Unlocking Db2 for z/OS data with IBM Data Gate for watsonx

Martin Schneider
Principal Product Manager, Data and AI on IBM Z
IBM Germany Research and Development

Keziah Knopp
Technical Enablement Specialist, Infrastructure
IBM USA

Agenda

Quick intro

Making Z data available for applications

What is a Lakehouse?

How Data Gate delivers Z data

How to make Z data available to downstream applications that are…

Hybrid Cloud

Public cloud
Virtual private cloud
Private cloud

Multi-Cloud

AWS
Azure
IBM Cloud
…

Distributed

Distributed platforms
Data from many sources

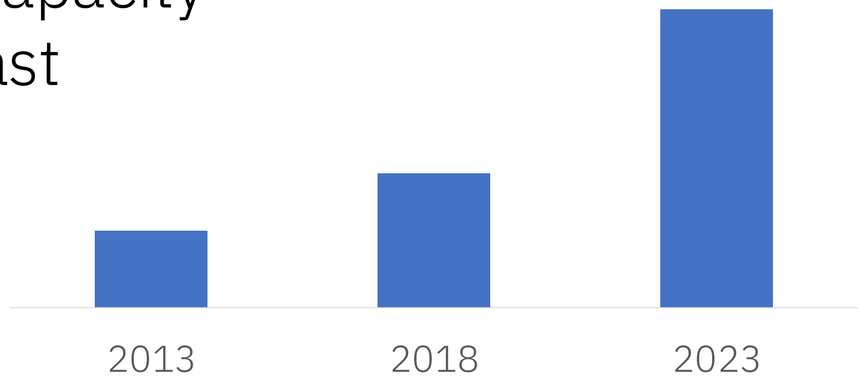
→ e.g., **data lakehouses**

Mainframe data - essential to any data strategy

- IBM mainframes run ~70% of all world transactions by value[1]
- 77 of the world's top 100 banks use an IBM mainframe[2]

~3x growth

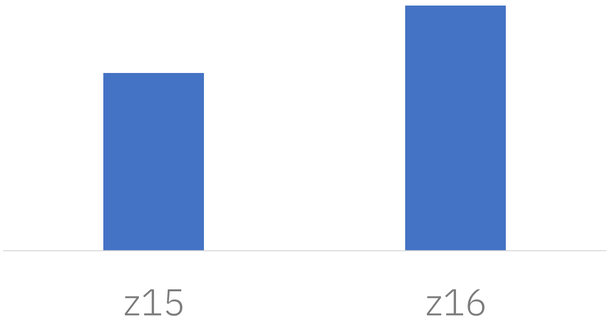
installed capacity over the last decade



Workload as measured by installed Million Instructions Per Second (MIPS)

Record setting IBM z16

shipped capacity fueled by hybrid cloud, Linux® and digital transformation



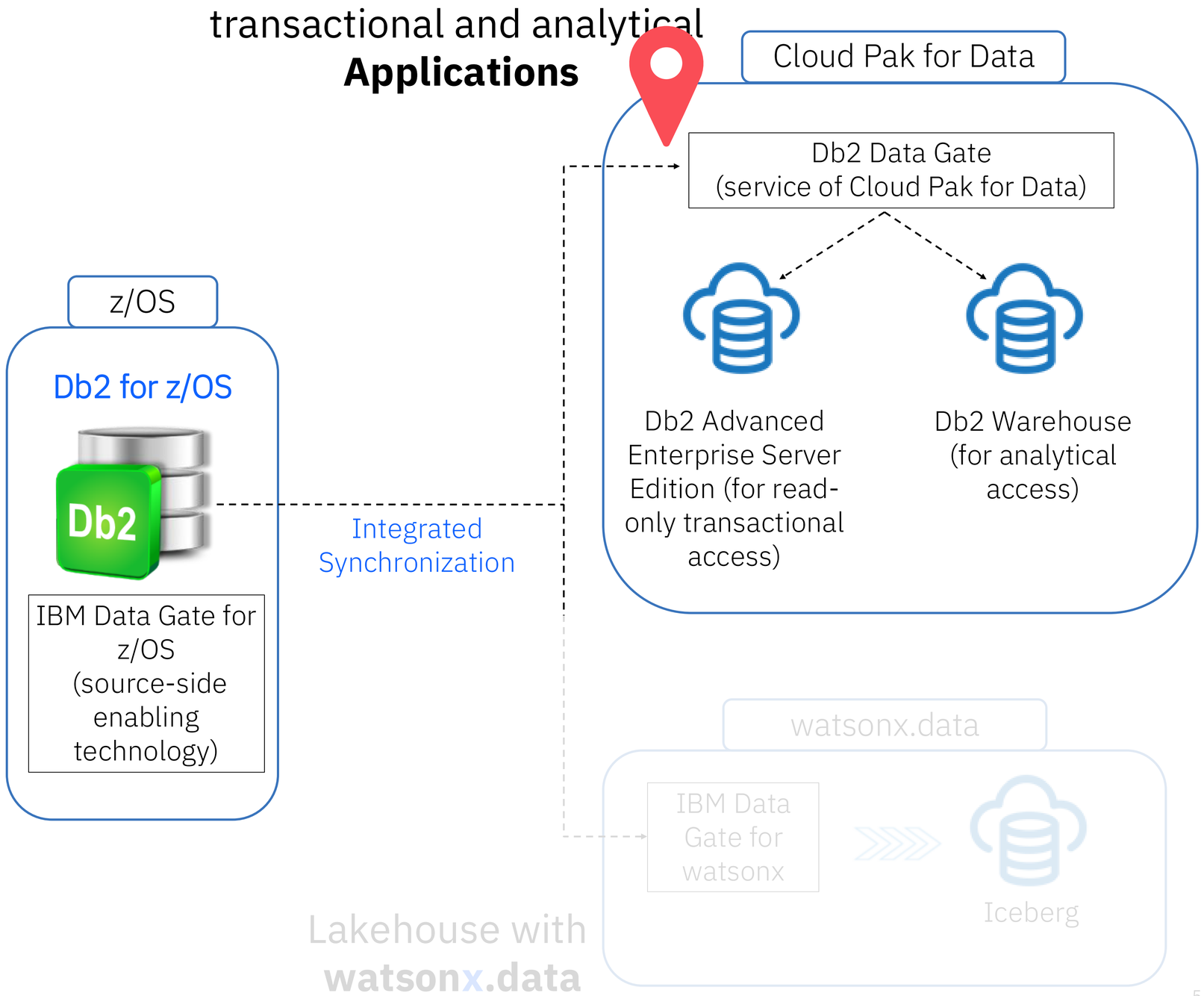
Shipped MIPS first 7 Qtrs of program

Source: IBM Financial data and IBM Z and IBM® LinuxONE shipments Feb 2024

IBM Data Gate Family

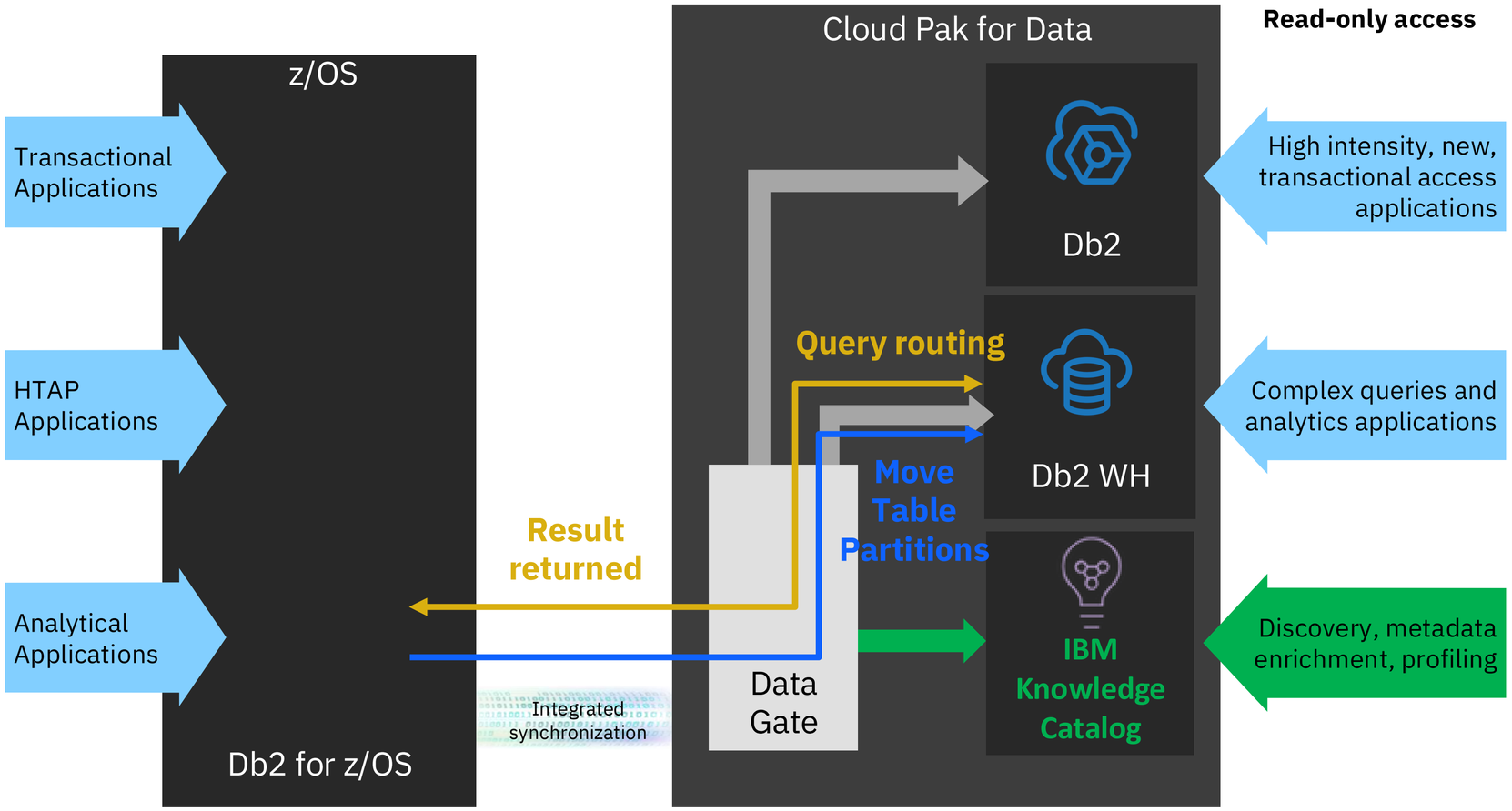
Data Gate has two different targets:

