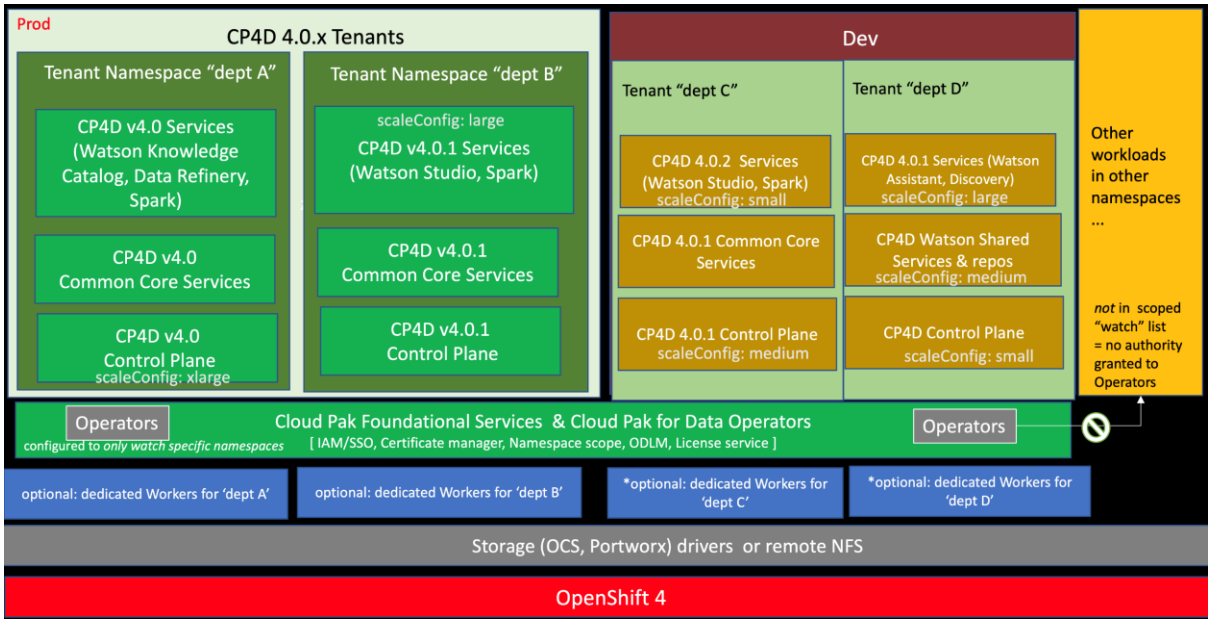
Chapter 12: Multi-Tenancy







Tenancy Aspects	Mechanism
Isolation	*Individual Kube Namespace quotas/OpenShift Projects for each Cloud Pak for Data instance*
1. Users – Authentication, Roles and Authorization 2. Compute Resources 3. Access to Instance 4. Storage Usage	✓ If needed, can support *different* LDAP/AD or SAML configurations *per* instance (each instance has its own unique user management service) ✓ Namespace quotas – set by the cluster admin. ✓ Unique DNS-based routes for each instance. ✓ Completely separated PVCs (in different namespaces) with different PVs – no sharing *across* instances.
Security	*RBAC privileges & Service Accounts scoped only to that namespace*
1. Restricting Kube Cluster-Level Privileges 2. Network Access 3. Auditing	✓ Service accounts and role binding – scoped to Project namespaces. ✓ Kube network policies for hardening *like a firewall around an RHOS Project namespace*. ✓ *Per-tenant* SIEM forwarding supported.
Management	*Administration privileges scoped to a namespace*
1. Monitoring/Metering 2. Ops – Backup/DR, Scale, Patch, Upgrade 3. Serviceability 4. Service Installations	✓ Different tenant "Admin users for different instances. Tenants are managed independently." ✓ Ops privileges scoped to that OpenShift project's namespace. ✓ Serviceability utilities only capture diagnostics that are appropriate for that instance. ✓ Different instances can have different sets of services at different versions scaled independently from other tenants.



Tenancy Aspects	Mechanism
Isolation	*Shared instance in one Kubernetes namespace* ⚠
1. Users – Authentication, Roles and Authorization 2. Compute Resources 3. Access to Instance 4. Storage Usage	✓ Authorization controls with groups and roles with only one LDAP/AD/SAML configuration possible. o Note: Some services are still shared, and some admin roles cannot be *exclusively* granted to one sub-tenant • No quotas for individual sub-tenants. • A single OpenShift route for all users; harder to support sub-tenant-specific DNS domain names. • No separation of persistent volumes.
Security	*RBAC privileges and service accounts scoped to that namespace* ⚠
1. Restricting Kube Cluster-Level Privileges 2. Network Access 3. Auditing	✓ No cluster-level privileges. • No network isolation between sub-tenant user workloads. ✓ Auditing covers *all* sub-tenant activities and individual users that have been uniquely identified, with only one SIEM configuration possible.
Management	*Administration and Ops privileges cannot be granted to individual users* ⚠
1. Monitoring/Metering 2. Ops – Backup/DR, Scale, Patch, Upgrade 3. Serviceability 4. Service Installations	✓ A primary tenant "admin" assignment can ensure security. "Sub-tenants" should not be granted and admin role. • Day 2 operations cannot be at a sub-tenant scope. • Serviceability and diagnostics cannot be done at the sub-tenant level. The primary tenant admin can view all logs from all sub-tenants. • The responsibility of the tenant admin (or cluster admin) - cannot be at a sub-tenant scope. Different service versions for different sub-tenants cannot be supported.