

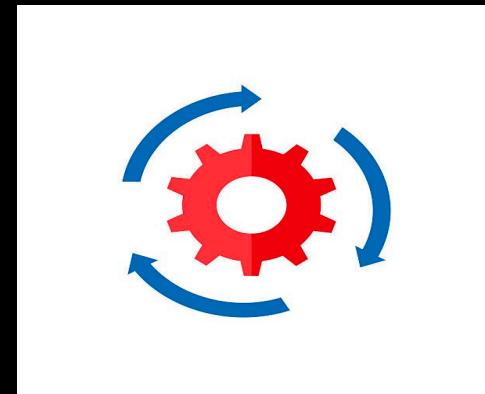
“The world is one big data problem.”

- Andrew Paul McAfee

IBM Cloud Pak for Data



Why use AI?



To improve business outcomes.

Predict - forecasts at macro and micro level

Automate - business processes that are often done manually

Decide – improve and optimize decision-making

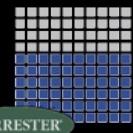
3 Main Challenges

The keys to turning aspirations in outcomes

DATA



The lifeblood of AI, but complexity slows progress



60%

Are challenge in
managing data quality

TALENT



AI skills are rare
and in high demand



62%

Are challenge to acquire
talent [and build skills]

TRUST



Skepticism of AI
systems & processes



62%

Need an approach to
AI production readiness

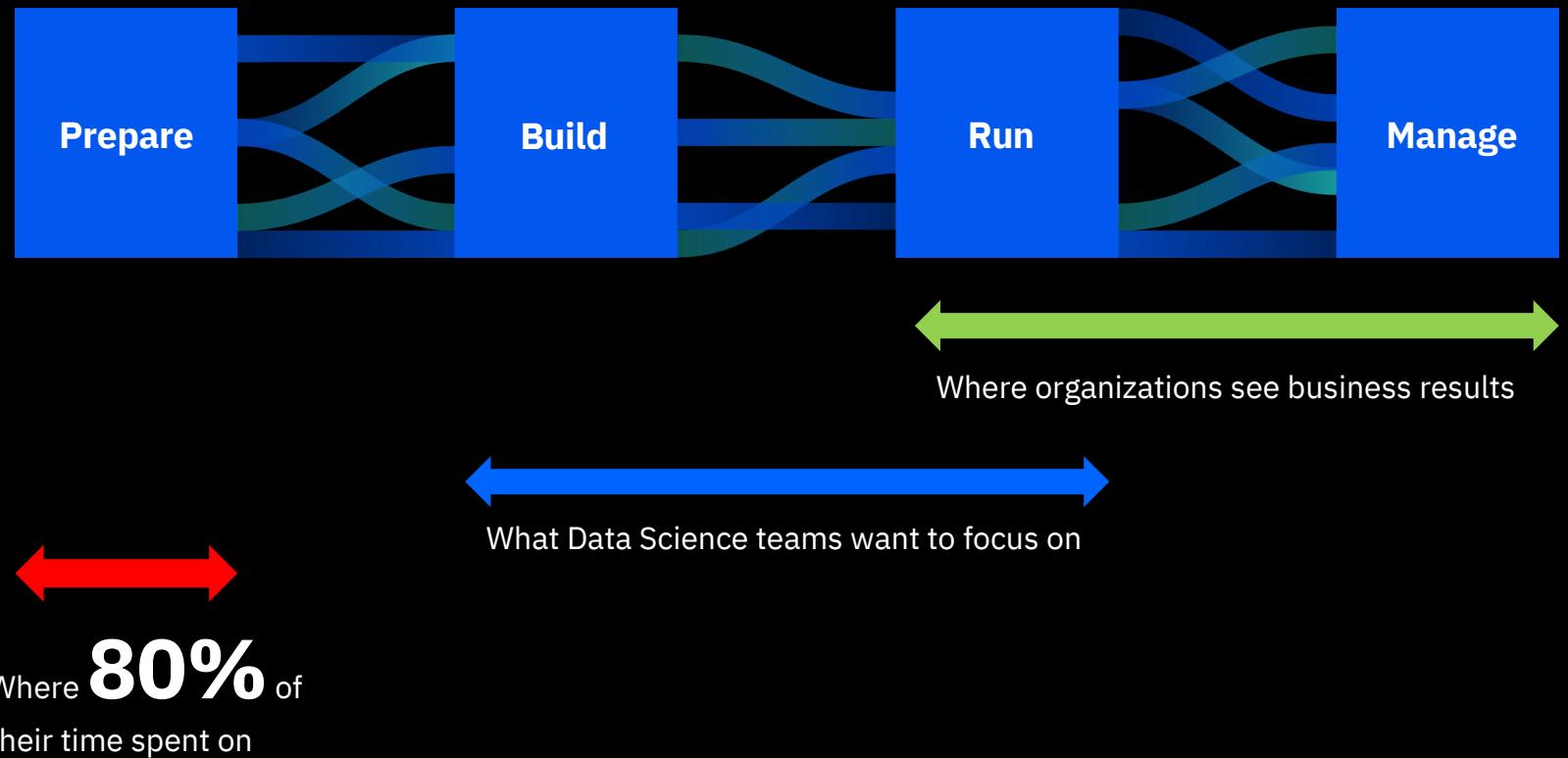
Stuck in
Experimentation

51%

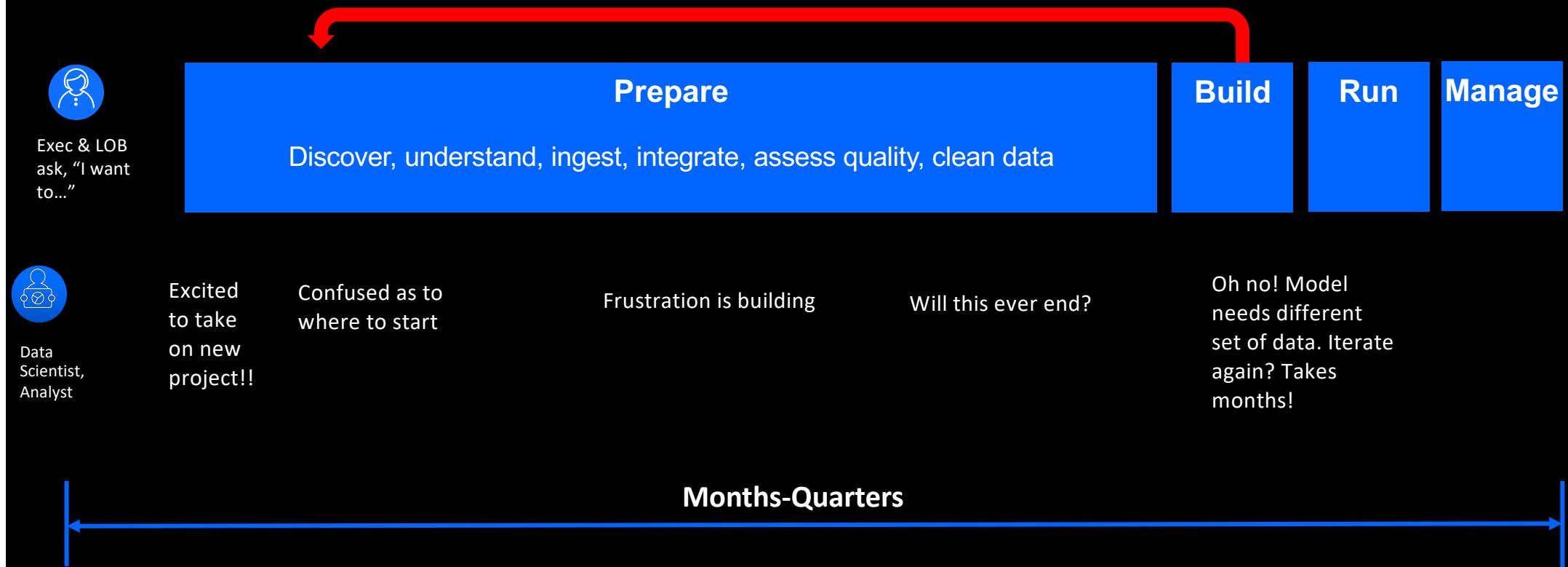
find operationalizing, sustaining
and scaling AI challenging

Challenge #1

Making your data ready for AI is hard...



Data preparation is often a frustrating bottleneck



#2 Talent: The Skills Gap



Two-thirds of AI decision makers struggle with finding and acquiring AI talent, and 83% struggle with retention.

– Forrester

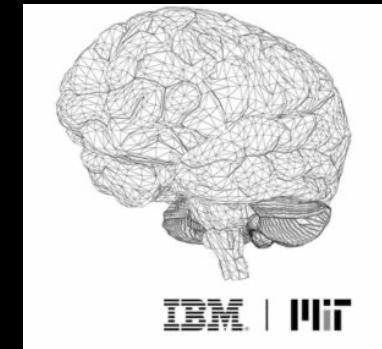
Finding a workforce with the right skills was cited as one of the most significant challenges to innovating.

– MIT Sloan

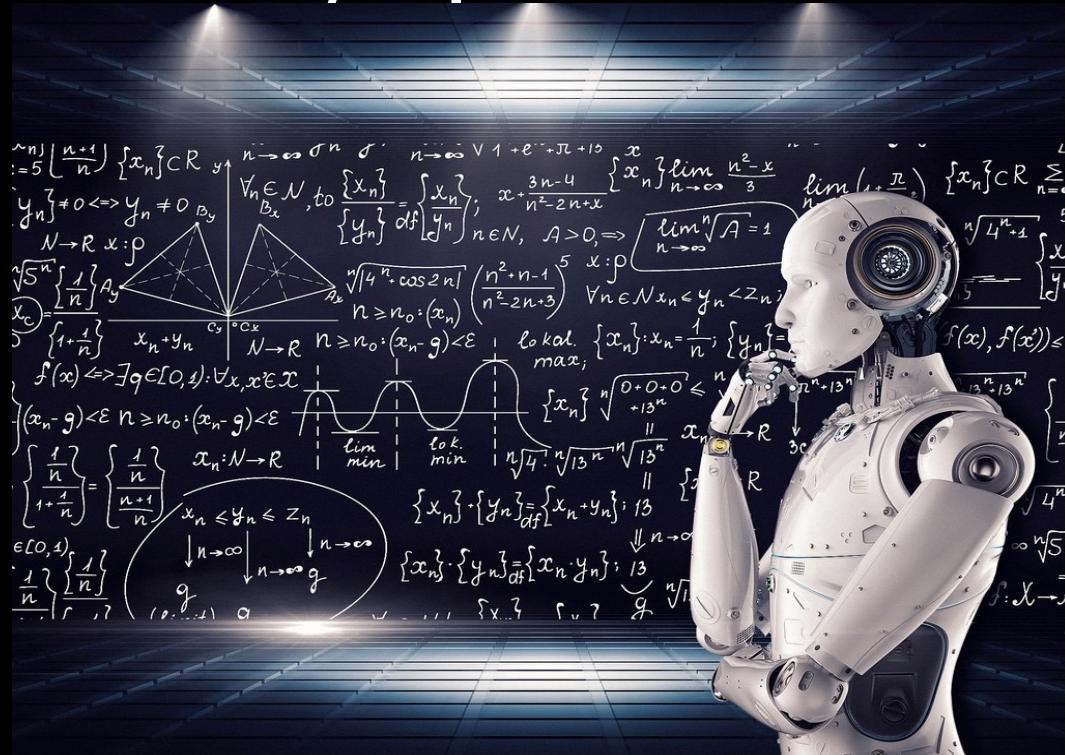
IBM AI Skills Academy
<https://ai-academy.w3bmix.ibm.com/>



the IBM AI Learning and Certification program,
<https://www.ibm.com/services/process/talent/ai-academy>



#3 Trust/Explain AI



8



DATA

There is no AI without an IA

[Information architecture]

81%

do not
understand the
data needed for AI

AI pioneers are

8x

more likely to
have a robust
data architecture

**“ No amount of AI algorithmic sophistication
will overcome a lack of data [architecture]”**

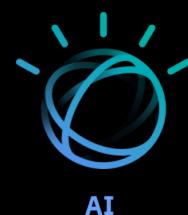
***Data collection & preparation is the most
time consuming and difficult part of AI***



Sources: 2018 MIT Sloan Report "Reshaping business with AI"

The AI Ladder

A prescriptive approach to accelerating the journey to AI



INFUSE – Operationalize AI with trust and transparency

ANALYZE - Scale insights with AI everywhere

ORGANIZE - Create a trusted analytics foundation

COLLECT - Make data simple and accessible

Data of every type,
regardless of where it lives



Talent &
Skills



One Platform,
Any Cloud



Cloud Paks – Enterprise-ready containerized software

A faster, more secure way to move your core business applications to any cloud through enterprise-ready containerized software solutions

IBM containerized software

Packaged with Open Source components, pre-integrated with the common operational services, and secure by design



Container platform and operational services

Logging, monitoring, security, identity access management



Google Cloud



Complete yet simple

Application, data and AI services, fully modular and easy to consume

IBM certified

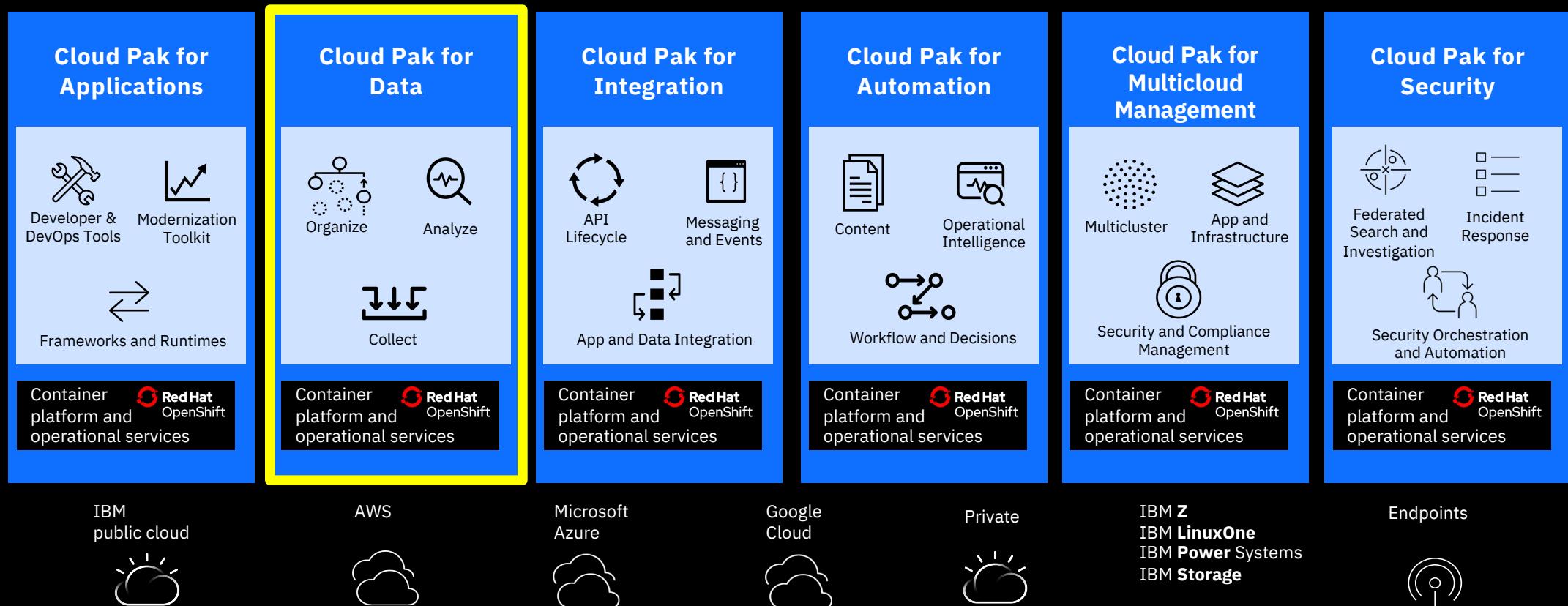
Full software stack support, and ongoing security, compliance and version compatibility

Run anywhere

On-premises, on private and public clouds, and in pre-integrated systems

Cloud Paks – Pre-integrated for cloud use cases

Today, IBM offers clients the first six Cloud Paks...



Circa 2005...

It was not convenient to carry a camera, cell phone, computer, contact book, etc. with you at all times.

Thus, most did not carry all of these tools with them. It was left to **specialists**.

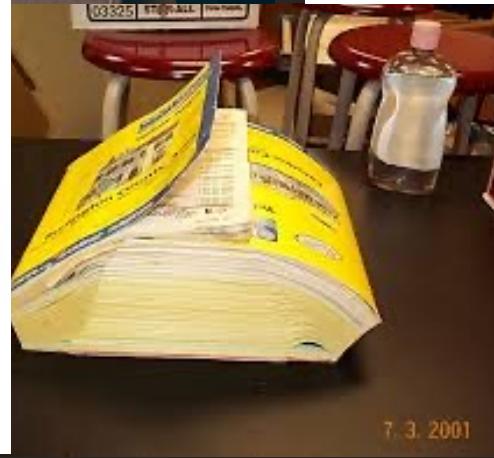


Circa 2013...

It was no longer about the tools, it was about a unified experience connecting our combined capabilities and perspectives.

Utilizing the combined capabilities became easier for **everyone** to use.



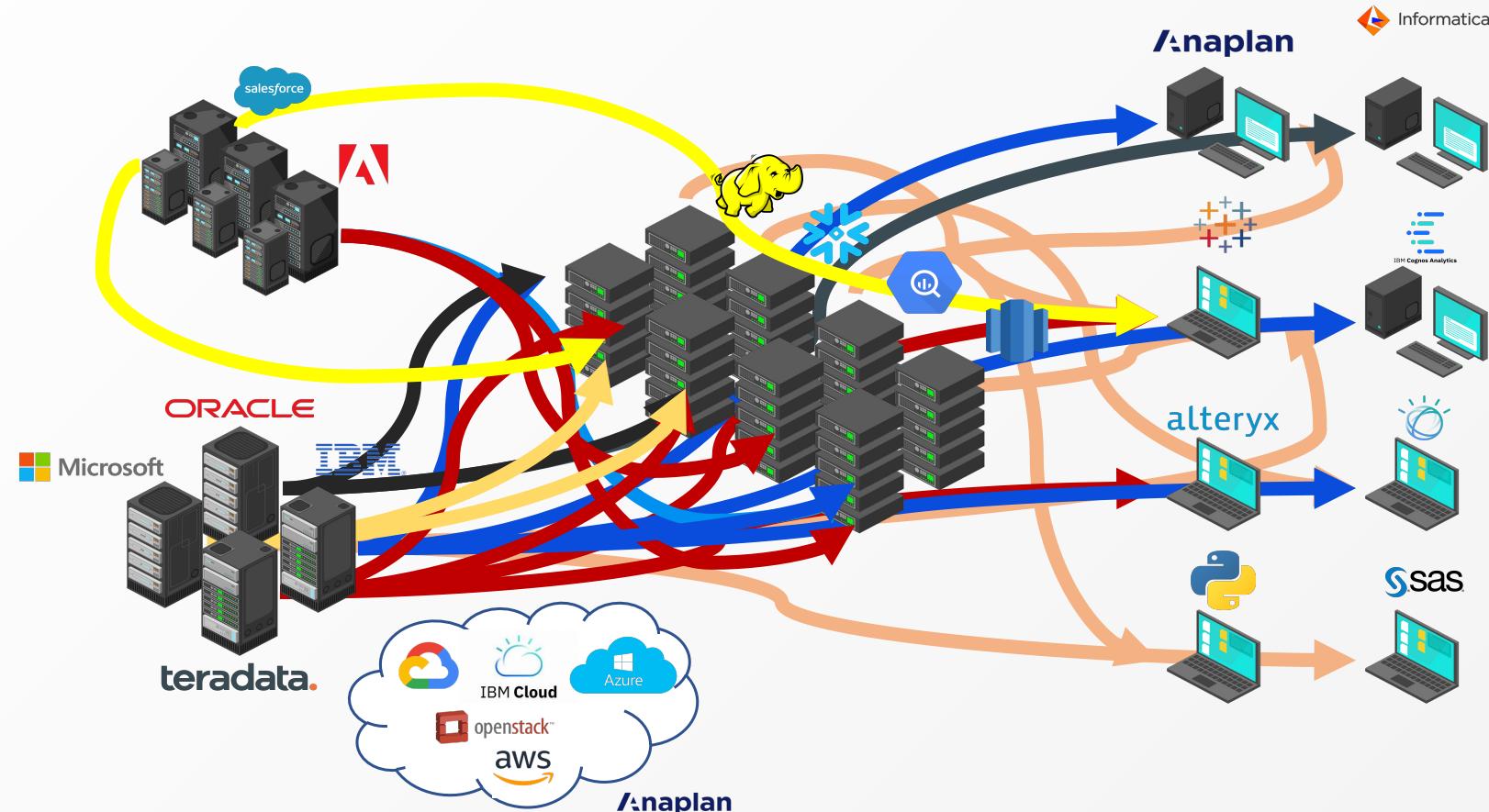


The sum of the PARTS is worth MORE than the whole.



The WHOLE is worth MORE than the sum of the parts.

Circa 2019.... Complex Data Environments



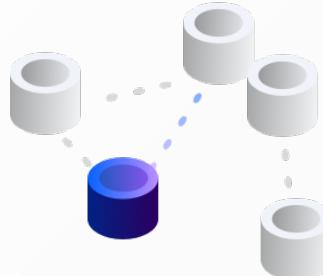
The sum of the PARTS is worth MORE than the whole.

Circa 2019.... IBM Cloud Pak for Data

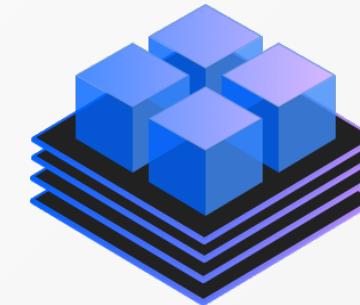


The WHOLE is worth MORE than the sum of the parts.

Our Approach



*How to simplify
your information
architecture “IA”*



*How to get data
working for you*

Find, connect to, govern, and leverage your data across multiple sources without needing to move or replicate.

