

Cloud Pak for Data v3.5

What's New – Level 200

Jonathan Entwistle – jentwistle@us.ibm.com



Agenda

1. Market Dynamics Guiding v3.5
2. Client Benefits of v3.5:
 - Cost Reduction
 - Innovation for Digital Transformation
3. Run Anywhere
4. Value Proposition of Cloud Pak for Data v3.5 for Existing Clients
5. Other Key Enhancements

IBM Cloud Pak for Data

Simplifies, unifies and automates the delivery of data and AI to the business

Foundational Services

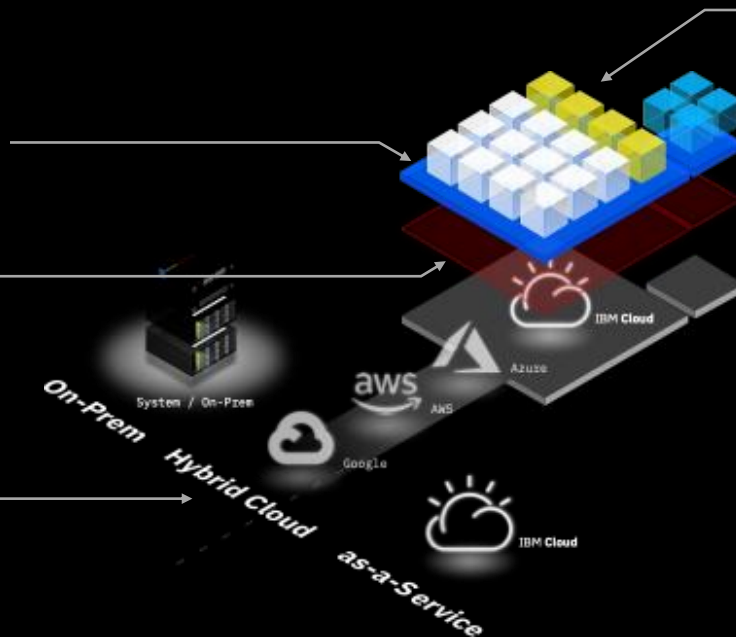
Essential to Cloud Pak for Data, Foundational Services provide a command-line interface, an administration interface, a services catalog, a central list of connections, and the central user experience.

Red Hat® OpenShift®

A streamlined multicloud container platform foundation that can run anywhere. OpenShift license dedicated to run Cloud Pak for Data.

A truly hybrid multicloud world

Don't let lock-in or major IT decisions affect how your Data and AI teams operate. Your work and data should be accessible in any hybrid multicloud strategy.



Services in the Platform Ecosystem

Base Services

Core to Cloud Pak for Data, these Base services fasttrack organizations on their **end-to-end data and AI journey** with comprehensive continuum of capabilities.

Extended Services

Beyond its Base Services, Cloud Pak for Data has a growing **ecosystem of Open Source, Partner, and IBM Extended Services** to expand the breadth of capabilities for teams.

<https://www.ibm.com/industries/federal/analytics>

Hands-on Training on Cloud Pak for Data on Red Hat Open Shift

February 25, 2021

9:00 AM - 4:30 PM US/Eastern

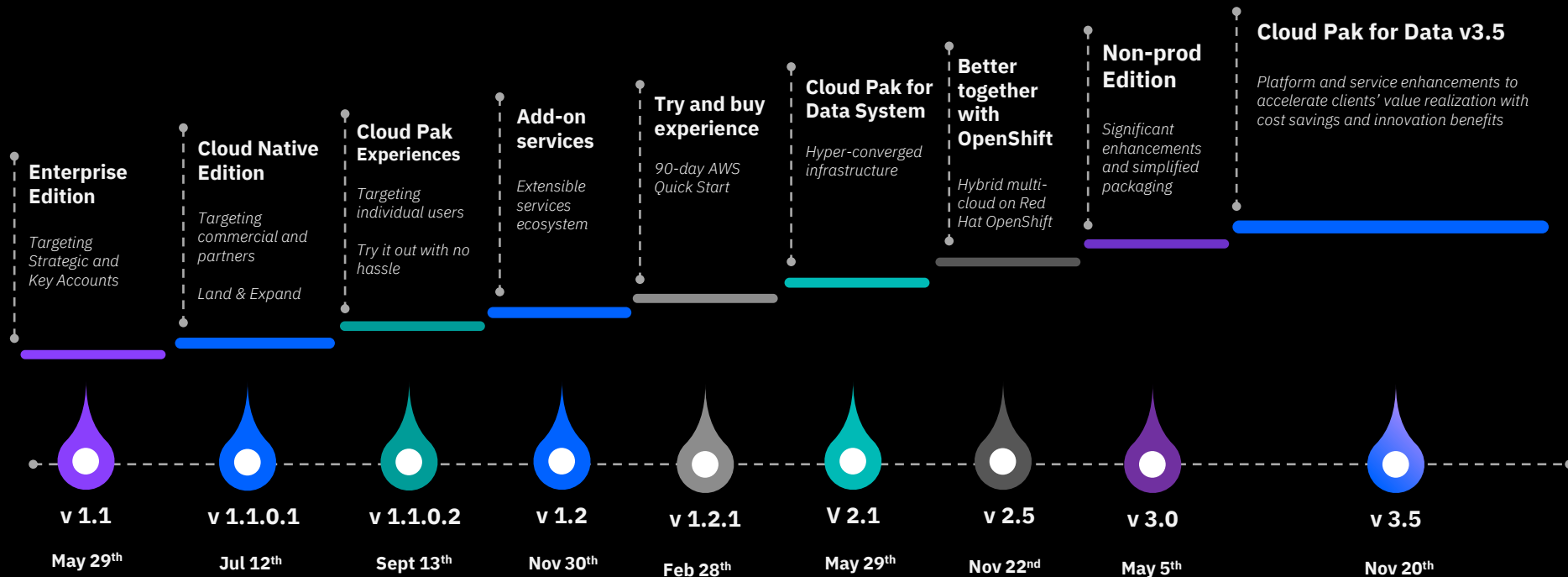
[Register Now](#)

If you have any questions regarding the event
please send an e-mail to a3center@us.ibm.com

Virtual workshop

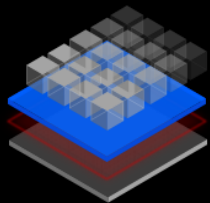
Set the stage through a customer analytics use case and walk through a dashboard detailing a customer churn problem. In the labs, users interact with a web application before adding artificial intelligence (the “before” view) and then interact with the same web application after adding artificial intelligence (the “after” view.) Users will analyze the resulting cost and benefit that an analytics modernization project with a deployed machine learning model will bring to the business over time.

Cloud Pak for Data Journey: 9th Release Since Launch



Cloud Pak for Data 3.5 Packaging

Details on Base Platform Services and Extensions



Cloud Pak for Data Base Platform Services

Db2 Warehouse

Data Virtualization

Db2 Eventstore

IBM Streams

Watson Knowledge Catalog *(including IGC)*

Information Analyzer *(included in WKC)*

Watson Studio *(includes Data Refinery)*

Watson Machine learning *(includes AutoAI)*

Watson OpenScale

Cognos Dashboards Embedded

Analytics Engine for Apache Spark

IBM Performance Server

InstaScan *(included in WKC)*

Master Data Connect

Db2 Big SQL

Guardium Integration

Data Management Console **NEW!**

Watson Machine Learning– Accelerator

Data Privacy (Tech Preview)

Federated Learning (Tech Preview)



Cloud Pak for Data Extensions

Db2 AESE

Informix

DataStage

Information Server

Watson Studio Premium
(includes SPSS Modeler, Decision Optimization and Hadoop Execution Engine)

Cognos Analytics

Planning Analytics

Watson Assistant

Watson Discovery

Watson API Kit *(includes WKS, STT, TTS & NLU)*

Financial Crimes Insights

Financial Services Workbench

Master Data Management

Virtual Data Pipeline (Actifio)

OpenPages

Open Data for Industries

Knowledge Accelerators
(Content Master Data Management)

MongoDB

EnterpriseDB Postgres

Advanced Data Prep

NEW!

Market dynamics in 2020

This year has been a catalyst for significant changes – but they haven't affected the global market in the same ways. As customers emerge in 2021, we'll see differing strategies in response to 2020.

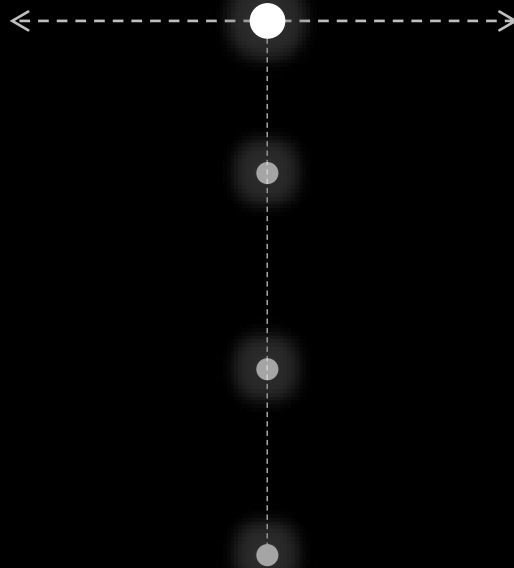
Cost Reduction Strategy

to persist through the downturn

30% of customers will pivot to technology cost reduction¹

Focused on savings through automation, cloud applications and infrastructure, and purchase of technologies with strong ROI¹

Likely in the public sector, manufacturing, retail and finance industries¹



Innovation Strategy

for transformational change

23% of customers expect an increase in technology spending¹

Focusing on big bets in business resiliency², enterprise risk management, data security, and advanced AI³

Likely in technology, business services and healthcare industries¹

Client Benefits of Cloud Pak for Data v3.5

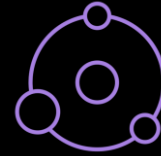
Enterprises are either thriving with exponential growth or in survival mode adapting to the new normal



Cost Reduction



Improved User Experience
Simplified Platform Management
Enhanced Automation



Innovation for Digital
Transformation



Accelerated AI
Improved Data Governance & Security
Expanded Ecosystem

Client Benefits of Cloud Pak for Data v3.5



Cost Reduction

Improved User Experience

- Navigation Improvements
- Data Management Console
- Platform Connections

Simplified Platform Management & Enhanced Automation

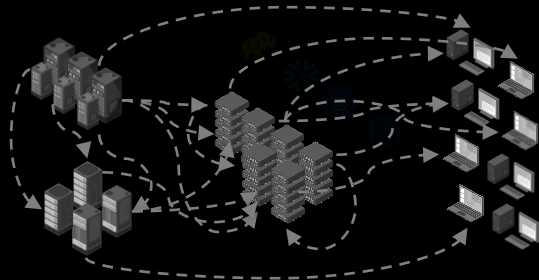
- Platform Management
- Deployment Spaces
- OpenShift v3.11 and v4.5
- OpenShift Operators

Navigation Improvements

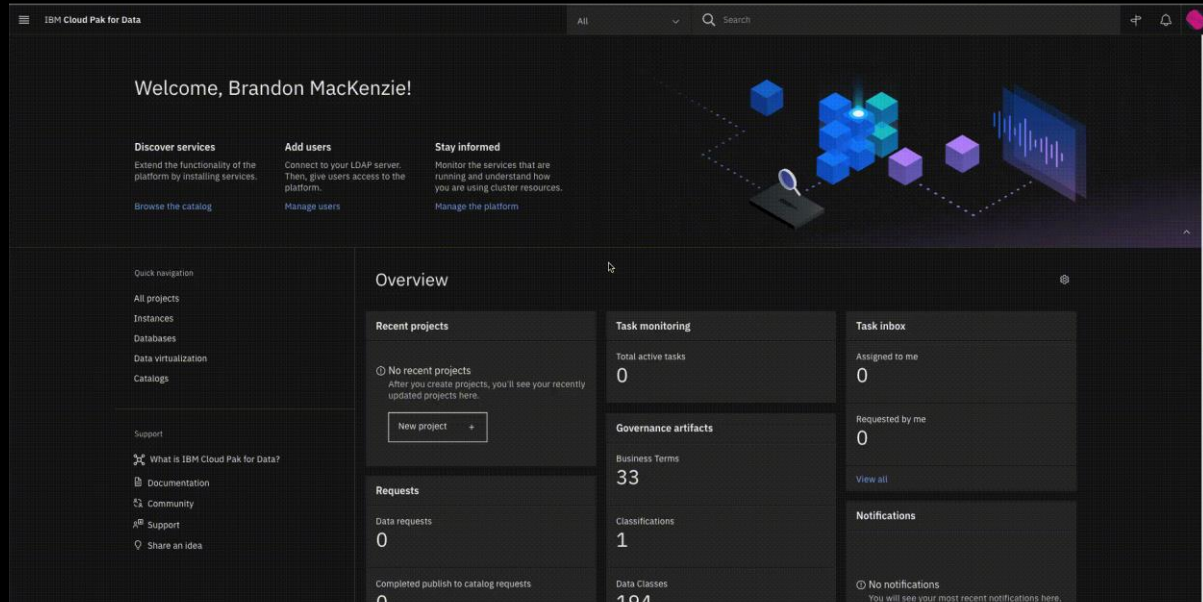
A unified experience built around our users



Cost reduction
All users



The **old** way:
chaos and pain
for users



The **new** way: a unified experience for all users

Navigation Improvements

A unified experience built around our users



Cost reduction
All users

Navigation Improvements

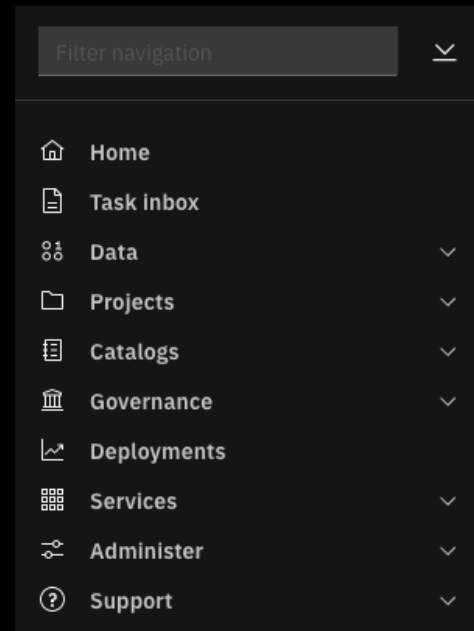
Unified user experience

- Designed to help users understand where they are, what they've found, what's around, and what to expect
- Improvements based on thousands of hours of user research studies

Problems Solved

The most efficient and elegant user experience

- Intuitive role-based navigation enables all users across the platform to get their jobs done more quickly
- Automatically expands and enriches the lives of our users as they expand their use of the platform