1. Db2 (for hybrid cloud applications)
2. watsonx.data (for Lakehouses)



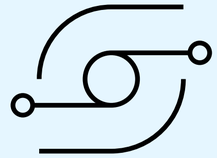
Db2 Data Gate – Capabilities – Overview with Db2 (distributed) as a target

- Make Db2 for z/OS data easily consumable by high intensity, new read-only transactional applications
- Provide Db2 for z/OS data/meta data with a path on journey to cloud
- Improve operational efficiencies
- Make Db2 for z/OS metadata available on IBM Knowledge Catalog
- Accelerate analytical queries from Db2 for z/OS
- Archive Db2 for z/OS data in the cloud



The secret sauce for Db2 for z/OS data within Data Gate - Integrated Synchronization

A game changing technology built into Data Gate and Db2



Custom-built synchronization mechanism

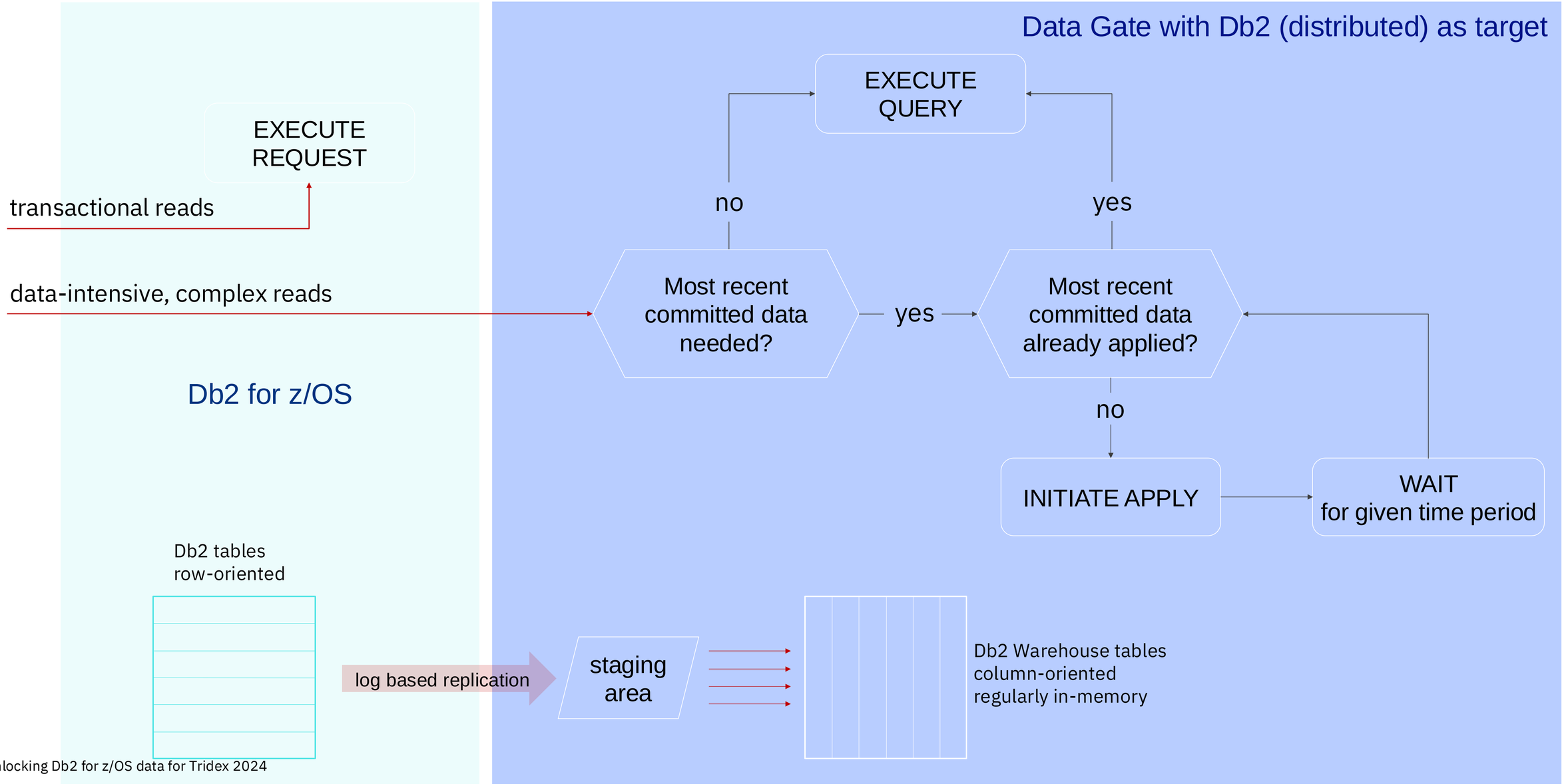
- **Simple.** Fully integrated into Db2 for z/OS. Nothing to install on the mainframe. Easier to maintain.
- **Isolated.** No impact to transaction workloads on IBM Z
- **Ultra-efficient.** Factors of improved overhead, throughput and latency compared to any other technology
    - Uses ½ the z/OS CPU of traditional replication
    - The remaining CPU is +96% zIIP eligible.

Use of a patent-pending algorithm that has in unit testing demonstrated low latency and significantly increased write-throughput to the Iceberg open-source table format

Real world customer experiences:

- Averaged a latency of 1.3 sec at a large Western European banking customer
- Measured a latency of 3 sec with Data Gate vs. 20 min with competitor replication product at large AP banking customer PoC
- Observed a peak throughput of 1.9 M updates/sec (11 TB/hour) at a European banking customer

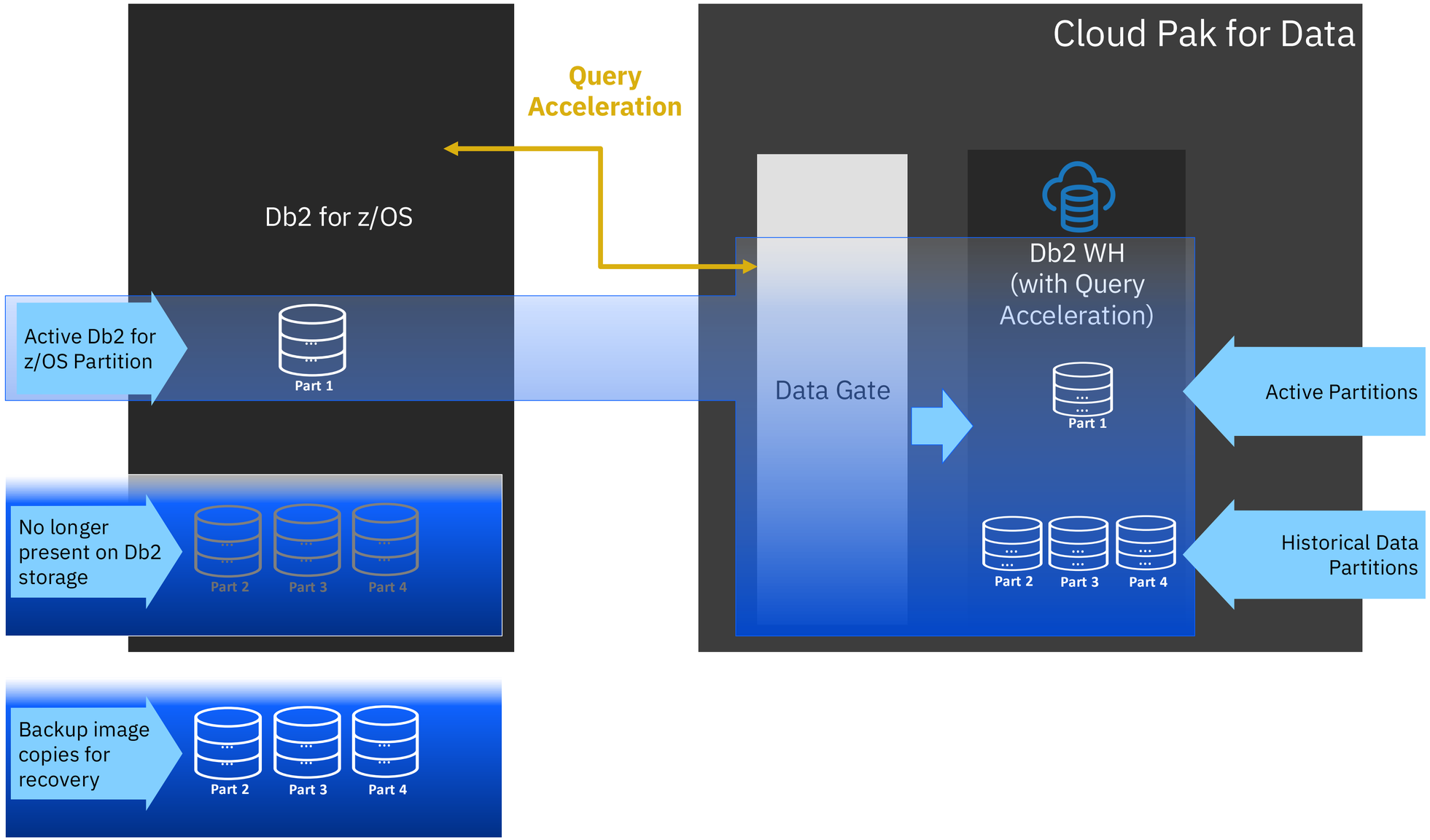
HTAP: Making sure that the data is current before the query runs



Storage Saver

Reduce disk storage costs for Db2 for z/OS by archiving data in Data Gate with excellent query performance.