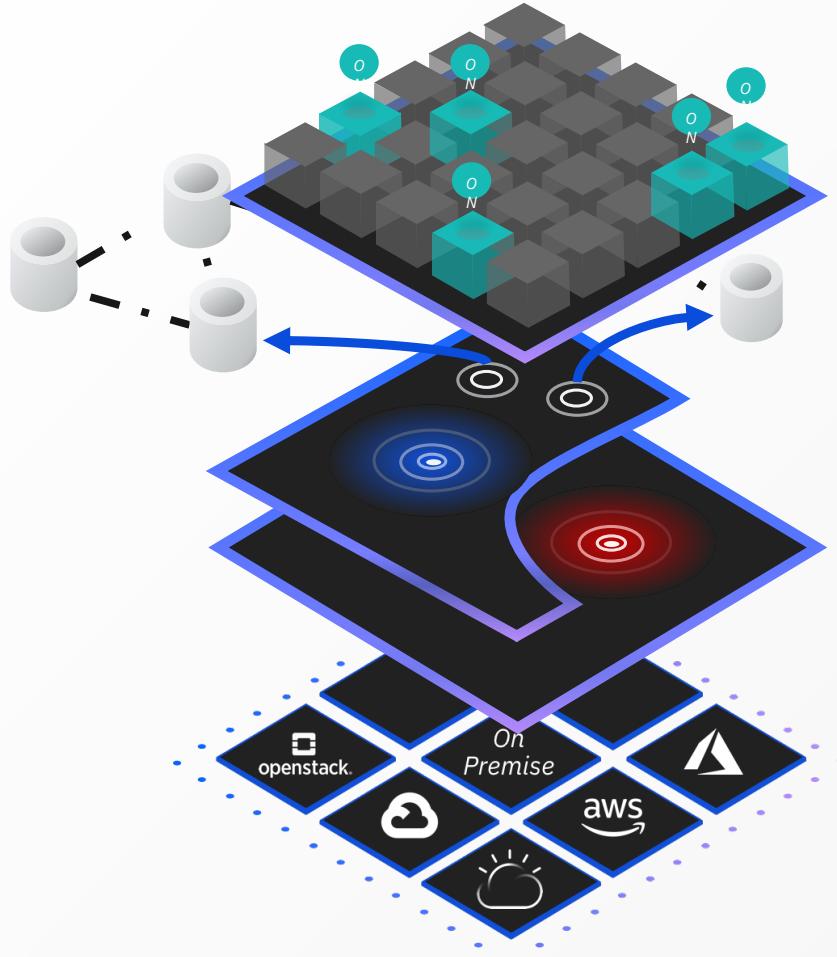
Automate many mundane and repeatable tasks like cleaning, matching, and metadata creation to reduce data prep time by 80%.

Leverage deployment flexibility amongst any Cloud, Hybrid Cloud or Private Cloud environment and provider with Red Hat OpenShift.

Eliminate working silos with a single unified experience allowing all data users to collaborate and connect to multiple analytics applications and models.

Centralize your teams’ workflow and operations management with an ecosystem of 45+ integrated services.

Enable your highly skilled and paid data teams to spend more of their time on business value generating innovations in big data and analytics.



The Platform

Add-on Ecosystem

Plug and play 45+ data, analytics and AI apps
Data & AI @ Scale.

Data Virtualization

Purpose built data requires purpose built repositories. Leave your data where it is.

Single Team Experience

Access, manage, govern and catalog data & AI.

OpenShift

Leverage the leading open source hybrid cloud platform to SCALE data & AI workloads. **The Future of AI is Flexible.**

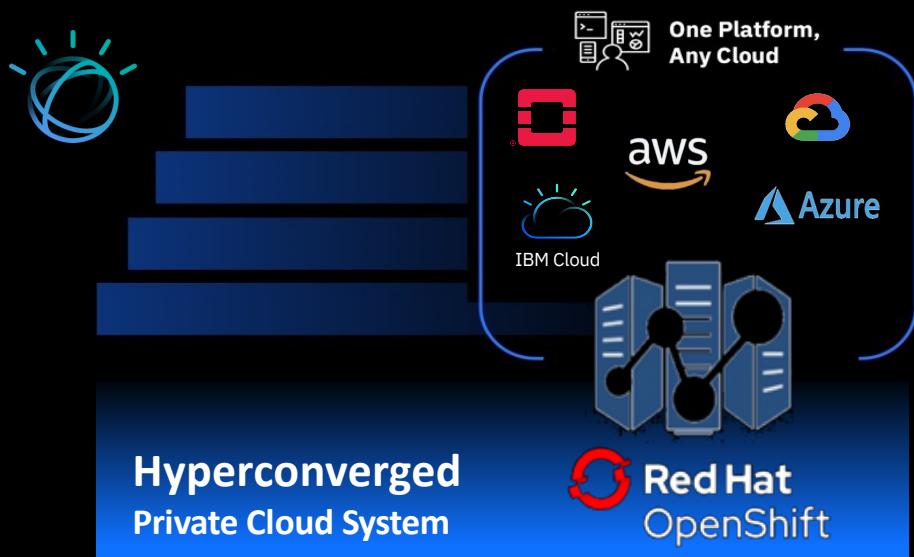
Any Cloud

Clouds are IT carriers. Decoupling enterprise data, analytics and AI will prevent lock in and accelerate polyglot eco-systems.
Any cloud, anywhere.

Cloud Pak for Data System

Enterprise Insights Platform – In a BOX

True plug-and-play enterprise data and AI in under 4 hours right out of the box, securely behind the firewall



■ An **all-in-one** system pre-integrated with all the necessary hardware and software components

Deploy a complete private cloud data and AI platform in hours, with no assembly required

Dynamically scale compute, storage and software with on-demand plug & play

Simplify management and optimization with a unified and intuitive dashboard

Cloud Pak for Data Strategy

The Ladder to AI

Modernize

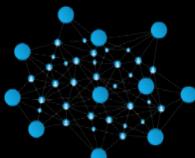
Infuse

Analyze

Organize

Collect

For data of every type, regardless of where it lives



Cloud Pak for Data - Advanced

Platform

HDM

DataOps

Data Science & AI



Unify on a Multicloud Data Platform

IBM Cloud Pak for Data

A one of a kind, pre-integrated set of data and AI services delivered within an open and extensive cloud native platform



Cloud-native container platform and operational services



IBM Cloud

AWS

MS Azure

Google Cloud



Edge



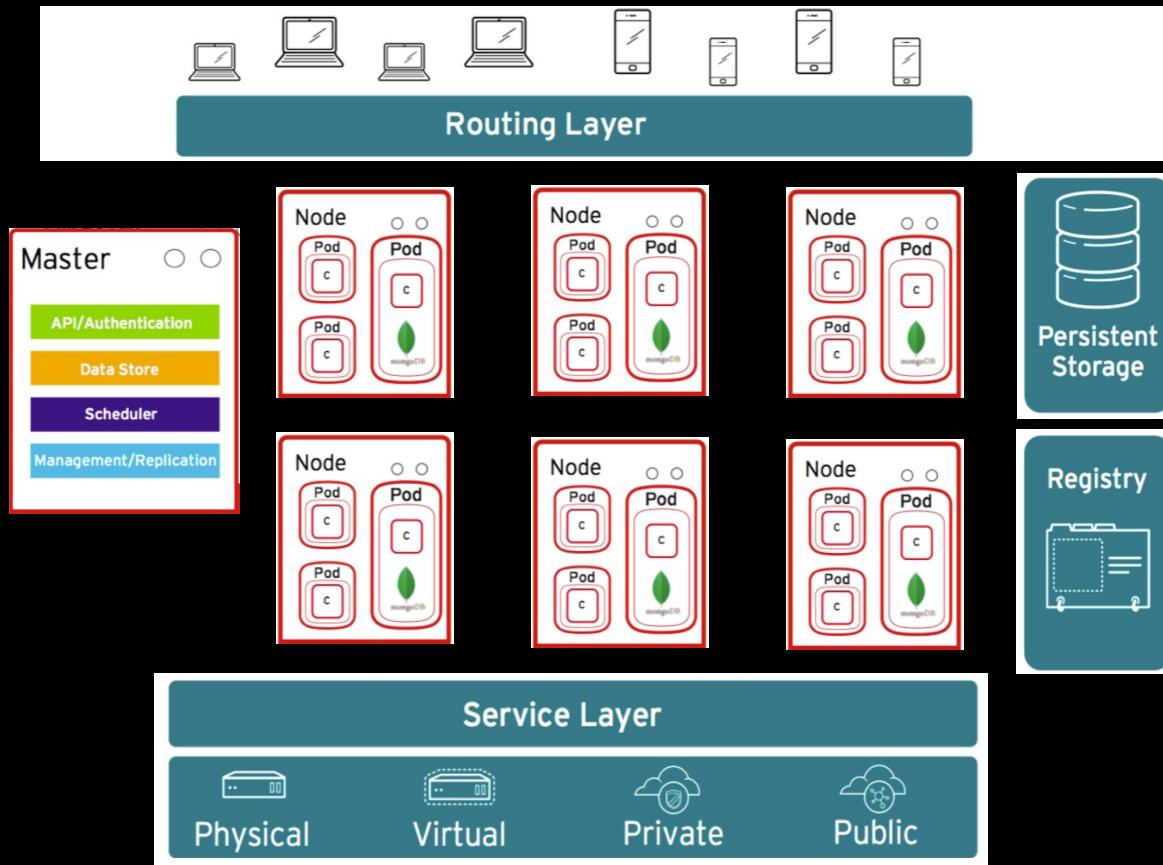
Private



Hyperconverged
Private Cloud System

How does OpenShift work ?

Kubernetes/OpenShift basics



Namespace/Projects – Entity used to group and isolate objects. It allows a community of users to organize and manage their content in isolation from other communities.

Pods – Smallest deployable unit of compute that can be made up of one or more containers.

Deployments – Application templates that bring together different components/pods of an application, best suited for stateless applications. Ensures the prescribed number of replicas are active at all times.

Statefulsets – Application templates that bring together different components/pods of an application, best suited for stateful applications. Ensures the prescribed number of replicas are active at all times.

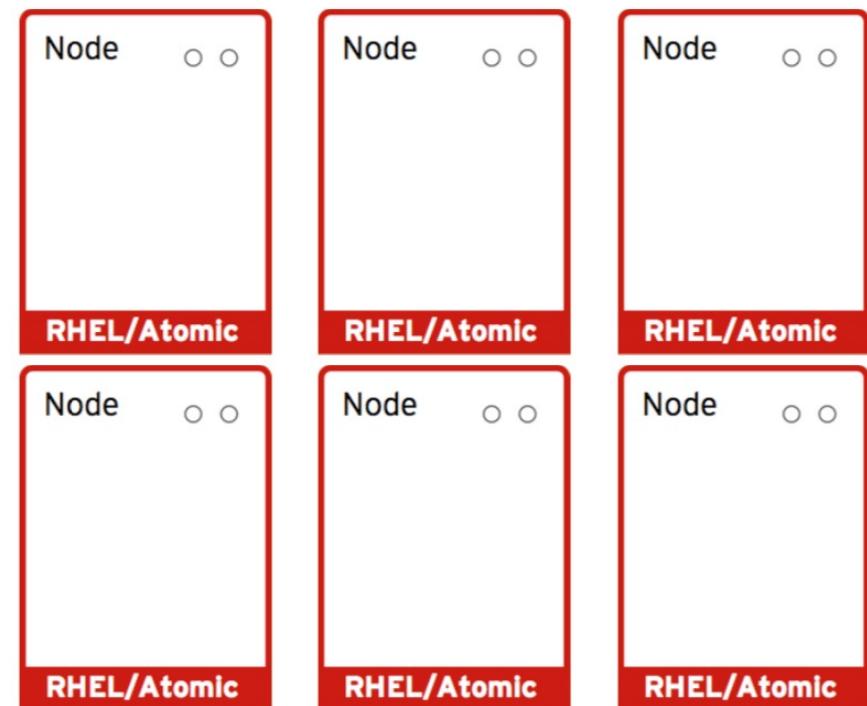
Services – Internal abstraction layer, load balancing within the cluster.

Ingress/Route – Mechanism to expose application to the outside world. Default mechanism is HAProxy, this looks at the request header and routes to appropriate service.

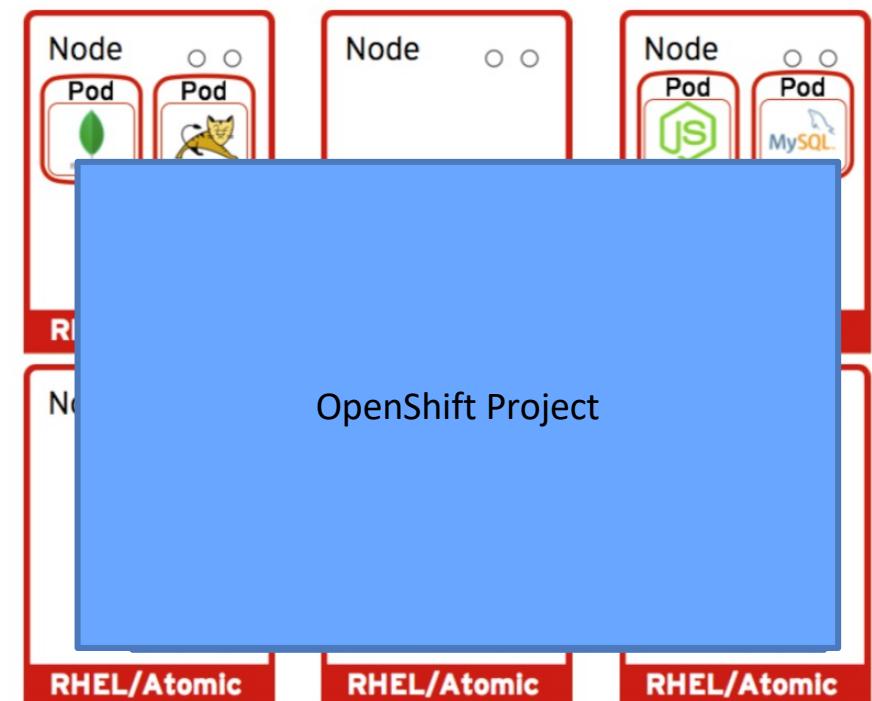
OpenShift runs on your choice of infrastructure



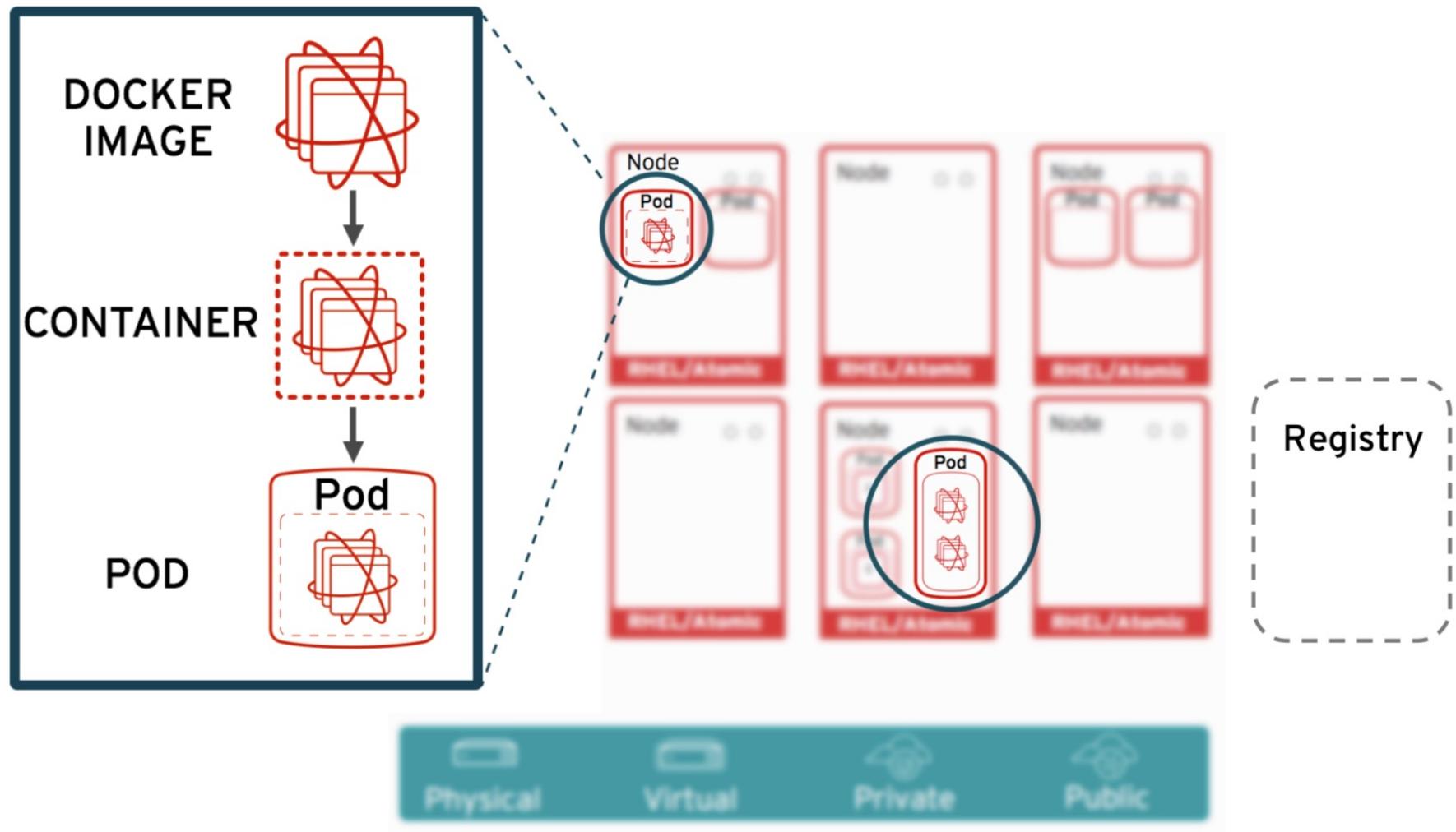
Nodes are instances of RHEL where applications will run



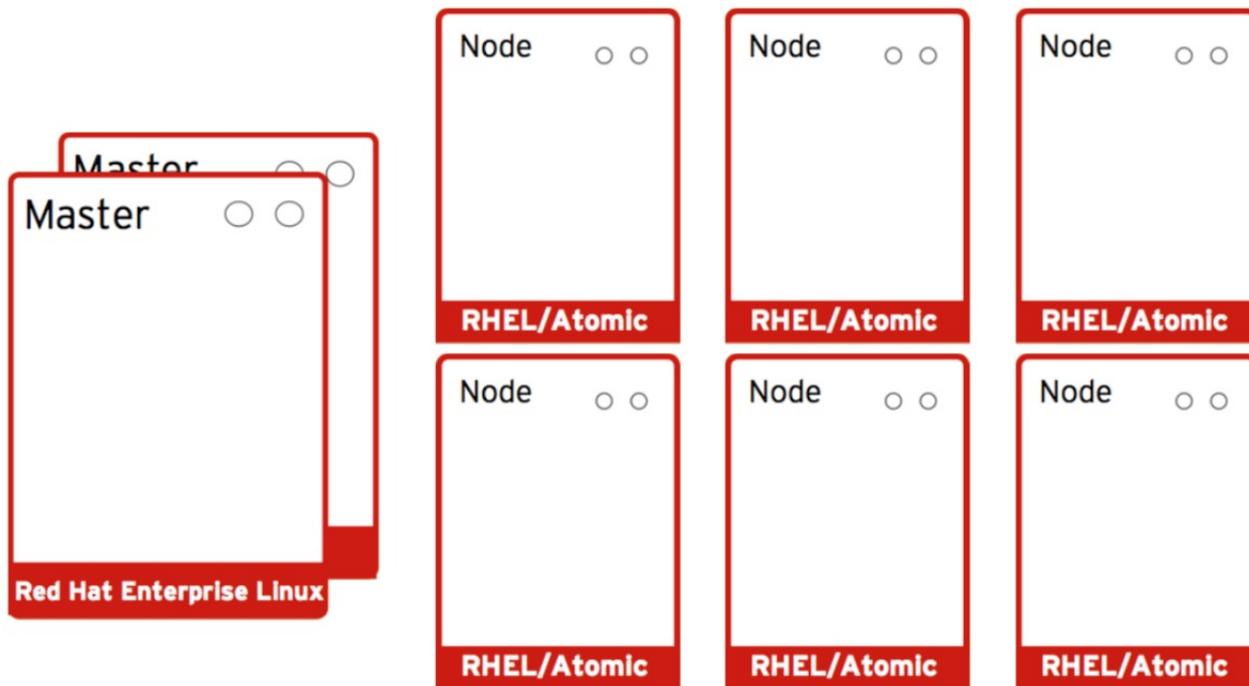
CP4D services run in containers (docker/CRI-O) on each node



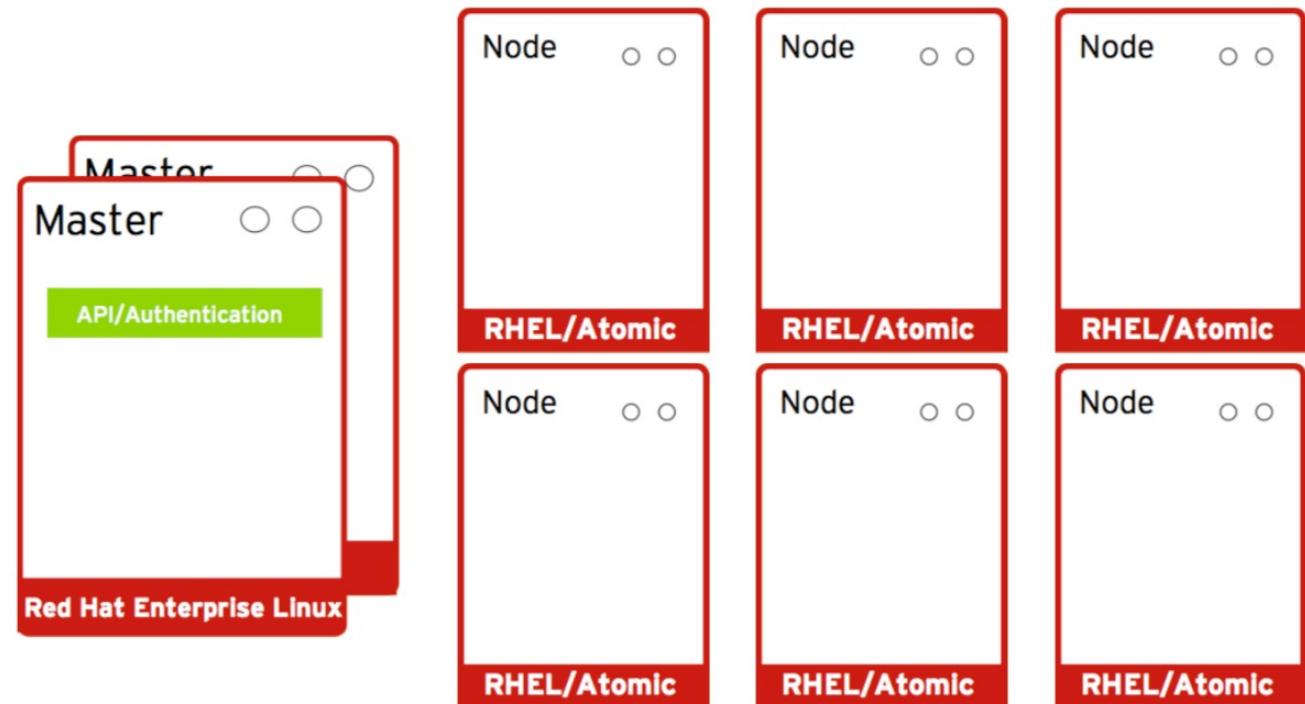
CP4D service Pods run one or more containers as a unit