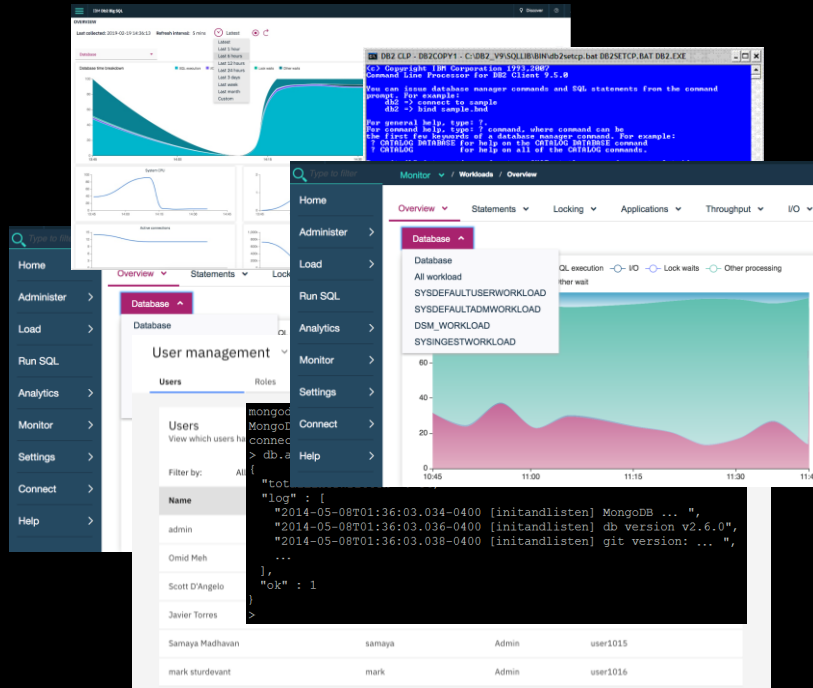
Data Management Console

A single place to manage all your databases

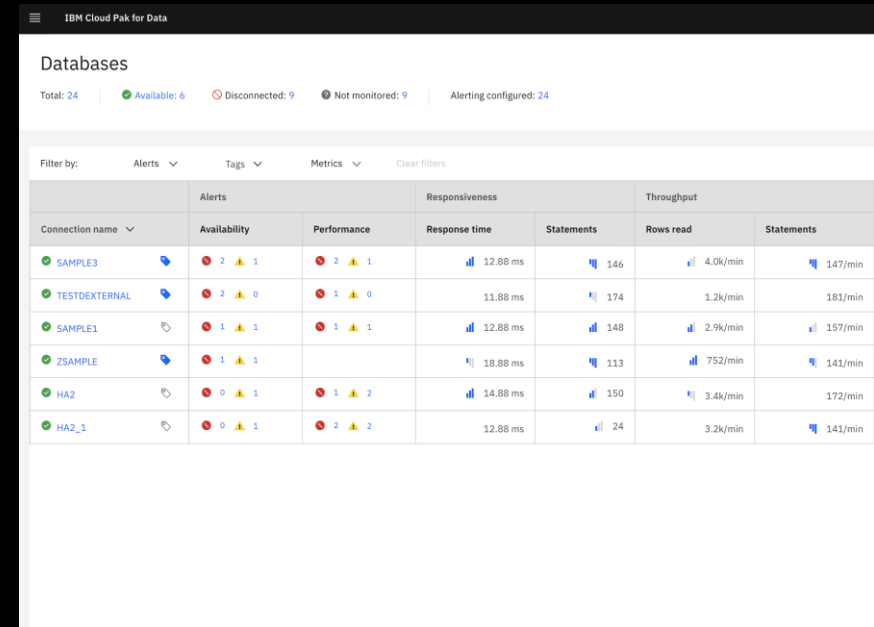


Cost reduction
Data Engineers

Without Cloud Pak for Data:



With Cloud Pak for Data:



Data Management Console

A single place to manage all your databases

IBM Cloud Pak for Data

Unified Experience

Manage **Db2 Advanced Edition, Db2 Warehouse, Db2 Big SQL, and Data Virtualization** from a single console

Business Value: design-driven unified experience - reduce training costs, ensure higher productivity and greater project success across your data enterprise

Enterprise View

Monitor 100s databases from a single screen

Use dynamic filtering by key KPIs and get extensive flyover details

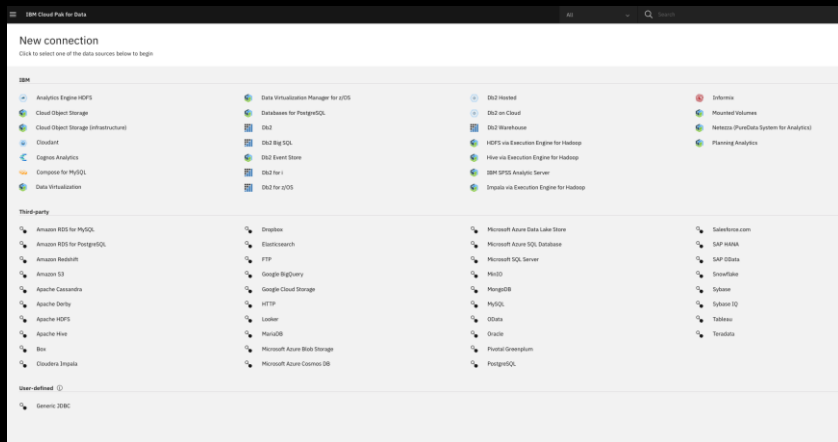
Business Value: get a high-level view of your enterprise and address potential issues before they impact performance

Monitor Database, Run SQL and Explore Database Objects

- Detailed historical monitoring across key database areas
- Fast and responsive SQL editor with integrated parsing; write and run SQL, use visual explain tool to display query implementation plans
- Manipulate database objects, such as tables, views, or stored procedures, and explore relationships between them

Business Value: access to key database functions from a single integrated interface

The simplest data connectivity experience for your users across their data and AI tooling



To collect, organize, analyze data to infuse AI, the first step is to define **connections** to data sources, in a multi-cloud environment

It can be difficult to define many connections, keep them up to date, and easily consumable

Define connections
and access control to your
data **once** with a unified
experience

Use your data connections seamlessly across Watson Studio, Watson Knowledge Catalog, Cognos, and Data Virtualization

- Define once, use everywhere
- Unified connections user interface for Watson Studio, Watson Knowledge Catalog, Data Virtualization, and Cognos Analytics
- 16 new connectors and enhancements (Azure Blob, SAP Hana, Elastic Search, etc.)
- Connections available in asset catalog – part of our common core services

Client Benefits of Cloud Pak for Data v3.5



Cost Reduction

Improved User Experience

- Navigation Improvements
- Data Management Console
- Platform Connections

Simplified Platform Management & Enhanced Automation

- Platform Management
- Deployment Spaces
- OpenShift v3.11 and v4.5
- OpenShift Operators

Cloud Pak for Data v3.5 Brings Simplified Platform and Operations Management



Cost reduction
System Admins

Jobs

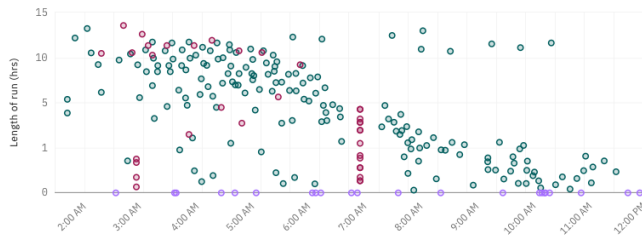
62 spaces

View: Finished runs within the last 8 hours

Run Type



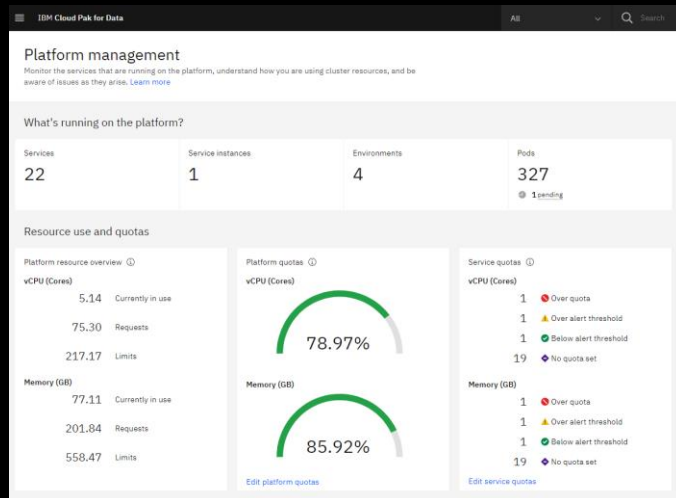
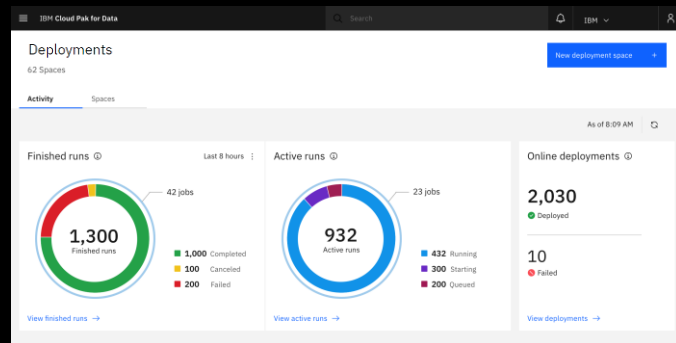
Runs by length and time



450 Completed 132 Failed 18 Canceled

Find jobs

Run ID	Type	Run state	Duration	Started by	Job name	Space	Last updated
DateTimeSt Amp	Model	Completed	20 min	Scheduler	AAAJob123	Bratislava	July 1 2020, 2:40 PM
DateTimeSt Amp	Script	Cancelled	2 hr 5 min	Brown	AAAJob123	Springfield	July 1 2020, 2:40 PM
DateTimeSt Amp	Model	Failed	12 hr 10 min	Sonali S.	Cosmo	Caracas	July 1 2020, 2:40 PM
DateTimeSt Amp	Script	Failed	12 hr	Scheduler	MusicBand	Yokoh Ama	July 1 2020, 2:40 PM



Platform Management

Automation designed by our clients for our clients



Cost reduction
System Admins

Platform Management

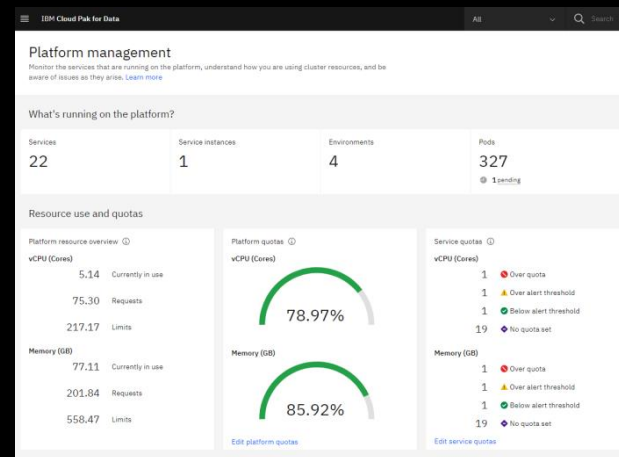
Now with advanced Resource Quotas

- Empowers System Administrators to monitor and control product usage and resources
- Gives System Administrators the ability to set and automate enforcement quotas on compute and memory thresholds

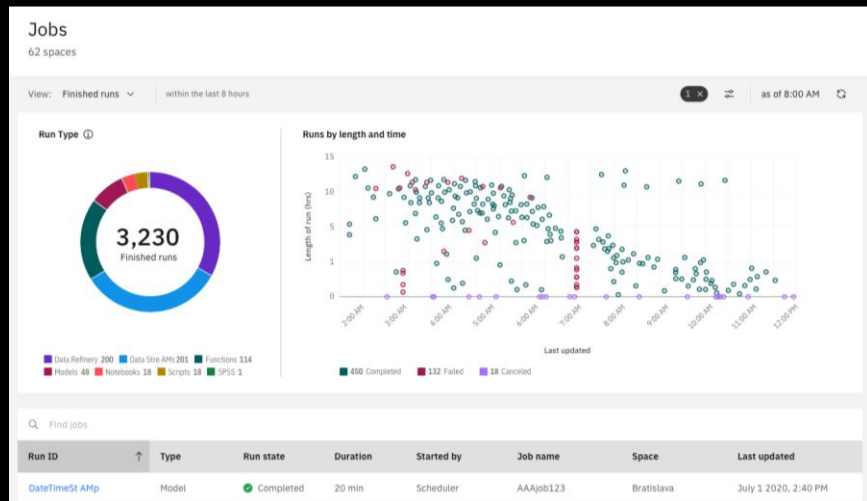
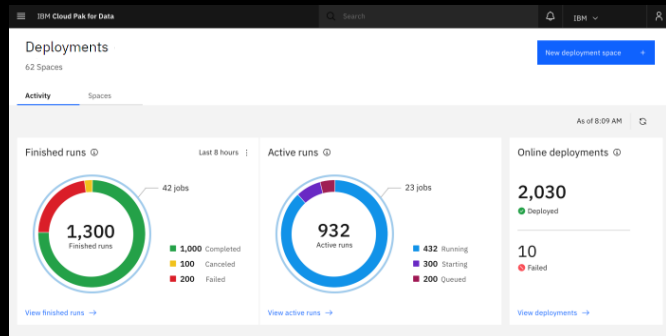
Problems Solved

Benefits of Resource Quotas

- Quickly determine where and how resources are being consumed, by which services and which users
- Quickly adjust resource allocations based on changing needs of the business
- Enables different teams to share a cluster without concern that one team might use too many resources



Deployment Spaces



Business Challenges

With many production workloads to monitor and manage, it can be tedious and time consuming for engineers to pinpoint relevant information

Business Value

Increased productivity for engineers managing production workloads such as jobs and online deployments

Quickly and efficiently assess status, identify issues, take actions, and drill down for more details

Accelerate problem resolution by quickly filtering across thousands of workloads to display only relevant information

Capabilities

- A unified operations dashboard to monitor and manage production workloads across multiple deployment spaces
- View, monitor, and manage deployment spaces and the assets and workloads they contain across all deployment spaces from a single screen
- Manage and compare jobs as they run and see issues surface across spaces, all from one cohesive dashboard
- Monitor's production job runs, and online deployments across multiple deployment spaces

Enhanced Automation with Red Hat OpenShift

Perfectly aligns with cost reduction strategies



Cost reduction
System Admins

OpenShift Container Platform

Support for v3.11 and v4.6

- Delivers full-stack automation
- 38% lower IT infrastructure and development platform costs per application¹
- 531% average five-year ROI¹