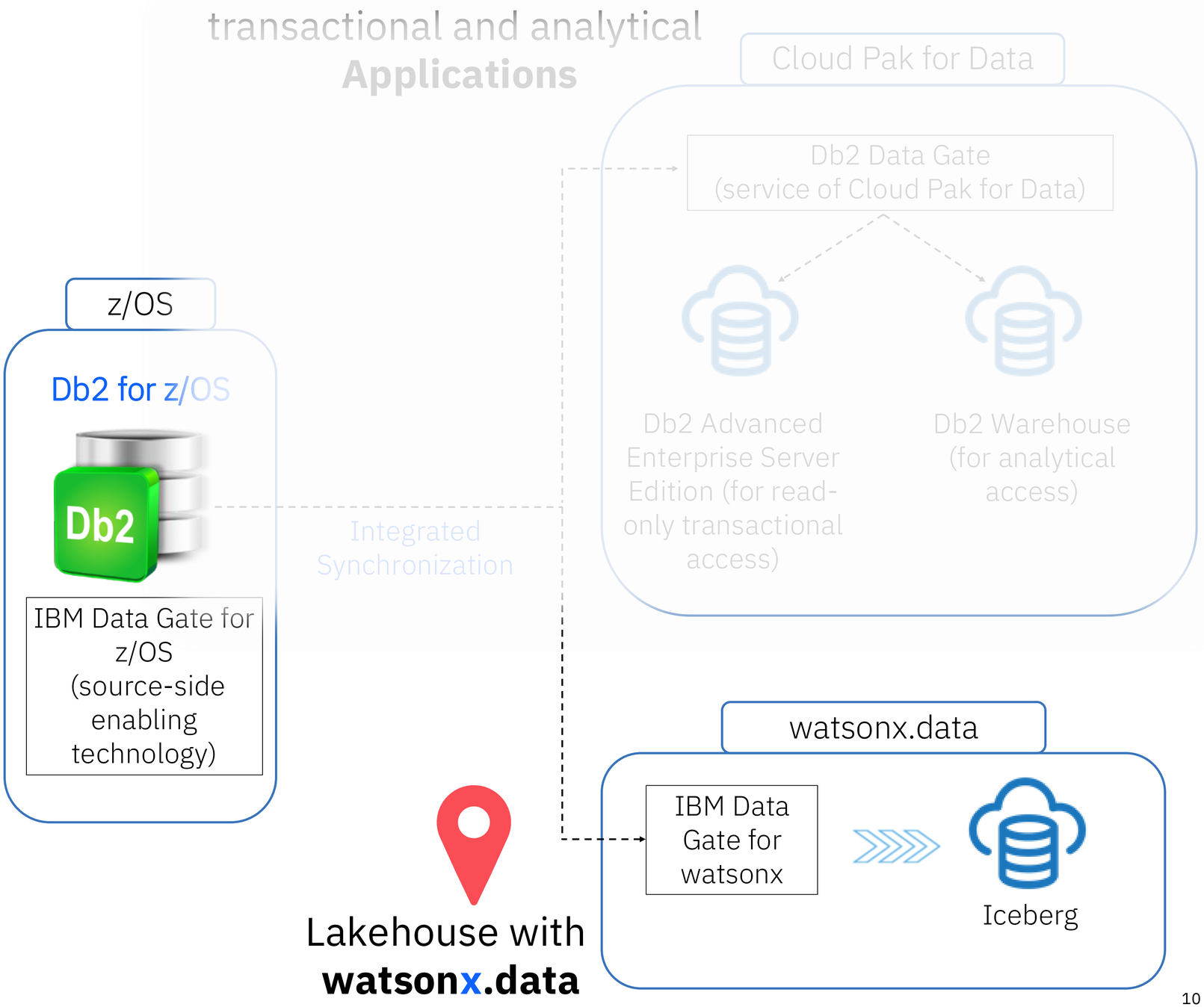
- Db2 Data Gate can manage archived Db2 for z/OS tables on Db2 (distributed)
- Storage saver enables to only have the recent partitions used and stored in Db2 for z/OS but have the entire history available for queries, e.g., for regulatory/audit purposes.

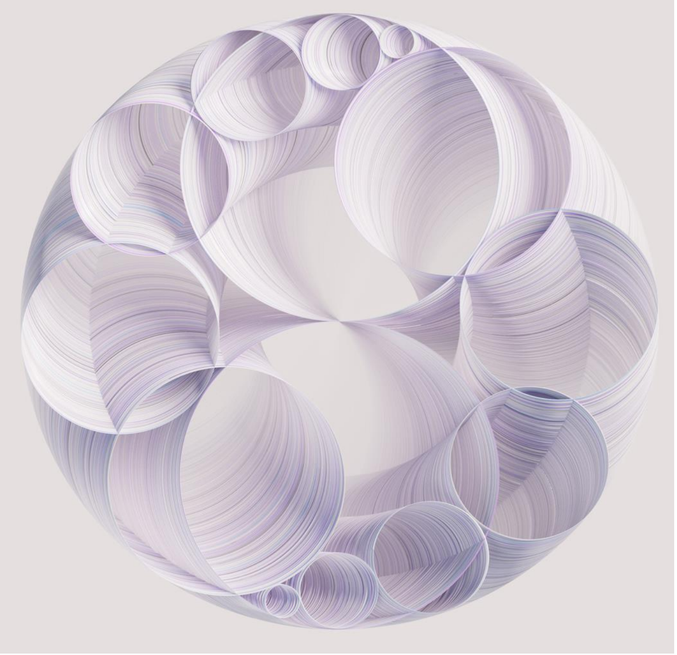


IBM Data Gate Family

Data Gate has two different targets:

1. Db2 (for hybrid cloud applications)
2. watsonx.data (for Lakehouses)





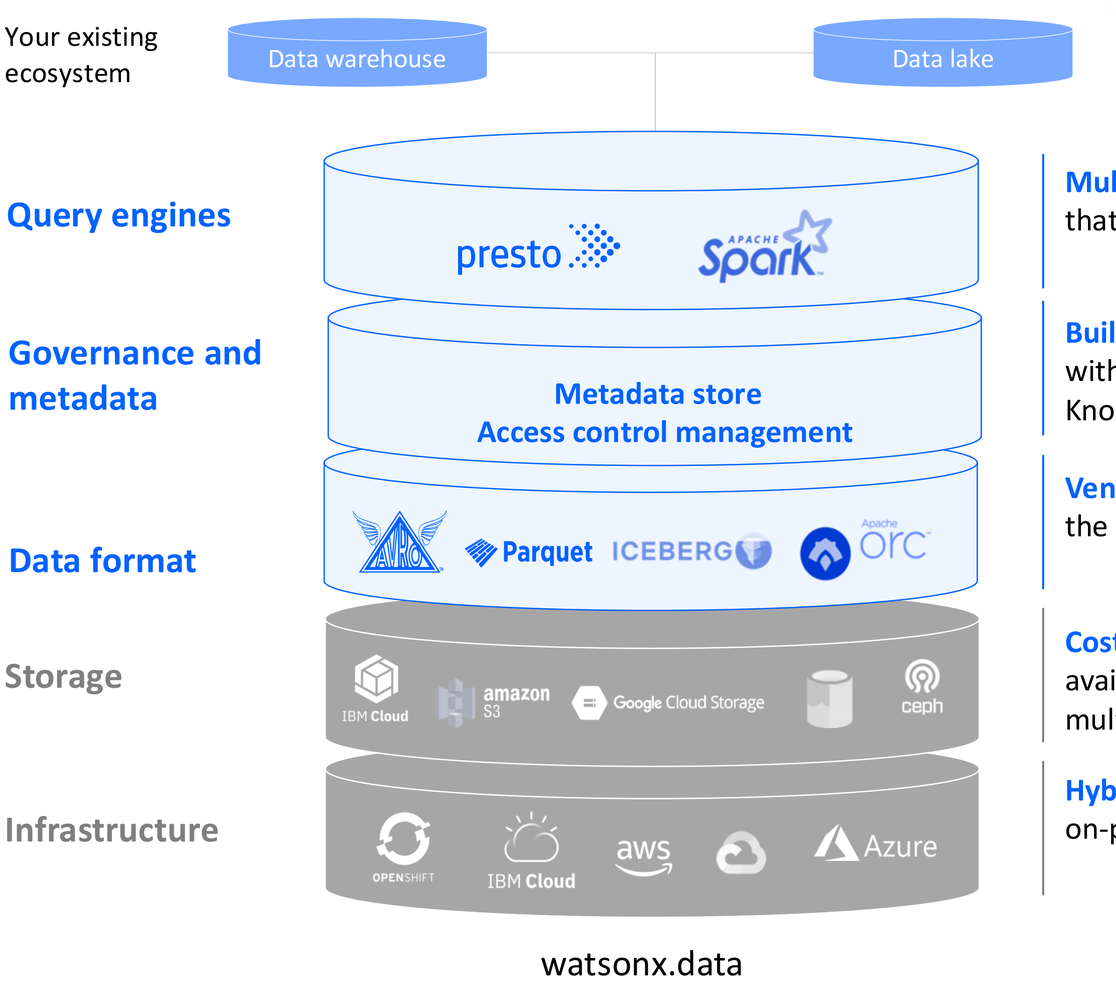
A data lakehouse combines the high-performance characteristics of a data warehouse with the cost-efficiency, flexibility, and scalability of a data lake

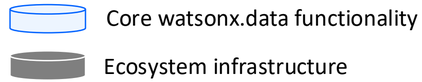
---

watsonx.data

11

Key components of IBM watsonx.data





Multiple engines such as Presto and Spark that provide **fast, reliable, and efficient processing of big data** at scale

Built-in governance that is compatible with existing solutions such as IBM Knowledge Catalog

Vendor agnostic open formats for analytic data sets, allowing different engines to access and share the same data, at the same time

Cost-effective, simple, object storage available across hybrid cloud and multicloud environments

Hybrid cloud deployments and workload portability across hyperscalers and on-premises with Red Hat OpenShift

watsonx.data

Scale AI workloads, for all your data, anywhere

A fit-for-purpose data store, based on an open lakehouse architecture, supported by querying, governance, and open data formats to access and share data.

Available as a managed service on IBM Cloud and AWS, and software-as-a-container

Access all your data through a single point of entry across all clouds and on-premises environments.

Get started in minutes with built-in governance, security, and automation.

Reduce the cost of a data warehouse by up to 50%* through workload optimization across multiple query engines and storage tiers.

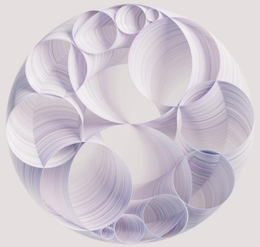
*When comparing published 2023 list prices normalized for VPC hours of IBM watsonx.data to several major cloud data warehouse vendors. Savings may vary depending on configurations, workloads and vendors.