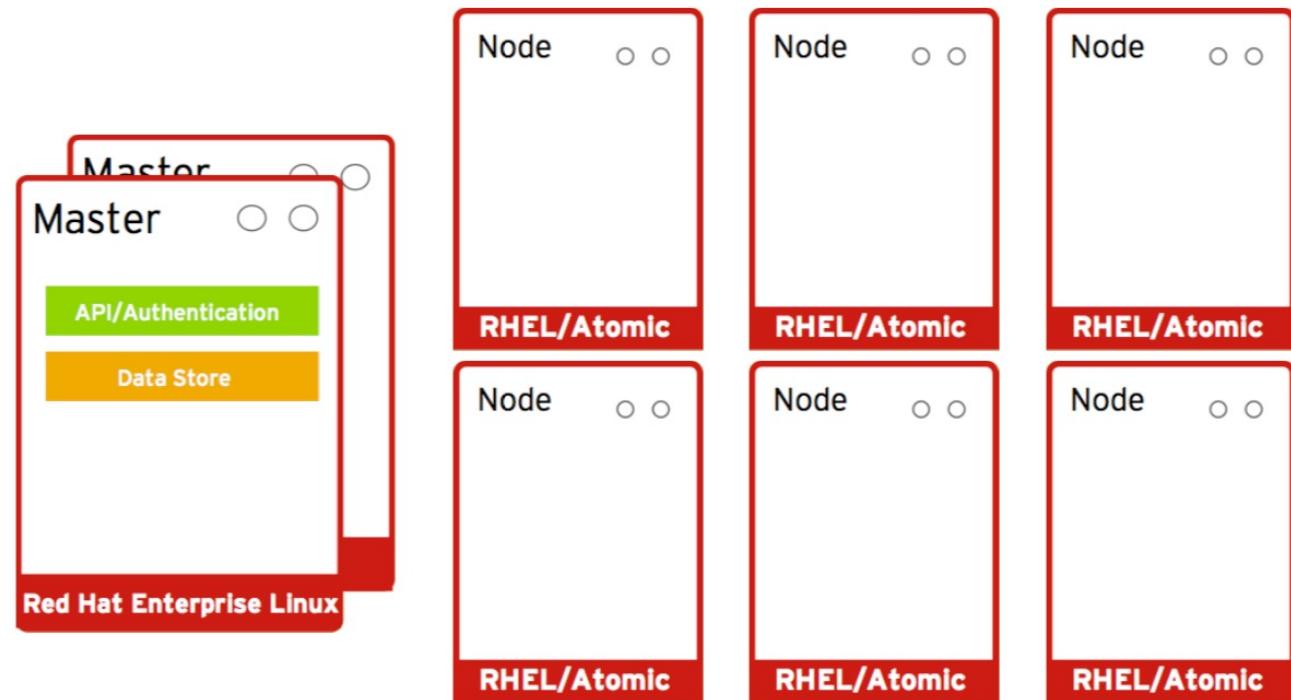
Master nodes leverage Kubernetes to orchestrate nodes/apps



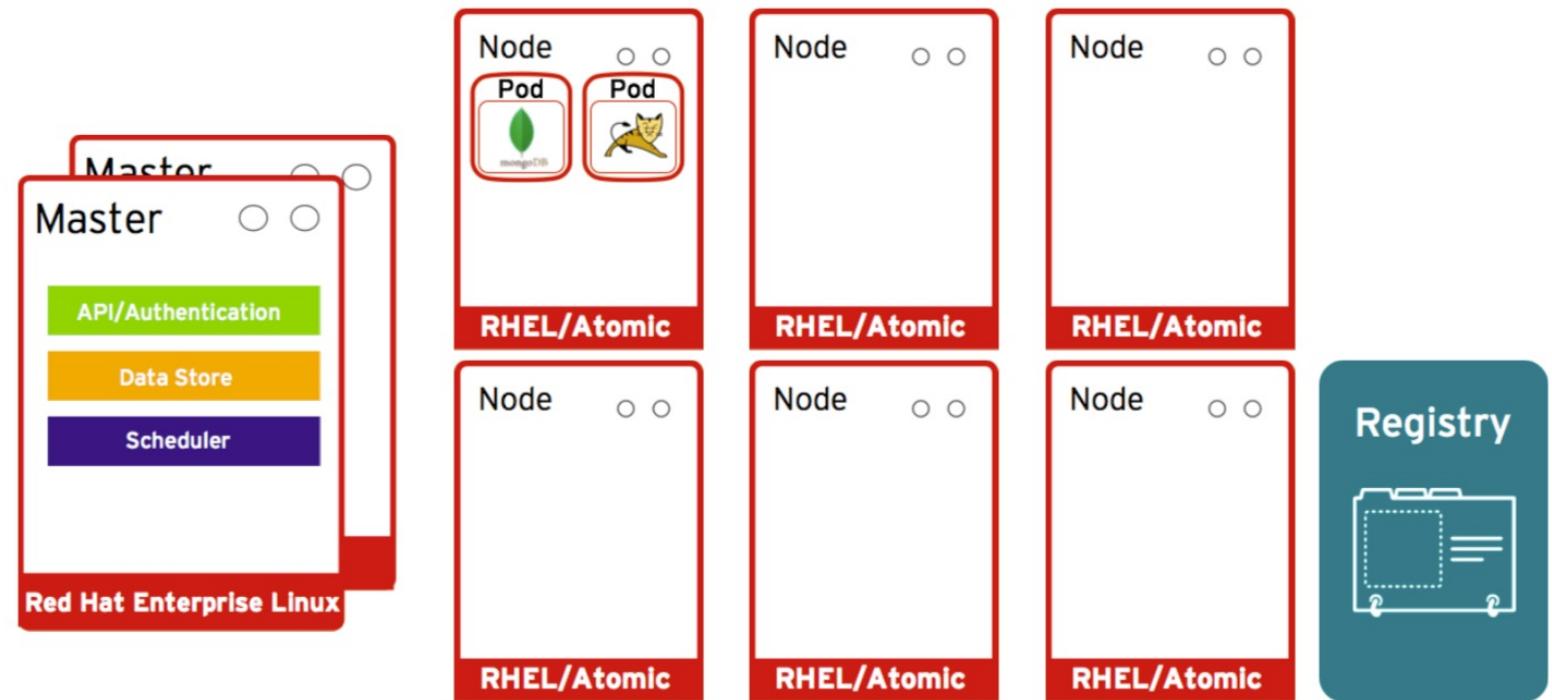
Master nodes provide authenticated API for users & clients



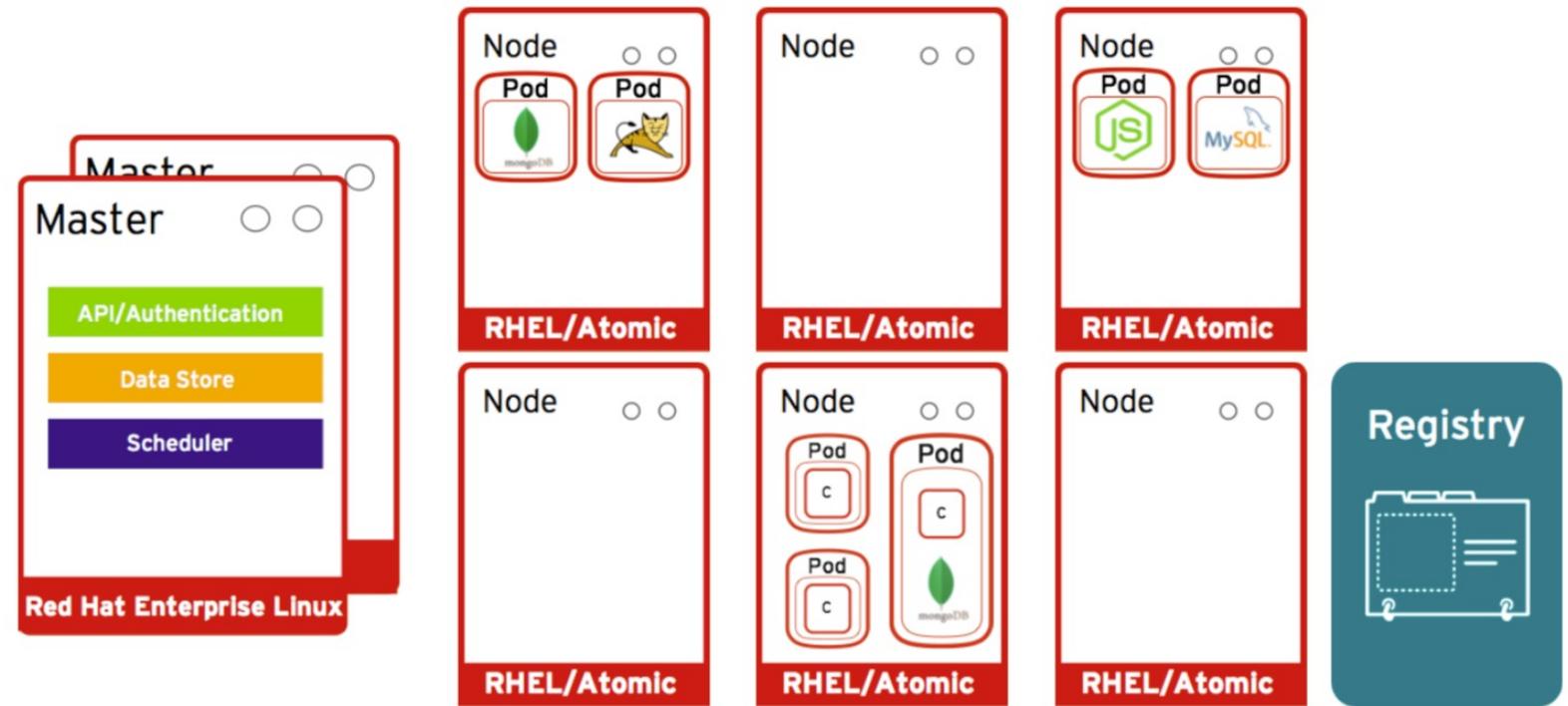
Master uses distributed key/value store (Etcd) for config, state data



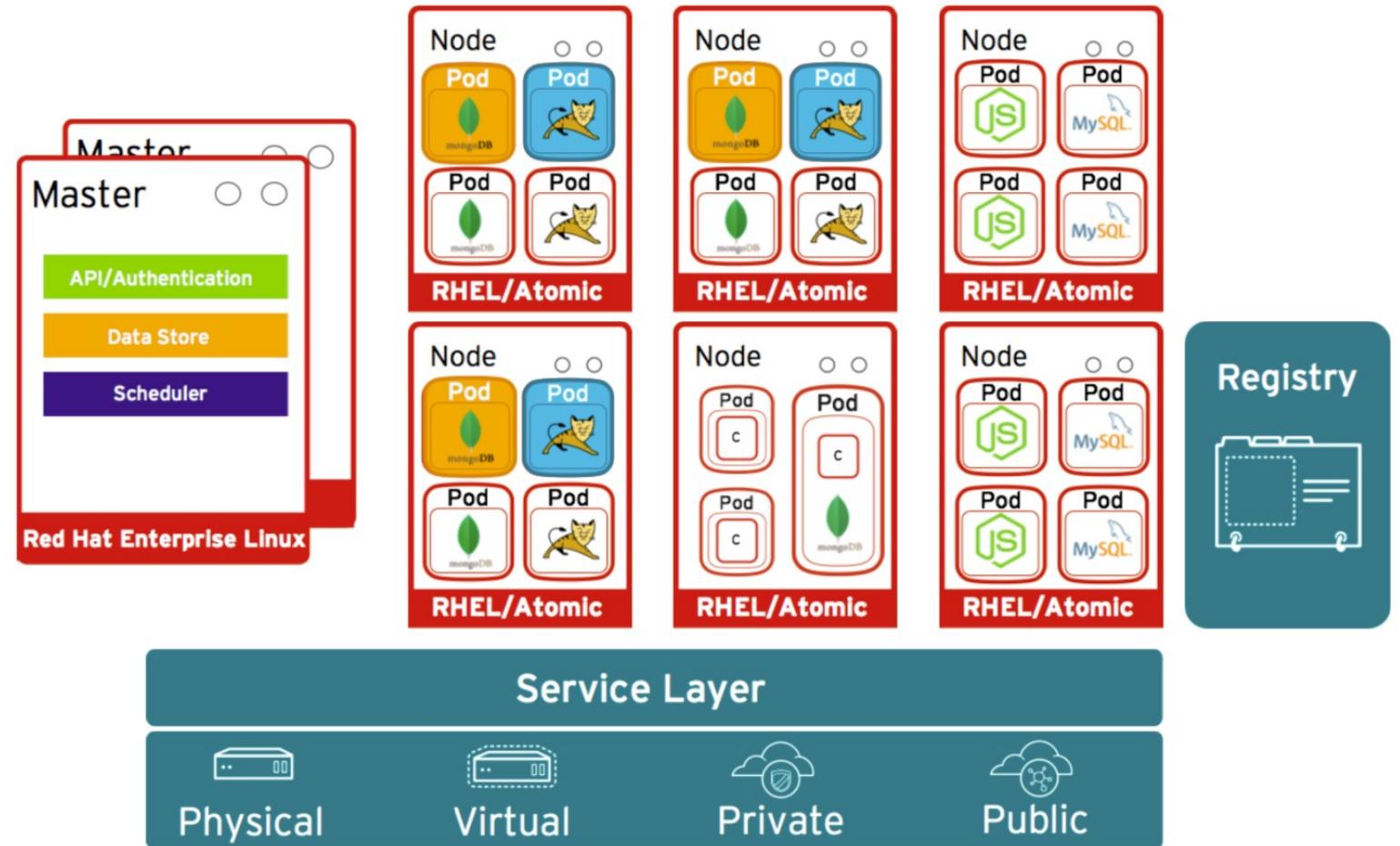
Master provides scheduler for pod placement on nodes



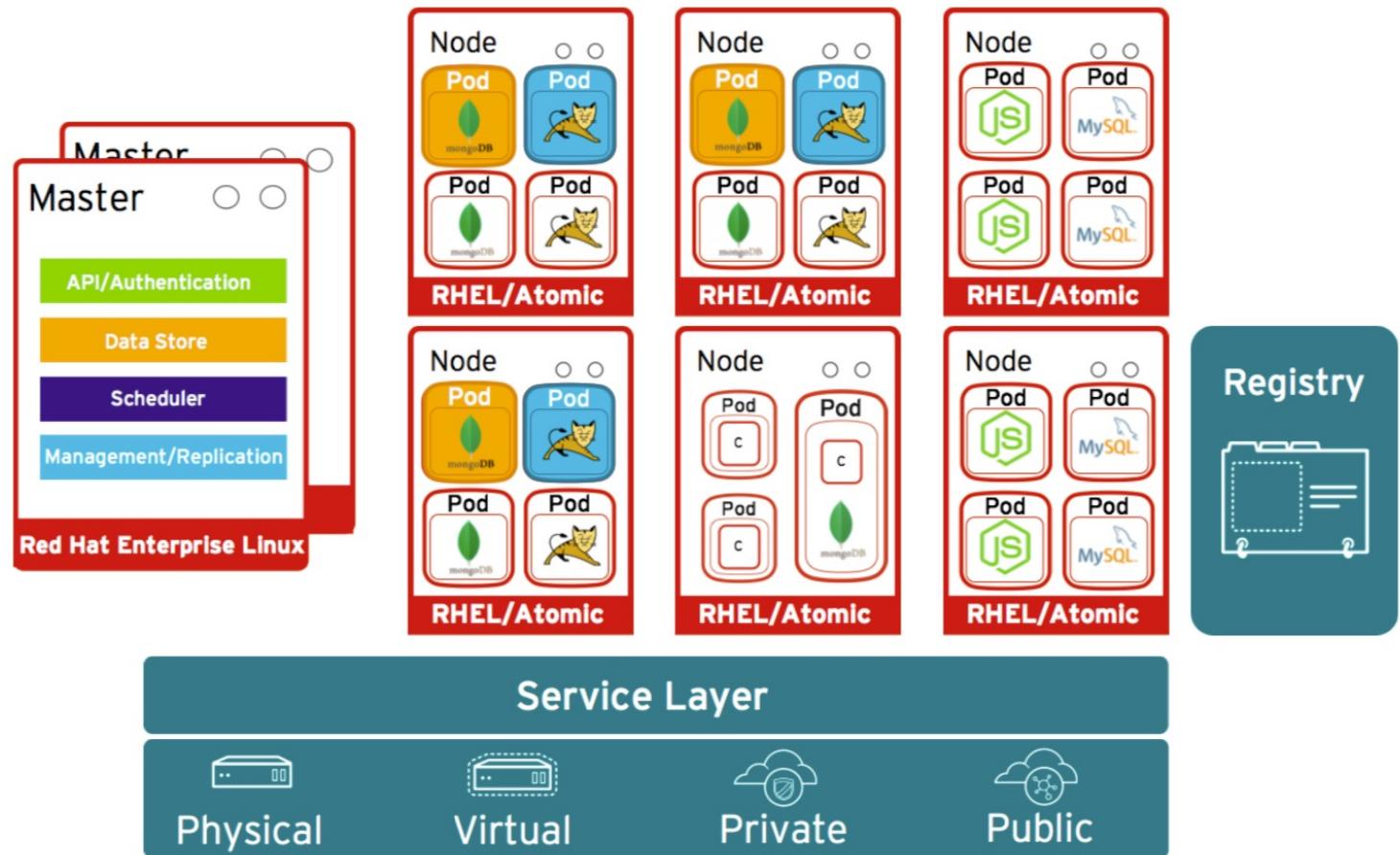
Policy-driven pod placement



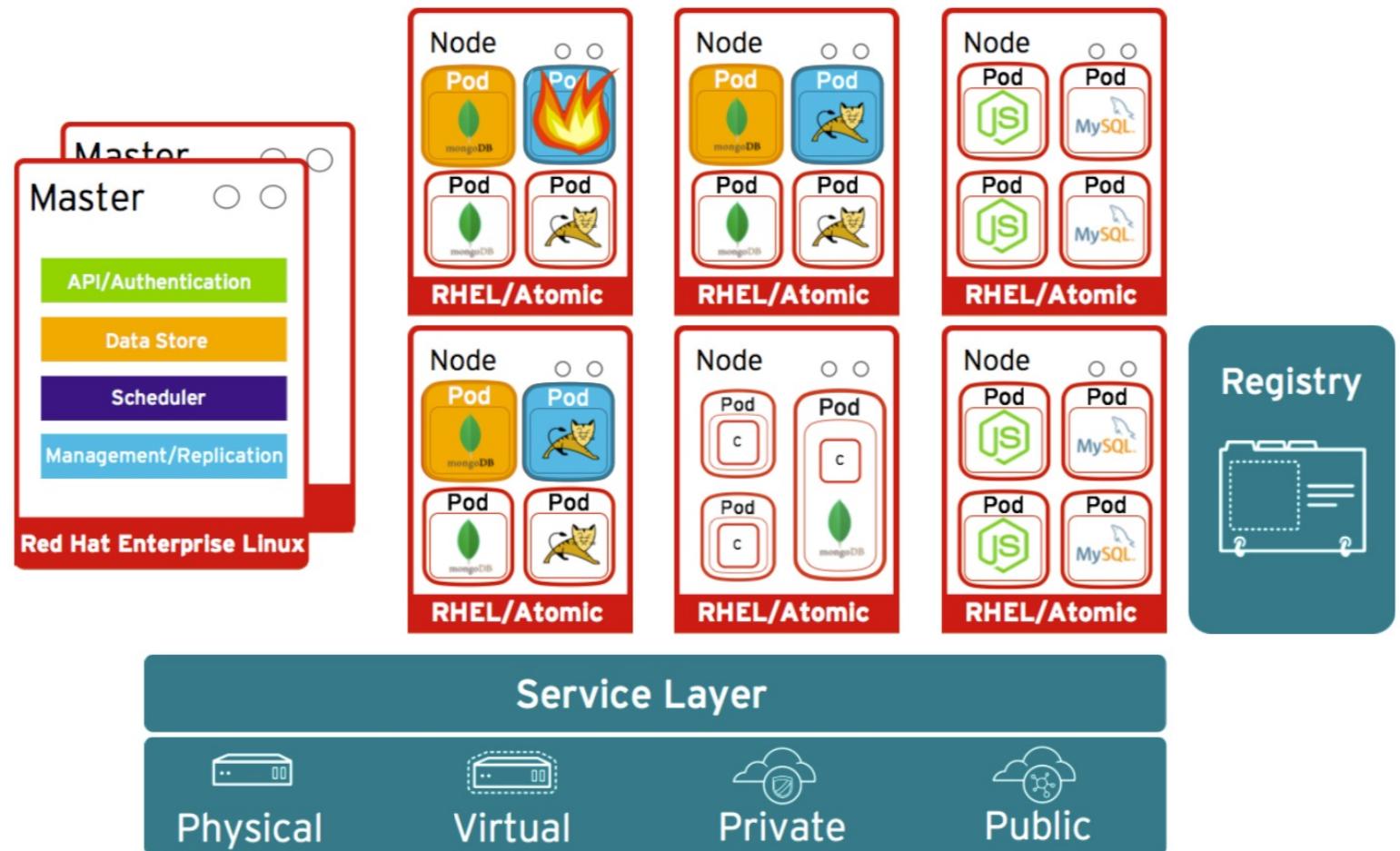
Services allow related pods to connect to each other