OpenShift Operators

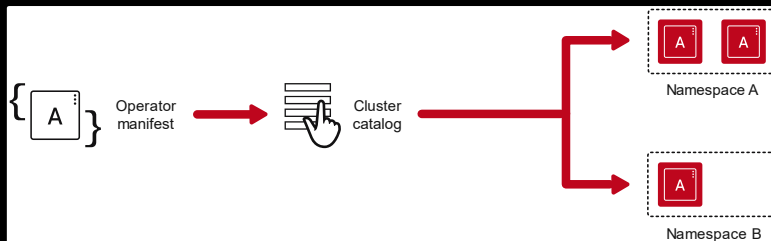
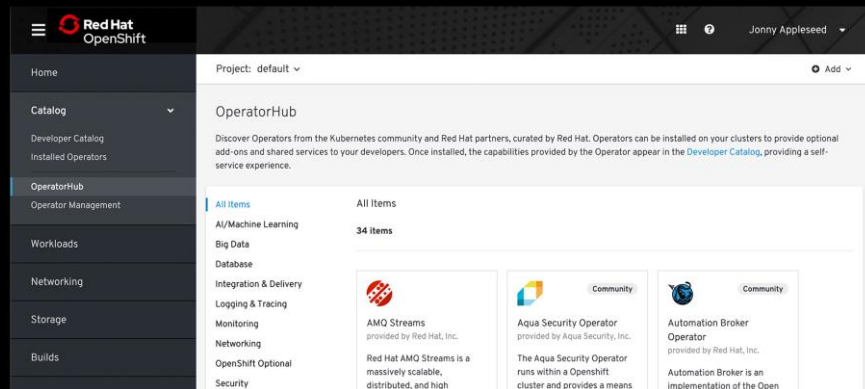
New Cloud Pak for Data Operator

- Cloud Pak for Data Operator simplifies installations and lifecycle management
- Provides automated provisioning, configuration management, seamless patching, upgrade and scaling with operational knowledge baked in, along with auto-pilot management



Cloud Pak for Data available as a Red Hat Operator

Automate the creation, configuration, and management of Cloud Pak for Data



Business Challenges

Significant effort and manual processes is required to update and maintain containerized applications

Human intervention makes managing applications at scale difficult, which can result in configuration errors and system vulnerabilities

Business Value

Lower costs, better governance verified interoperability: using Red Hat OpenShift Certified Operators improves time to value expediting deployment and configuration

Faster Operator development and testing: standardize on the Operator Framework to manage the lifecycle of a broad range of applications and infrastructure in a consistent manner across private and public clouds

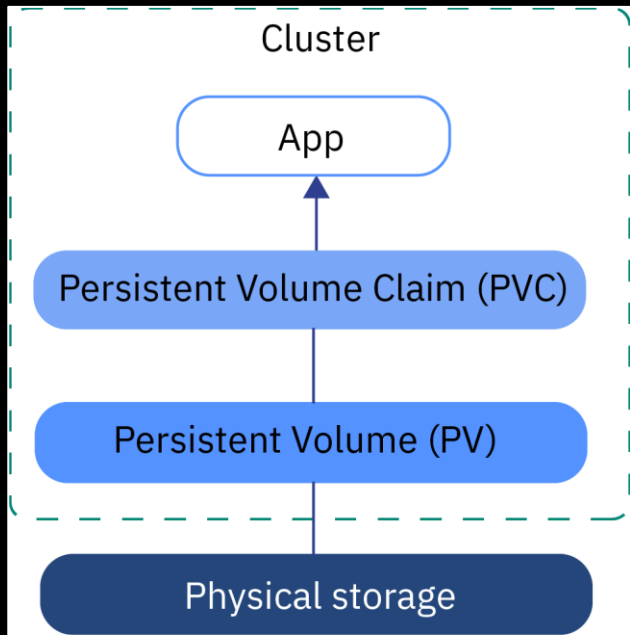
Capabilities

- Operators codify operational knowledge and workflows to automate lifecycle management of containerized applications with Kubernetes
- The Cloud Pak for Data operator (custom controller) streamlines and automates quick installation, uninstallation, patching, scaling and seamless upgrading of the container-based services using a Custom Resource (CR)
- The operator will be available on the Red Hat Marketplace and compatible with the Red Hat Operator Lifecycle Manager

Link: [Red Hat OpenShift Operators](#)

Storage

Support for OpenShift Container Storage (OCS), NFS, and Portworx



NFS

Majority of clients already have investments in Network File System (NFS) and already have in-house NFS skills

Requires an external NFS high-availability setup to ensure no single point of failure

Container-ready storage

Portworx

Acquired by Pure Storage in September 2020

Has advanced functionality such as snapshot backups, backup to cloud, replication to disaster recovery site, etc.

Container-native storage

OpenShift Container Storage

- Fully integrated with and optimized for OpenShift Container Platform
- Dynamic, stateful, and highly available container-native storage that can be provisioned and de-provisioned on demand as an integral part of the OpenShift administrator console
- Container-native storage

IBM License Service

Collects and measures the license usage of the IBM Cloud Pak portfolio

Passport Advantage > Licensing > Container Licenses >

IBM Container Licenses

Passport Advantage Overview Passport Advantage Online **Licensing** Subscription and Support

About Eligible Public Cloud BYOSL policy Processor Value Unit Container Licenses Sub-capacity Term Licenses

Use the links below to learn more about IBM Container Licenses and reporting requirements for IBM Containers and IBM Cloud Paks

- ↓ Overview
- ↓ vCPU Capacity Counting Methodology
- ↓ Additional License Counting guidance
- ↓ IBM License Service

Business Challenge

Clients need a simple to use license management tool to collect and monitor their license usage on Cloud Paks

Capabilities

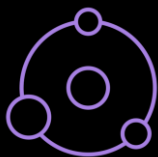
Collects and measures the license usage of Virtual Processor Core (VPC) metric at the cluster level

Provides the API that you can use to retrieve an audit snapshot that lists the highest license usage values for the requested period for products that are deployed on a cluster

Use Case

- Automatically collects and measures license usage
- Simple way to retrieve data for monitoring, audit, and compliance
- IBM License Service is mandatory for the “Container Licensing” (sub-capacity for containers) option, and not valid for traditional sub-capacity or full-capacity licensing options

Client Benefits of Cloud Pak for Data v3.5



Innovation for Digital
Transformation

Accelerated AI

- Watson Machine Learning Accelerator
- Federated Learning (Tech Preview)

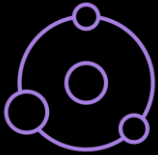
Improved Data Governance & Security

- Data Privacy (Tech Preview)
- Knowledge Accelerators
- OpenPages Cartridge

Expanded Ecosystem

- New 3rd Party Extensions
- IBM Open Data for Industries Cartridge
- Z Support for Db2 Databases

Client Benefits of Cloud Pak for Data v3.5



Innovation for Digital
Transformation

Accelerated AI

- Watson Machine Learning Accelerator
- Federated Learning (Tech Preview)

Improved Data Governance & Security

- Data Privacy (Tech Preview)
- Knowledge Accelerators
- OpenPages Cartridge

Expanded Ecosystem

- New 3rd Party Extensions
- IBM Open Data for Industries Cartridge
- Z Support for Db2 Databases

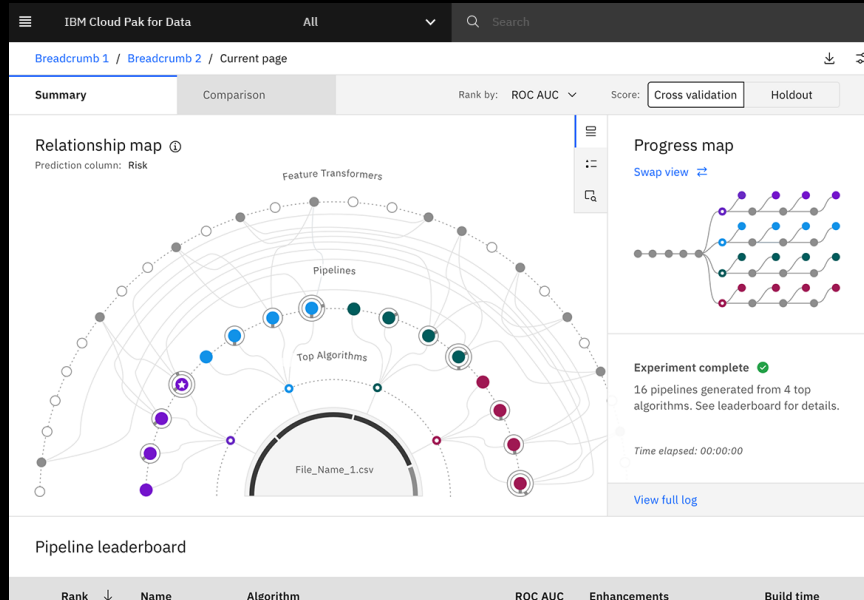
Data Science Year 1: I need GPUs!

Data Science Year 2: I need the right question

Data Science Year 5: I just need good data

Watson Studio and Watson Machine Learning on Cloud Pak for Data 3.5

AI development and deployment



Custom Runtimes

Users can bring in libraries of their choice via custom images to analyze data, build models in notebooks or scripts and deploy in WML.

Business Value: provides the extensibility and flexibility required by data science teams to create AI solutions effectively.

Auto AI

Use SDK to run AutoAI experiment through programming without UI

Tech Preview Feature: AutoAI support multiple input datasets with configurable join relationships

Business Value: New features make AutoAI suitable for automated and customized workflow through programming, saving users time for data preparation when dealing with multiple datasets

Watson Studio and Watson Machine Learning

- Use Python 3.7.9 version with Notebooks and Scripts to build model and deploy in production with Watson Machine learning.
- Bitbucket server and self-signed certificates support for Git integration
- Introducing Multi-Cloud Machine Learning to Cloud Pak for Data (Tech Preview)
- **Business Value:** Keep up to date with the latest innovations in open source for your AI lifecycle. Use enhanced git integration for collaboration on your data science projects. Train your machine learning models by utilizing the data distributed across multiple parties or locations.

Deep Learning use cases used today by leading companies



Provide **personalized** recommendations with up-to-date context



Increase accuracy of **fraud detection** and mitigate risks



Advance **image analysis**, scientific research, and drug discovery



Adapt to changing conditions for autonomous vehicles



Predict health problems and **disease** symptoms



Protect **critical** infrastructures and speed response



Accelerate and drive new insights from **genomic** sequencing



Optimize offers and **claims** management while detecting potential abuse

Watson Machine Learning Accelerator

Get more from your deep learning investments

IBM Watson Machine Learning Accelerator Workload Monitoring Resource Plan

Applications

Filter by Type Status

Filter table

ID & Name	Status	Type	CPU	GPU	Pack ID	User	Submit
bjwila04-31 pytorch-mnist-edt-gpu	Pending	Training	- / 0	- / 4	-	wmla-user	10/20/2020, 03:06 PM
bjwila04-30 pytorch-mnist-edt-gpu	Running	Training	0.10 / 0.1	4 / 4	-	wmla-user	10/20/2020, 03:05 PM

IBM Watson Machine Learning Accelerator Workload Monitoring Resource Plan

Applications

Filter by Type Status

Filter table

ID & Name	Status	Type	CPU	GPU	Pack ID	User	Submit
bjwila04-31 pytorch-mnist-edt-gpu	Running	Training	0.10 / 0.1	2 / 4	-	wmla-user	10/20/2020, 03:06 PM
bjwila04-30 pytorch-mnist-edt-gpu	Running	Training	0.10 / 0.1	2 / 4	-	wmla-user	10/20/2020, 03:05 PM

IBM Watson Machine Learning Accelerator Workload Monitoring Resource Plan

Applications

Filter by Type Status

Filter table

ID & Name	Status	Type	CPU	GPU	Pack ID	User	Submit
bjwila04-31 pytorch-mnist-edt-gpu	Running	Training	0.10 / 0.1	3 / 4	-	wmla-user	10/20/2020, 03:06 PM
bjwila04-30 pytorch-mnist-edt-gpu	Running	Training	0.10 / 0.1	1 / 4	-	wmla-user	10/20/2020, 03:05 PM

Highlights

Accelerate deep learning as part of AI lifecycle on a single platform

Dynamically scale training without disruptions

Accelerate training runs with unique performance optimization

Auto-scale inferencing for faster insights

Share GPUs across data scientists for high system efficiency.

Business Value

94%

Scaling efficiency for training from 6 to 48 GPU¹

11x

Faster GPU performance with accelerated ML library¹

45%

Inference throughput enhancement vs. open source¹

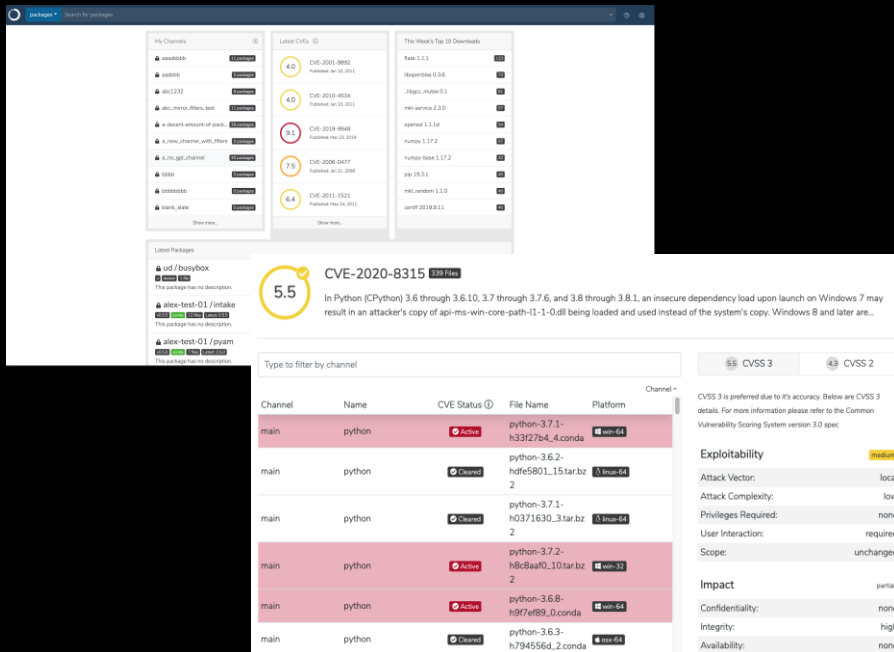
¹ Based on IBM and client laboratory testing.