IBM Data Gate for watsonx

First-class integration of IBM Z data into watsonx.data

Modernize access to mainframe data for analytics and AI by making IBM Z data readily available to watsonx.data

Provides easy and efficient access to data for the lakehouse, with lower overhead, latency and higher throughput compared to most other replication options



- Seamlessly synchronize Db2 for z/OS, IMS DB, and VSAM to Iceberg in watsonx.data
- Use the most up-to-data transactional data for analytics and AI
- Reduce the time it takes to access, analyze and score IBM Z data

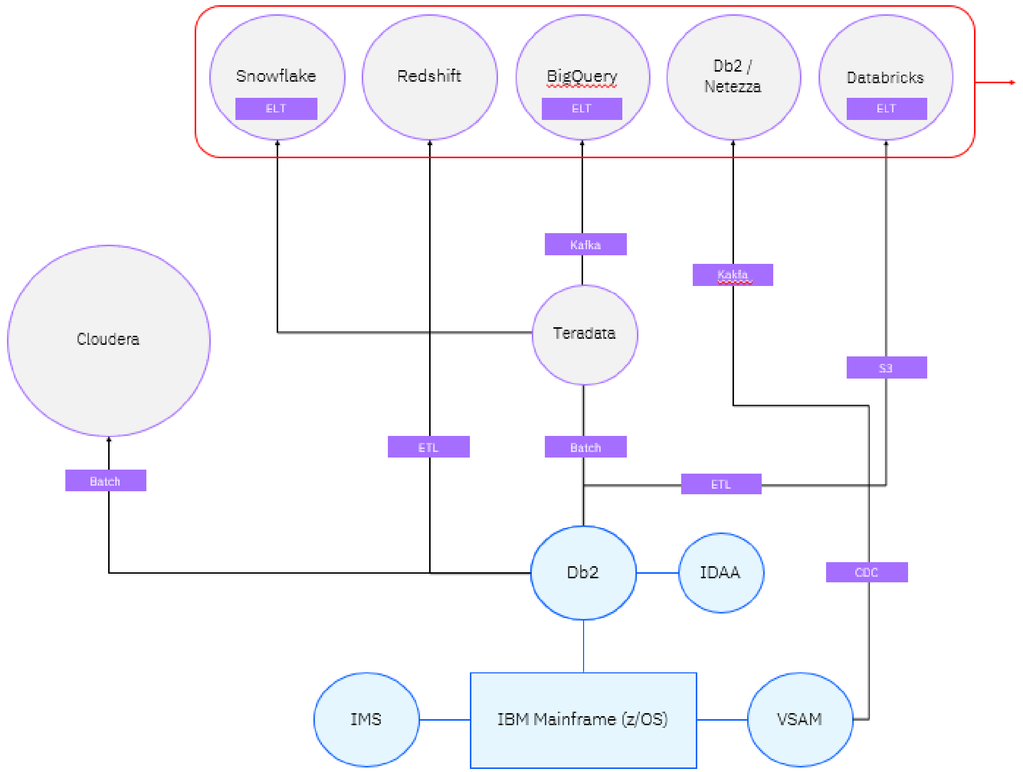
Yesterday - complex integrations with IBM Z data lead to expensive solutions

Cloud data warehouses / lake solutions become extremely expensive if treated like on-prem appliances

Previously had fixed cost models and now charged based on consumption

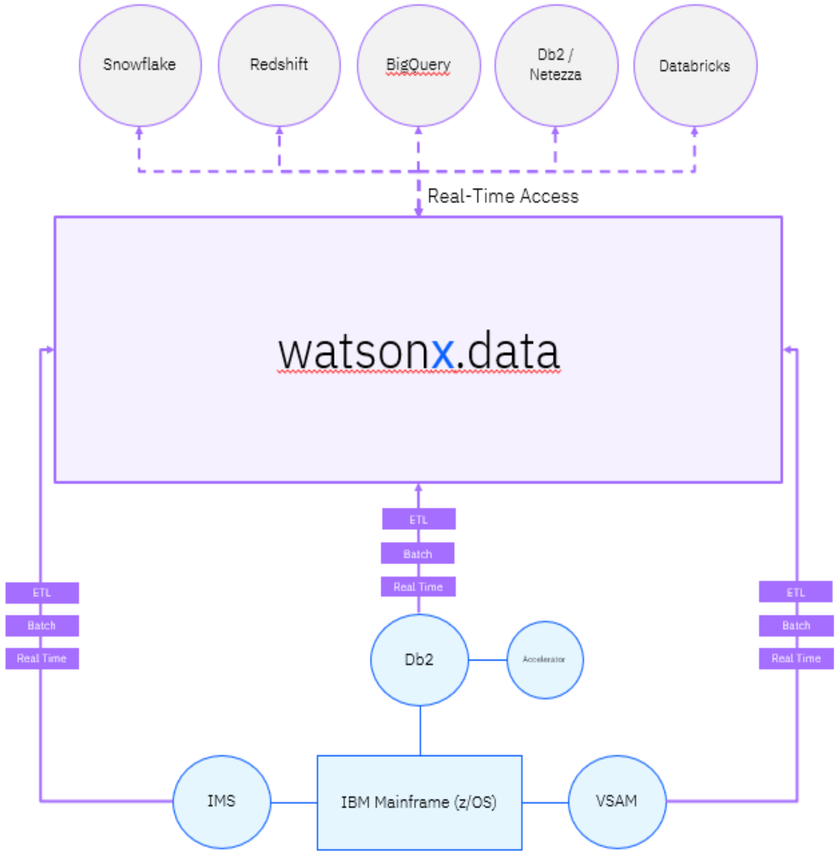
Cloud warehouses are not optimized for “write” / “load” operations which drive a huge amount of wasted consumption

~70% of all warehouse processing time is “write” oriented meaning clients spend most of their cloud warehousing spend on loading data vs. querying data for business applications



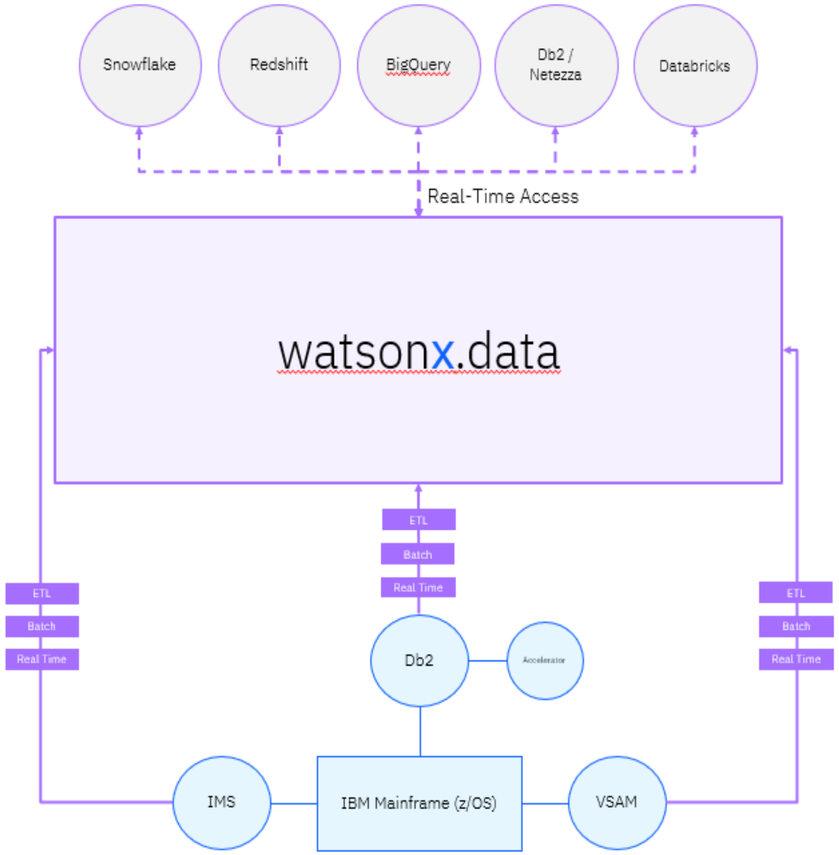
watsonx.data

Builds a lakehouse to optimize access to IBM Z workloads and integrates with analytics & AI database systems



watsonx.data

Builds a lakehouse to optimize access to IBM Z workloads and integrates with analytics & AI database systems



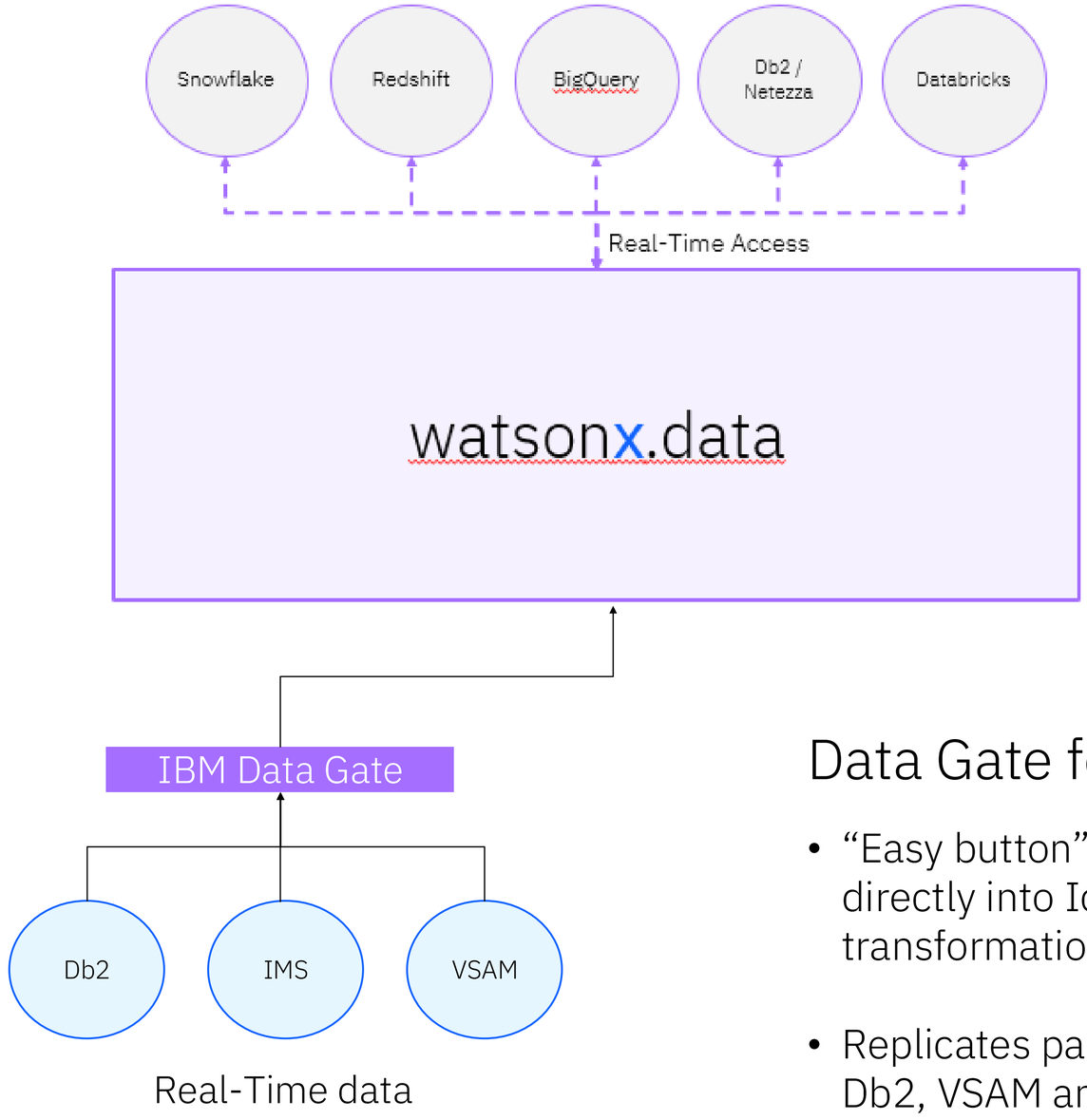
Technical Value

- Store ONE copy of data
- Write data into an open format (Iceberg)
- Multiple analytics warehouses can read from Iceberg using the object storage of your choice
- Built on open-source (Spark, Presto, Iceberg, Hive)
- Access enable distributed access to VSAM, IMS and Db2 data with a single solution
- Realtime access