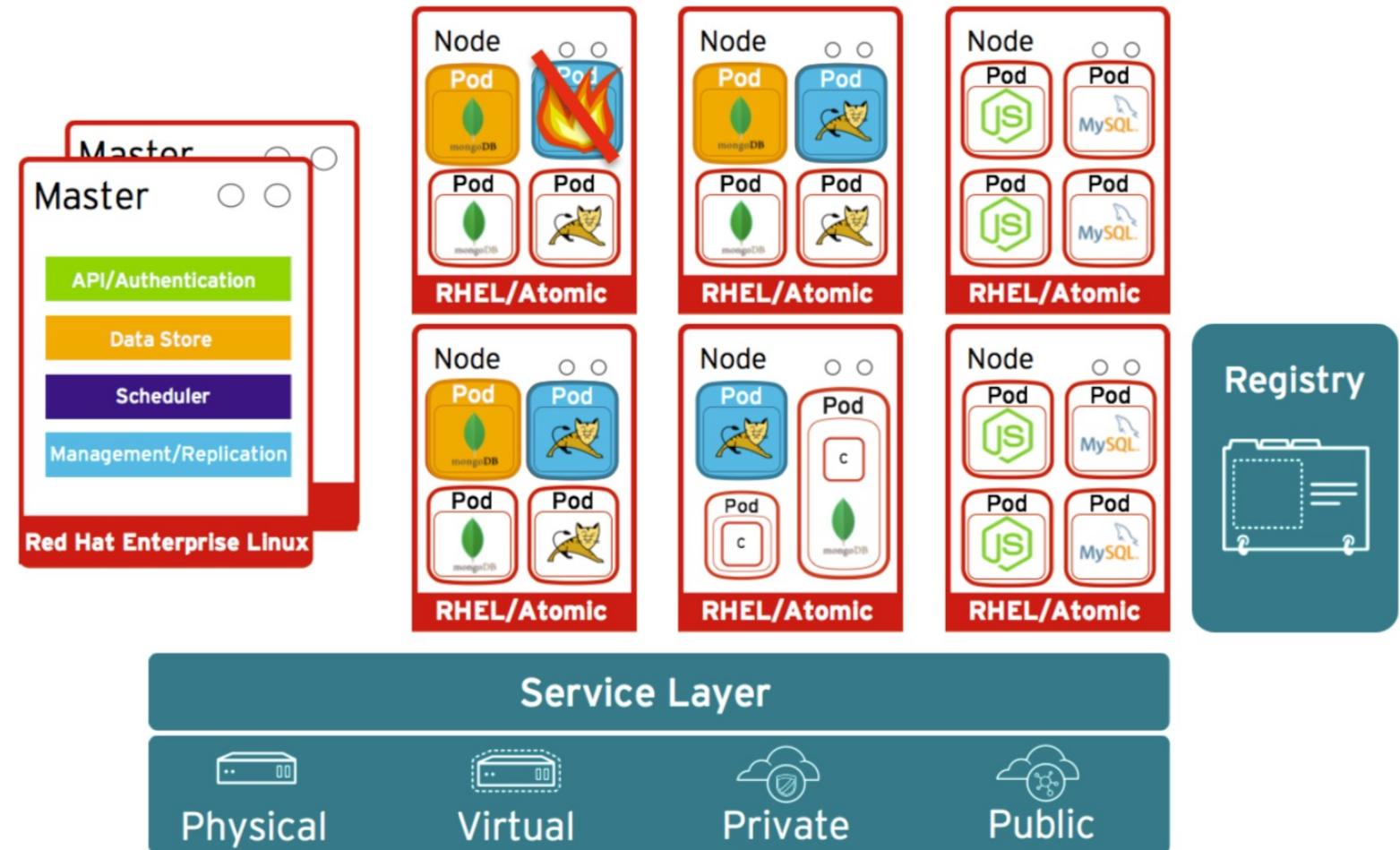
Management/replication controller manages the pod lifecycle



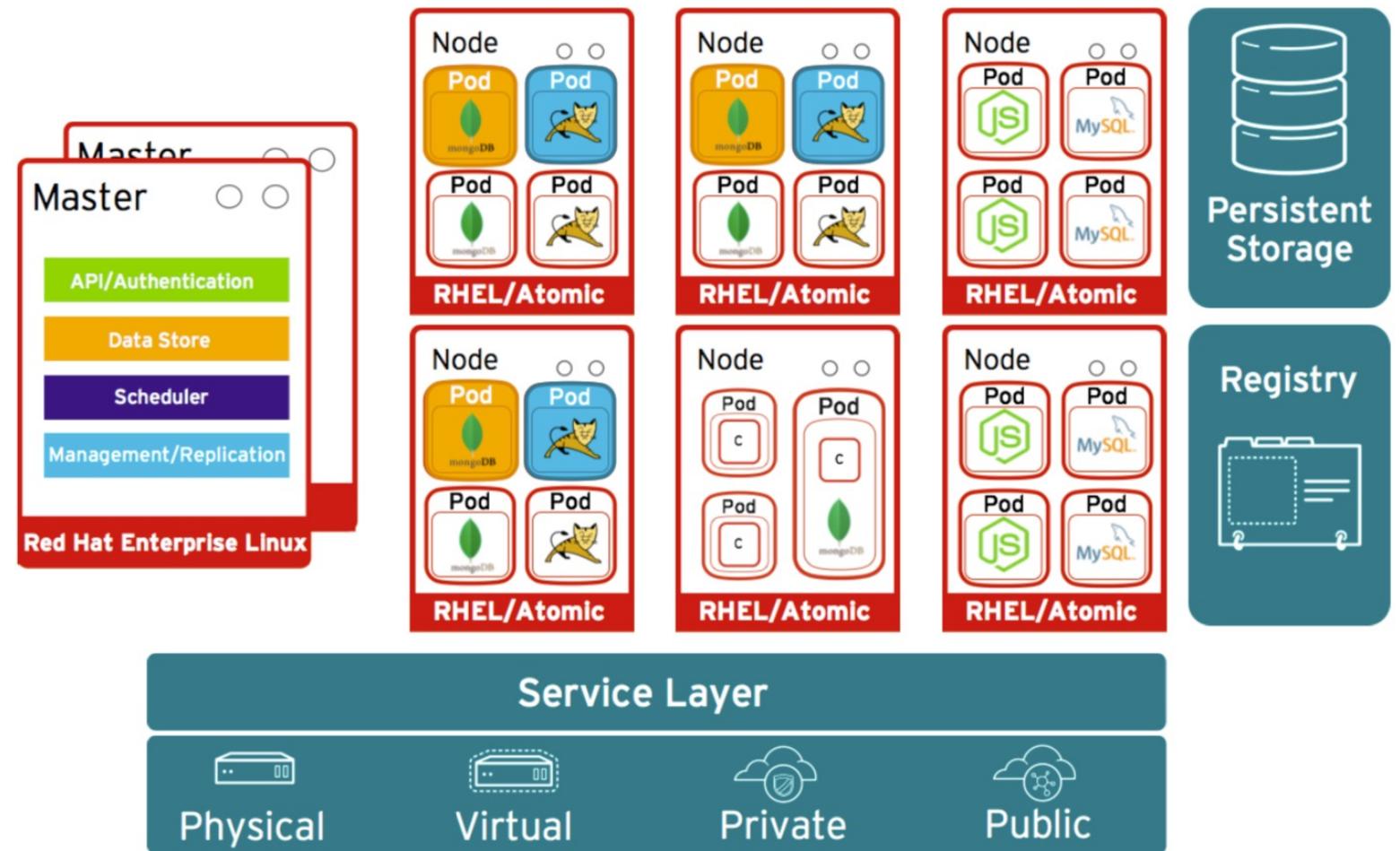
What happens when a pod goes down ?



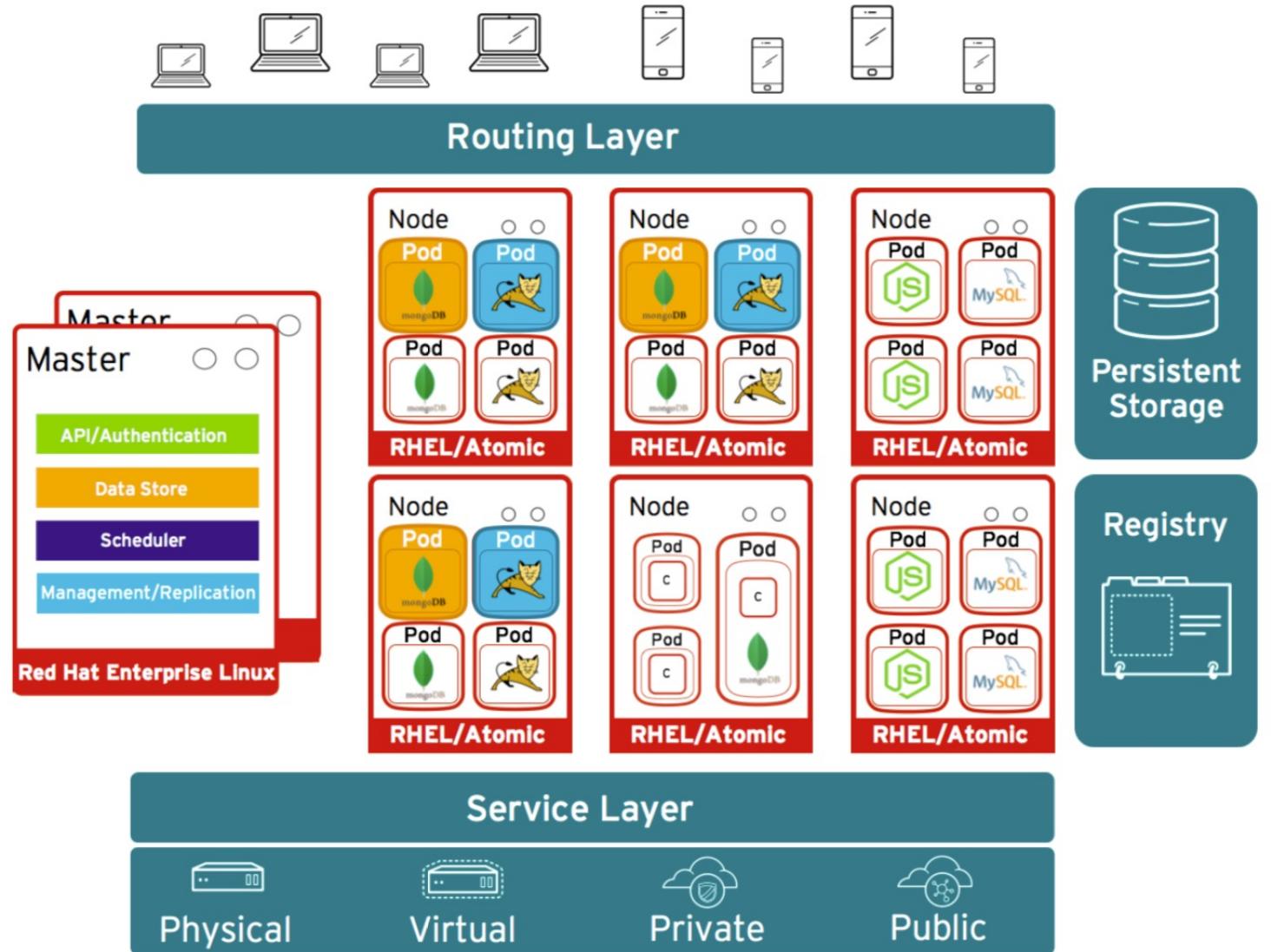
Openshift automatically recovers and deploys a new Pod



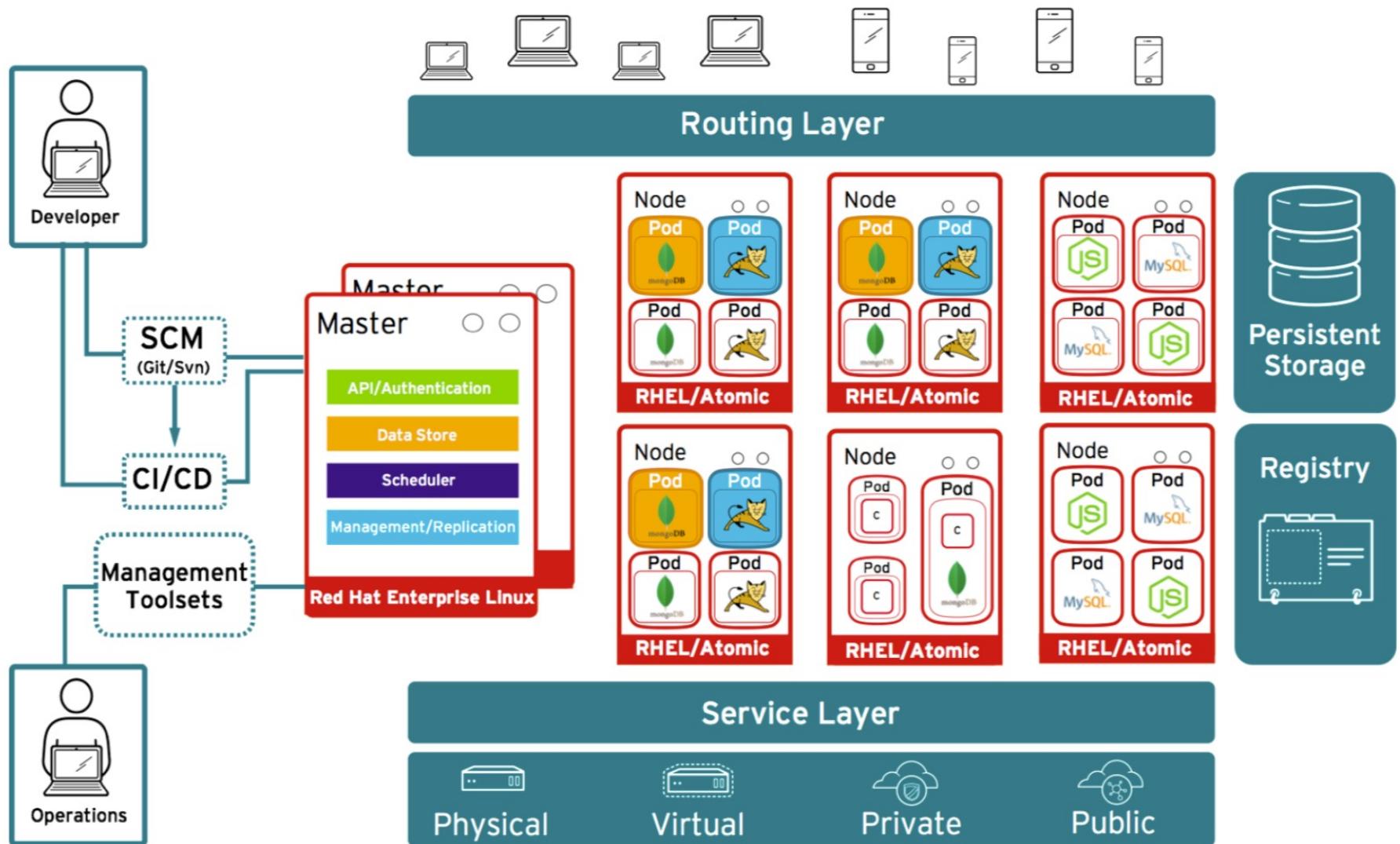
Pods can attach to shared storage for stateful services



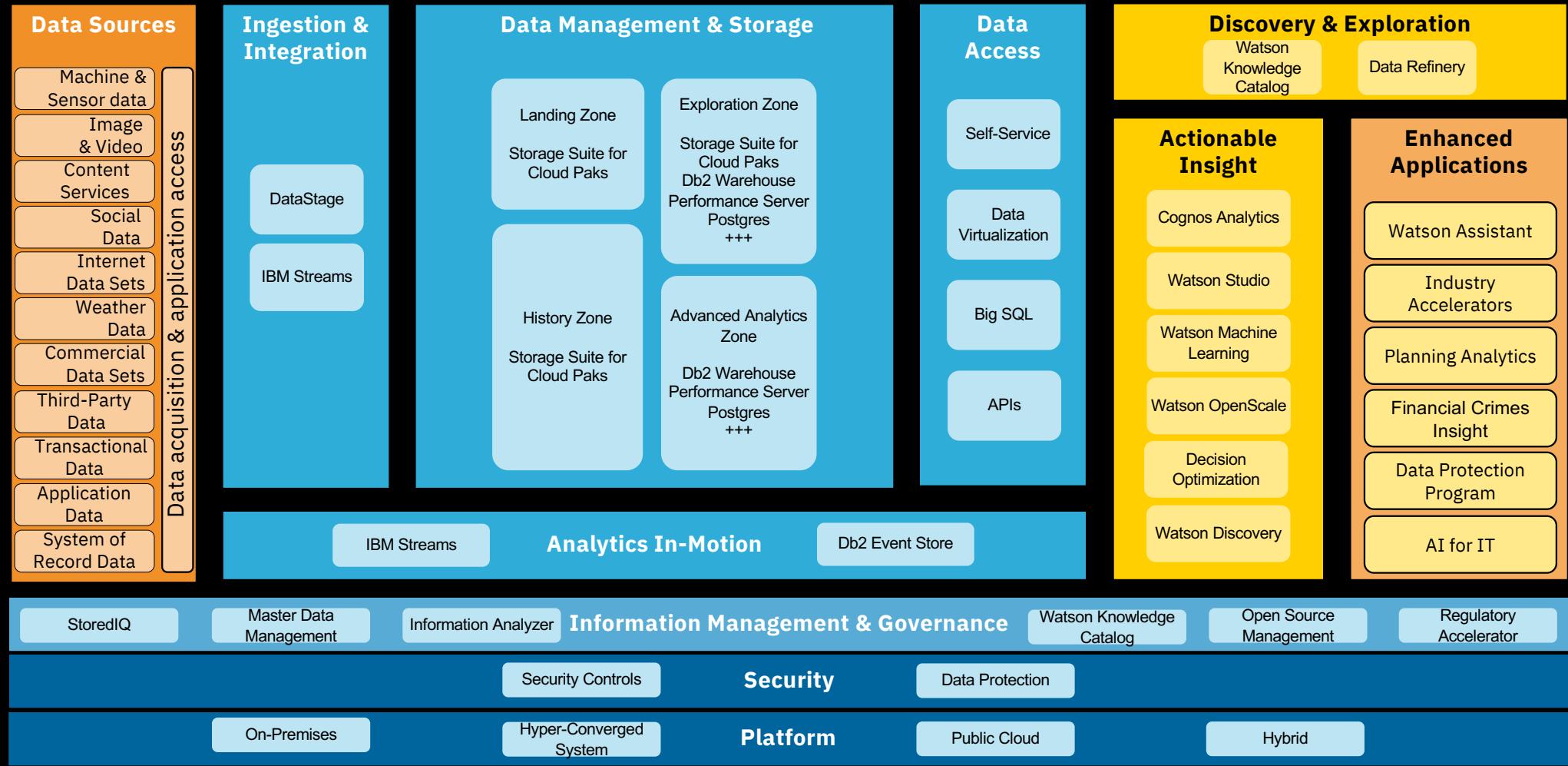
Routing layer routes external app requests to Pods



Developers can access Openshift via web, CLI or IDE



Reference Architecture – IBM Cloud Pak for Data





SAME

public cloud

hybrid cloud

private cloud

on premises

redhat

SAME

A large, bold, black word "redhat" is positioned in the center. Above it, the word "SAME" is written twice in a smaller, bold, dark red font, once above the "h" and once below the "d". To the left of the "redhat" word, four horizontal lines extend from the text "public cloud", "hybrid cloud", "private cloud", and "on premises" towards the center. These lines are connected by a vertical line that also extends downwards, ending at the second "SAME" word.

Reference Architecture – Journey to Cloud & AI Requirements

Journey to Cloud & AI - Requirements	Legacy Architecture	Modern Architecture
Licensing Model Economics	Point products, each with their own licensing metrics; licensing model often does not translate well to cloud	Single, simple licensing metric across all capabilities; aligned with multi-cloud deployment (Virtual Processor Core)
Secure Container Platform	Point products containerized, each with their own security engineering practices and security controls	Enables a consistent secure software supply chain, deployable anywhere (Red Hat OpenShift)
Unified Data & Analytics Services	Point products manually integrated, increasing likely of misconfiguration and security threat exposures; lack seamless user experiences and process execution workflows	“No assembly required” automatic service component integration upon provisioning; seamless look-and-feel and user experience across all services
Cloud Data Ecosystem Topology	Tethered to a given technology or vendor; data service deployment options, egress, coupling of compute and storage, etc.	“Query anything, anywhere, safely” regardless of where the data lives, as if it is a single data source (Data Virtualization). Optimize data topology economics (DataStage)
AI Augmented User Experience	All user tasks in the job-to-be-done are manually performed	User tasks become progressively easier through embedded AI, augmenting the user experience (e.g., Watson Knowledge Catalog, AutoAI, OpenScale, etc.)

Cloud Pak for Data: Delivery & Install

Introducing the concept of an “assembly”

Problems:

- Packaging of CPD functionality is complex & we need to allow for a *mix-n-match* of Services.
- *Shared* components and pre-requisite dependencies that need to be setup in a specific order
- Our *terminology* interferes with simplicity: “base”, “lite”, “add-ons”, Services extensions, cartridges etc.

An **Assembly** = (a package of) capabilities or even use case patterns that gets deployed.

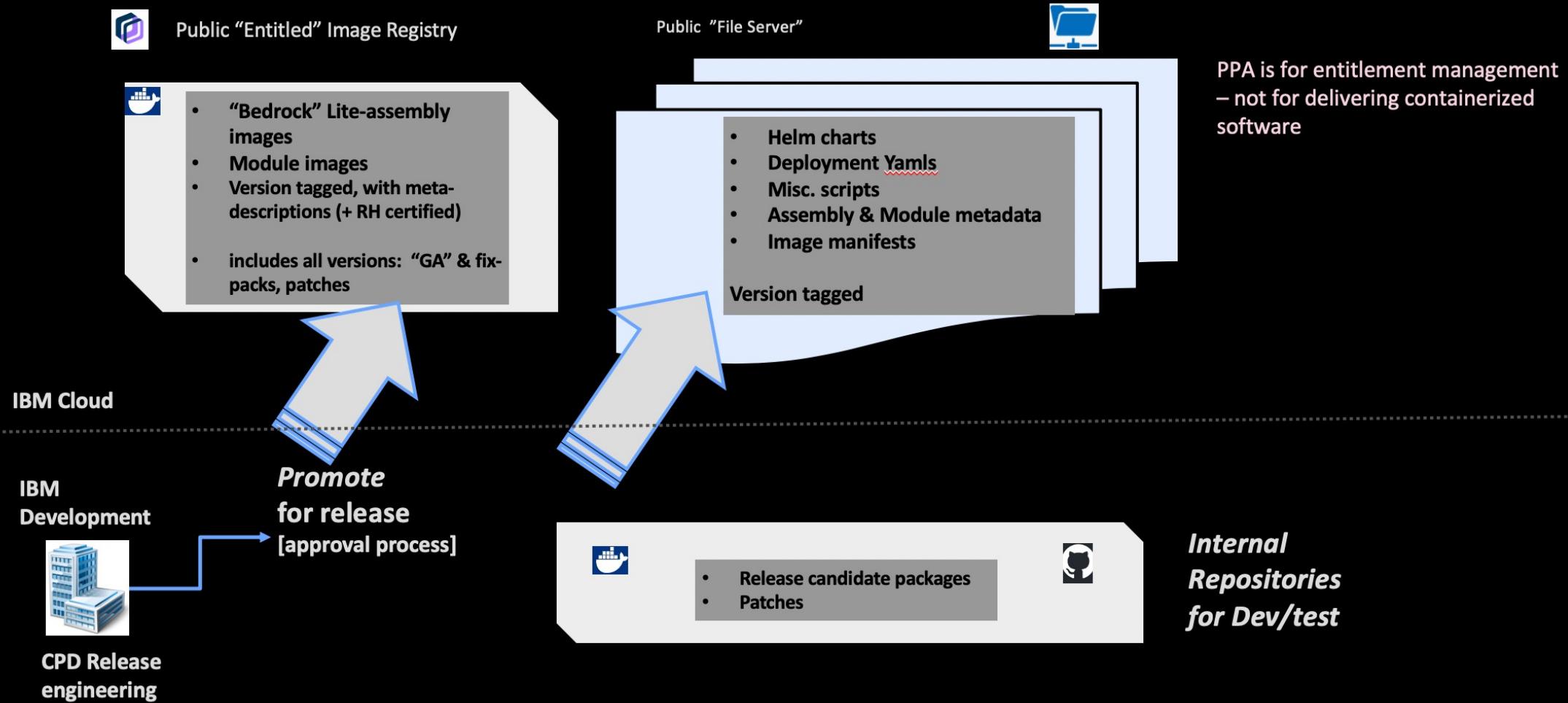
Examples:

WKC
Watson Studio
WML
Data Virtualization
Watson Assistant
Watson Discovery
Watson Open Scale
:
Watson Financial Crimes Insights ...