Capabilities and Use Case

- Use GPU optimized deep learning to understand customer sentiment, model financial risk, enhance diagnostics, identify defects or improve operational safety and efficiency
- Define GPU sharing policies across data scientists and projects to ensure the highest priority work doesn't have to wait
- Improve deep learning model accuracy faster with parallel, hyper-parameter optimization using your preferred search algorithm
- Reduce IT costs and improve utilization by dynamically sharing GPU optimized servers across multiple line of business data science teams

Anaconda Repository for Cloud Pak for Data 3.5

Enterprise-grade package manager for data science and machine learning packages



Highlights

Innovation: Access to a curated repository of open source packages sourced from experts and updated in real time

Management: Easily put open source and proprietary packages in hands of your team

Security: Ensure your teams use only approved packages according to enterprise standards

Business Values

Leverage the wisdom and innovation of the crowd with confidence

Enable your teams to be more productive and efficient with easy to distribute data science packages

Reduce risk by getting steps ahead of vulnerabilities and threats

Anaconda Repository for IBM Cloud Pak for Data

Anaconda Repository brings comprehensive, curated repository of more than 7,500 powerful open-source data science and machine learning packages to a multi-cloud, governed data and AI platform

Federated Learning Use Cases: a new technology that enables multiple organizations to train ML models collaboratively without sharing data



Improve **anti-money laundering** predictions by multiple banks collaborating in building the models



A bank with **outsourced IT operations** improving AIOps without sharing data



Mobile apps personalization without needing to share private information



Self-driving cars limiting volume of data transfer and accelerating learning processes



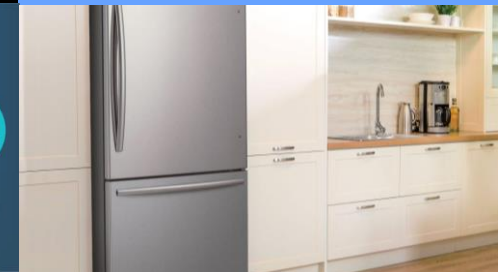
Companies training ML models on data stored in several **geographically distant** locations



Insurance models developed through collaboration of multiple insurers and reinsurers



Aligning with regulations such as HIPAA for **digital health**



Edge and embedded to limit data transfer and processing power needed at the edge

IBM Federated Learning

Multiple organizations training ML models collaboratively without sharing data

IBM Federated Learning

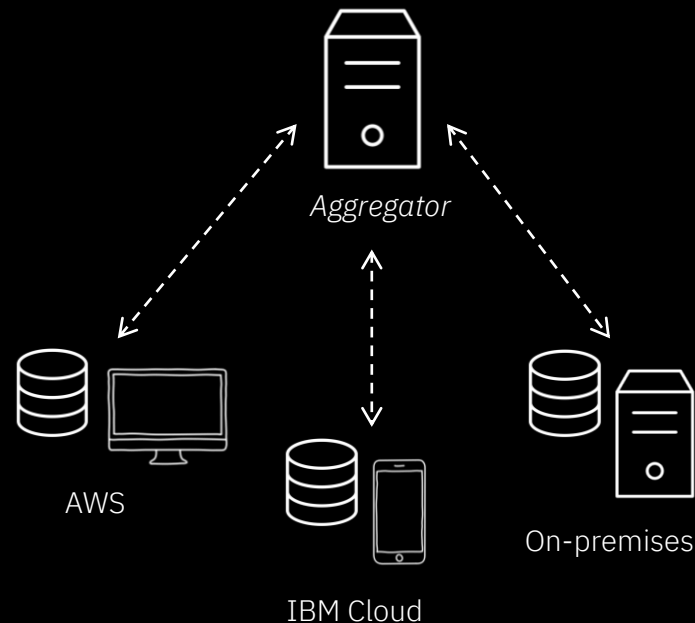
Tech Preview

- Machine learning without moving the data
- Python package of federated machine learning algorithms
- Developed by IBM Research

Use Cases

Reasons to use federated learning

- Data privacy, confidentiality, and security concerns
- Regulatory compliance requirements
- Data volumes too large to move (e.g., minimize egress charges)



Federated Learning

Example an Anti-Money Laundering use case

Use Case: Anti-Money Laundering

Improving detection of money laundering using Federated Learning to reduce compliance costs

Business Challenge:

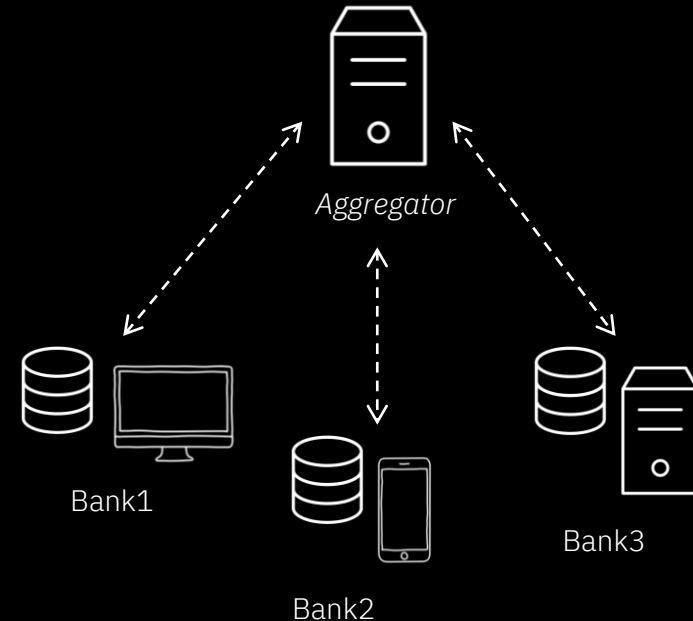
- Without Federated Learning banks must train ML models on their own, resulting in suboptimal models.

Goals:

- Three banks team up using Federated Learning to improve detection of money laundry patterns beyond individual banks performance
- Reduce compliance costs

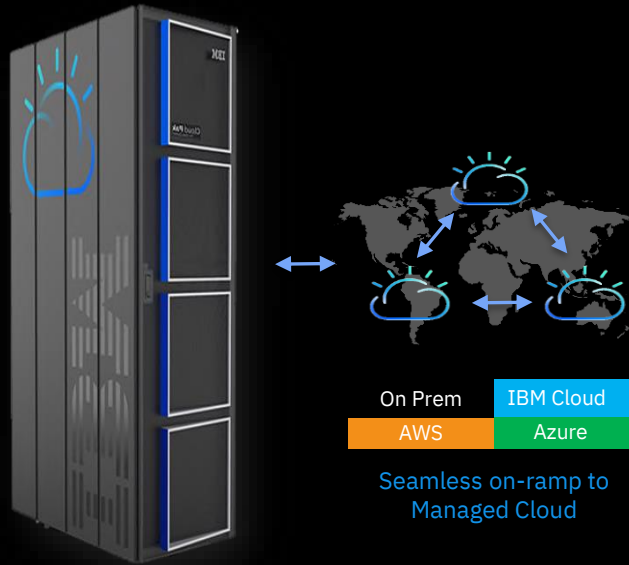
Results:

- Over 20% reduction in ML model false positive rate. This ML model improvement can save banks numerous person-months in unnecessary money laundering investigations



Netezza Performance Server for v3.5

Hybrid, multi cloud data warehouse and analytics platform



Highlights

Simplicity

Minimal administration and tuning

Scalable Hybrid Analytics

petabyte scaling, independently scale compute and storage in cloud

Seamless Data Integration

Built-in Data Virtualization to in-place connect, manage and query data as one

Resiliency for Business Continuity

Infrastructure resiliency, backup to object storage, replicated to multiple availability zones

Business Value

Build once, Run anywhere

Flexible deployment options, no vendor lock-in, 100% compatibility, risk free frictionless migration

Faster Actionable insights

Blazing speeds, up to 3X performance and 2X concurrency improvement over legacy Netezza systems

Data Science & ML at Scale

Operationalize AI with built-in Watson DS/ML and 200+ in-database algorithms

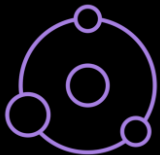
Use cases

On-premise, Cloud or Hybrid environments: Flexibility in choosing deployment and consumption model with License Flexibility that best suit the business needs

Hybrid scenarios:

- **Dev/test environments on Cloud:** Easily spin up/down a cloud instance, seamlessly move data from on-premise to cloud instance with a single command
- **Disaster recovery on Cloud:** Backup to cloud object storage and restore to your Cloud data warehouse, switch over if disaster strikes
- **Make the move to Cloud on your own terms:** If Cloud is your strategic direction, start small, scale storage and compute independently, when you're ready

Client Benefits of Cloud Pak for Data v3.5



Innovation for Digital
Transformation

Accelerated AI

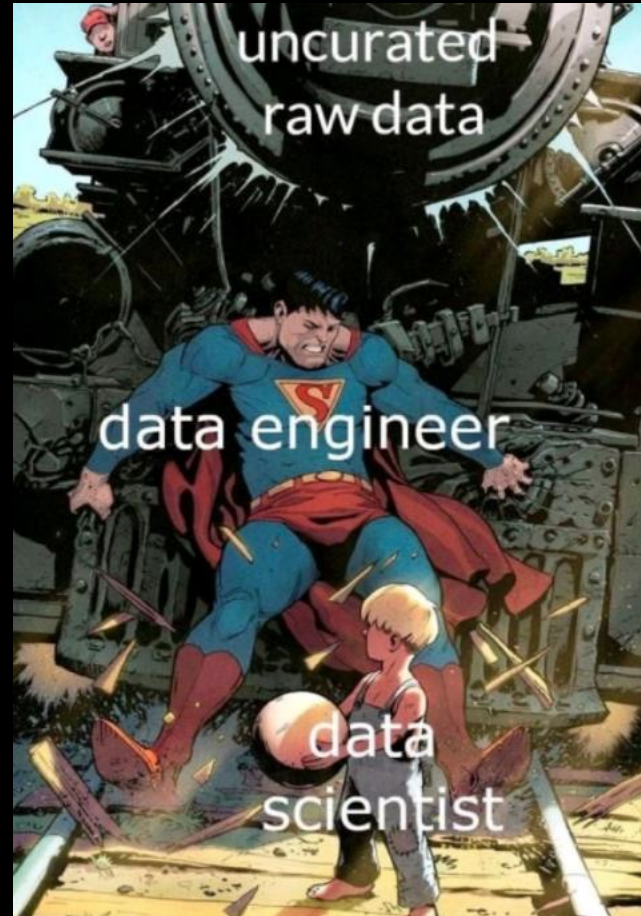
- Watson Machine Learning Accelerator
- Federated Learning (Tech Preview)

Improved Data Governance & Security

- Data Privacy (Tech Preview)
- Knowledge Accelerators
- OpenPages Cartridge

Expanded Ecosystem

- New 3rd Party Extensions
- IBM Open Data for Industries Cartridge
- Z Support for Db2 Databases



All the capabilities for DataOps in a single experience

Data governance

Business
glossary

Policy
management

Reference data
management

Data lineage

Classification

Data quality

Data
discovery

Business term
suggestions

Data profiling
and analysis

Data quality
issue detection

Knowledge
catalog

Data consumption

Policy
enforcement

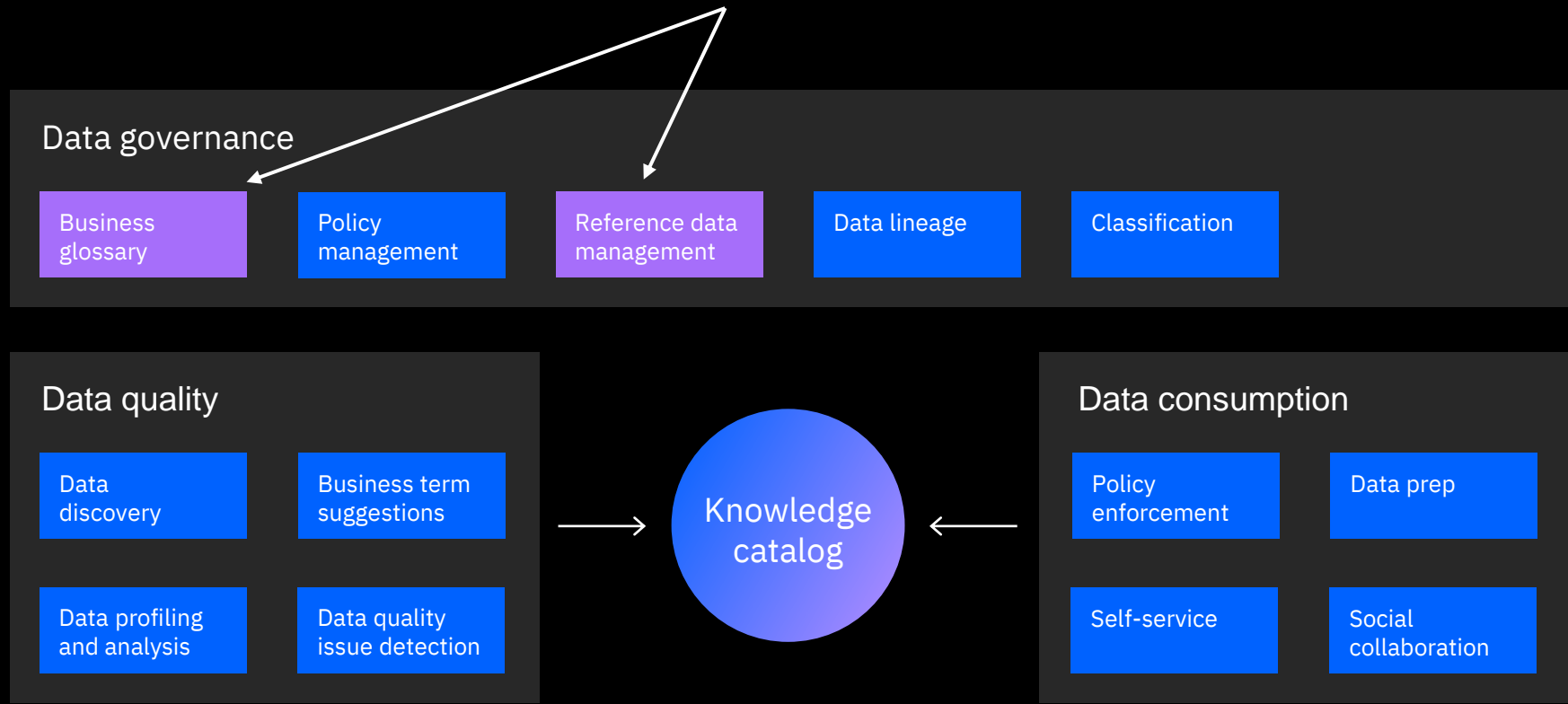
Data prep

Self-service

Social
collaboration

All the capabilities for DataOps in a single experience

Now with prebuilt industry-specific Knowledge Accelerators



Knowledge Accelerators

Industry-specific content for Watson Knowledge Catalog



Digital transformation
Data Stewards

Knowledge Accelerators

Content offering

- Aligns concepts from industry regulations and standards with your business data to accelerate regulatory compliance
- Pre-created curated industry-specific content for Energy and Utilities, Healthcare, Insurance and Financial Services

Problems Solved

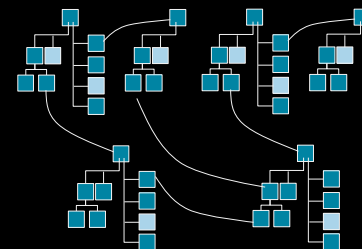
Reasons to use Knowledge Accelerators

- Out-of-the-box compliance support for multiple regulations and industry standards
- Streamline data governance workflows with an industry-specific enterprise vocabulary
- Simplify self-analytics by connecting data assets with business users

Watson Knowledge Catalog

Knowledge Accelerators

Single integrated repository of industry-specific business knowledge to fully support Data Governance



Data Privacy

Cloud Pak for Data Base Service

Provision de-identified data for data science, business analytics and testing.

