

Data Gravity and IBM Z: Bringing Processing to the Data

Mehmet Cuneyt Goksu
Executive IT Specialist,
IBM Z Data Platform, IBM
mehmet.goksu@ibm.com

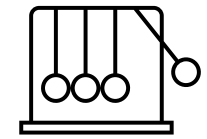
Vishal Gupta
Senior Solution Architect,
IBM Z Ecosystem, IBM
vishal.gupta16@ibm.com

Agenda

- ✓ Introduction – Why Data Gravity Matters
- ✓ Understanding Data Gravity Parameters
- ✓ The Cost of Fighting Data Gravity
- ✓ Data Gravity and IBM Z
- ✓ Future-Ready Approaches to Supporting Data Gravity
- ✓ IBM Solutions That Support Data Gravity

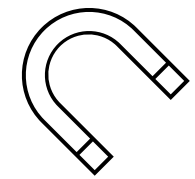
What is Data Gravity (The Core Concept)

As data grows, it attracts applications



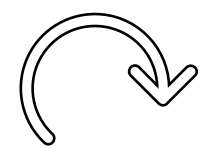
Physics Analogy

- Gravity is related to mass
- The greater the mass, the stronger the gravitational pull.



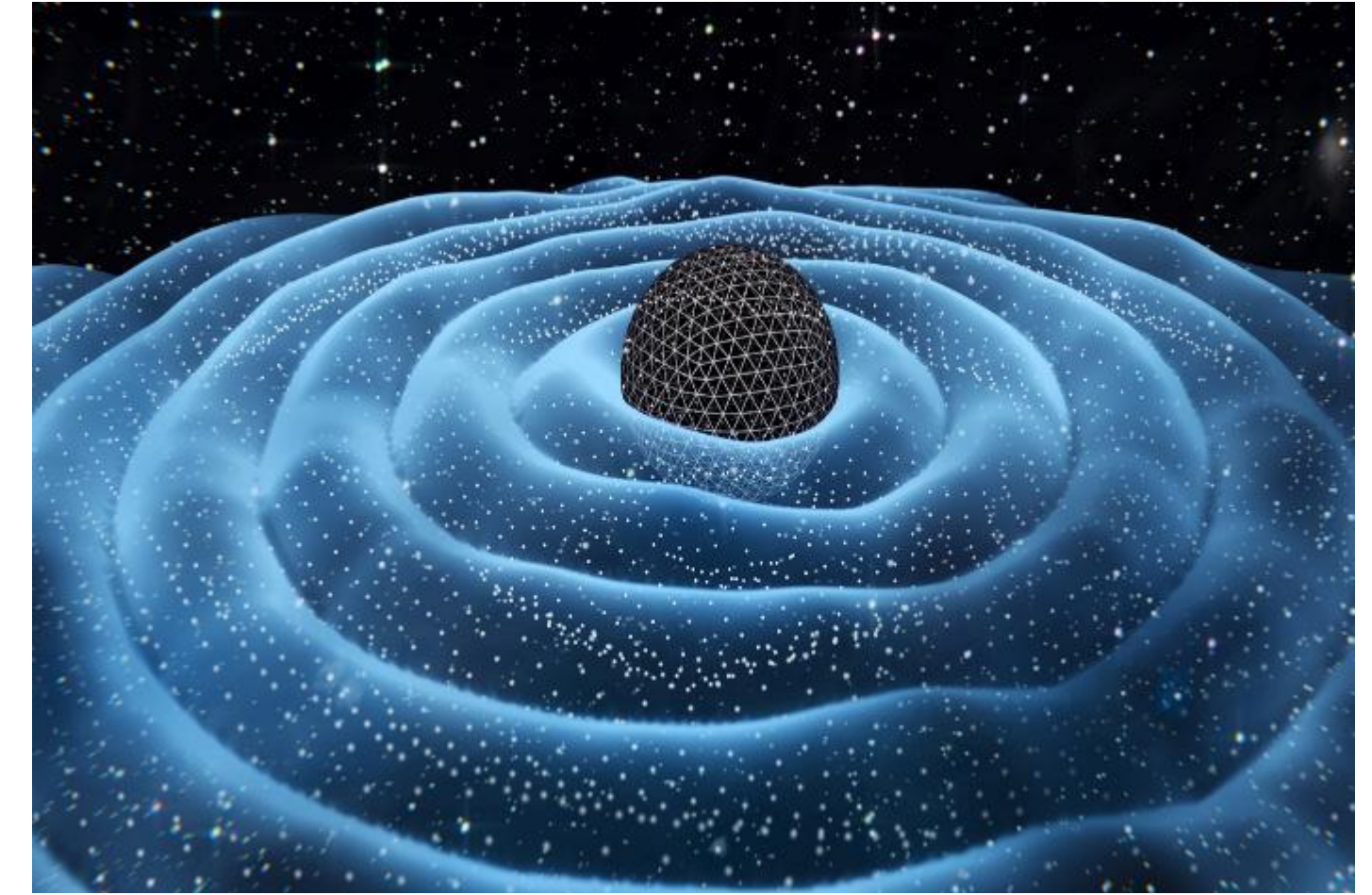
Data Gravity

- The greater the amount of data in a “system of origin”, the stronger the “pull” of that data



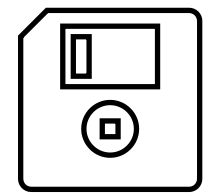
Architectural Shift

- Instead of moving data, bring processing to the data



Why Data Gravity Matters in Today's Hybrid Cloud

Data growth and hybrid cloud are making data movement inefficient



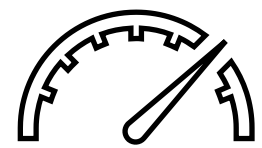
Data Explosion

- Massive data growth
- Continuous generation



Hybrid Cloud Complexity

- Data spread across systems
- Integration challenges



Performance Expectations

- Real-time insights
- Low latency required



Cost & Governance

- Data movement is costly
- Security & compliance risk

Moving data everywhere is no longer scalable, efficient or sustainable

Data Gravity: Key Factors

Data gravity depends on more than just data size

- **Data mass** - The larger the dataset, the stronger the pull to keep processing near it.
- **Application mass** - Complex, tightly coupled applications increase data gravity
- **Acceptable Latency** - If low latency is critical, processing must stay close to the data.
- **Frequency of Access** - When data is used frequently, processing naturally moves closer to it.