- Assembly meta-data defines what *all* is to be deployed
 - lists all the included “*modules*” & versions
 - Identifies all the pre-req module dependencies and *order* of installation.
- Goal: Assemblies as a consistent way to *automate* deployments on a target cluster

Cloud Pak for Data adopts a Continuous Delivery model

Public repositories (replicated)



Install approach overview

Step 1. Install RH OpenShift Container Platform on target IaaS or on-prem or provision an OCP Managed Service

Step 2. Deploy & Configure storage manager/ Dynamic Provisioners.

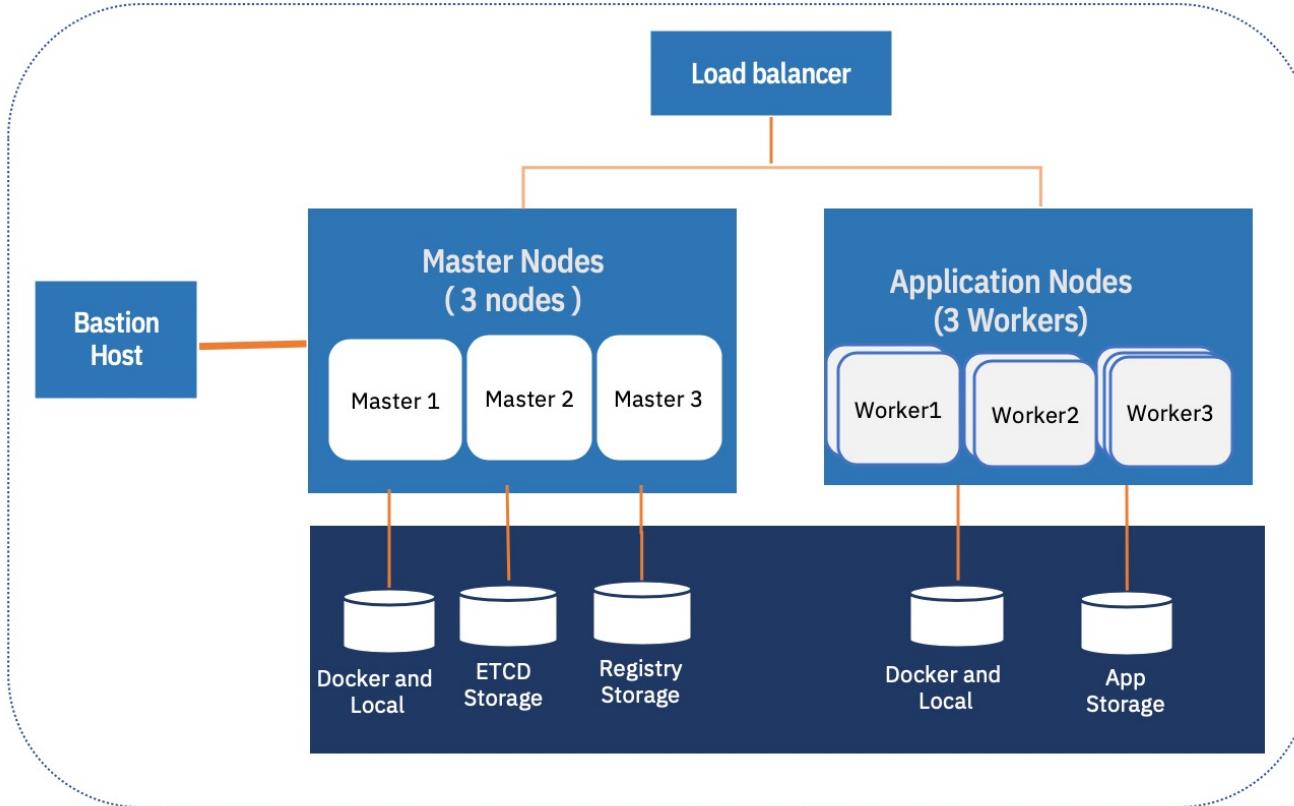
- Most of the *common* ones will be validated over time.
 - Options: Portworx , NFS, Openshift Container Storage v4.x*, NetApp OnTap/Trident*, IBM Spectrum Scale/ESS (via NFS & directly*), IBM Cloud storage classes like ibmc-file-gold-gid

Step 3. Provision Cloud Pak for Data and any add-on Services on this target RH OpenShift Cluster

- Typically from a laptop or workstation – *not* from any OpenShift host.
- Cluster Admin creates Project, sets up Security Context Constraints (and in some cases elevated privileges etc.) -- you can also deploy a *different* copy in a *different* namespace.
- Namespace / Project Admin (or Cluster Admin) “bootstraps” the CPD Control Plane into that namespace.
- Whenever needed, a Namespace/Project Admin deploys one or more add-on Services

* in progress

Cloud Pak for Data 2.5 on OpenShift 3.11



Node	Count	CPU (VPC)	Memory(GB)	Disk storage(GB)
Load Balancer	1	4	8	100
Bastion Host (Local OSE yum Repo, OCP Registry for AIR GAP install)	1	4	16	Root 100 Docker 200 Yum Repo 500
Master / Infra	3	16	64	Root 100 Docker: 500
Worker	3	16	64	100 Docker: 500 NFS: 1 TB

Installable:

Cloud Pak for Data, DV,
Watson Knowledge Catalog

Notes:

- For Air Gap only, on bastion node – we need Docker and Yum Repo, else we just need root.
- Load Balancer – If client does not have an external load balancer and can be optional. External load balancer preferred.
- For Prod env, Master and Infra should be separate with the same specs (Performance impact)
- NFS – 1TB is a guideline and will need adjustment based on workloads and data set sizing.

The ‘cpd’ utility

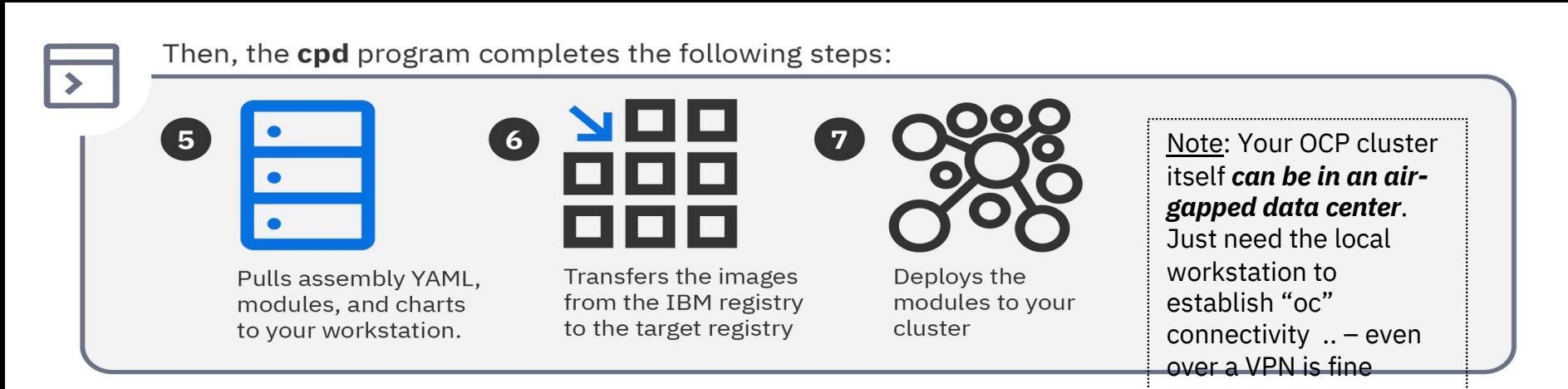
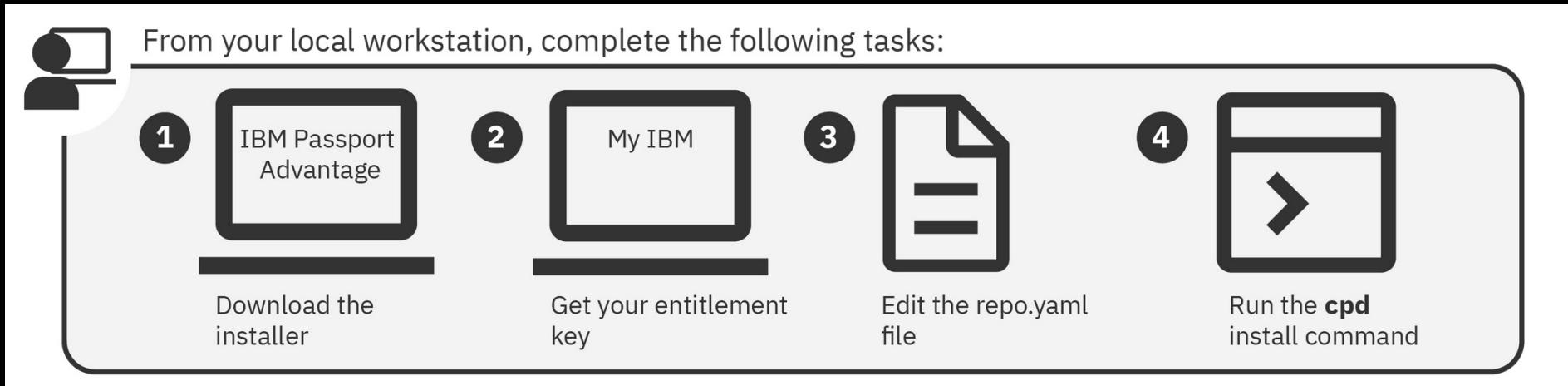
- for installs, upgrades, patches, scaling etc..

executable:-
cpd-linux
cpd-darwin
cpd-windows [in progress]

- A ‘**cpd**’ client-side utility, figures out *what* images, yaml, charts are needed, based on *the “assembly”* that needs to be deployed.
 - It works in tandem with the ‘cpd-install-operator’ inside the cluster.
- cpd *copies* Docker images from the IBM public registry to the target registry. No Docker daemon required on client.
- Metadata helps ‘cpd’ to figure out if shared service modules referenced by different add-ons are already deployed or if upgrade to a newer version of the module is needed.
- The cpd utility also pushes yaml, charts and other orchestration artifacts for the operator to use.
- The operator watches for CRD changes & triggers helm chart installs or other kube actions as needed.
 - The ‘cpd’ cli can also perform some actions directly via the kube API or invoking oc commands

Installing Cloud Pak for Data Services onto OpenShift

When your local workstation can connect to the OpenShift Cluster & to IBM's repositories



Example: Install the Watson Studio assembly

- 1). Connect to the RHOS cluster

```
oc login -u ocadmin -p password https://9.30.210.151:8443
```

repo.yaml - has the co-ordinates and access keys to connect to the IBM Public Registry & file server

- 2). Inspect what is needed to setup the RHOS Project for an “assembly”

```
./cpd-linux adm --repo ./repo.yaml --assembly wsl --namespace cpd1 --dry-run
```

- inspect what commands would be triggered to run to create the RHOS Project/namespace, service accounts etc.
- run the commands independently (different authorized user) if needed.

Approve setup after validation of security privileges etc.

- 3). Run by Cluster Admin to perform setup

```
./cpd-linux adm --repo ./repo.yaml --assembly wsl --namespace cpd1 --apply
```

- run the commands to setup the project

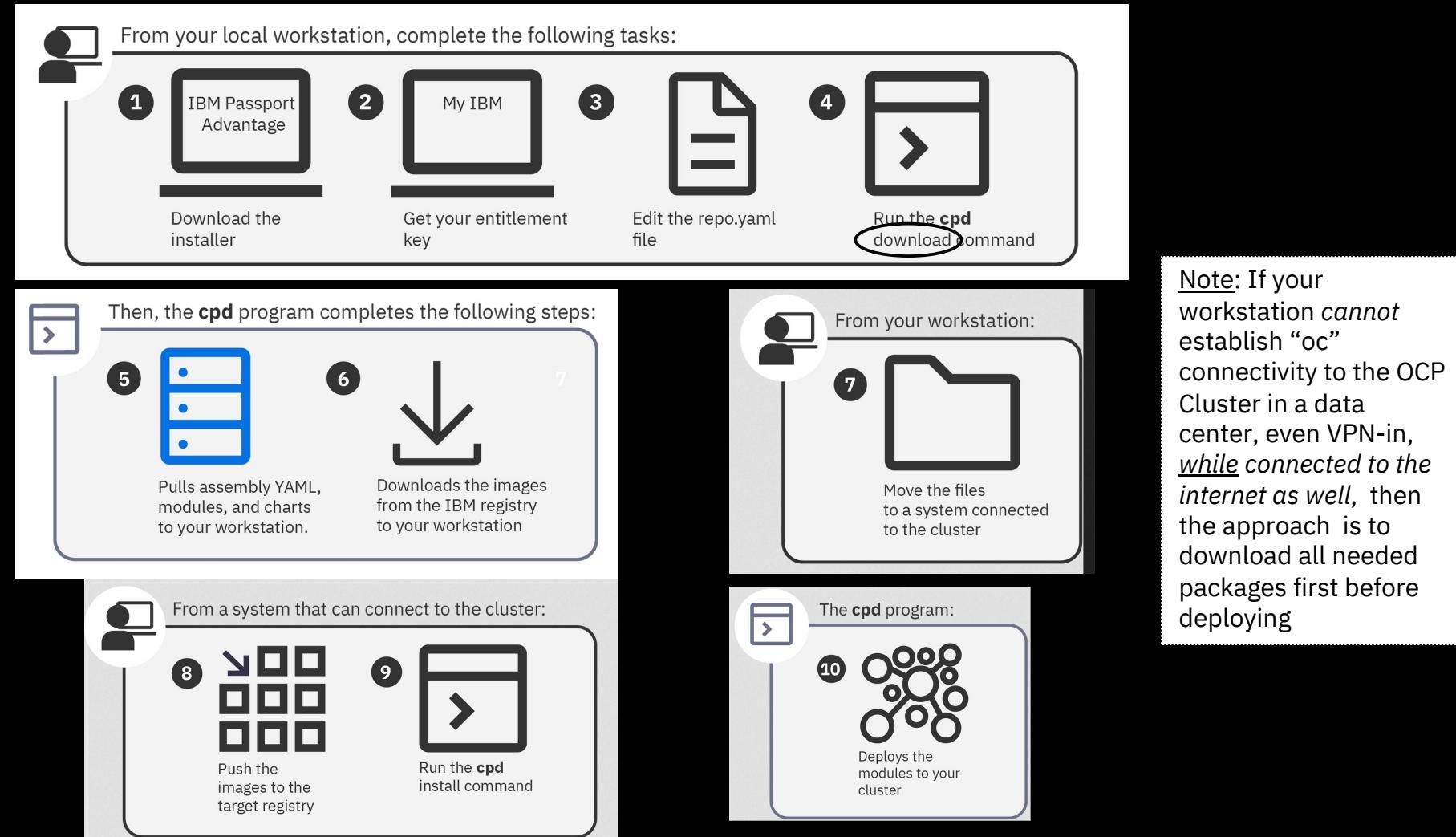
images are in the in-cluster registry in this example
- since the cluster is typically unable to pull from the internet.

- 4). Run by Project Admin to install an Assembly

```
./cpd-linux -a wsl -n cpd1 \
-c <storage-class> -s ./repo.yaml \
--transfer-image-to=docker-registry-default.9.30.51.28.nip.io/cpd1\
--target-registry-username=ocadmin --target-registry-password=$(oc whoami -t) \
--cluster-pull-prefix=docker-registry.default.svc:5000/cpd1
```

Air-gapped: Installing Cloud Pak for Data Services on OpenShift

When your local workstation cannot connect to *both* the OpenShift Cluster & to IBM's repositories



Air-gapped environment Example - Download as .tgz first

- Download all content on-premises & install from .tgz files
 - a). On a machine that has internet access (e.g. your laptop) -
./cpd-linux preloadImages -a lite --action **download** -r ./repo.yaml --download-path=.
 - b). Copy, move or transfer the downloaded files into the air-gapped environment. Make sure the folder structure is preserved
 - c). Push the image to the air-gapped registry (Using Openshift's internal registry as example)

```
oc login https://mycluster.example.com:8443  
oc new-project myproj  
./cpd preloadImages --action push --load-from=. --transfer-image-to=docker-registry.default.svc:5000/myproj --target-  
registry-username=$(oc whoami) --target-registry-password=$(oc whoami -t)
```

52

- d). Install:
(Note the version flag is required when installing in air-gapped environment)
./cpd-linux -a WatsonStudio -n namespace -c <storage-class> --load-from=. \
--cluster-pull-prefix=docker-registry.default.svc:5000/myproj \
--cluster-pull-username=username --cluster-pull-password=password \
--version=2.5.0.0

Air-gapped Alternative- with replication of repositories

- Replicate the registry & file server on-premises

- a). copy images to another registry, download other files locally

```
./cpd-linux preloadImages --download-path ~/work/cpd-build/ -s ./repo.yaml --action transfer --assembly wsl --transfer-image-to dsxl-dev-build1.fyre.ibm.com:5000 --target-registry-username testuser --target-registry-password testpassword
```

- b). Serve the files: -example with nginx

```
podman run --name file-server -p 8080:80 -v ${HOME}/work/cpd-build:/usr/share/nginx/html:ro -d nginx
```

- c). Run by Project Admin to install an Assembly

```
./cpd-linux -a wsl -n cpd1 \ -c <storage-class> -s ./custom-repo.yaml
```

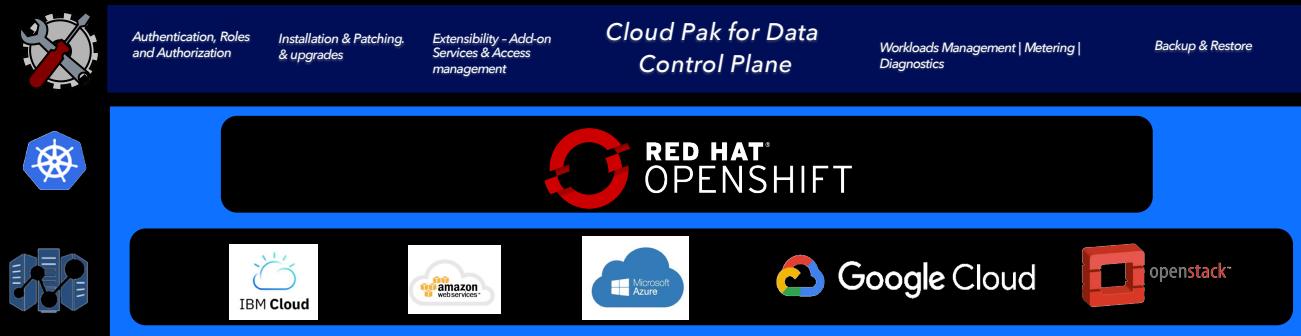
repo.yaml is updated to point
to the target registry & file
server

Cloud Pak for Data: add-on services

- integrated & managed by one control plane

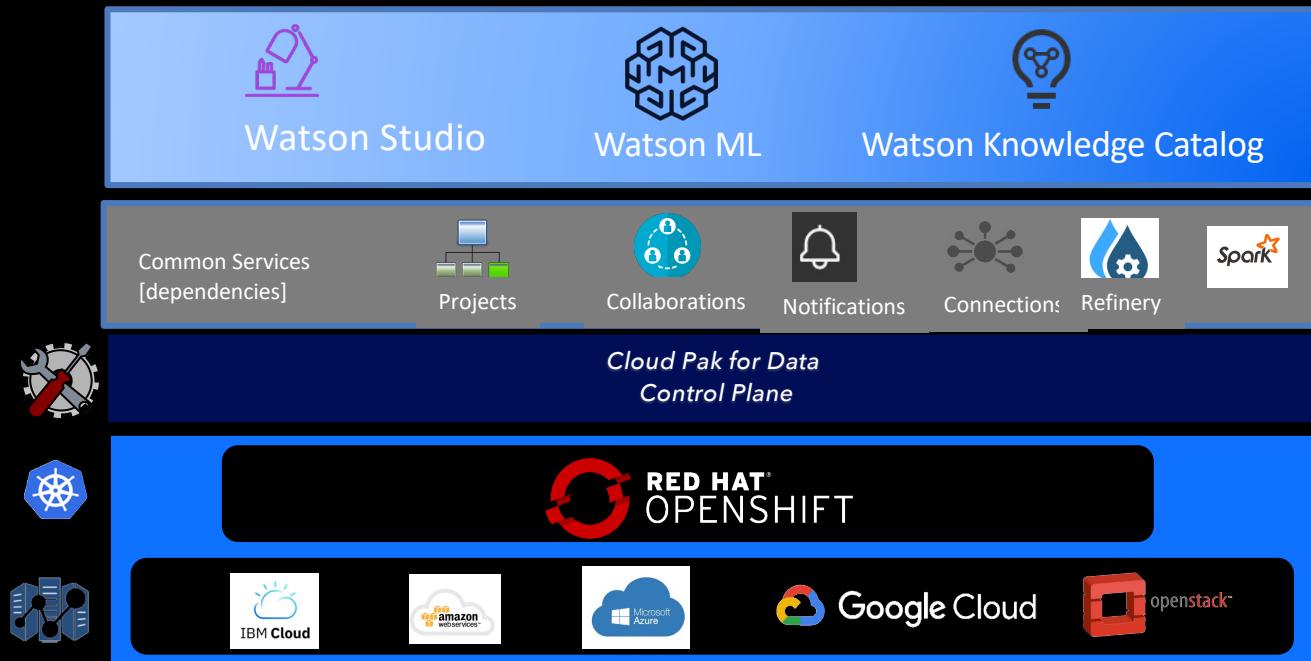
The CPD Control Plane (“Bedrock”) enables a layered approach and plug-n-play