Tech Preview Feature, GA planned for Q1 2020

IBM Cloud Pak for Data

Breadcrumb 1 / OnlineOrders /

Template: OnlineOrders

Run OnlineOrders1 Job

Cancel Back Next

Review data Set target options

Tables	OPTIM_SALES
5 tables	Find columns
OPTIM_CUSTOMERS Driver	Column name Data type Data protection rule
OPTIM_SALES Related	SALESMAN_ID CHAR(6) No
OPTIM_DETAILS Related	Conditions: Sales_ID: 3294
OPTIM_ITEMS Related	FIRST_NAME VARCHAR(15) Yes
OPTIM_ORDERS Related	LAST_NAME VARCHAR(15) Yes
	NATIONALITY VARCHAR(30) No
	NATIONAL_ID VARCHAR(30) Yes

Protect Data, Maintain Utility

Create data protection policies with **advanced, format-preserving de-identification options** to apply to all enterprise data.

Business Value: Easily comply with data privacy regulations (GDPR), while maintaining data usefulness for **data modeling, business analytics and testing**.

Combine Governance with Data Protection

Leverage **IBM Watson Knowledge Catalog** to auto-discover data classes, define custom data classes and manage data protection policies.

Business Value: Ensure that data is **consistently protected across the enterprise** and aligned with governance strategy.

Self-Service Data Provisioning Experience

Simplified **data subset & movement process**, enabling end users to provision data in minutes.

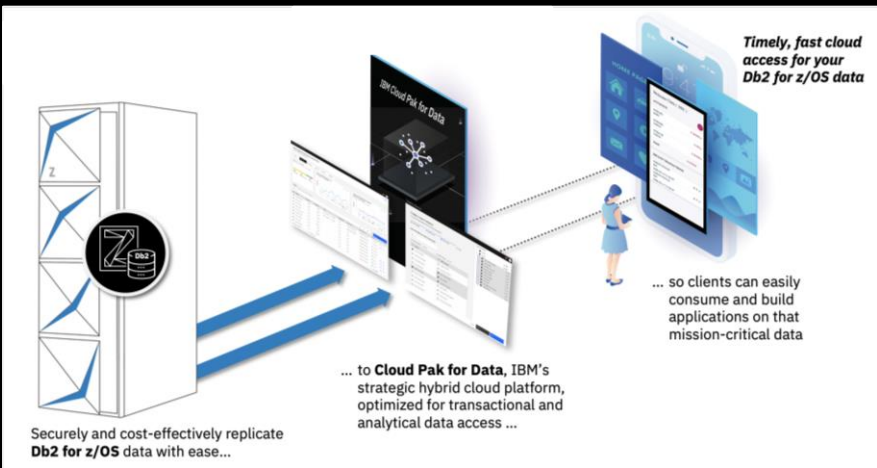
- Search for data by Natural Language Query.
- Define reusable template jobs to share with other users.
- Create data jobs to load **masked data copies and subsets** into environments for immediate use.

Business Value: Increase productivity of end data users with self-service data extraction and keep up with fast-paced project cycles.

Db2 for z/OS Data Gate on Cloud Pak for Data v3.5

Easy access to mainframe data

Extension of Db2 for z/OS into the Cloud



Easier to operate

Improved **ease** of installation and configuration through simplified security setup and certificate generation on z/OS

Streamlined UI process for **provisioning** and **configuring** a new instance of the Db2 Data Gate service

Log Reader now supports High Availability (HA)

Even faster

Improved performance in terms of **increased throughput** and **reduced latency** in both data loading and synchronization from the Db2 for z/OS source on Z to the target database on Cloud Pak for Data

Additional Platforms, including Public Cloud

Clients can now deploy the Db2 Data Gate service on **IBM Cloud** or **Amazon Web Services**

Clients can deploy the Db2 Data Gate service on Red Hat OpenShift clusters that run **Linux on Z**

OpenPages Extension

A market leader in risk management

OpenPages

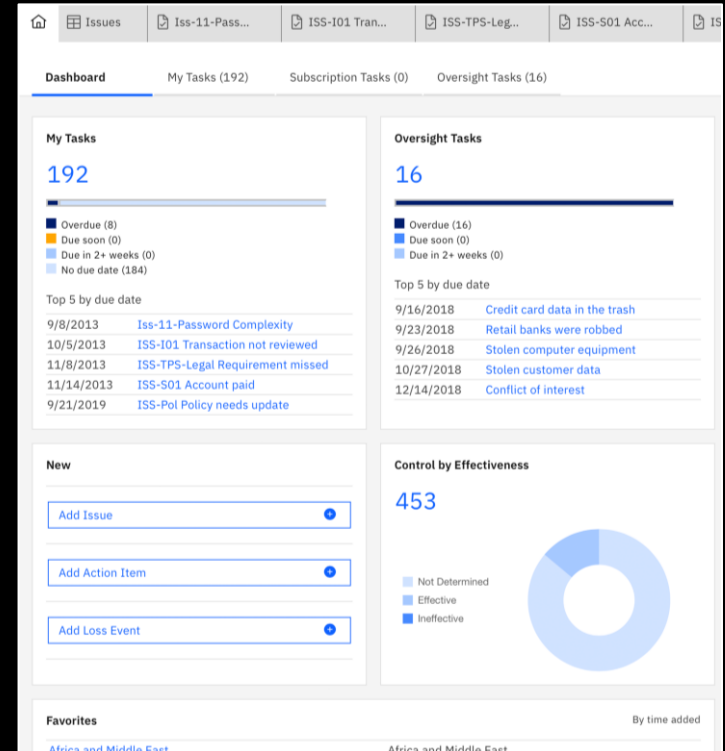
Governance, risk and compliance

- Integrated governance, risk, and compliance (GRC) suite
- Empowers managers to identify, manage, monitor, and report on risk and compliance initiatives
- A leader in Gartner Magic Quadrant for “IT Risk Management” and “Integrated Risk Management”

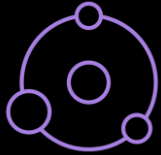
Use Cases

Reasons to use OpenPages

- Operation risk management
- Financial controls management
- Internal audit management
- Regulatory compliance management



Client Benefits of Cloud Pak for Data v3.5



Innovation for Digital
Transformation

Accelerated AI

- Watson Machine Learning Accelerator
- Federated Learning (Tech Preview)

Improved Data Governance & Security

- Data Privacy (Tech Preview)
- Knowledge Accelerators
- OpenPages Cartridge

Expanded Ecosystem

- New 3rd Party Extensions
- IBM Open Data for Industries Cartridge
- Z Support for Db2 Databases

Expanded Ecosystem



Digital transformation
All users

NEW!



hazelcast®

TRILI



Prolifics.



OA4ERP
Operational Analytics for ERP



mongoDB



portworx

Senzing



PostgreSQL

figure
eight



wand®
Prebuilt Taxonomies and Ontologies



Lightbend

ANACONDA.

GUAVUS
a Thales company



EVERSTRING

personetics®



crunchydata

TAVANT



People Data Labs



**CAMBRIDGE
SEMANTICS**

YASKAWA



EQUINIX
WHERE OPPORTUNITY CONNECTS

precisely



EUROTECH
Imagine. Build. Succeed.



EQUIFAX

**FINDABILITY
SCIENCES**

Cloud Pak for Data v3.5

Deployment Options

Deployment Options – Run Anywhere



Cloud Pak for Data System

A true plug-and-play all-in-one enterprise Data and AI platform pre-integrated with all the necessary hardware and software components

Cloud Pak for Data as a Service Starter

A starter set of IBM Cloud Pak for Data services delivered as-a-Service. Modernize how you collect, organize and analyze data and infuse AI with no installation, management or updating required

Red Hat OpenShift (ROKS) on IBM Cloud

Managed service providing secure and resilient OpenShift Container Platform. Clients buying Cloud Pak for Data can run one instance on IBM Cloud **free** for up to 1 year - learn more [at this link](#)

Cloud Pak for Data Initial Support for IBM Z and LinuxONE

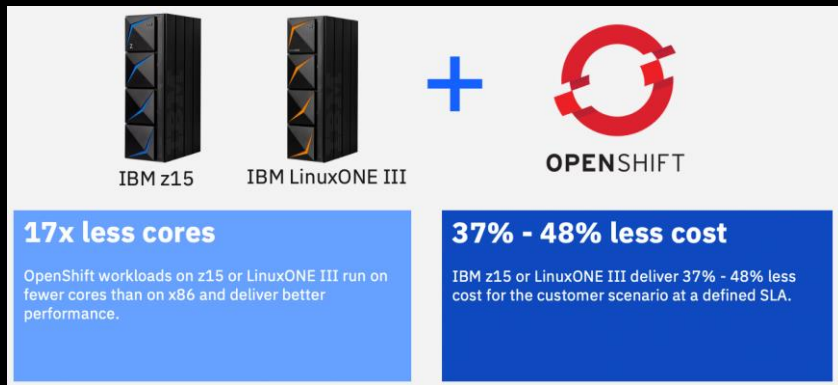
Rapidly modernize, provision, deploy, and consolidate Db2 distributed databases on an OpenShift-based private cloud on IBM Z and LinuxONE

Cloud Pak for Data on Power Systems

Simplify your information architecture and put your data to work on an enterprise system built for data and AI workloads – learn more [at this link](#)

Z Support

Z support for Db2 databases



*Consolidation and Cost Benefits of running OpenShift on **z15** over **x86** using real banking customer scenario*

Results highly dependent on workload and environment

**Analysis provided by the Competitive Project Office*

**Banking workload running Java application microservices in WebSphere Liberty*

Business Challenges

- Complicated deployment and management for database deployments
- Increased operational cost and decreased agility
- Higher IT infrastructure and development platform costs

Business Value

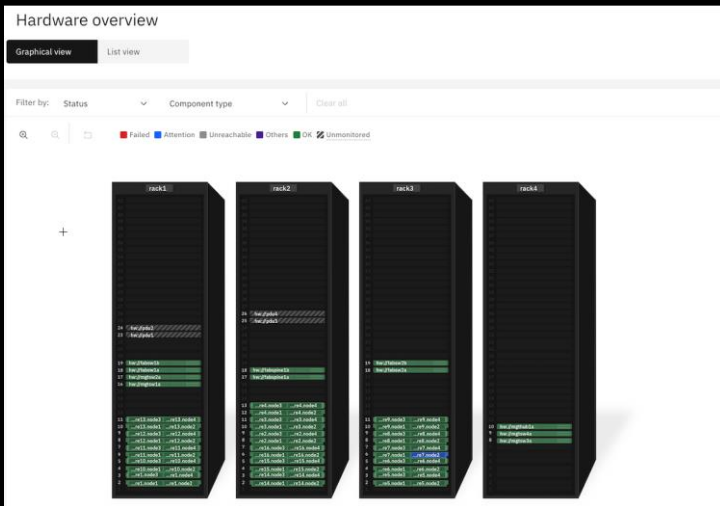
- Modernize data architecture at your own pace with time to value
- Consolidate infrastructure on Z for better TCO and availability
- Stay secure and compliant with superior security features of Z
- Co-location with z/OS backed to achieve lower latency

Capabilities

- Rapidly deploy and provision Db2 databases (*incl. Data Management Console*) and Db2 for Z/OS Data Gate (*Beta*) on Linux for IBM Z
- Modernize with cloud native for support on multi-cloud and hybrid environments with the scalability and resiliency of Z
- Rapidly deploy and provision Db2 databases on a private and public clouds on OpenShift on Z
- Eliminate need for manual provisioning
- Upgrade standalone products to cartridges for workload modernization

Cloud Pak for Data System

A true plug-and-play all-in-one enterprise Data and AI platform pre-integrated with all the necessary hardware and software components



Benefits of Cloud Pak for Data System:

High
performance

Security

Experience
and support

Cost
efficiency

Highlights What's New

Cloud Pak for Data v3.5 support (with OCP 4.X support)

Connector nodes to expand your storage options

Rack simplifications that combine rack and power in a single part

New intuitive multi-rack Hardware Visualization for easy troubleshooting

Ability to ship to Russia as-is

Support for "Data Exploration" and "Data Science" Use Case

Business Value

IBM will install and setup your Cloud Pak for Data System on-site as a "Cloud-in-a-Box" combining storage, compute, networking, and software into a single system

Start out small and scale your Cloud Pak for Data System depending on your need

No lengthy hardware purchase and verification process

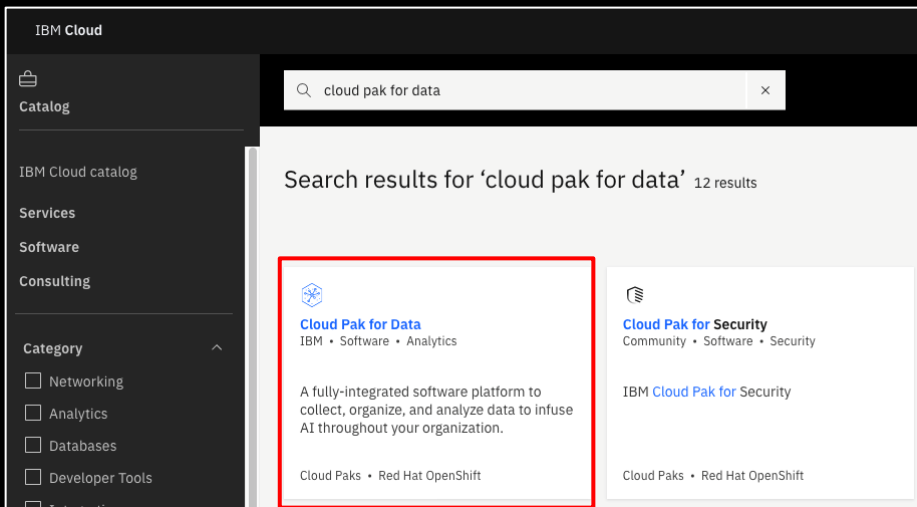
New proactive hardware maintenance support