Higher mass → stronger gravity | Higher latency → weaker gravity



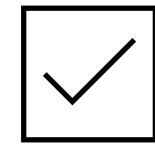
The Cost of Fighting Data Gravity

Moving or copying data creates hidden costs and risks



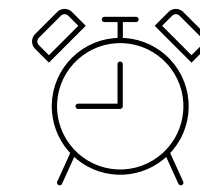
Data Security

Duplicating sensitive data increases exposure



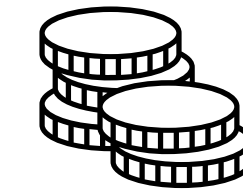
Data Consistency

Keeping data accurate across systems is difficult



Data Latency

Data copies to another system can be hours behind real time



Cost

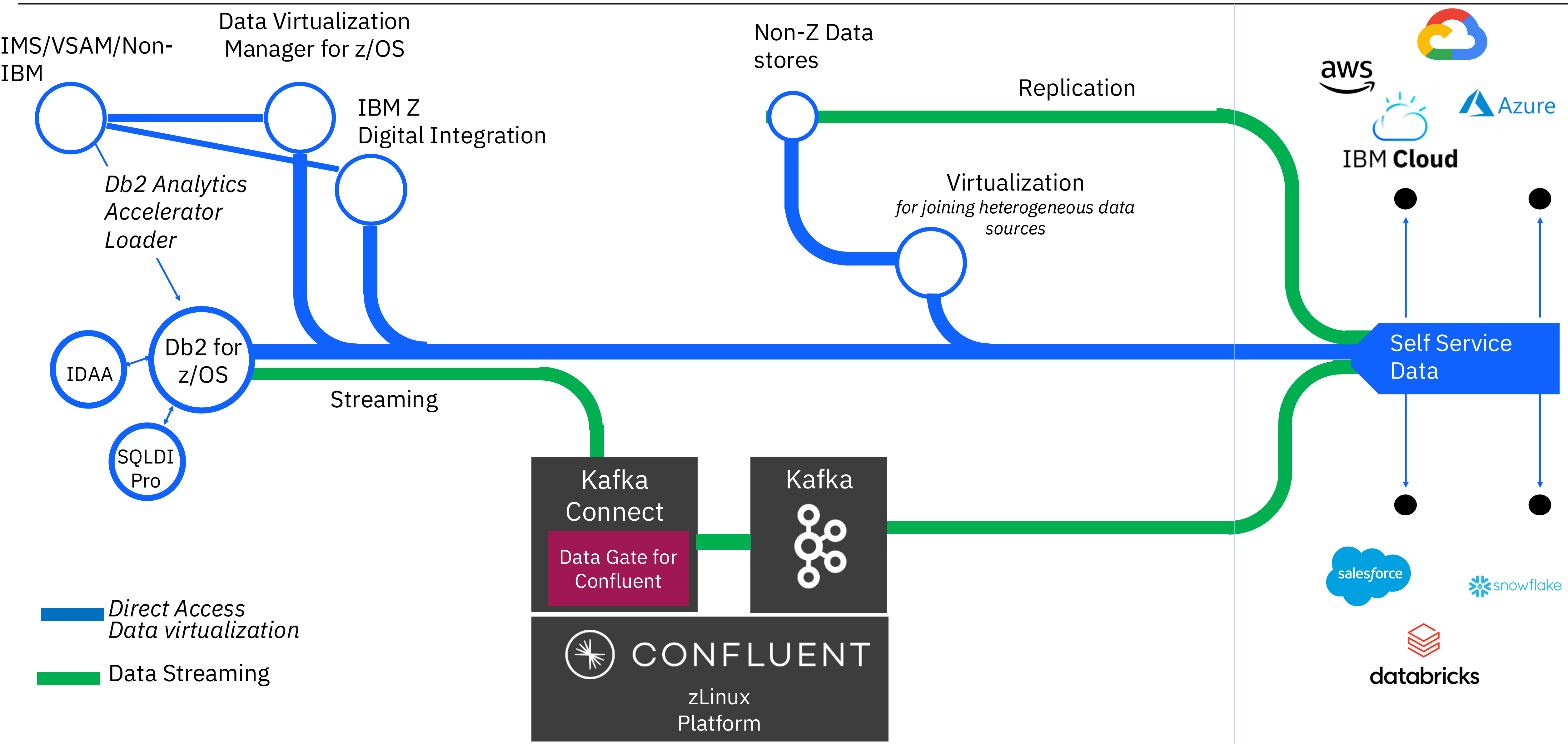
Continuous replication and storage can be expensive

Moving data creates more problems than it solves

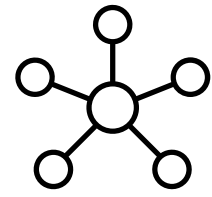
Enterprise data is created and resides on **IBM Z**

So the question is — how do we **bring processing closer** to it?

IBM Z Data and Application Architecture

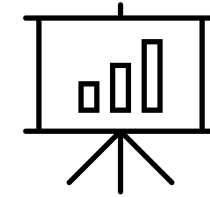


Bringing Processing Closer to Data



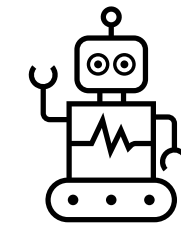
Data Virtualization

Access data where it resides without moving it



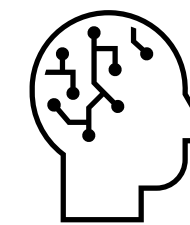
Analytics on Platform

Run analytics directly where the data is created



AI Near Data

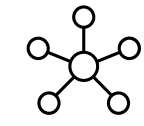
Enable real-time AI by bringing models closer to data



AI Embedded in Data

Make data intelligent with built-in semantic understanding

From Approaches to IBM Z Capabilities



Data Virtualization



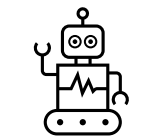
IBM Data Virtualization Manager for z/OS