A consistency in delivery, Deployments & management



Plug-n-play with add-on services

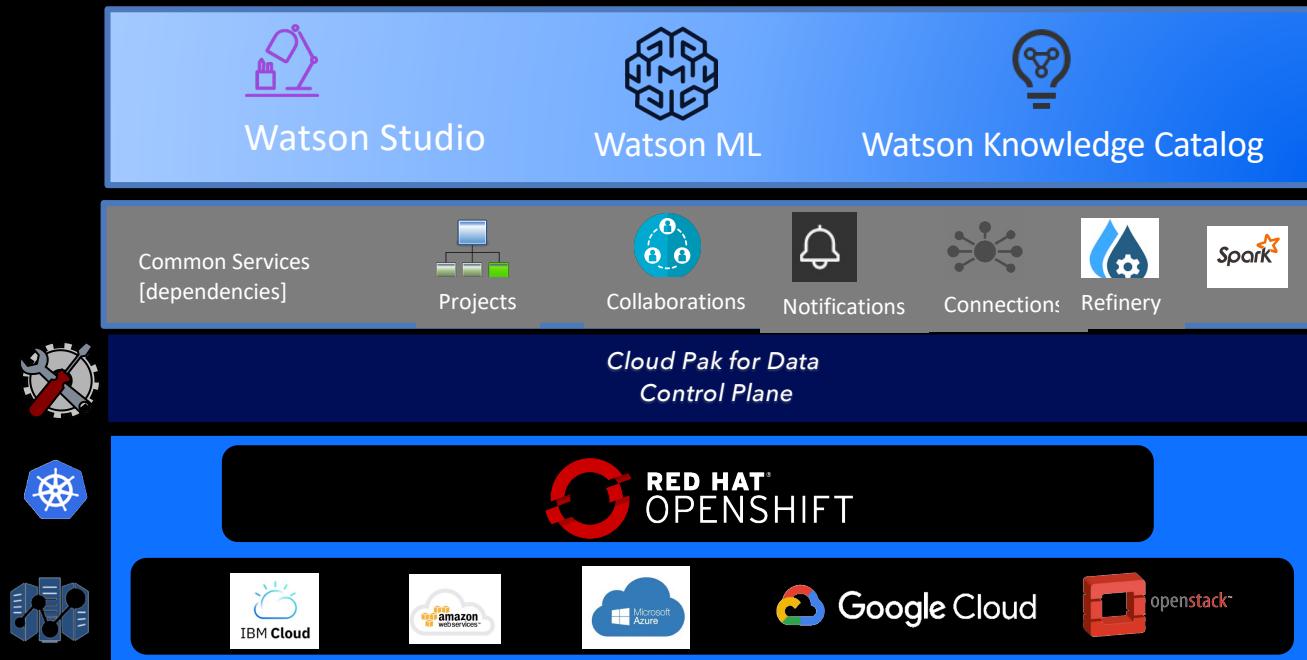
example: Governed Data Science Pattern



Plug-n-play with add-on services

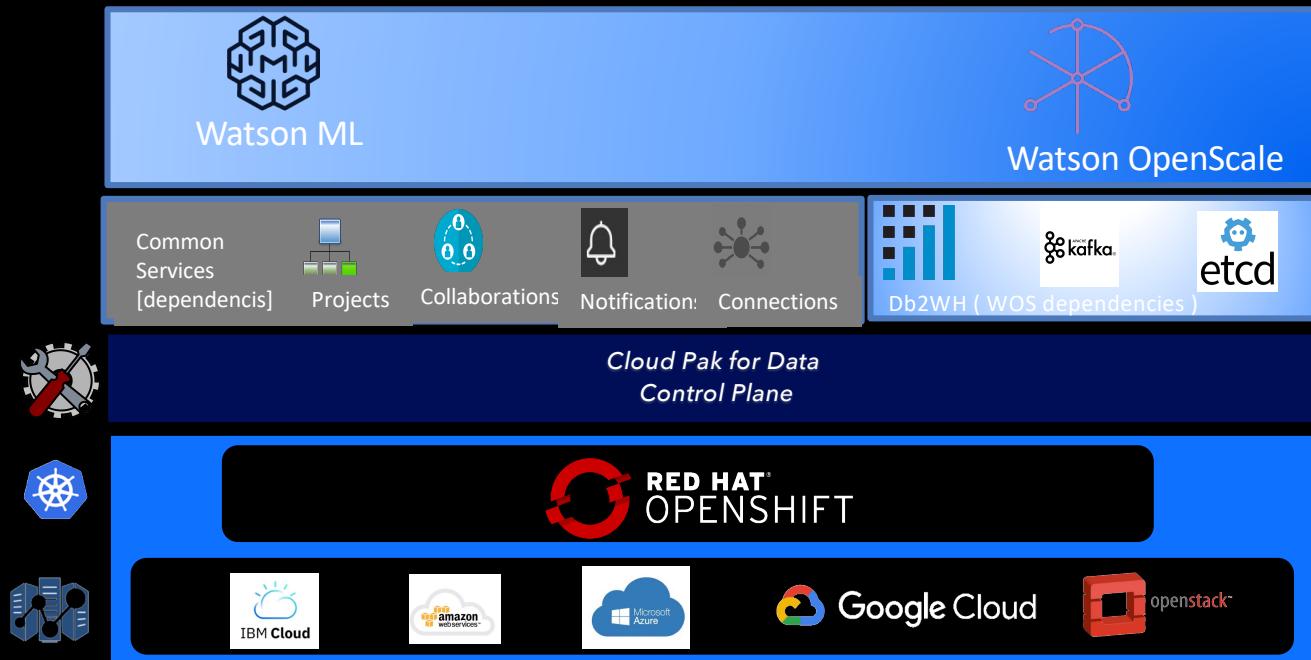
deploy in whatever order you need to

and scale out



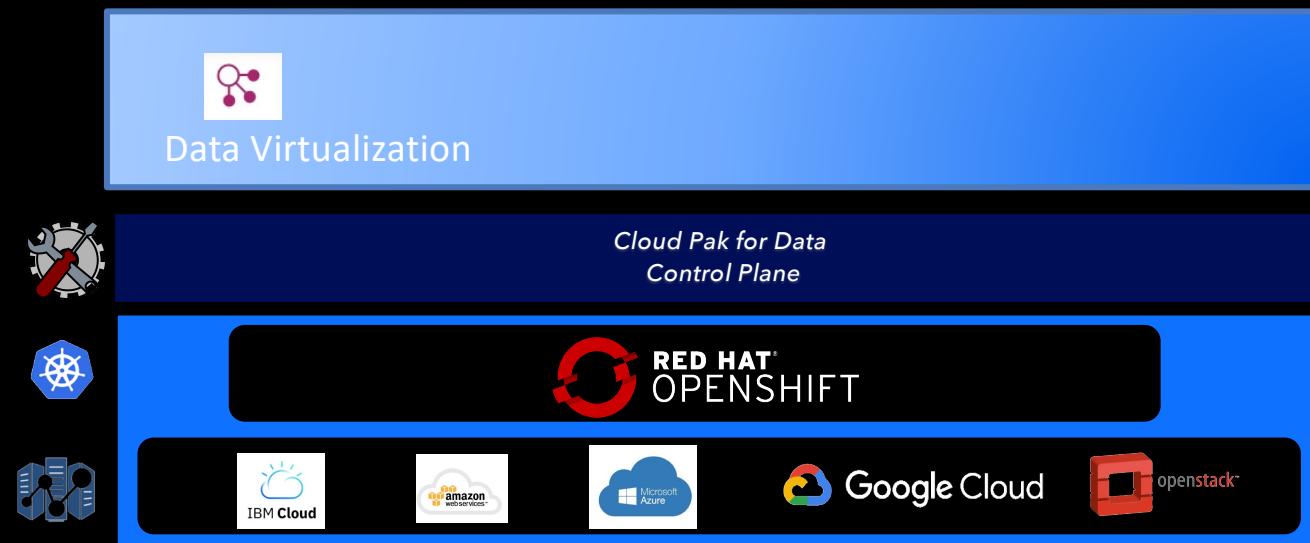
Plug-n-play with add-on services

example: Operationalize ML Pattern



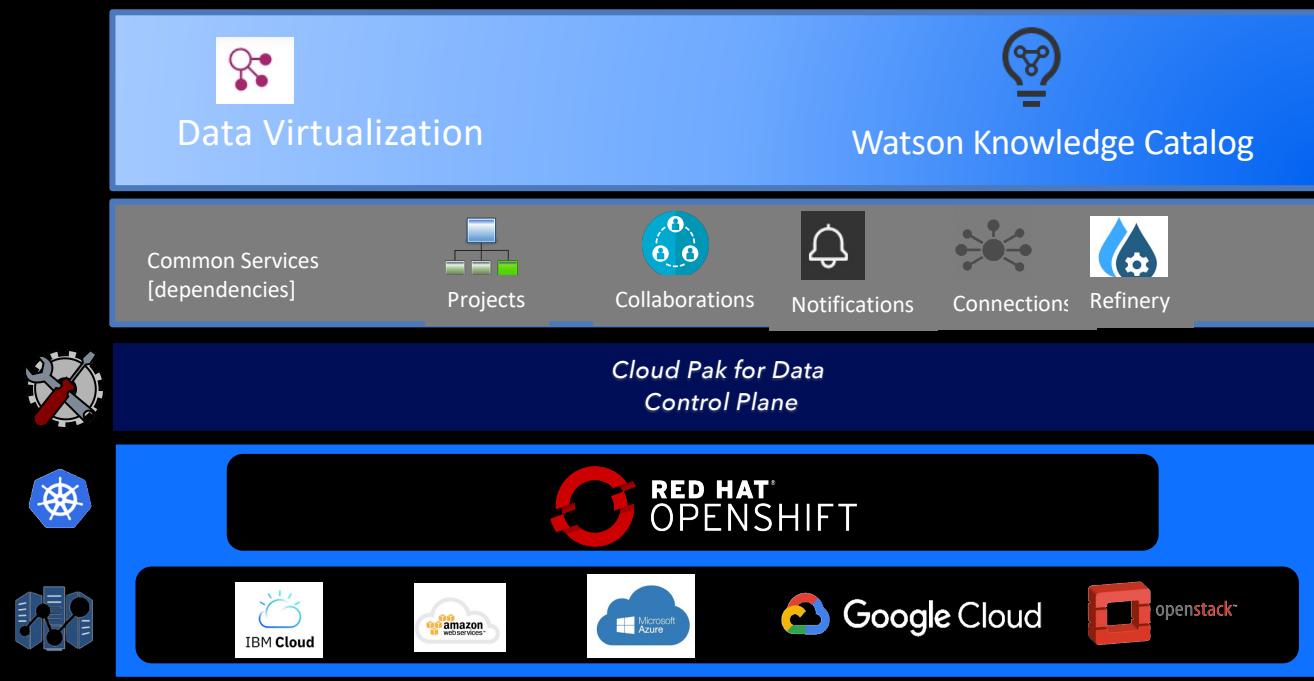
Plug-n-play with add-on services

example: Data Virtualization



Plug-n-play with add-on services

example: Governed Data Access Plane Pattern



The control plane facilitates “standalone” Services

such as Data Stage

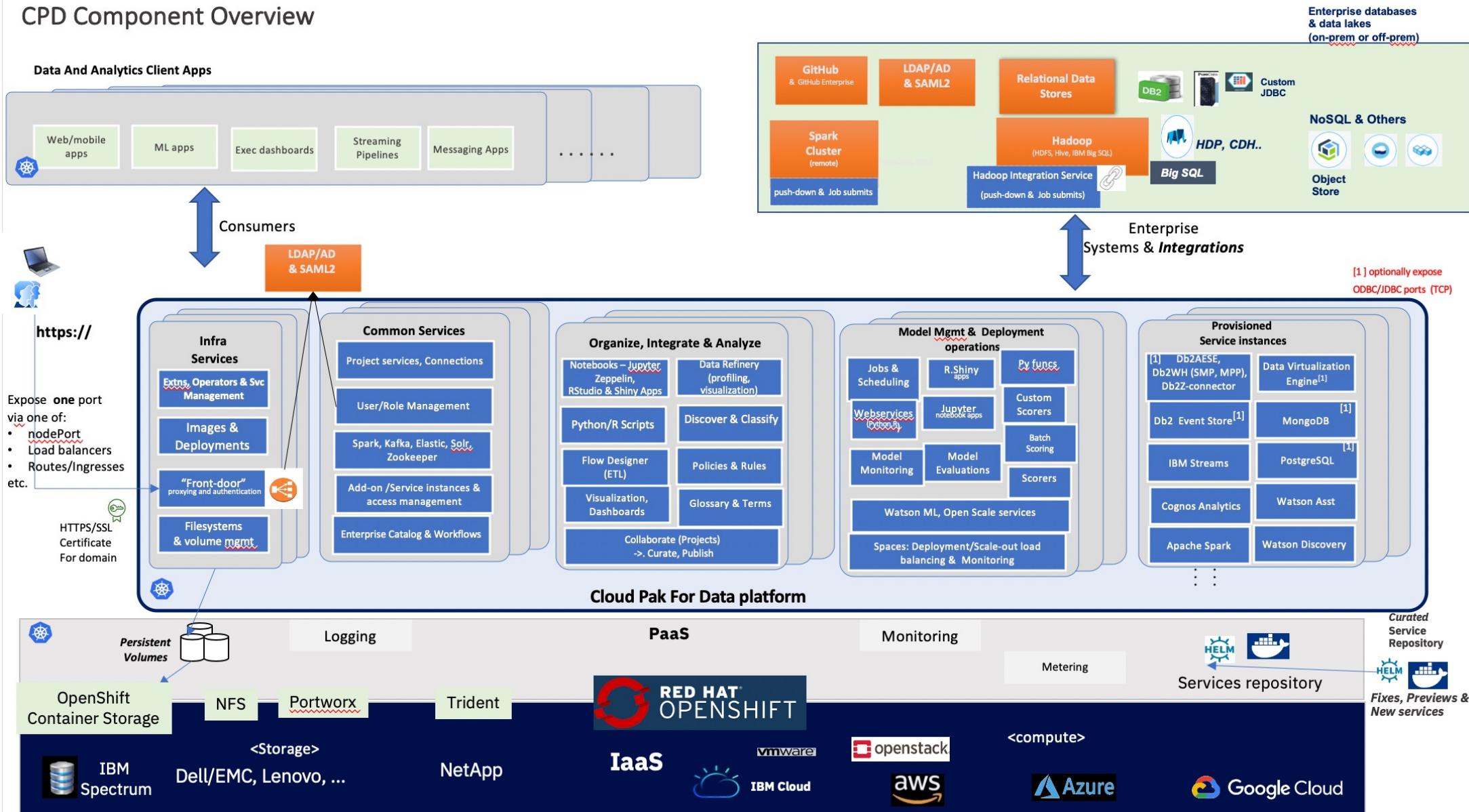


The control plane facilitates “standalone” Services

and Watson AI services



CPD Component Overview



IBM Cloud Pak for Data *Extensions*

- *What is an Extension?*
 - A stand-alone offering combining IBM Cloud Pak for Data Enterprise Edition with additional premium function

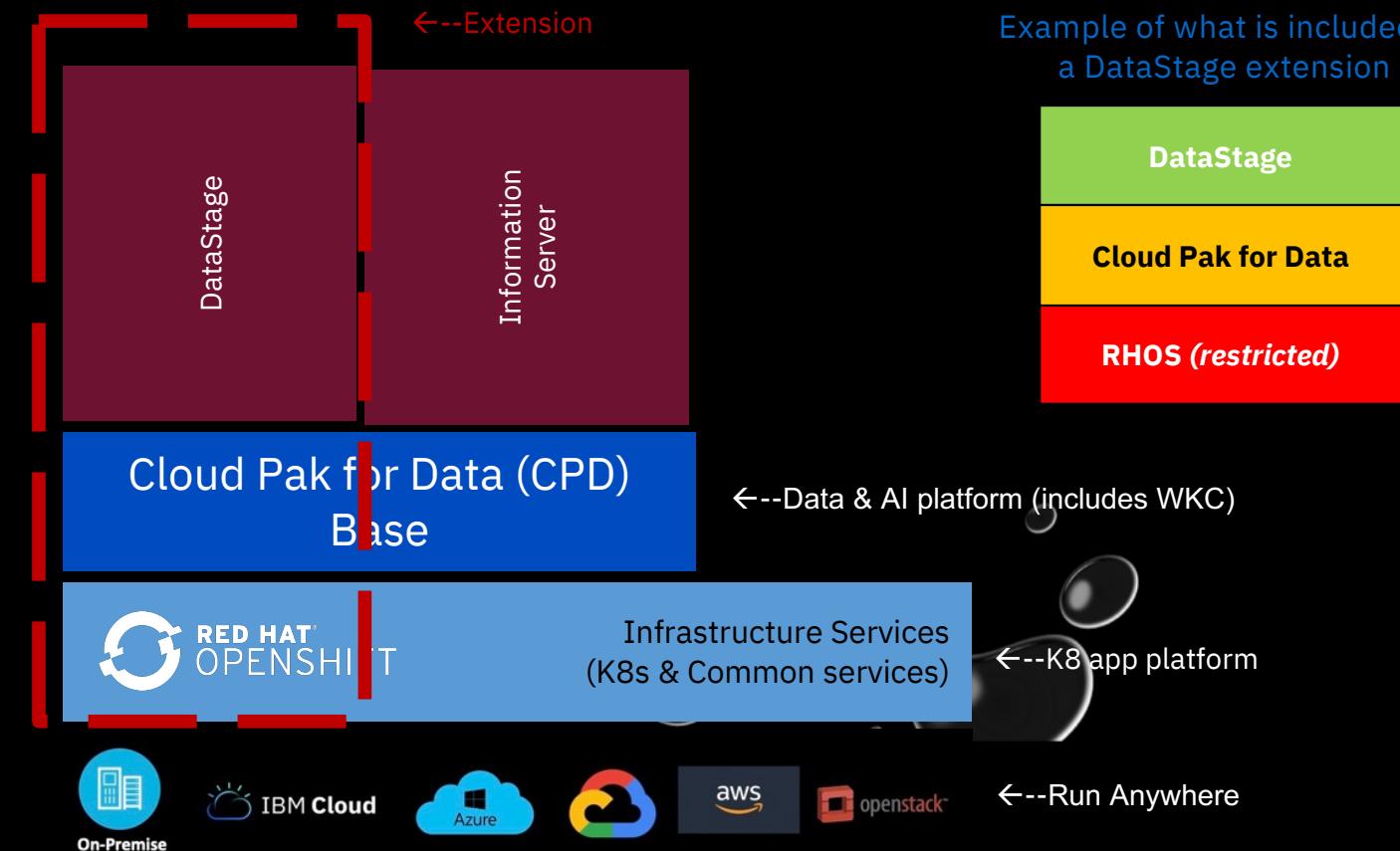
Available Extensions

- ✓ DataStage
- ✓ Information Server
- ✓ Db2
- ✓ Watson Studio Premium (includes SPSS Modeler & DO)
- ✓ Cognos Analytics
- ✓ Watson Assistant
- ✓ Watson Discovery
- ✓ Watson API Kit
- ✓ Watson Financial Crimes Insights

DataStage Extension

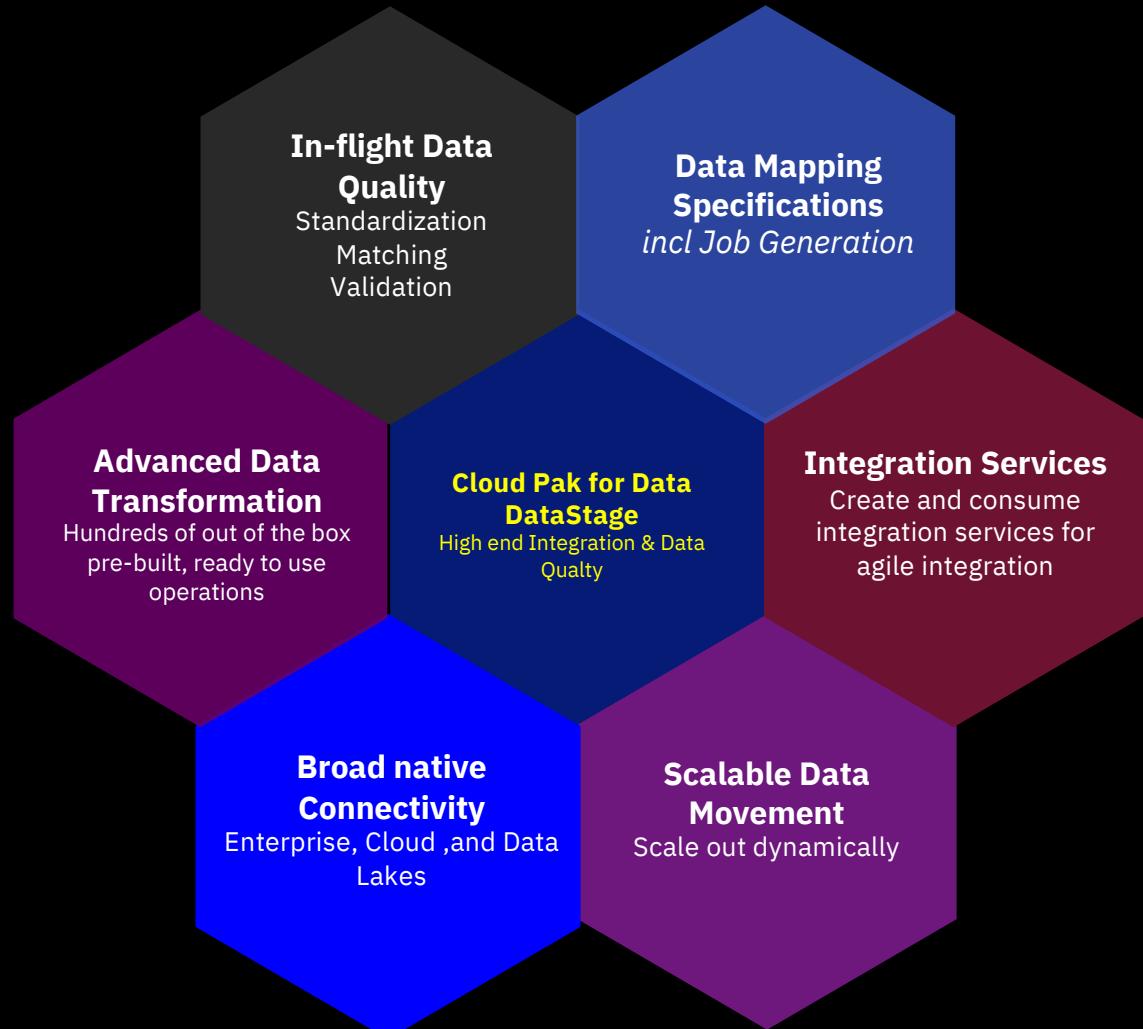
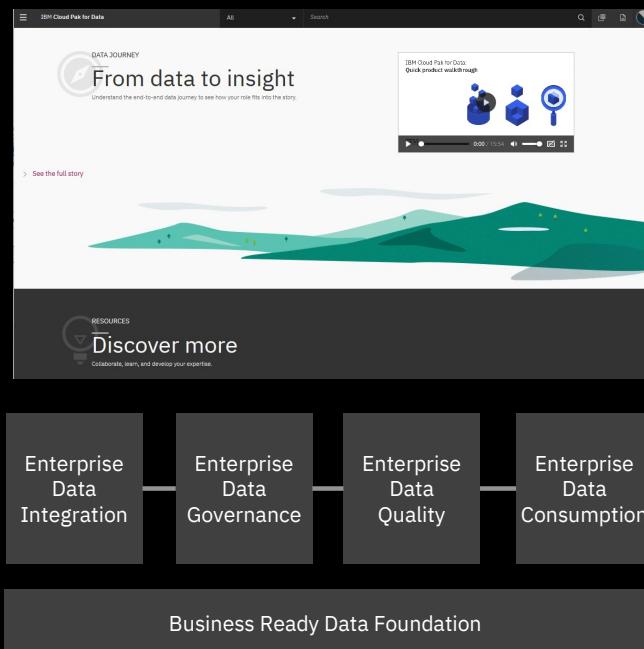
Introduced Oct 15th, 2019