Use Cases and Capabilities

- **Brand new hardware and rack visualization** making it more intuitive when managing your System
- **Automated AI Lifecycle Use Case:** an end-to-end streamlined data science experience from preparation to development to deployment to correction in a single system. System will be equipped with the necessary services for a data science-ready workflows
- **Data Modernization Use Case:** a flexible plug-and-play data exploration experience equipped for low/no-code and code-based exploration environments. System will be equipped with the necessary services for a data exploration-driven workflows

Red Hat OpenShift on IBM Cloud (ROKS)

Deploying Cloud Pak for Data on managed OpenShift on IBM Cloud in minutes



For more details: [link](#)

Highlights

- The easiest way to run and manage Cloud Paks software on IBM Cloud
- Easy-to-use provisioning guide gets you up and running in minutes

Business Value

- Reduce the need for in-house OpenShift Container Platform skills
- Run securely on IBM Cloud

Supported Configurations with v3.5

Deploying on Classic infrastructure

- PoC/MVP use IBM File Storage (Gold) storage
- Production use Portworx Enterprise storage

Deploying on VPC Gen2 infrastructure

- Non-Production and Production use Portworx Enterprise storage

Cloud Pak for Data

BYOL Multi-Cloud Deployments

Trial Signup: <https://www.ibm.com/account/reg/us-en/signup?formid=urx-42212>

Deployment Options: https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/cpd/plan/deployment-t-

IBM Cloud

60-day trial

Managed OpenShift
Container Platform

IBM Catalog

Getting Started

➤ Quick Start

<https://cloud.ibm.com/catalog/content/ibm-cp-datacore-6825cc5d-dbf8-4ba2-ad98-690e6f221701-global>

➤ Deployment Guide

<https://cloud.ibm.com/docs/cloud-pak-data>

➤ Terraform Scripts:

<https://github.com/IBM/cp4d-deployment/tree/master/ibmcloud>

➤ Sales Play:

<https://ibm.seismic.com/Link/Content/DC4lWTvh5YYE-UEYEqSBWGZQ>

AWS

60-day trial

OpenShift
Container Platform

Quick Start

Using AWS Cloud Formation Templates

Getting Started

➤ Quick Start

<https://aws.amazon.com/quickstart/architecture/ibm-cloud-pak-for-data/>

➤ Deployment Guide

<https://aws-quickstart.s3.amazonaws.com/quickstart-ibm-icp-for-data/doc/ibm-cloud-pak-for-data-on-the-aws-cloud.pdf>

➤ Terraform Scripts:

<https://github.com/IBM/cp4d-deployment/tree/master/aws>

➤ Sales Play (for Infrastructure credits):

https://ibm.seismic.com/Link/Content/DC_G2xH2WTDUW8Y-qIT-qAAw

Azure

60-day trial

OpenShift
Container Platform

Quick Start and Marketplace

Using Azure Resource Manager Templates

Getting Started

➤ Quick Start

https://azuremarketplace.microsoft.com/en-us/marketplace/apps/ibm-alliance-global-1560886.cloud_pak_for_data?tab=overview

➤ Deployment Guide

<https://github.com/Azure/azure-quickstart-templates/tree/master/ibm-cloud-pak-for-data>

➤ Terraform Scripts:

<https://github.com/IBM/cp4d-deployment/tree/master/azure>

Resiliency & Performance in Cloud Pak for Data v3.5

Site Reliability Engineering (SRE) Testing

- Extensive and continuous reliability and resiliency on long-running clusters
- 10,000+ automated testing scenarios running daily
- Daily tests running across all supported deployment options and storage options

Chaos Monkey Testing

- Significant improvement in pod restarts and node reboots
- Enhanced chaos scenarios for testing CPU, memory, network, and storage stress and failures
- Automatic limit testing and resource constrained testing

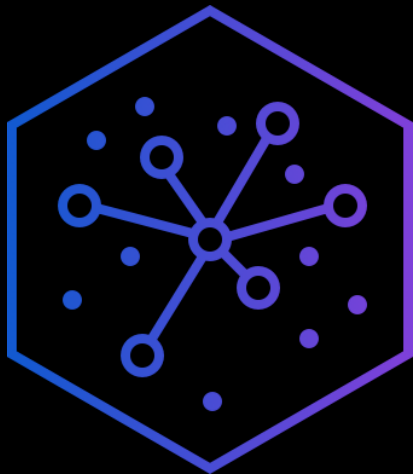
Performance

- Added 100+ new performance testing scenarios
- Significant improvement in time to install by several services

Our clients benefit significantly from the millions of dollars we invest every release into ensuring product quality, resiliency, and performance

Cloud Pak for Data 3.5

Version specs



OpenShift Version

v3.11, v4.5

Deployment Options

Azure, AWS, Google Cloud, IBM Cloud, Cloud Pak for Data System, Power, Z

Storage

OpenShift Container Storage v4.6, NFS, Portworx v2.5.5

Third party ISVs

50+

IBM Data and AI services

35

Industry Accelerators

31

User Personas

Data scientists, data engineers, data stewards, business analysts

Cloud Pak for Data v3.5: *Transform your business to survive and thrive*

1. Cost Reduction *to persist through the downturn*

Enhanced Automation

- Cloud Pak for Data “Operator” automates installation and lifecycle management
- Automated discovery and quick scan improvements, enhancements to governance workflows
- Automated query optimization in data virtualization for better performance

Simplified Administration

- Platform management with advanced resource quotas
- Manage production workloads with Deployment spaces
- CLI commands and APIs to manage services, backup, restore projects, export/import metadata and access data virtualization data

Improved User Experience

- Guided walk me flows and product tours to get started
- Navigation improvements and customization to enhance productivity
- Platform Connections to streamline data access
- Data Management Console: integrated database management tool

2. Innovation *for digital transformation*

Accelerated AI

- Watson Machine Learning Accelerator
- Federated Machine Learning (*Tech Preview**)

Improved Governance and Security

- Enhanced data privacy capabilities (*Tech Preview**)
- User groups for more granular access management
- Store security secrets in a secret vault: Tokens, API Keys etc.

Expanded Ecosystem

- New IBM cartridges: OpenPages and Open Data for Industries
- 31+ Industry Accelerators and 50+ ISV Services
- Support for Z (Db2), Red Hat and Tech Data Marketplaces

Definition of “Tech Preview”*: Licensee may only use the following components as part of Licensee's internal development and test environment for internal non-production activities ("Non-Production")

Next Steps

IBM A3 Center

<https://www.ibm.com/industries/federal/analytics>

Kickstart your data science project

<https://www.ibm.com/community/datascience/elite/>

Reach out: jentwistle@us.ibm.com

[linkedin.com/in/jonathan-entwistle-a860315/](https://www.linkedin.com/in/jonathan-entwistle-a860315/)

The AI Ladder

A prescriptive approach to the journey to AI



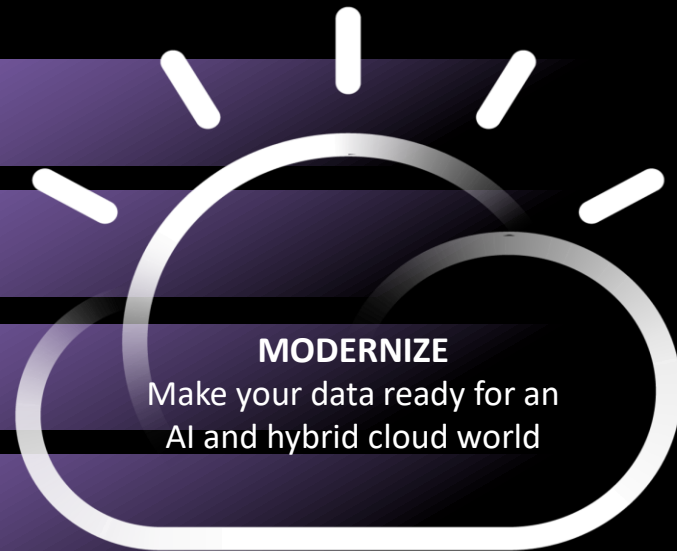
AI

INFUSE - Operationalize AI throughout the business

ANALYZE - Build and scale AI with trust and transparency

ORGANIZE - Create a business-ready analytics foundation

COLLECT - Make data simple and accessible



MODERNIZE

Make your data ready for an
AI and hybrid cloud world

Talent &
Skills



**One Platform,
Any Cloud**

Modernize

-

Security Vaults

User Groups

Security Vaults

Business Challenges

Credentials and secrets need to be safeguarded with access control and audit logging

Passwords and sensitive information should never be stored in plain text within code

Business Value

Securely store and tightly control access to *secrets* (such as passwords, tokens, API keys, encryption keys, and certificates) to protect access to sensitive data.

Store credentials such as passwords, API keys, tokens, certificates, or other sensitive information within a vault.

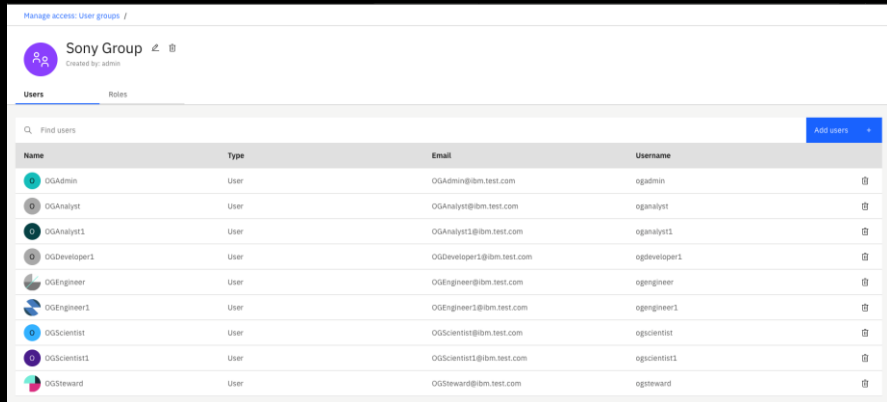
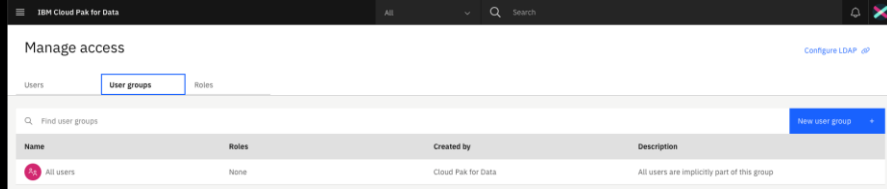
Access vaults using Python from within notebooks, scripts, and other code to store and retrieve secrets.

Capabilities

- A unified interface to manage any secret, providing tight access control and recording a detailed audit log
- Encrypt/decrypt data that is stored elsewhere. For example, an application may use vault to encrypt data that is then stored in a database
- Secret revocation for individual secrets and for trees of secrets, for example revocation of all secrets read by a specific user, or all secrets of a particular type. Revocation assists in key rolling (rotation) as well as locking down systems in the case of an intrusion
- API-driven default platform Vault
- Python SDK



User Groups



Business Challenges

Organizations have numerous business units and divisions who need to group their users based on roles and responsibilities

It is tedious to manually add multiple users one-by-one at different access levels across different data and AI services

Business Value

Efficiently manage users in bulk using groups for multiple departments or different roles/responsibilities in a scalable and sustainable manner

Save time with easier management of users and improved collaborations by extending these groups to all the services on the platform promoting reusability

Capabilities

- A user group is a collection of different users that are assigned a shared set of permissions
- User groups can be created from existing local users and/or LDAP users and groups
- Add user groups and assign roles so that users can collaborate across their organization with the same level of access
- Default “All Users” group used in the global catalog or to be available for services to be able to pick from

Collect

-

Netezza Performance Server

Data Virtualization

Db2

Db2 Big SQL

Db2 for zOS Data Gate

Data Virtualization

Self-Serve, AI/ML Enabled for
Business Users and Citizen
Data Scientists

Highlights

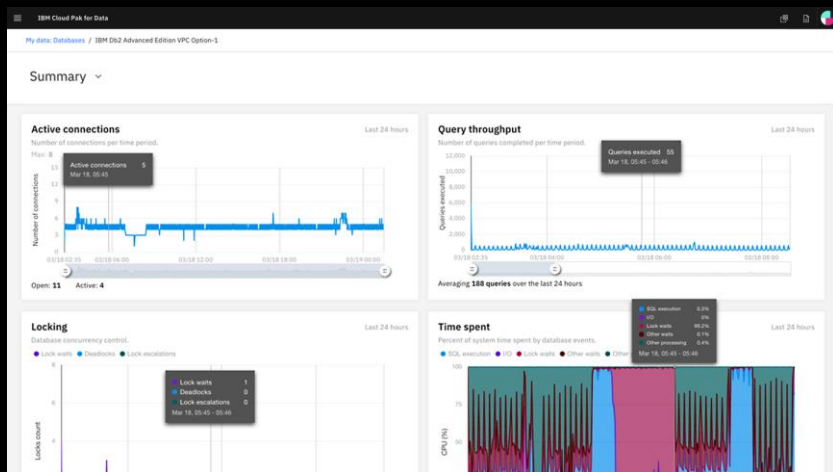
- High-end scalability
- Distributed processing – further optimization for query processing
- Autonomous Cache Recommendation – focuses on further simplification and automation
- API for Virtual Object creation and consumption
- Data Virtualization Monitoring Capability
- IBM Governance/security – continue to extend the framework from DV object discovery to Policy
- Support for SAP S4/HANA and Salesforce

Use Cases / Capabilities

- Real-time unified data reporting
- BI / Integrated analytics– across various silos
- Various dashboard – customers, partners, etc.
- Hybrid-and-multi-cloud data silos reporting
- Microservices applications

IBM Db2 for Cloud Pak for Data v3.5

Database for any workload running everywhere



Link: [Webinar Deep Dive on Db2 enhancements](#)

Highlights for v 3.5

Significant storage cost savings with support for wide range of Software Defined Storage solutions

Leverage high performance, availability and scalability with support for Linux on IBM Z