Financial Value

- Reduce Cloud Data Warehousing costs by +50%
- Eliminate legacy lake, warehouse and replication solutions costs
- Reduce MIP utilization attributed to data replication

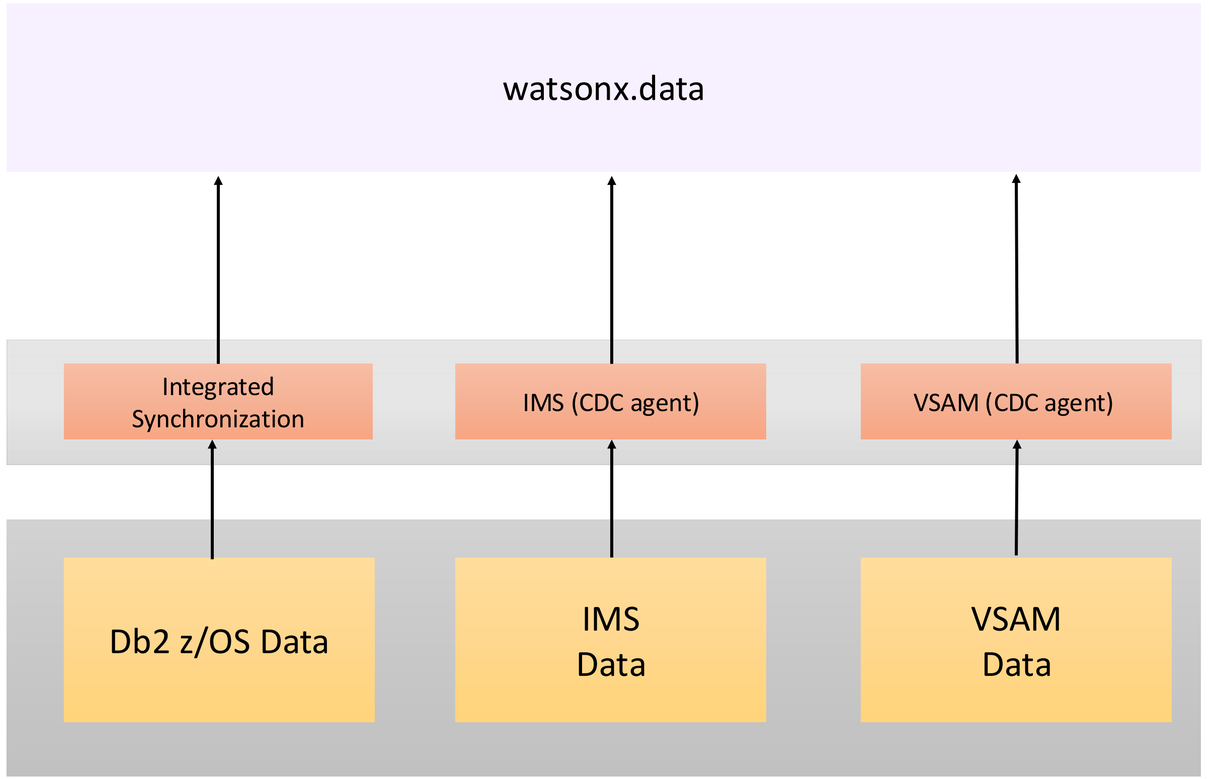
How do we get data from IBM Z to watsonx.data



Data Gate for watsonx

- “Easy button” to replicate Db2, VSAM, IMS data directly into Iceberg tables (without additional transformations)
- Replicates partial or full tables, transactions, sets of Db2, VSAM and IMS data
- Minimal transformation of data beyond format changes from proprietary to open table

Data Gate for watsonx architecture

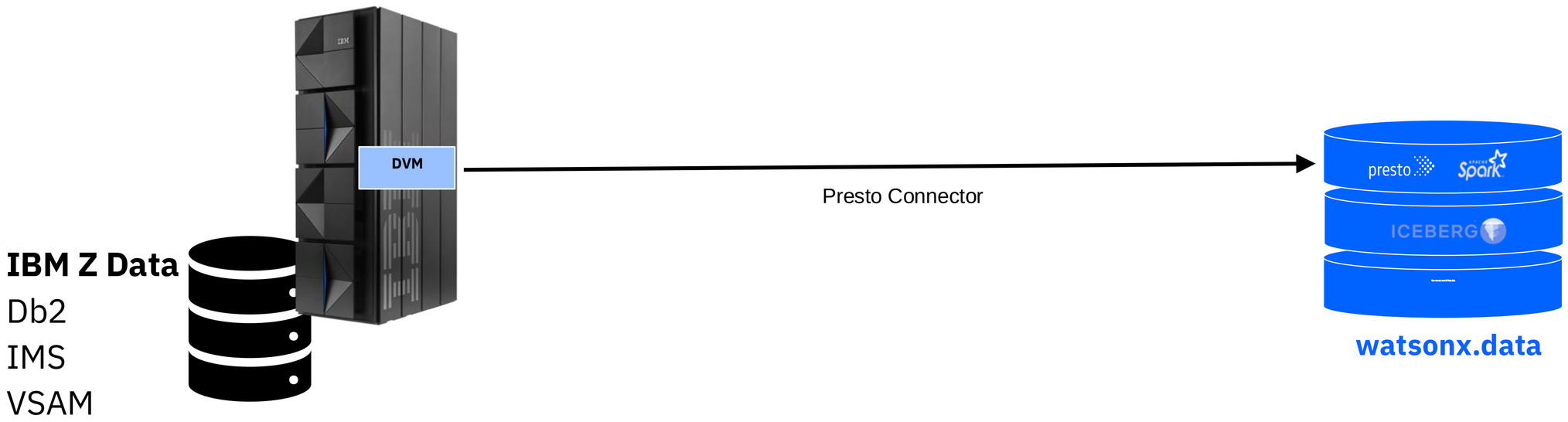


IBM Lakehouse

Data Gate running on OpenShift Container Platform (OCP)

Client data source on IBM Z

Data Virtualization (DVM) can also be used to access IBM Z data from watsonx.data



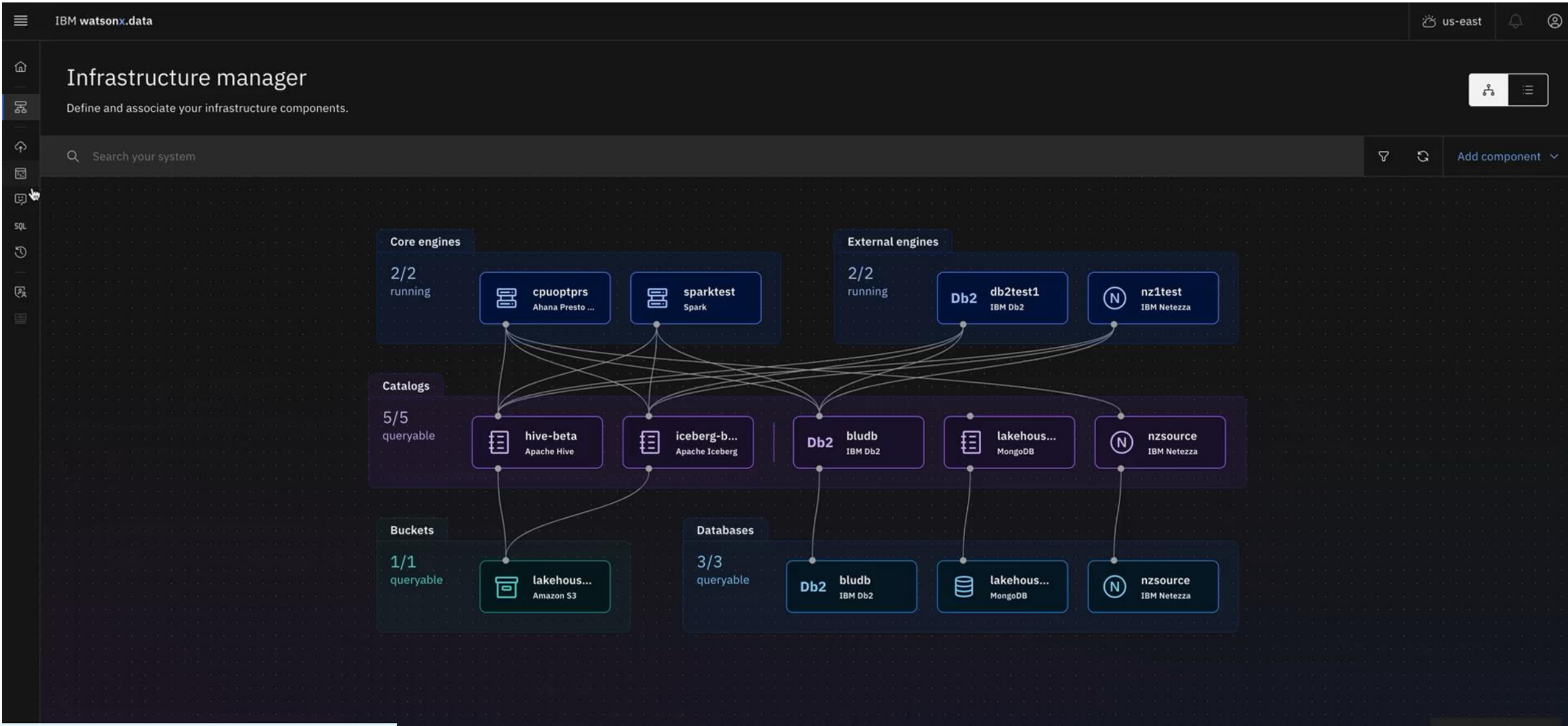
Data Virtualization Manager (DVM)