DataStage Offering Delivering Integration



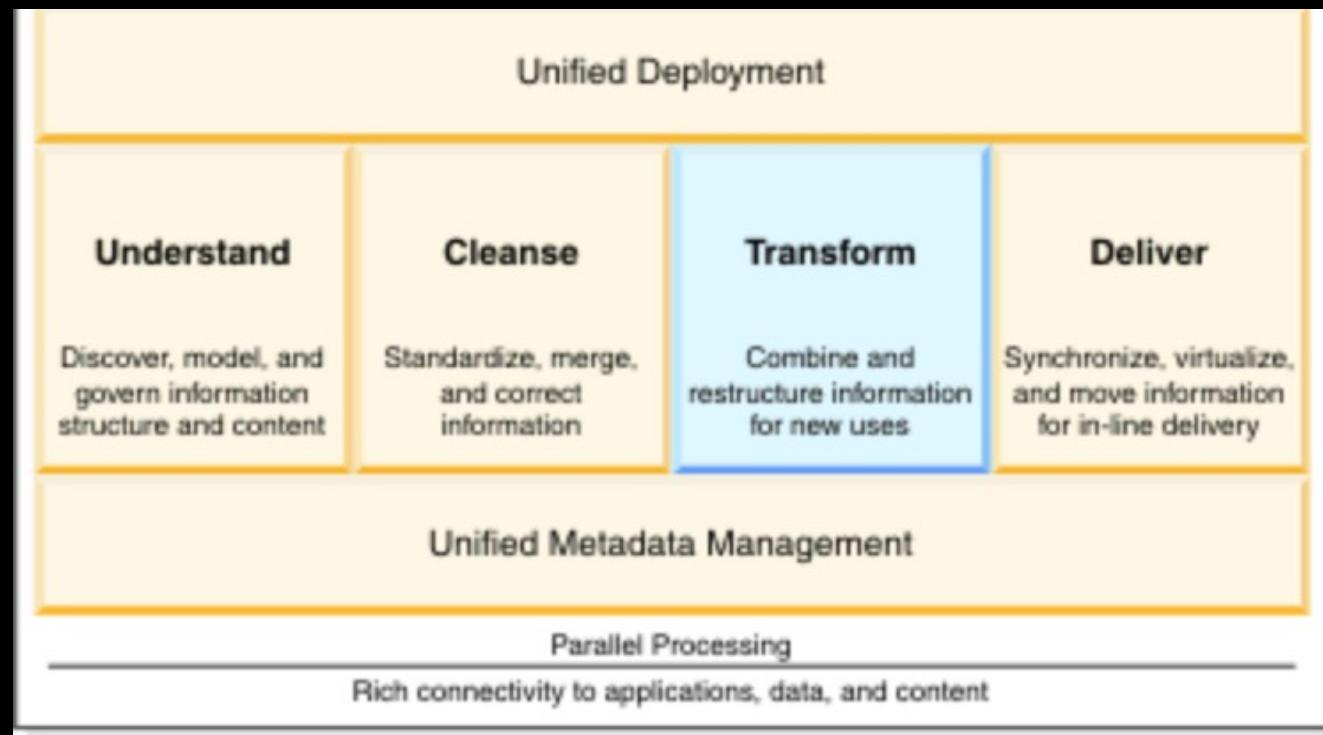
IBM Cloud Pak for Data DataStage

Delivering trusted data to the Enterprise.



IBM Cloud Pak for Data DataStage

Where DataStage fits in the overall business context?



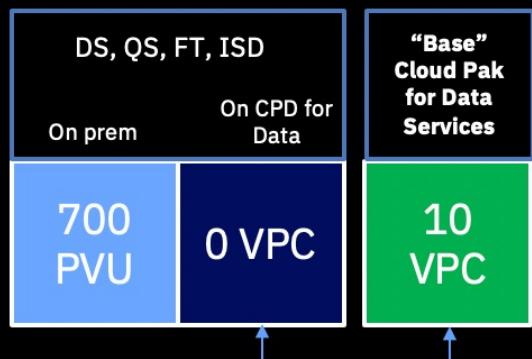
Transformation as part of the integration process

How can IBM Cloud Pak for Data DataStage be used?

Example: Client buys **10 VPC** of IBM Cloud Pak for Data DataStage

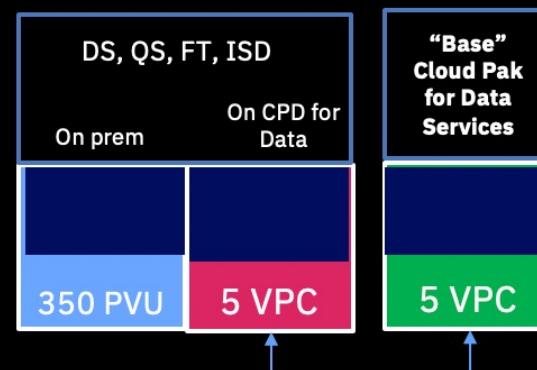
Option 1:

Running CPD DataStage
“on premise”



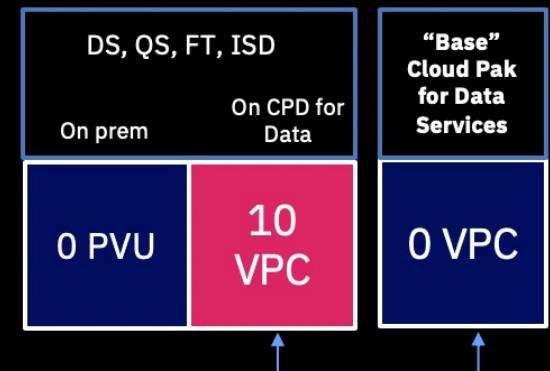
Mix of Option 1 and 2:

Partial use of the Program on
Cloud Pak for Data
(e.g. 50%)



Option 2:

Full use of the Program ON
Cloud Pak for Data



IBM Cloud Pak for Data DataStage

Deploy

The screenshot shows the IBM Cloud Pak for Data Services catalog. On the left, there's a sidebar with categories like All categories, AI, Analytics, Dashboards, Data governance (which is selected), Data sources, Developer tools, Industry solutions, and Storage. Below that is an 'Industry accelerators' section with a link to jump-start analysis of common business problems. The main area displays a grid of services under 'Data governance'. One service, 'DataStage Edition', is highlighted with a tooltip showing its icon, name, IBM status, and a 'Deploy' button. Other visible services include 'Regulatory Accelerator' (by Senzing, Partner, Premium), 'Senzing' (by Senzing, Partner, Premium), and 'Watson Knowledge Catalog' (by IBM, Enabled). Each service card has a brief description and a three-dot menu icon.

Command to deploy

```
cpd-darwin --repo repo.yaml --assembly ds --namespace zen-dv-acc --storageclass ibmc-file-gold-gid --transfer-image-to=$(oc get route -n default docker-registry|tail -1|awk '{print $2}')/zen-dv-acc --target-registry-username=$(oc whoami) --target-registry-password=$(oc whoami -t) --cluster-pull-prefix docker-registry.default.svc:5000/zen-dv-acc --accept-all-licenses --insecure-skip-tls-verify
```

IBM Cloud Pak for Data DataStage

Deployed

IBM Cloud Pak for Data

All

Search

Filter ▾

Find services

Services

All categories

- AI
- Analytics
- Dashboards
- Data governance**
- Data sources
- Developer tools
- Industry solutions
- Storage

Data governance



DataStage Edition

IBM Enabled Premium

Effortlessly deliver data at the right time to the right place with integration, transformation, and delivery of data in batch and real time.



Regulatory Accelerator

IBM Enabled

Streamline the process of complying with regulations by harnessing the power of machine learning.



Senzing

Partner Premium

Senzing is the first real-time, purpose-built artificial intelligence for entity resolution. Discover who is who and who is related to who in your data in real-time



Watson Knowledge Catalog

IBM Enabled

Organize and govern data. Automatically discover, classify, profile, and protect your data so data scientists can find trusted data fast.

Industry accelerators ▾

Jump-start your analysis of common business problems with sample data science assets.

IBM Cloud Pak for Data DataStage – Create a new Job

The screenshot displays the IBM Cloud Pak for Data DataStage interface. On the left, a navigation sidebar lists various categories under 'Organize' and 'Management'. The main area shows a project named 'ds2' with sections for 'Connections', 'Table definitions', 'Parameter sets', and 'Jobs'. A search bar at the top right is set to 'All'. The 'Jobs' section shows two parallel jobs: 'Job_1' and 'vkDSjob1', both of which are currently locked.

In the bottom right corner, a context menu is open from the 'Create' button, offering options to 'Parallel job' or 'Sequence job'.

IBM Cloud Pak for Data DataStage – Connector

The screenshot shows the IBM Cloud Pak for Data DataStage interface. At the top, there's a navigation bar with tabs for "Connections", "Table definitions", "Parameter sets", "Jobs", and "Job_2*". The "Jobs" tab is currently selected. Below the navigation bar, the main area displays a "Project: ds2" and a "PARALLEL JOB Job_2*". A search bar shows "Search (87 shown)". On the left, there's a sidebar titled "Connector" containing a grid of icons for various data connectors. The connectors listed are:

- Row 1: Connection, Amazon S3, Azure Storage, BDFs, Big Query
- Row 2: Cassandra, Classic Federation, Cloud Object Storage, DB2, DRS
- Row 3: Data Set, Distributed Transaction, External Source, External Target, FTP
- Row 4: File, File Set, Google Cloud Storage, Greenplum, HBase
- Row 5: Hierarchical Data, Hive, ISD Input, ISD Output, Informix Enterprise
- Row 6: Informix Load, JDBC, Java, Kafka, Lookup File Set
- Row 7: Netezza, ODBC, Oracle, Salesforce, Sequential File
- Row 8: Snowflake, Sybase, Sybase IQ 12 Load, Sybase OC, Teradata
- Row 9: WebSphere MQ

IBM Cloud Pak for Data DataStage – Stages

IBM Cloud Pak for Data

All

Search

Project: ds2

PARALLEL JOB

Job_2*

Stages

Connections Table definitions Parameter sets Jobs Job_2*

Aggregator Annotation BloomFilter Change Apply Change Capture

Checksum Column Export Column Generator Column Import Combine Records

Compare Compress Copy Decode Difference

Encode Expand External Filter Filter Funnel

Generic Head Join Lookup Make Subrecord

Make Vector Merge Modify Peek Promote Subrecord

Pivot Enterprise Remove Duplicates Row Generator Sample Sort

Split Subrecord Split Vector Surrogate Key Generator Switch Tail

Transformer Wave Generator Range Map

General

Job_2*

The screenshot shows the IBM Cloud Pak for Data DataStage interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data' and a search bar. Below it, a header bar includes tabs for 'Connections', 'Table definitions', 'Parameter sets', 'Jobs', and 'Job_2*'. The main area is titled 'Project: ds2' and 'PARALLEL JOB Job_2*'. It features a grid of icons representing various data processing stages, such as Aggregator, Annotation, BloomFilter, Change Apply, Change Capture, Checksum, Column Export, Column Generator, Column Import, Combine Records, Compare, Compress, Copy, Decode, Difference, Encode, Expand, External Filter, Filter, Funnel, Generic, Head, Join, Lookup, Make Subrecord, Make Vector, Merge, Modify, Peek, Promote Subrecord, Pivot, Remove Duplicates, Row Generator, Sample, Sort, Split Subrecord, Split Vector, Surrogate Key Generator, Switch, Tail, Transformer, Wave Generator, and Range Map. A 'General' section is also present at the bottom. On the right side, there are several small icons for filtering and searching.

Basic Sizing Information

	Small	Medium	Mid - Large
Minimum Entitlement of Cloud Pak for Data DataStage (in VPC) → <i>This is what a client would purchase</i>	6 (better 8)	12	24
Min Number of cores on system → <i>This is how they would setup their environment</i>	Total: 18 (20) 6 CPD Control Plane 6 Services tier 6 (8) Engine Tier E.g. 2 x 3(4) VPC per compute node	Total: 30 6 CPD Control Plane 12 Services tier (depending on # of users and activity, 6 VPCs could be enough) 12 Engine Tier E.g. 3 x 4 VPC per compute node	Total: 42 6 CPD Control Plane 12 Services tier (depending on # of users and activity, 6 VPCs could be enough) 24 Engine Tier E.g. 4 x 6 VPC per compute node
Min RAM	16	32	64

Others:

Min Storage: 100 GB
 Min root file system space on nodes: 30 GB
 Min number of VPC per engine compute node: 2

IBM Cloud Pak for Data Cognos Analytics



Smarter BI with Cognos Analytics on Cloud Pak for Data

Invest in a modern containerized system that built for the future!

Save time with easy installation, maintenance and upgrades

Grows with your business - Instantly add more resources as needed

Enable new services quickly using the power of an integrated Data & AI platform



Smarter BI, Superior Results, Driven by AI

Cognos Analytics on IBM Cloud Pak for Data

You should have detailed conversations with existing customers!

Supported

- Dynamic Query – JDBC
 - Db2 (via the IBM JCC driver datasources)
 - Microsoft SQL Server *
 - MySQL *
 - Oracle *
 - PostgreSQL
 - CPD Virtualization service (via the IBM JCC driver)
- FM provided as client tool
- Data Modules
- Reporting, Dashboarding, Storytelling, Explore
- Content Store database (Db2 base) will be provided with the solution
- Authentication source - CPD authentication service
- Linux x86 conformance only
- Integration Points:
 - Integration with CPD Logging services
 - Installation/onboarding experience (Tiles)

Not Supported/Roadmap

Near Term Roadmap (2H 2020)

- Jupyter Notebooks
- Save to a file system for bursting and output
- MS Office Connection
- Authentication via Cognos (AD,LDAP, OIDC, Siteminder, SAP and CJAP)
- Server routing (routing rules)
- Other supported JDBC datasources

Long Term Roadmap

- Legacy studios (Analysis . Query Studio, Cognos Workspace)
- PowerPlay Studio
- Transformer

No Planned support

- Classic Query
- File system access
- Dynamic Cubes

* See Cognos Analytics Supported Environments page

IBM Cloud Pak for Data – Cognos Analytics

Considerations – Upgrading Cognos Analytics

How to upgrade Cognos Analytics:

- Cognos export deployment of all content (Full or Partial deployment)
 - Retrieve Cognos deployment file from the system
- Deprovision the Cognos instance in Cloud Pak for Data
- Deploy the new version of Cognos Analytics (new tile)
- Provision a new instance of Cognos Analytics
- Import Cognos deployment into new Cognos Analytics instance

IBM Cloud Pak for Data – Cognos Analytics

Sizing considerations

Cognos Analytics Cartridge Scaling

- Cognos starts up with 11 cores to run all needed containers.
- As load increases, Cognos will scale (in seconds) to the purchased configuration's max core count

Cognos Analytics **Cartridge** Sizing

- A Cognos Analytics Sizing SME will use the sizing PDF
- The SME will determine the appropriate number of cores needed.
- When translating cores to VPC: 1 core = 1 VPC

Sizing Summary

Small: min 11 - max 16

Medium: min 11 - max 21

Large: min 11 - max 25

IBM Cloud Pak for Data – Cognos Analytics Cartridge

*Sizing considerations – Cores**

Minimum Number of Cores

- Red Hat OpenShift – See their documentation for system requirements
 - There is no additional charge for RHOS with the Cognos Cartridge
- Cloud Pak for Data – Minimum 6 cores required
 - There is no Cognos Analytics cost for these 6 cores
- Cognos Analytics – Customers must purchase the maximum of cores required for their configuration
 - Sizing Summary
 - Small min 11 - max 16
 - Medium min 11 - max 21
 - Large min 11 - max 25

Notes

Why Maximum

CPD today cannot meter additional VPC above 11 or handle the concept of a pool of shared VPC's that can be used by any application under load (scale out).

What does this mean for Cognos Analytic on CP today:

Customer cannot buy minimum number of VPC's because at scale, they will be out of compliance.

The customer must buy the maximum including Cognos Content Manager which is measured in the Cloud-Pak model.

* 1 Core = 1 VPC (70:1 Legacy PVU to VPC conversion)

IBM Cloud Pak for Data – Cognos Analytics Cartridge

Considerations – Installation of multiple environments

Installation Experience for Cloud Pak for Data package

- Customer is entitled to Red Hat Openshift and must install and configure it themselves
- Customer then installs Cloud Pak for Data
- Customer then deploys Cognos Analytics into Cloud Pak for data which enables it to be provisioned

IMPORTANT:

- Each instance of Cognos Analytics will require an instance of Cloud Pak for Data today. **This includes Non-Production as well**
- This is currently being worked on so that in the future a single instance of Cloud Pak for Data can manage multiple instances of Cognos Analytics

User Experience Flow

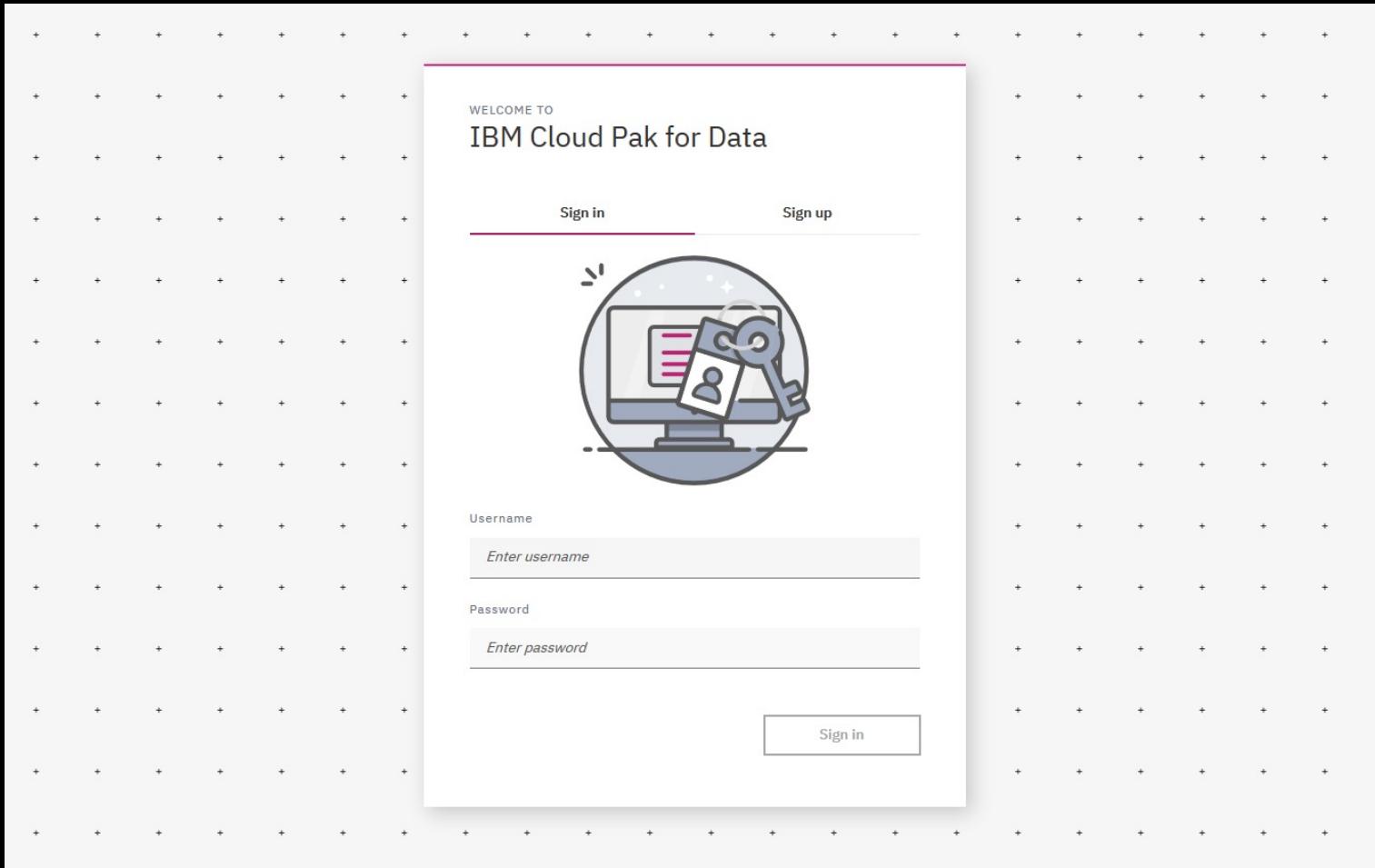
IBM Cloud Pak for Data – Cognos Analytics

What will the Experience be

- Customer will launch CPD UI and have the Cognos Analytics premium tile available
- Customer will select Getting Started option from the tile
- Customer is presented with
 - Marketing material on Cognos Analytics
 - How to purchase the Cognos Analytics Add-on
- Customer goes to Passport Advantage and downloads the Cognos Analytics package
- Customer then deploys the Cognos Analytics package into CPD
 - Once deployed it is then ready to be provisioned
- Customer can then select premium tile for Cognos Analytics and provision
- CPD customer gives users permission to access the tile through the CPD catalog
- Approved users can now access Cognos Analytics

IBM Cloud Pak for Data – Cognos Analytics

Before Provisioning



IBM Cloud Pak for Data – Cognos Analytics

Before Provisioning

The screenshot shows the 'IBM Cloud Pak for Data' interface. At the top, there's a navigation bar with a menu icon, the title 'IBM Cloud Pak for Data', and user profile icons. Below the header, a large circular graphic with a pink-to-black gradient is displayed next to the text 'WELCOME, admin!'. A prominent call-to-action button says 'Let's get started!'. Below this, a sub-header reads 'Use these resources to make the most of your IBM Cloud Pak for Data experience.' On the right side, a section titled 'Administer and monitor' is highlighted with a red underline. It contains a video thumbnail for 'IBM Cloud Pak for Data: Overview' and a list of administrative tasks:

- Set up an LDAP server
- Manage users
- Install services
- Enable email notifications
- Track your resource usage

IBM Cloud Pak for Data – Cognos Analytics

Before Provisioning

The screenshot shows the main landing page of the IBM Cloud Pak for Data interface. At the top right, there is a navigation bar with icons for search, refresh, and user profile. Below the navigation bar, a blue callout bubble points to the 'Add-ons' icon, which is located next to the user profile icon. The callout contains the text 'Presents the Add-On Pages'. The main content area features a large pie chart graphic and the text 'Let's get started!'. Below this, there is a section titled 'Administer and monitor' which includes a video player for an 'Overview' video and a list of administrative tasks.