Single Data Management Console to manage multiple instances of databases

Business Value relative to non-containerized Db2

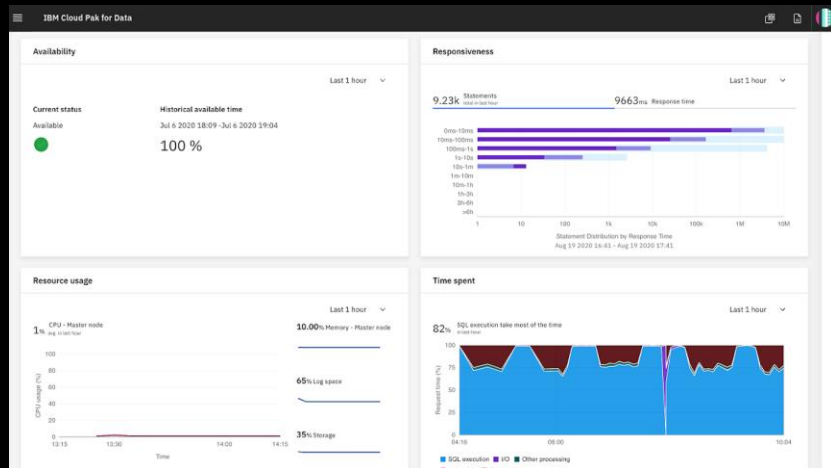
- 26% operations cost savings with automated updates and backups
- 40% less infrastructure costs with Cloud Native database and elastic scaling
- 50% less storage costs with Software Defined Storage

Use cases/Capabilities

- Collect data from all types of data sources and virtualize it with increased time to value and lower costs
- Cloud-native containerized database solution with elastic scalability and portability in hybrid cloud environment
- Install a database in less than 5 minutes and easily manage with automated patches and backup/restore and single data management console.
- Pre-configured database created to work seamlessly with rest of AI services in Cloud Pak for Data

Db2 Big SQL for v3.5

Query data on Hadoop or Object Stores instantly



Highlights

Fast, integrated, secure access layer for data engineering and data science

Uncover deep insights from data stored in Big Data stores like Hadoop or object stores

This compute only SQL engine allows easy portability of traditional applications when data is offloaded to archival data stores

Business Value

- **Immediate access** to data in Parquet, ORC, etc. for actionable insights
- Significantly reduce the query execution of multi-table join (like 12 tables) **from 13mins to 79secs**
- **Increased performance** allowing analytics and data provisioning **from monthly to daily**

Use cases/Capabilities

- Quickly adapt to modern information architecture by having **access to data instantly**
- **Infuse the power of Db2** to data stored open source file formats in Hadoop or object store
- Extend Cloud Pak for Data's use cases to **AWS GovCloud to exploit Watson use cases**
- **SQL power users** can query data where it resides and scaling up/down based on workload or concurrency needs

Organize

-

Watson Knowledge Catalog

InstaScan

DataStage and Information Server

Master Data Management

Watson Knowledge Catalog

Deliver business-ready data for AI and analytics with intelligent data catalog, automated metadata enrichment, data quality analysis and active policy enforcement

New request

Breakglass

Label

Description (optional)

Choose an option

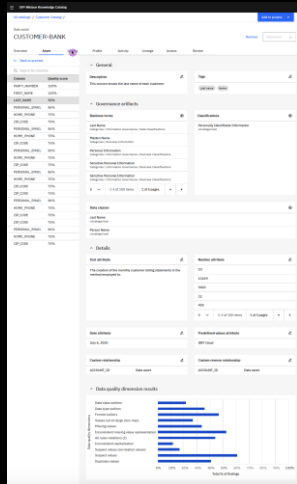
Request title

Request description

Category 33

Collaborators

Name	Email address	Category role	Date added
Sammy Schuckert	sammy.schuckert@ibm.com	Owner of Category 1	Apr 17, 2019
Albert Hauer	albert.hauer@ibm.com	Admin of Category 1	Apr 17, 2019
Andreas Reich	andreas.reich@ibm.com	Editor of Category 1	Apr 17, 2019
Group 1	Email address (0)	Viewer of Category 1	May 4, 2019



Highlights

Automated profiling and classification to scale understanding of your data landscape

View and manage assets, including custom assets, and their relationships

Manage governance artifacts by category users and roles

Manage hierarchies and relationships in reference data

Business Value

- Speeds up metadata classification time for regulations by **90%**.
- Infrastructure needs reduced 50% with end-to-end DataOps services on Cloud Pak for Data and up to **158% ROI**
- Productivity increased by **95%** when Watson Knowledge Catalog and other CP4D services are deployed.

Use Cases

- End-to-end data governance** - Single integrated solution to serve customers' data needs from data ingestion, governance, quality and consumption
- Self-service access to trusted data for analytics** - Enable data consumers to use a self-service, integrated experience to search through catalogs, collaborate with other users and visualize, shape & analyze data
- Support regulatory compliance** - Quickly discover and inventory assets into the catalog, automatically classify and tag them with business terms to detect sensitive data

What's Coming in Watson Knowledge Catalog in CPD v3.5

Categories

- Restrict access to Categories and assign roles like “Owner” and “Admin”
- Build workflows based on Category and Category roles then further refine that workflow by action and artifact type
- Tree navigation and activity history
- Custom attributes on categories

Workflow

- Import custom Flowable templates
- Create new global workflows based on custom forms and templates built in Flowable

Reference Data (Available in a patch for v3.0)

- Manage more complex reference data sets with any number of columns and column types
- Establish hierarchies of values within a reference data set and between reference data sets
- Map reference data values to values in other reference data sets

Catalog

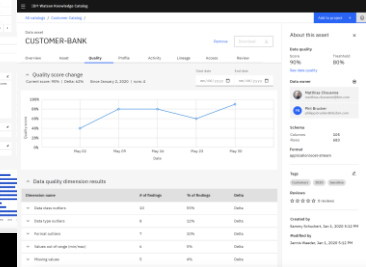
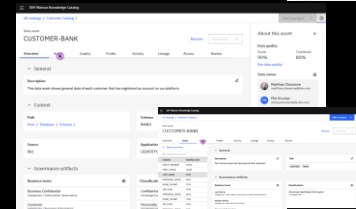
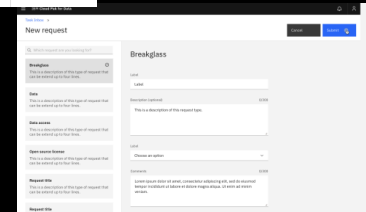
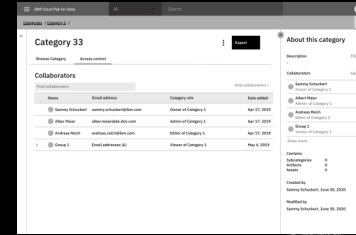
- Create custom assets in the catalog
- Manage relationships between assets in the catalog
- New look and feel for all assets in the catalog that connects data quality scores to the assets

Infrastructure

- Support Disaster Recovery
- Support Backup and Restore
- Support for OpenShift v4.5+
- Import/Export all artifacts and assets via command line

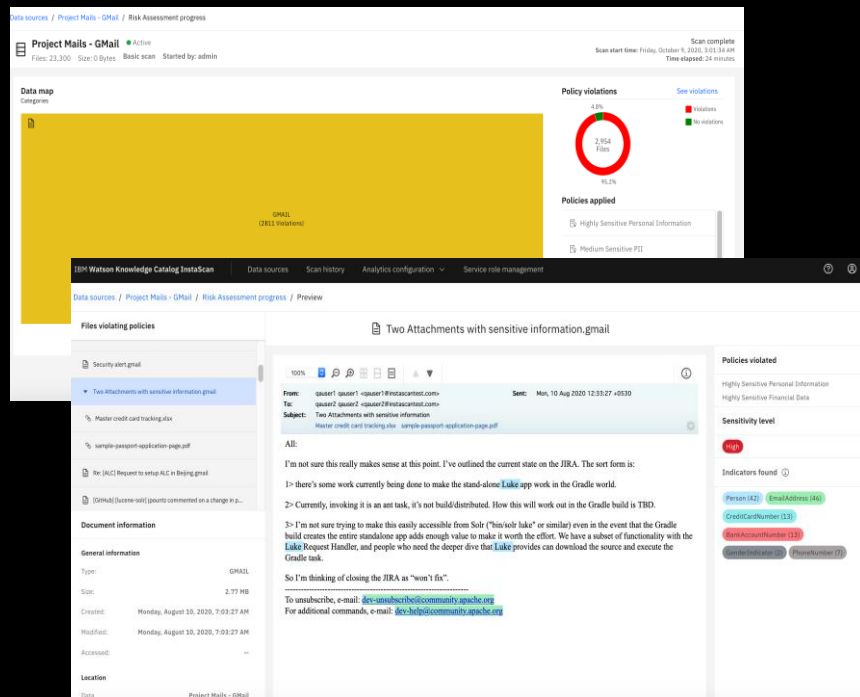
Integrations

- Test Data Management



WKC InstaScan in Cloud Pak for Data v3.5

Assessing Regulatory Risk in Gmail Environments



Highlights

Assess regulatory risk footprint in Gmail environments across the enterprise

Understand which Gmail folders and users have the highest risk footprint to prioritize remediation efforts

Reduce overall regulatory risk to show 'reasonable steps' towards compliance

Business Value

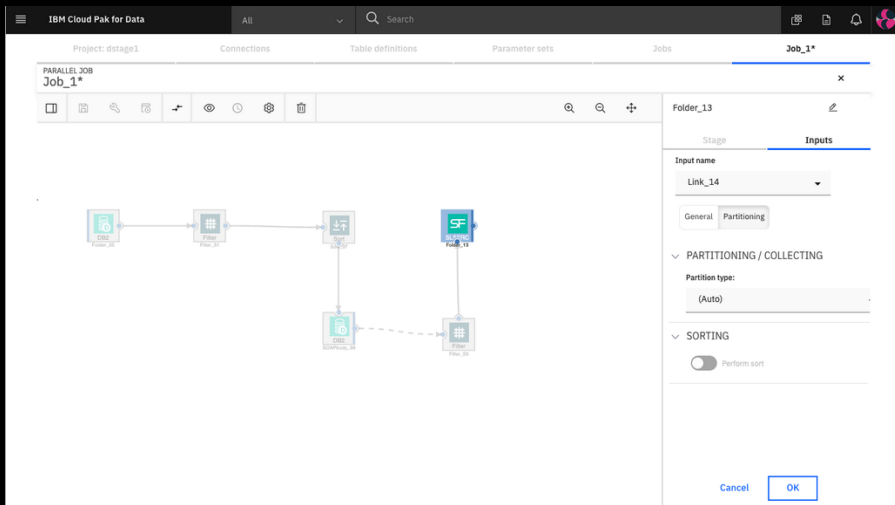
- 90% time saved via random sampling of Gmail data compared to full indexing of each file

Use cases/Capabilities

- Govern unstructured cloud data sources.** Apply your organization's data management and compliance policies to your Gmail data.
- Strengthen data discovery and privacy.** Discover what types of sensitive data are most prevalent within your Gmail instances and prioritize data clean-up activities.
- Gain insights fast.** Use native indexing to get started quickly and run assessments in a fraction of the time.

DataStage and Information Server for Cloud Pak for Data v3.5

High performance data integration



More information [here](#)

Highlights

Dynamic configuration for DataStage and QualityStage jobs

Modern Flow Designer design interface with improved performance

Include unstructured data as part of data integration flow design

Enhancement to Classic Designer (Windows Client) support

Value

- **Up to 30% performance improvement** when executing flows due to dynamic resource allocation
- Design environment performance improvements that **boost user productivity**
- Leverage existing job designs with **benefits of containerized deployment**

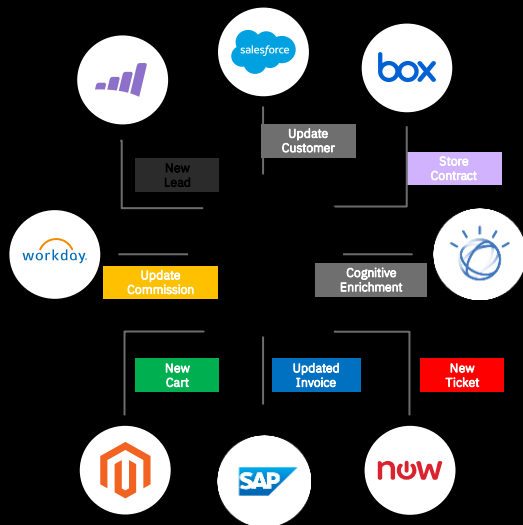
What's new:

Improved performance in a modern interface on key aspects

- **Slowly Changing Dimension (SCD) stage** – very critical data movement task to track history of dimension records or structured data – **save operation performance improved by 52% in new interface**
- **Leading and common targets / sources (Db2, Snowflake, Salesforce)** save operation performance improved between 31 to 48%.
- **Add Azure Data Lake Store and Redshift Connector** to Flow Designer in Cloud Pak for Data
- **SAP Packs supported** via classic designer (windows client)
- Removal of requirement of NodePort entries and addition of end to end encryption from Windows Client to CP4D cluster

Master Data Management 3.5

Bridging customer information via AI/ML to provide a transparent and accurate *view of your data*



A modernized MDM isn't just an investment in MDM's premium capabilities but an investment in the **Next Generation of MDM**, this means operating on a platform with:

- New AI/ML driven entity match engine
- Auto configuration
- Data driven data model definition
- Completely new UX
- Tight integration with Watson Knowledge Catalog*

Highlights

Utilize native REST APIs or IBM App Connect connectors to accelerate application integration

Deploy one or more cache instances to service consumers

Search across entities and traverse relationships for new insights or push data to your data warehouse

Business Value

- Majority of MDM workload are read/inquiry type (**typical 80-95%**)
- Accelerated processing time: **3,000 read side TPS rate** compared to 1,000 TPS rate (~ 200% increase)
- Easy to deploy, offers agility, and provides more scalability than on-premises software at an affordable cost

Use cases/Capabilities

- Provide application developers with an instance of master data for **faster** time to market
- Support mobile and online applications requiring **extremely low latency & high availability**
- Utilize master data for downstream analytics
- Support applications needing more **local access to global Master Data**
- Set up data filters and enforce publishing policies to users and geographies as required