Virtualize (access) to data in place on IBM Z

- Can access/update live Db2, VSAM, IMS, etc.
- Very efficient virtualization on the mainframe
- SQL-view on mainframe data through the DVM Presto Connector

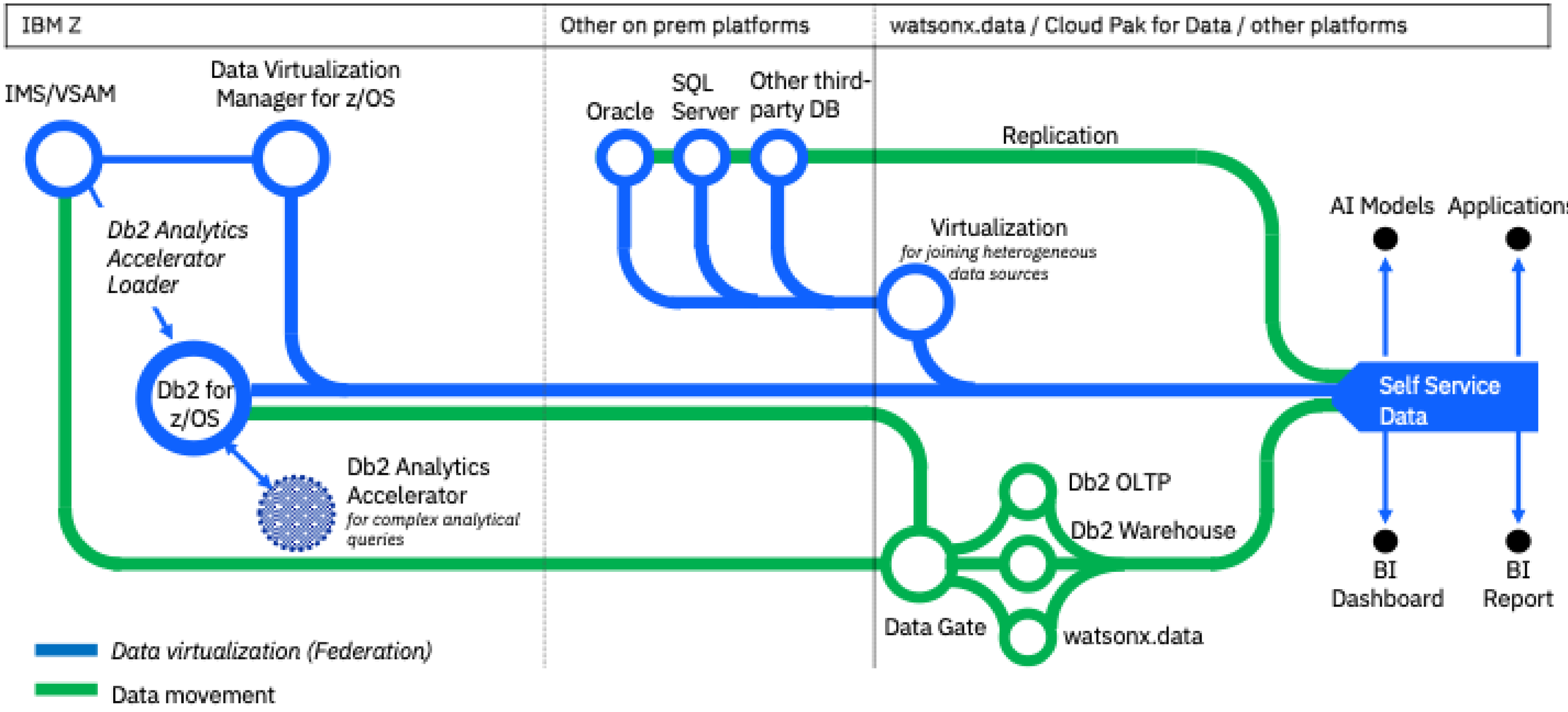
Two main techniques to access data: Push or Pull the data. The use case will drive the pattern

Data Virtualization Manager	Data Gate for watsonx
Leaves data in place	Creates a physical copy of the data
SQL-view on all sorts of data, structured and unstructured	Data Gate: 1:1 copy only
Direct real time data access	Low latency
Need PrestoDB connector	Direct integration with watsonx.data



watsonx.data

Integrate zSystems data with a data fabric

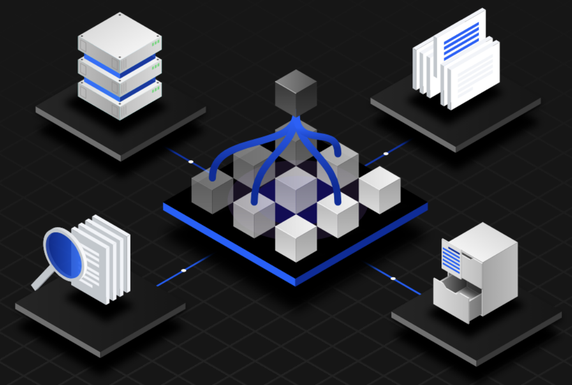


Q&A

IBM **Cloud Pak for Data**

Welcome, kubeadmin.

You've been logged in for a few seconds.



Architect your lakehouse

Define and associate infrastructure components to make your data queryable for you and other users.

Infrastructure manager →

Work with your data

Build and run queries against and across your data, monitor their progress, and save them for reuse.

Query workspace →

Welcome to IBM watsonx.data.

Browse the recommended resources below to get up to speed quickly, catch up on what's new, and discover what you can achieve through integrations with watsonx.data.

Recommended resources

- Release notes
- Ingestion CLI docs

View full documentation

Infrastructure components 9

Engines	Catalogs
1	4
Buckets	**Databases**
4	0

Go to infrastructure manager

Recent tables 0

No recent tables.

Use the data manager to explore and curate tables across catalogs.

View more data objects

Recent ingestion jobs 0

No recent ingestion jobs.

Create an ingestion job to move external data into watsonx.data.