WELCOME, admin!

Let's get started!

Use these resources to make the most of your IBM Cloud Pak for Data experience.

Administrator and monitor

IBM Cloud Pak for Data: Overview

Get IBM Cloud Pak for Data set up so that your users can get to work.

- Set up an LDAP server
- Manage users
- Install services
- Enable email notifications
- Track your resource usage

https://tom-cpd-tom.apps.obi-ocp1-lb-1.fyre.ibm.com/zen/#/addons

IBM Cloud Pak for Data – Cognos Analytics

Before Provisioning

The screenshot shows the 'Services' section of the IBM Cloud Pak for Data catalog. A blue callout bubble points to the 'Analytics' category in the sidebar, which is highlighted with a blue arrow. The main content area displays various analytics services:

- ShareInsights**: Accelerate ShareInsights. Partner. Premium. Explore, transform, and visualize large, complex data sets with ease.
- Cognos Analytics**: Self-service analytics, infused with AI and machine learning, enable you to create stunning visualizations and share your findings through dashboards and reports.
- Data Refinery**: Simplify the process of preparing large amounts of raw data for analysis.
- Datameer**: Data exploration and preparation platform that enables enterprises to unlock and analyze all their raw data.
- Decision Optimization**: Evaluate millions of possibilities to find the best solution to any given problem.
- Figure Eight**: Figure Eight combines human and machine intelligence to provide high-quality annotated training data that powers the world's most innovative machine learning and business solutions.
- Intel Distribution of Python**: Popular Python libraries, including analytics and machine learning libraries, accelerated for Intel architecture.
- SPSS Modeler**: Create flows to prepare and blend data, build and manage models, and visualize the results.
- Streams**: Derive real-time insight from your in-flight data by developing and deploying real-time streaming applications.

The sidebar also lists other categories like AI, Dashboards, Data governance, Data sources, Developer tools, Industry solutions, Storage, and Industry accelerators. A note at the bottom encourages checking out available industry accelerators.

IBM Cloud Pak for Data – Cognos Analytics

Before Provisioning

The screenshot shows the 'Services' page in the IBM Cloud Pak for Data interface. The left sidebar lists categories like All categories, AI, Analytics, Dashboards, Data governance, Data sources, Developer tools, Industry solutions, Storage, and Industry accelerators. The main area displays a grid of service cards under the 'Analytics' category. One card for 'Cognos Analytics' is highlighted with a blue callout bubble containing the text: 'Getting started: information needed to download the add on'. A mouse cursor is hovering over the 'Get started' button on the Cognos Analytics card. A modal window titled 'Get started' provides instructions: 'To use this add-on, you must download the installation package.' and 'To get started, see [Downloading the installation package](#)'. Other visible cards include ShareInsights, Data Refinery, Datameer, Decision Optimizer, Intel Distribution of Python, SPSS Modeler, and Streams.

Getting started:
information needed to
download the add on

Get started

To use this add-on, you must download the installation package.

To get started, see [Downloading the installation package](#)

ShareInsights Accelerite ShareInsights Partner Premium

Cognos Analytics IBM Get started

Data Refinery IBM Premium

Datameer DataRefiner Partner Premium

Decision Optimizer IBM Premium

Intel Distribution of Python Partner Premium

SPSS Modeler IBM Premium

Streams IBM

Installing Cognos Analytics

1. Download from Passport Advantage



2. Deploy CA Add-On to CPD



The screenshot shows the 'Add-ons' page in the CPD interface. The 'Analytics' section is highlighted, displaying the 'Cognos Analytics' add-on. Other categories shown include 'Industry accelerators' (Contact Center Optimization, Customer 360 Degree View, Customer Churn Management) and 'Databases' (Db2 Warehouse). A search bar at the top right is labeled 'Search for Add-ons'.



3. Cognos Analytics is available in CPD Catalog

IBM Cloud Pak for Data – Cognos Analytics Provisioning

The screenshot shows the IBM Cloud Pak for Data Services catalog interface. The left sidebar lists categories: All categories, AI, Analytics (selected), Dashboards, Data governance, Data sources, Developer tools, Industry solutions, Storage, and Industry accelerators. The main content area displays the 'Analytics' section with the following services:

- Accelerite ShareInsights** (ShareInsights, Partner, Premium): Explore, transform, and visualize large, complex data sets with ease.
- Cognos Analytics** (IBM, Available, Premium): Self-service analytics, infused with AI and machine learning, enable you to create stunning visualizations and share your findings through dashboards and reports.
- Data Refinery** (IBM, Premium): Simplify the process of preparing large amounts of raw data for analysis.
- Datameer** (Datameer, Partner, Premium): Datameer is a data exploration and preparation platform that enables enterprises to unlock and analyze all their raw data.
- Decision Optimization** (IBM, Premium): Evaluate millions of possibilities to find the best solution to any given problem.
- Figure Eight** (Figure Eight, Partner, Premium): Figure Eight combines human and machine intelligence to provide high-quality annotated training data that powers the world's most innovative machine learning and business solutions.
- Intel Distribution of Python** (intel, Partner, Premium): Popular Python libraries, including analytics and machine learning libraries, accelerated for Intel architecture.
- SPSS Modeler** (IBM, Premium): Create flows to prepare and blend data, build and manage models, and visualize the results.
- Streams** (IBM): Derive real-time insight from your in-flight data by developing and deploying real-time streaming applications.

IBM Cloud Pak for Data – Cognos Analytics Provisioning

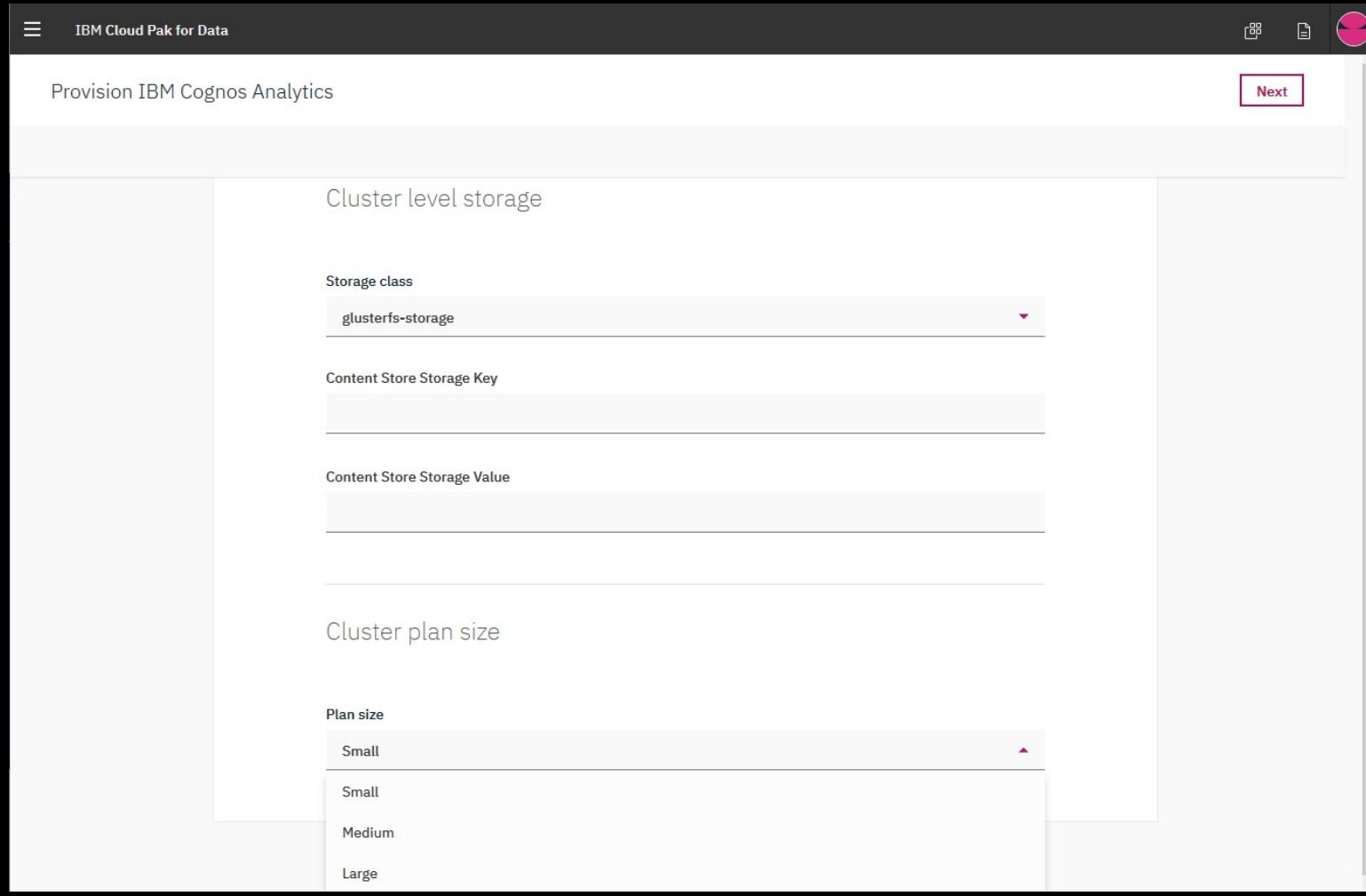
The screenshot shows the 'Services' section of the IBM Cloud Pak for Data interface. On the left, a sidebar lists categories like AI, Analytics, Dashboards, etc. A blue callout bubble points to the 'Analytics' category. In the main area, under the 'Analytics' heading, there are several service cards:

- ShareInsights**: Accelerate ShareInsights. Partner. Premium. Explore, transform, and visualize large, complex data sets with ease.
- Cognos Analytics**: Self-service analytics, infused with AI and machine learning, enable you to create stunning visualizations and share your findings through dashboards and reports. This card is highlighted with a yellow box and has a 'Provision instance' button. A blue callout bubble points to this button with the text 'Click here to initiate provisioning of CA instance inside ICP4D'.
- Data Refinery**: Simplify the process of preparing large amounts of raw data for analysis. IBM. Premium.
- Datameer**: Datameer is a data exploration and preparation platform that enables enterprises to unlock and analyze all their raw data. Partner. Premium.
- Decision Optimization**: Evaluate millions of possibilities to find the best solution to any given problem. IBM. Premium.
- Figure Eight**: Figure Eight combines human and machine intelligence to provide high-quality annotated training data that powers the world's most innovative machine learning and business solutions. IBM. Premium.
- Intel Distribution of Python**: Popular Python libraries, including analytics and machine learning libraries, accelerated for Intel architecture. Partner. Premium.
- SPSS Modeler**: Create flows to prepare and blend data, build and manage models, and visualize the results. IBM. Premium.
- Streams**: Derive real-time insight from your in-flight data by developing and deploying real-time streaming applications. IBM.

At the top right of the main area, there are 'Filter' and 'Find services' buttons.

IBM Cloud Pak for Data – Cognos Analytics

Provision Instance



IBM Cloud Pak for Data – Cognos Analytics

Provision Instance

The screenshot shows a provisioning interface for an IBM Cloud Pak for Data instance. At the top, there's a navigation bar with the title "IBM Cloud Pak for Data" and a user icon. Below the title, the sub-page title is "Provision IBM Cognos Analytics". On the right side of the header are "Back" and "Provision" buttons. The main content area has a heading "You're nearly ready to go!" followed by a sub-instruction: "Review the full summary to make sure everything is correct, and then hit Provision to provision the application." Below this, there's a "Summary" section with various configuration details:

Storage Class	glusterfs-storage
Plan Size	Small
Content Storage	40Gi
Minimum Memory	24.39Gi
Maximum Memory	57.56Gi
Minimum Cpu	10
Maximum Cpu	18
Content Store Storage Label Key	
Content Store Storage Label Value	

IBM Cloud Pak for Data – Cognos Analytics

The screenshot shows the IBM Cognos Analytics home page. The left sidebar includes links for Home, Search, My content, Team content, Recent, Manage, and New. The main area features a "Welcome to IBM Cognos Analytics" message with a "Get started by opening a dashboard, report or story!" button. A central placeholder says "Your content will appear here. Add some data or try our samples to get started." Below it is a "Drag and drop files, open Quick launch or Browse" button. The right sidebar has a "Quick reference" section with links for Get started (Overview, Get started videos), Sample data, and Support. A blue callout bubble points to the "Get started" link in the "Quick reference" section with the text "SSO into CA using token from IC4PD".

IBM Cognos Analytics

Welcome

Home

Search

My content

Team content

Recent

Manage

New

... ●

Quick reference

Get started

- Overview
- Get started videos

Sample data

Support

SSO into CA using token from IC4PD

Welcome to IBM Cognos Analytics

Get started by opening a dashboard, report or story!

Your content will appear here.

Add some data or try our samples to get started.

Drag and drop files, open Quick launch or [Browse](#)

Data Sources

IBM Cloud Pak for Data – Cognos Analytics

Data Sources

The screenshot shows the 'Add-ons' section of the IBM Cloud Private for Data interface. On the left, a sidebar lists categories: All Categories, AI, Analytics, Dashboards, **Data sources**, Developer tools, and Data governance. The 'Data sources' category is selected. The main area displays several add-on cards:

- Db2 Event Store** (IBM Premium): In-memory data store capable of extremely high speed ingest and deep, real-time analytics.
- IBM Db2** (IBM Premium): Relational database that delivers advanced data management and analytics capabilities for transactional and warehousing workloads.
- Db2 Advanced Enterprise Server Edition** (IBM Premium): Relational database that delivers advanced data management and analytics capabilities for transactional and warehousing workloads.
- Db2 Warehouse** (IBM): Data warehouse designed for high-performance, in-database analytics.
- IBM Db2 for z/OS** (IBM): Create databases in Db2 for z/OS and work directly with the data from IBM Cloud Private for Data.
- Data Virtualization** (IBM): Query many data sources as one.
- MongoDB** (Partner): Scalable, open source NoSQL database.

A large blue arrow points from the 'Data Virtualization (Db2)' card to a blue box labeled 'Db2 Instance'. To the right of the 'Db2 Instance' box is a green 3D cylinder icon with the word 'DB2' on it.

CI/CD pipelines - implementation overview and example using Tekton

<https://ibm.seismic.com/Link/Content/DCPvVIJFde40OIBrjrFP2OFw>

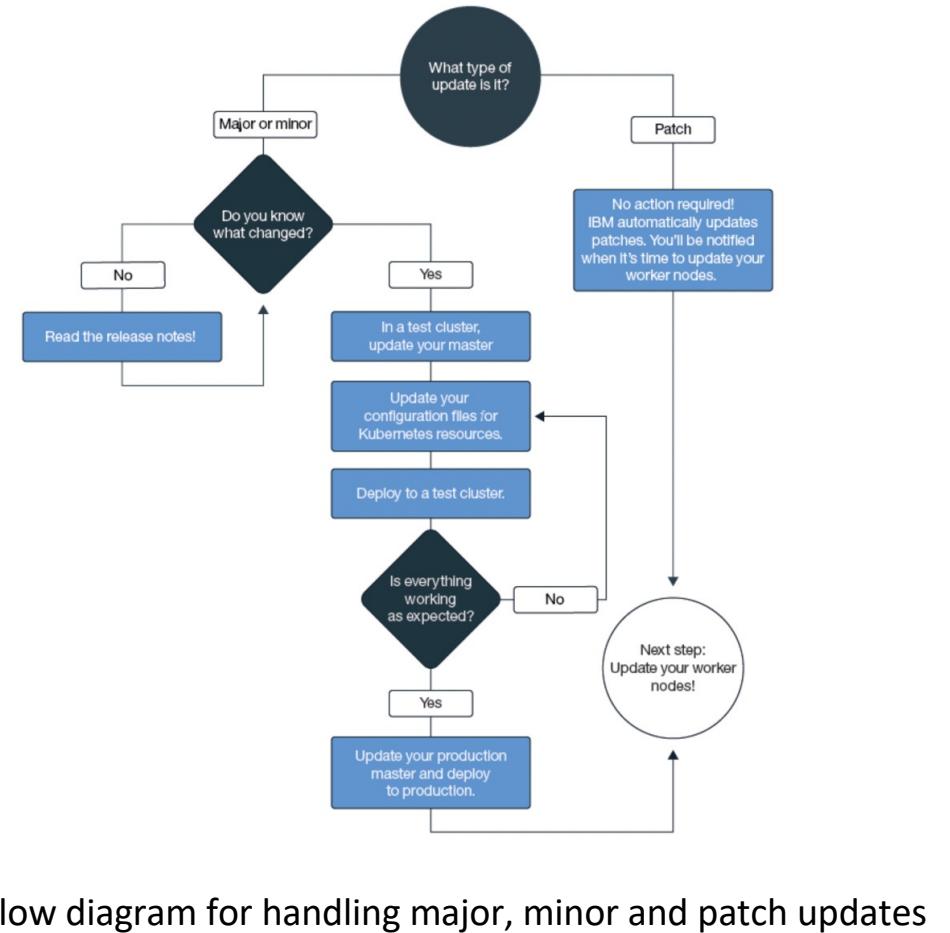
Updates to the OpenShift platform (masters, workers)

Platform updates happen!

- Security fixes (typically patches)
- New features
- API (e.g. Kubernetes) updates

Patch updates are automatically rolled out by IBM to master nodes

Cluster admins choose when to implement master updates for major/minor updates



Source: <https://cloud.ibm.com/docs/openshift?topic=openshift-update#master>

Updating clusters, worker nodes, and cluster components

	Name	Status	Worker Pool	Zone	Private IP	Public IP	Version	
▼	0000025d	Normal	default	wdc07	10.191.61.69	169.62.47.221	3.11.153_1529	
▼	00000136	Normal	default	wdc07	10.191.61.67	169.62.40.202	3.	Update available

Items per page: 10 ▾ 1–2 of 2 items 1 ▾ 1 of 1 pages ◀ ▶

Updating clusters, worker nodes, and cluster components

Updating worker nodes (masters already updated)

You received a notification to update your worker nodes. What does that mean? As security updates and patches are put in place for the API server and other master components, you must be sure that the worker nodes remain in sync.

➤ **What happens to my apps during an update?**

If you run apps as part of a deployment on worker nodes that you update, the apps are rescheduled onto other worker nodes in the cluster. These worker nodes might be in a different worker pool, or if you have stand-alone worker nodes, apps might be scheduled onto stand-alone worker nodes. **To avoid downtime for your app, you must ensure that you have enough capacity in the cluster to carry the workload.**

➤ **How can I control how many worker nodes go down at a time during the update?**

If you need all your worker nodes to be up and running, consider resizing your worker pool or adding stand-alone worker nodes to add more worker nodes. You can remove the additional worker nodes after the update is completed. In addition, you can create a Kubernetes config map that specifies the maximum number of worker nodes that can be unavailable at a time during the update. Worker nodes are identified by the worker node labels. You can use IBM-provided labels or custom labels that you added to the worker node.

➤ **What if I choose not to define a config map?**

When the config map is not defined, the default is used. By default, a maximum of 20% of all of your worker nodes in each cluster can be unavailable during the update process.

Updating clusters, worker nodes, and cluster components

Updating the masters (IBM does this for patches, users control for major.minor)

➤ How do I know when to update the master?

You are notified in the IBM Cloud console and CLI when updates are available and can also check the [supported versions](#) page.

➤ Can my worker nodes run a later version than the master?

Your worker nodes cannot run a later major.minor Kubernetes version than the master. First, [update your master](#) to the latest Kubernetes version. Then, [update the worker nodes](#) in your cluster. Worker nodes can run later patch versions than the master, such as patch versions that are specific to worker nodes for security updates.

➤ What happens during the master update?

Your master is **highly available** with three replica master pods. The master pods have a rolling update, during which only one pod is unavailable at a time. Two instances are up and running so that you can access and change the cluster during the update. Your worker nodes, apps, and resources continue to run.

➤ Can I roll back the update?

No, you cannot roll back a cluster to a previous version after the update process takes place. Be sure to use a test cluster and follow the instructions to address potential issues before you update your production master.

Running updates from the CLI/Web UI (sufficient worker capacity)

- Schedule the rolling update of workers (default behavior will have max 20% workers offline)

```
ibmcloud oc worker update --cluster <cluster_name_or_ID> --worker <worker_node1_ID> --worker <worker_node2_ID> ...
```

- Confirm that the update is complete

```
oc get nodes
```

- Web UI – navigate to the Worker Nodes, select desired nodes to update and click on "Update"

	Name ↑	Status	Worker Pool	Zone	Private IP	Public IP	Version
<input checked="" type="checkbox"/>	0000025d	Normal	default	wdc07	10.191.61.69	169.62.47.221	3.11.153_1529 ⓘ
<input type="checkbox"/>	00000136	Normal	default	wdc07	10.191.61.67	169.62.40.202	3.11.153_1529 ⓘ

Temporarily adding capacity through a worker pool

- **Create a worker pool with the number of worker nodes that you want to replace**

```
ibmcloud oc worker-pool create classic --name <pool_name> --cluster <cluster_name_or_ID> --flavor  
<flavor> --size-per-zone <number_of_workers_per_zone>
```

```
ibmcloud oc worker-pool ls --cluster <cluster_name_or_ID> # Verify that the worker pool is created.
```

```
ibmcloud oc zone add classic --zone <zone> --cluster <cluster_name_or_ID> --worker-pool <pool_name> --  
private-vlan <private_VLAN_ID> --public-vlan <public_VLAN_ID>
```

#Add the zone to your worker pool that you retrieved earlier. When you add a zone, the worker nodes that are defined in your worker pool are provisioned in the zone and considered for future workload scheduling. If you want to spread your worker nodes across multiple zones, choose a multizone-capable zone.

- **Cleaning up a worker pool:**

```
ibmcloud oc worker-pool rm --worker-pool <pool_name> --cluster <cluster_name_or_ID>
```

#Remove the worker pool with the old machine type. Removing a worker pool removes all worker nodes in the pool in all zones. This process might take a few minutes to complete.

```
ibmcloud oc worker-pool ls --cluster <cluster_name_or_ID>
```

#Verify that the worker pool is removed.

Useful Resources/Bookmarks

- IBM Knowledge Center Cloud Pak for Data -
https://www.ibm.com/support/knowledgecenter/SSQNUZ_2.5.0/cpd/overview/welcome.html
- IBM Cloud Pak for Experiences –
<https://www.ibm.com/cloud/paks/experiences/cloud-pak-for-data>