Analyze

-

Watson Studio and Watson Machine Learning

Watson OpenScale

Watson Studio Premium

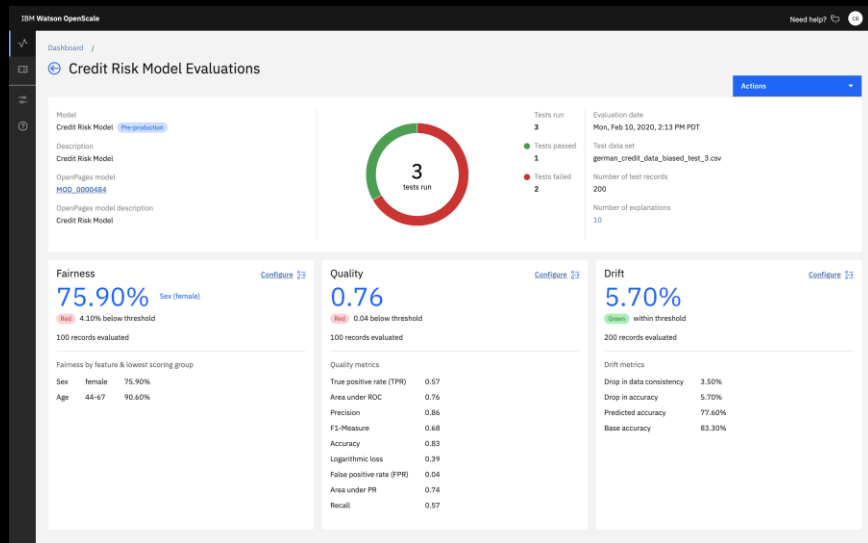
Anaconda Repository

Analytics Engine powered by Apache Spark

IBM Streams

Watson OpenScale on Cloud Pak for Data v 3.5

Operating Trusted AI



Highlights

Role based user access

Explainability enhancements including “what if” interactivity and improved understandability

Monitor models for indirect bias

Business Value

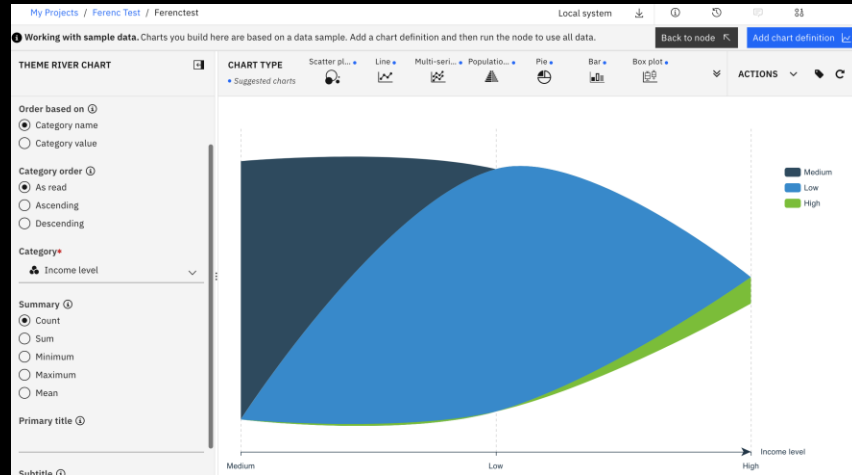
- Share Watson OpenScale across multiple teams of business users and data scientists with appropriate content and function visible to each
- Enable end users of AI applications to better understand model decisions
- Detects potential fairness issues due to unseen correlations in input data

Use Cases

- Monitor production models to ensure accuracy, fairness and explainability
- Ensure models continue to preform as expected over time by detecting and evaluating impact of inputs drifting from data used to build models
- Enable model validators and risk managers to run tests, compare candidates, document results and determine when AI/ML models are ready for production.

Watson Studio Premium for Cloud Pak for Data 3.5

SPSS Modeler, Decision Optimization, and Hadoop Execution Engine



New Chart Builder within SPSS Modeler adds dozens of new chart types and a theme builder for attractive and repeatable visualizations

Highlights

SPSS Modeler includes Extension nodes, new charts, easier migration, and now works with Jobs

Decision Optimization includes engine statistics while running the model as well as with the solution

Hadoop Execution Engine adds support for Spectrum Conductor, and now works with Python scripts

Business Value

- SPSS Modeler unlocks non-coders and business users for data understanding, transformations, visualizations, and model building
- Decision Optimization advances model builder for Operations Research users
- Hadoop Execution Engine expands supported use cases for Spark and Python users

How New Capabilities Translate into Use Cases

- SPSS Modeler opens data science to the non-coders in an organization with repeatable data analysis and model building
- SPSS Modeler also executes all v17 and later SPSS Modeler Streams for centralized production and collaboration
- Decision Optimization engine statistics helps in introspecting the optimized solution without any coding or APIs calls. It also helps in parameter tuning, speed of convergence and improving quality of the solution
- Hadoop Execution Engine now integrates with Spectrum Conductor that brings large scale computing cluster management capability to advanced enterprise machine learning projects

Analytics Engine powered by Apache Spark in v3.5

Custom ML development and deployment

Servers catalog / Analytics Engine powered by Apache Spark /

Analytics Engine powered by Apache Spark

History Server

Running history: https://sparkhistory.ibmcloud.com

Last updated: 2019-10-01 21:08:19

Click here to view more details.

App ID	App Name	Started	Completed	Duration	Spark User	Last Updated	Event Log
app-20191001-10-01-0002	PythonWordCount	2019-10-01 21:04:20	2019-10-01 21:05:00	39 s	spark	2019-10-01 21:05:00	View Log
app-20191001-10-01-0003	PythonWordCount	2019-10-01 21:05:10	2019-10-01 21:05:37	25 s	spark	2019-10-01 21:05:38	View Log

Showing 1 to 2 of 2 entries

Automatically spin up lightweight, dedicated Apache Spark clusters to run a wide range of workloads.

Version: 3.5.0

Description

You can use Analytics Engine powered by Apache Spark to run a variety of workloads on your IBM Cloud Pak for Data cluster:

- Watson Studio notebooks that call Apache Spark APIs
- Spark applications that run Spark SQL
- Data transformation jobs

IBM Cloud Pak for Data

Instances / instance12 /

instance12

About this instance

Created on: October 8, 2020

Status: Running

Access information

Deployment space	Instance12_space
History server endpoint	https://hijobui-cpd-hijobui.apps.hb-ocp...
Spark jobs endpoint	https://hijobui-cpd-hijobui.apps.hb-ocp...
View deployment space	https://hijobui-cpd-hijobui.apps.hb-ocp...
View history server	https://hijobui-cpd-hijobui.apps.hb-ocp...

Storage

Claim name: vol12

Highlights

Capability to consume **Serverless Spark**

Analytics Engine service is **100% open source** which enables faster time to market to provide the support latest versions of Apache Spark

Support for **time series** libraries

Additionally, **columnar parquet file encryption** and **Data Skipping** when reading from COS

New Features in 3.5:

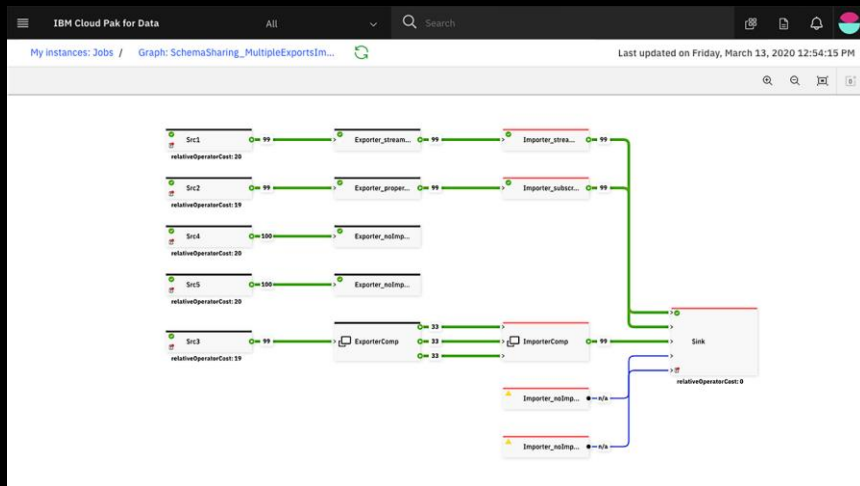
- **Support** for Apache Spark 3.0 and Python 3.7
- **Integrate** with external Kerberized HDP 2.6.5 cluster
- **Customize** Spark configuration through instance manager payload
- **Automated** approach to create Hadoop and Hive Metastore setup
- **Unified jobs** experience

Benefits

- Connect to external data sources like HDFS (HDP 2.6.5), IBM Cloud Object Storage or Red Hat OpenShift Container Storage to run Spark compute workloads
- Optimized performance with IBM Research assets such as Stocator and Data Skipping
- Better integration and a single vendor to deal with from a support perspective
- Platform layer to support logging, monitoring, user management, access control, metering etc.

IBM Streams for v3.5

Acquire, Analyze, Act



Highlights

Custom application resource templates

Edge Analytics (beta) to analyze and act on data where its created

Auto-creation of Cloud Pak for Data service(s) from any point in a Streams application

Business Value

Improved resource utilization by specifying vCPU and memory requirements for each application

Millisecond latency can be achieved to act in the moment

Immediate creation of OpenShift services for discovery and access to analytic applications over standard REST interfaces

Use Cases and Capabilities

- Intelligently Collect data to analyze, filter and summarize real time data before landing it in persistent stores
- Agent Assist to convert speech to text and perform natural language processing to provide recommendations to call center agents
- Real-time situational awareness infused with AI for hospital patients, manufacturing devices and automobiles to improve operations
- Geospatial analytics to create alerts when people or things enter an area of interest for marketing or safety concerns

Link: [Streams Seismic Sales Kit](#)

Infuse

-

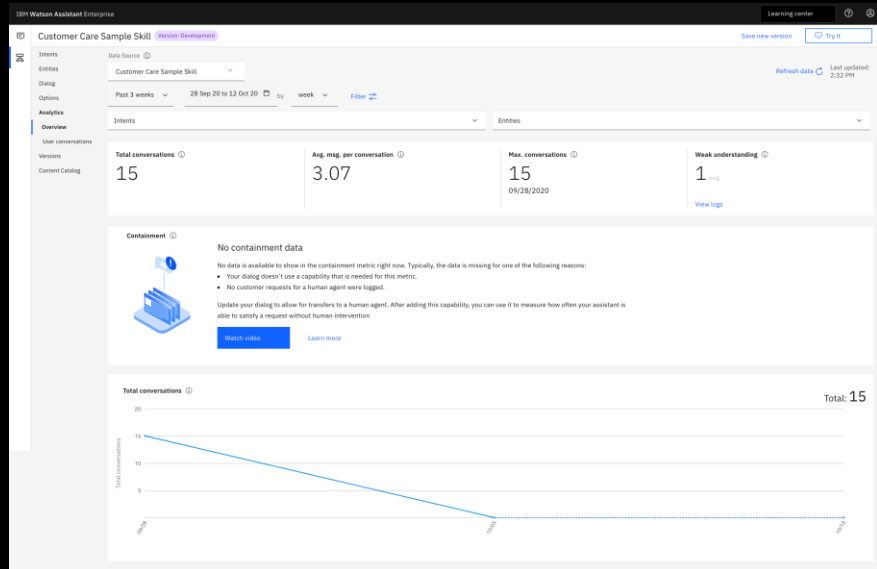
Cognos Analytics

Planning Analytics

Watson Assistant

Watson Assistant for Cloud Pak for Data v3.5

AI Powered Chat With Watson



Web Chat Integration

Users will be able to easily deploy their virtual agents using the web chat integration. Using a snippet of code they add to their website, they will have a fully functional front end

The Preview link integration is also included in this release. It provides a webapp ready to share for testing and validating internally.

Conversation Logging and Analytics

End-user conversations will be logged (message API calls). Clients will be able to see the conversations in the tooling and will be able to download the raw data.

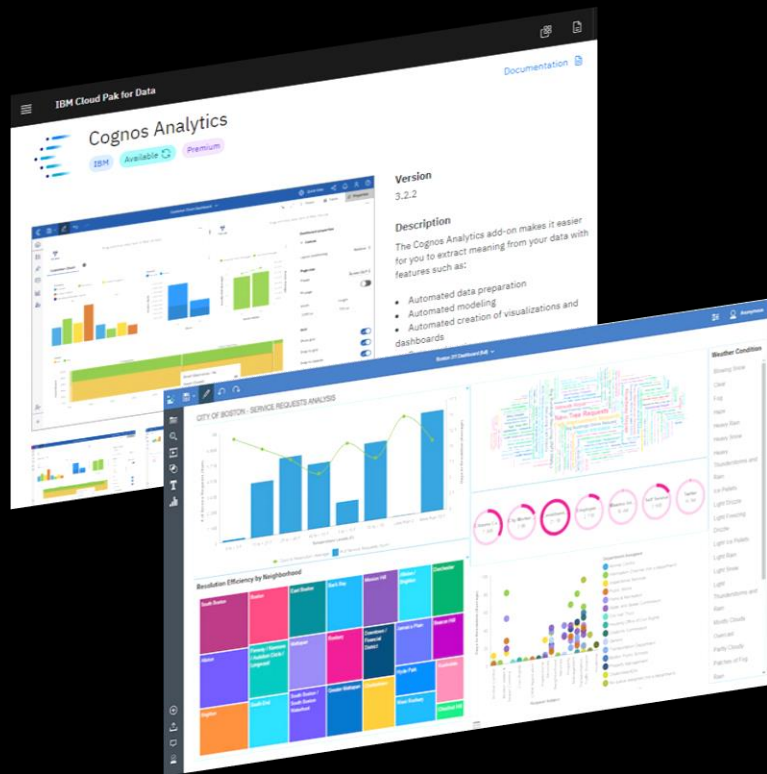
Also, the same Analytics available on public cloud will be available on Cloud Pak for Data.

Additional feature parity

- Full-text search of a dialog skill: Clients can search where a text string appears in their intents, entities and dialog nodes.
- Improved algorithms for detecting conversations that are not covered by a skill. This accuracy improvement will ensure conversations are not incorrectly directed to intents.
- New system entities: This introduces the new and improved system entities that we have on public cloud to our Cloud Pak clients.

Cognos Analytics for Cloud Pak for Data v3.5

AI Driven Business Intelligence



Highlights

Save time with automated data preparation, data discovery and dynamic visualizations and find deeper insights faster

Execute complex queries faster and analyze data where it resides using data virtualization

Create an analytics foundation by integrating business intelligence predictive, prescriptive and planning analytics

Business Value

Empowers users with **AI-infused self-service** capabilities

Easily visualize data and share insights across your team to drive confident decisions

Reduce the complexity of deploying and managing a BI environment to meet your users business needs

Use cases/Capabilities

- Get the self-service you expect, the data governance you require, and the reporting you trust, with a secure business intelligence platform
- Deploy one AI-infused business intelligence platform for all analytics use cases, from marketing campaign performance to human resources analysis, customer sentiment analysis to sales pipeline analysis
- Ensure Managed Reporting Production workload SLAs are met with confidence using modern container architecture
- Enable the most secure and compliant strategy with data governance for managed reporting and data exploration