View more ingestion jobs

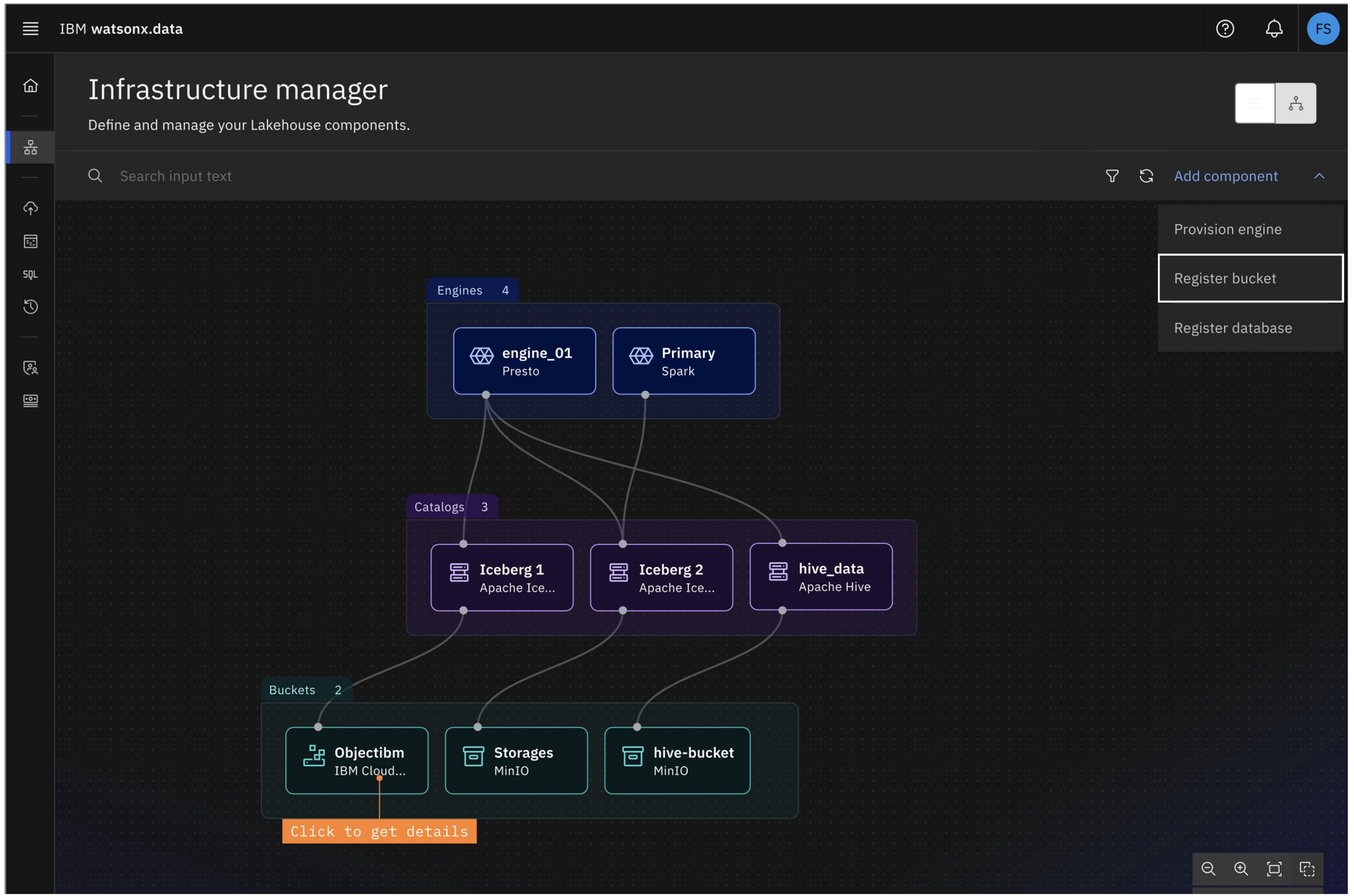
Saved worksheets 0

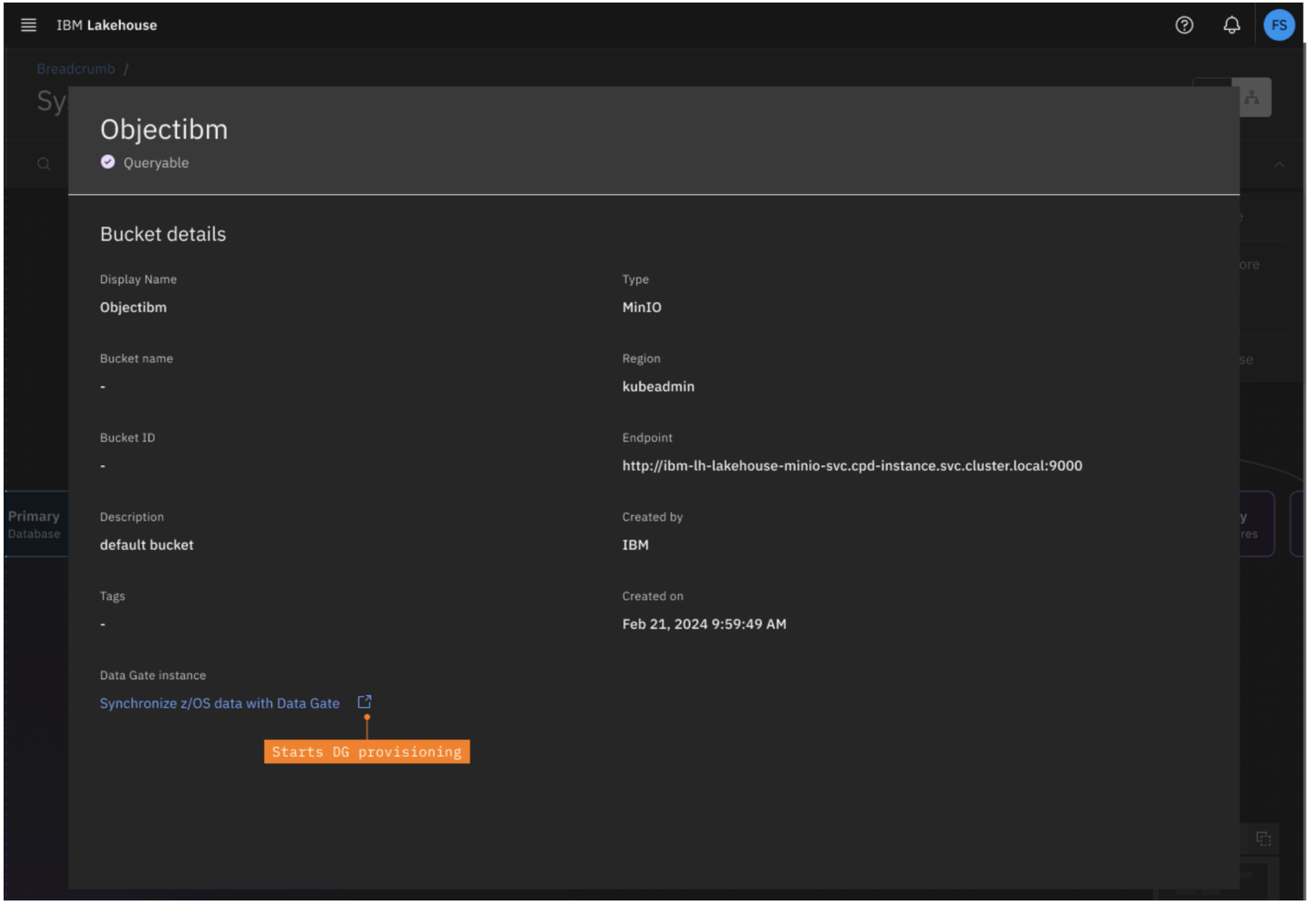
Recent queries 10

Mar 12, 2024 10:46:42 AM

```
select * from system.runtime.queries order by query_id desc
```

Mar 12, 2024 10:39:05 AM


IBM watsonx.data
FS
Infrastructure manager
Define and manage your Lakehouse components.
Search input text
Add component
Provision engine
Register bucket
Register database
Engines 4
engine_01
Presto
Primary
Spark
Catalogs 3
Iceberg 1
Apache Ice...
Iceberg 2
Apache Ice...
hive_data
Apache Hive
Buckets 2
Objectibm
IBM Cloud...
Storages
MinIO
hive-bucket
MinIO
Click to get details


IBM Lakehouse
FS
Breadcrumb /
Objectibm
Queryable
Bucket details
Display Name
Objectibm
Type
MinIO
Bucket name
-
Region
kubeadmin
Bucket ID
-
Endpoint
http://ibm-lh-lakehouse-minio-svc.cpd-instance.svc.cluster.local:9000
Description
default bucket
Created by
IBM
Tags
-
Created on
Feb 21, 2024 9:59:49 AM
Data Gate instance
Synchronize z/OS data with Data Gate
Starts DG provisioning
Primary
Database

IBM Cloud Pak for Data

localhost:3004/#/createDatagate/data-gate/5.0.0

Services catalog / Db2 Data Gate / Create data gate instance

Create data gate instance

Source data

Select source data type and instance. You will be able to provide source connection information after the Data Gate instance is created.

- ◯ Db2 for z/OS
- ◯ IMS
- ◯ VSAM

Target data

Select target data type and instance. Create one if none exists on the cluster. Your Cloud Pak for Data userid is used to access the target data, which requires the admin role.

- ◯ Transactional (Db2)
- ◯ Analytics (Db2 Warehouse)
- ◯ Iceberg

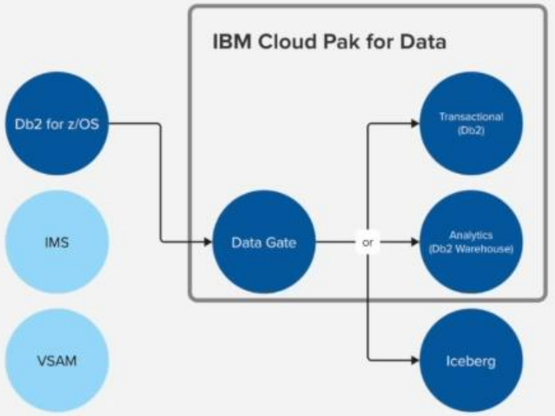
Compute resources

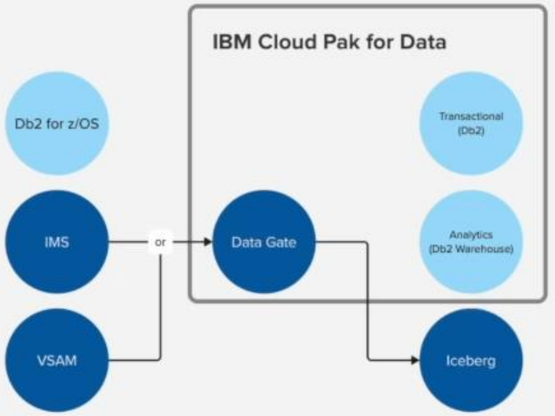
CPU: 6 (2 – 32)

Memory: 16 Gi (13 Gi – 32 Gi)

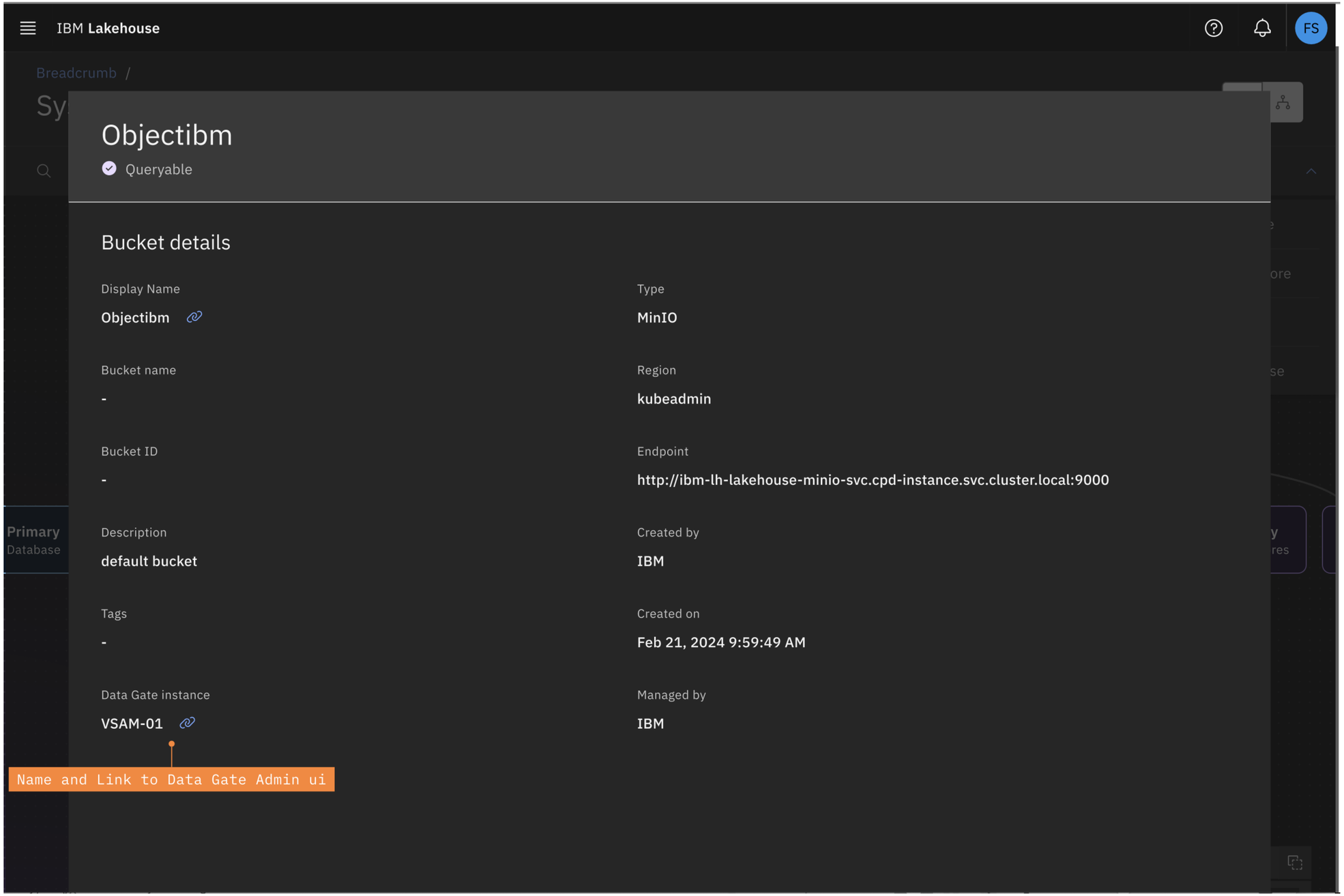
Namespace

zen





Cancel Review


IBM Lakehouse
FS
Breadcrumb /
Objectibm
Queryable
Bucket details
Display Name
Objectibm
Type
MinIO
Bucket name
-
Region
kubeadmin
Bucket ID
-
Endpoint
http://ibm-lh-lakehouse-minio-svc.cpd-instance.svc.cluster.local:9000
Description
default bucket
Created by
IBM
Tags
-
Created on
Feb 21, 2024 9:59:49 AM
Data Gate instance
VSAM-01
Managed by
IBM
Name and Link to Data Gate Admin ui

DC1A11

Overview | Monitor | Table | Query

DC1A11

Target: Db2_wh_svt

Source: STEC2@idaadb2z.mycompany.com:691

Status

- Data gate server active
- Synchronization enabled — On
- Query rounting disabled — Off

Data Catalog Integration

Status: Available

Last Asset Synchronization: N/A

Synchronize now

Table

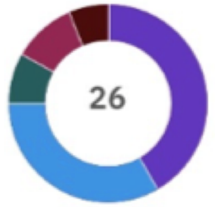


Table: 26 total

Active	xxx	Pending	xxx
Loaded	xxx	Error	xxx
Loading	xxx		

Query

Query: 35 total

Successful	xxx	Queued	xxx
Executing	xxx	Failed	xxx

How to position Data Gate for watsonx and CDC - with watsonx.data

	Data Gate for watsonx	Change Data Capture
One-liner	Simple and efficient synchronization of 3 major IBM Z sources with watsonx.data	General purpose, low impact, real time data delivery from many sources and many targets.
Sources	Db2 z/OS, IMS, VSAM only	Many
Targets	Iceberg in watsonx.data	Many, not Iceberg
Superpower	Simplicity (purchase, setup, operation) performance and efficiency	Performance, versatility and capability
Additional differentiatiors	Transactionally consistent replication	Transactionally consistent data delivery, bidirectional replication, simple transformations, maturity

Replication and Syncronization with Data Gate differences

	Data Gate	Change Data Capture
One-liner	Simple and efficient synchronization of 3 major IBM Z sources with watsonx.data and Db2 in Cloud Pak for Data	General purpose, low impact, real time data delivery from many sources and many targets.
Sources	Db2 z/OS, IMS, VSAM only*	Many
Targets	Iceberg in watsonx.data Db2 LUW / WH in Cloud Pak for Data	Many, not Iceberg
Superpower	Simplicity (purchase, setup, operation) performance and efficiency	Performance, versatility and capability
Additional differentiatiors	Transactionally consistent replication	Bidirectional replication, simple transformations

*) Future enhancements – see disclaimer

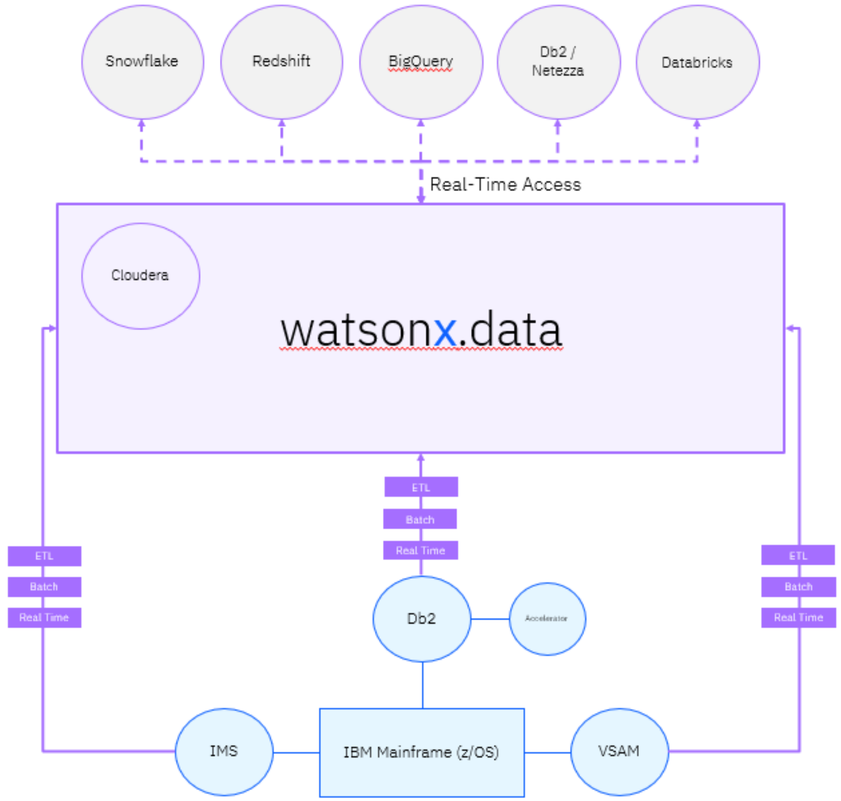
What if the client wants IBM Z data in Iceberg format only?

Once data is moved from IBM Z to iceberg it must be managed

We want that management of the data to be via watsonx.data with our Metadata store that has a catalog of the s3 compatible data

The customer can technically move the S3 data outside of watsonx.data but then they are in a DIY (Do It On Your Own) scenario and would not be supported by IBM

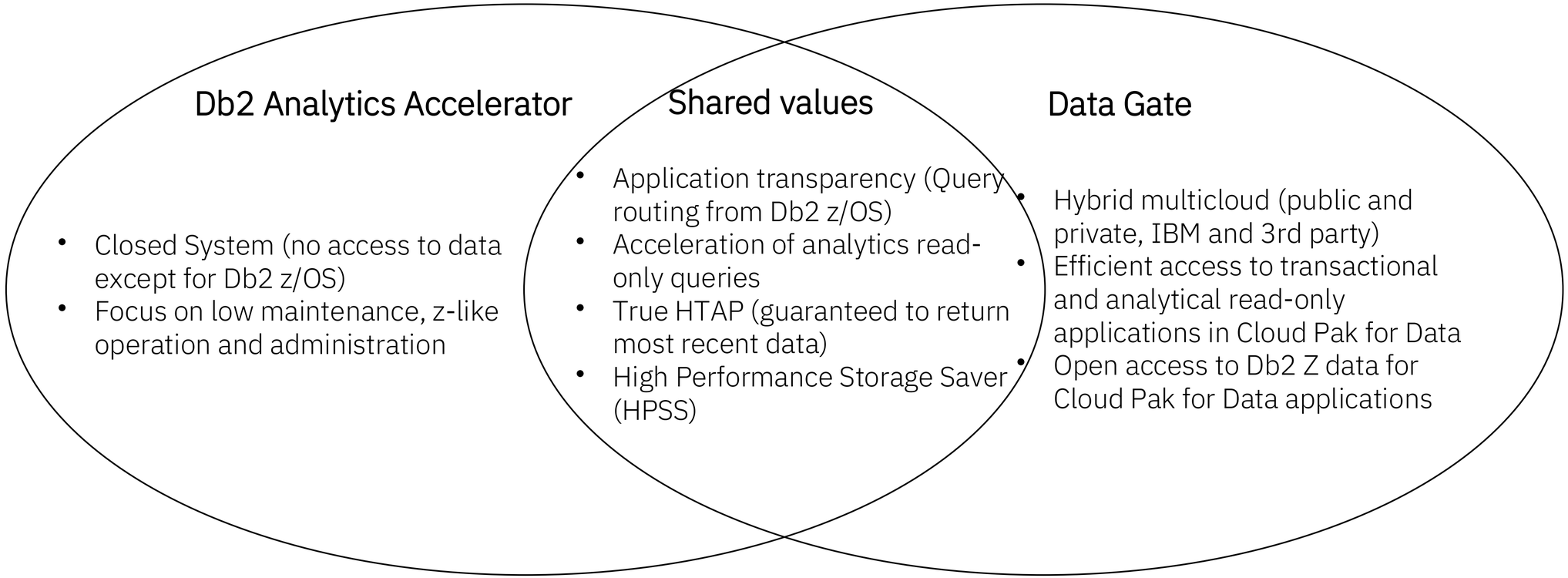
We want the data to stay with IBM



Two main techniques to access data: Push or Pull the data.
The use case will drive the pattern

Data Virtualization Manager	Data Gate
Leaves data in place	Creates a physical copy of the data
Direct real time data access	Low latency
A metadata catalog to keep track of data location and availability	Feeds the metadata store of the platform
Support for transactions that write back to the original data sources (except for sequential files)	Unidirectional – no writeback

Key values and differentiation



Data Gate for watsonx pre-requisites

watsonx.data productive install

- Red Hat OpenShift Container Platform (entitlement contained in wx.data license)
- Minimum HW requirements:
  - Master and infrastructure servers: 3 master + 3 infrastructure servers on the same nodes (8 vCPU 32 GB per node)
  - Worker/compute nodes: 3+ worker / compute nodes (16 vCPU per node, 64-128 GB per node)
    - They can co-host the watsonx.data instances below, in that case increase to 20 vCPUs 92 GB RAM per node)
  - Load balancer: 2 nodes (4 vCPU per node)
- Detailed prereqs here

IBM Z

- Db2 z/OS V12 or higher, z/OS 2.4 or higher
- Distributed data facility (DDF) with a secure port, configured for network encryption through AT-TLS
- Detailed prereqs here

Data Gate for watsonx

- Minimum HW requirements per S-sized instance: 4 vCPU, 28 GB RAM
- A typical Data Gate for watsonx.data will require multiple instances (one for every source system, more for high workloads, so multiply this with the # of instances required)
- Presized configurations: Small = 16 vCPUs, Medium = 42 vCPUs, Large = 96 vCPUs

Note: Always refer to the system requirements documentation of the respective product for detailed planning!

Mainframe data - essential to any data lakehouse strategy

- Transactional data is essential to understand the current state of the business
- Transactional data is the most valuable data for predicting business outcomes
- For most enterprise organizations, transactional data is kept on IBM Z

Data Gate – 3 flavors Review

	Data Gate with CP4D	Data Gate on Cloud	Data Gate for watsonx NEW
Usage	Analytics/AI with CP4D; Transactional cache for Db2z data	Analytics	Analytics/AI
Capture side part	5698-DG3 Data Gate for z/OS 3.1	5698-DG3 Data Gate for z/OS 3.1	5698-DG3 Data Gate for z/OS 3.1
Apply side part	Data Gate Service on CP4D (not priced separately from CP4D)	Data Gate on Cloud	Data Gate for watsonx
Captures data from	Db2 for z/OS	Db2 for z/OS	Db2 for z/OS, IMS DB, VSAM
Data applied to	Db2 LUW (Db2 AE cartridge) and/or Db2 WH	Db2 WHaaS	Iceberg
Pre-reqs	CP4D and one or both types of Db2s	Db2 WHaaS on AWS	watsonx.data
Notes	“Classic” Data Gate	No CP4D pre-req. Currently only supported on AWS	