Analytics on Platform



IBM Db2 Analytics Accelerator



AI Near Data



IBM Machine Learning for z/OS



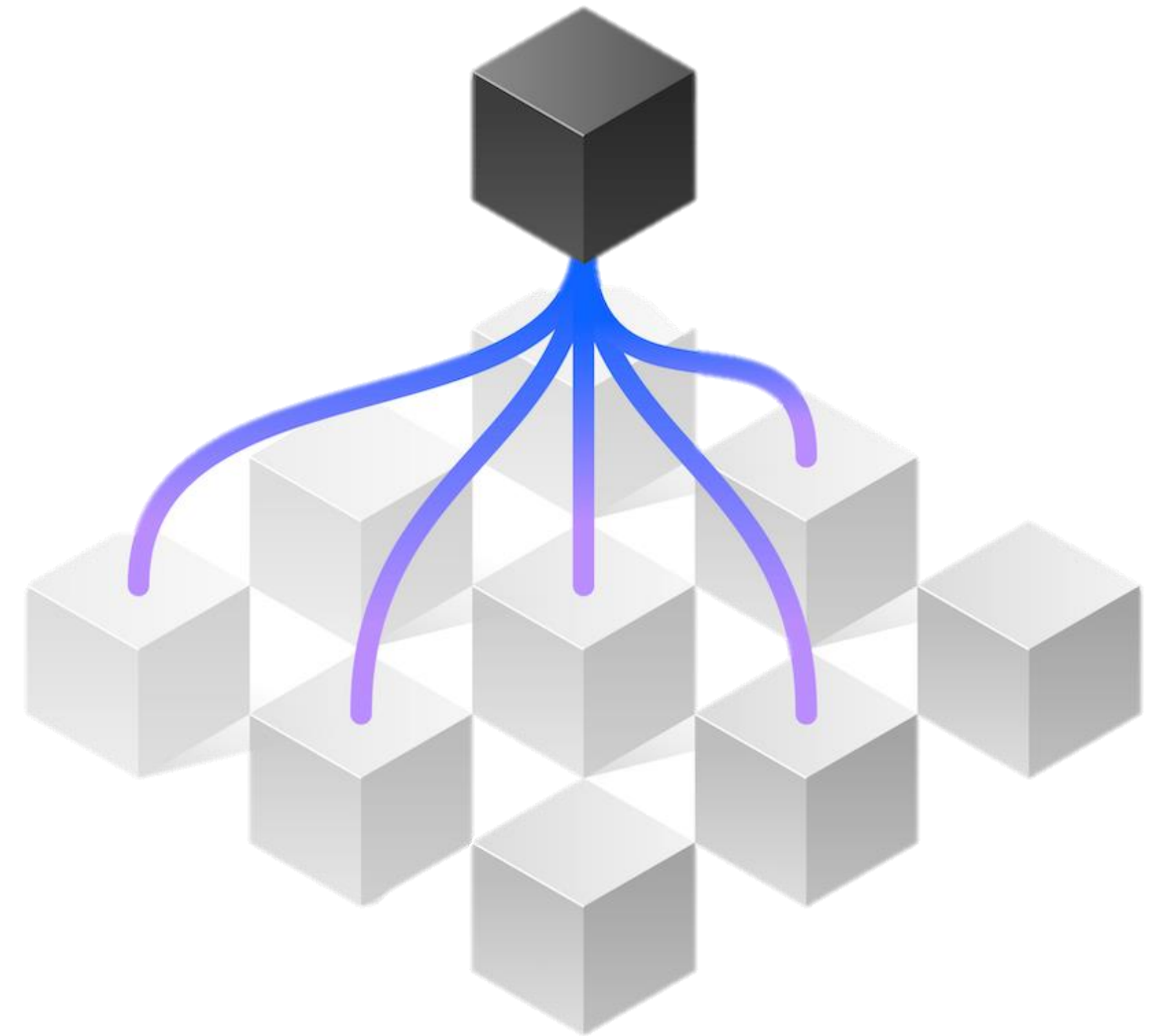
AI Embedded in Data



IBM SQL Data Insights Pro

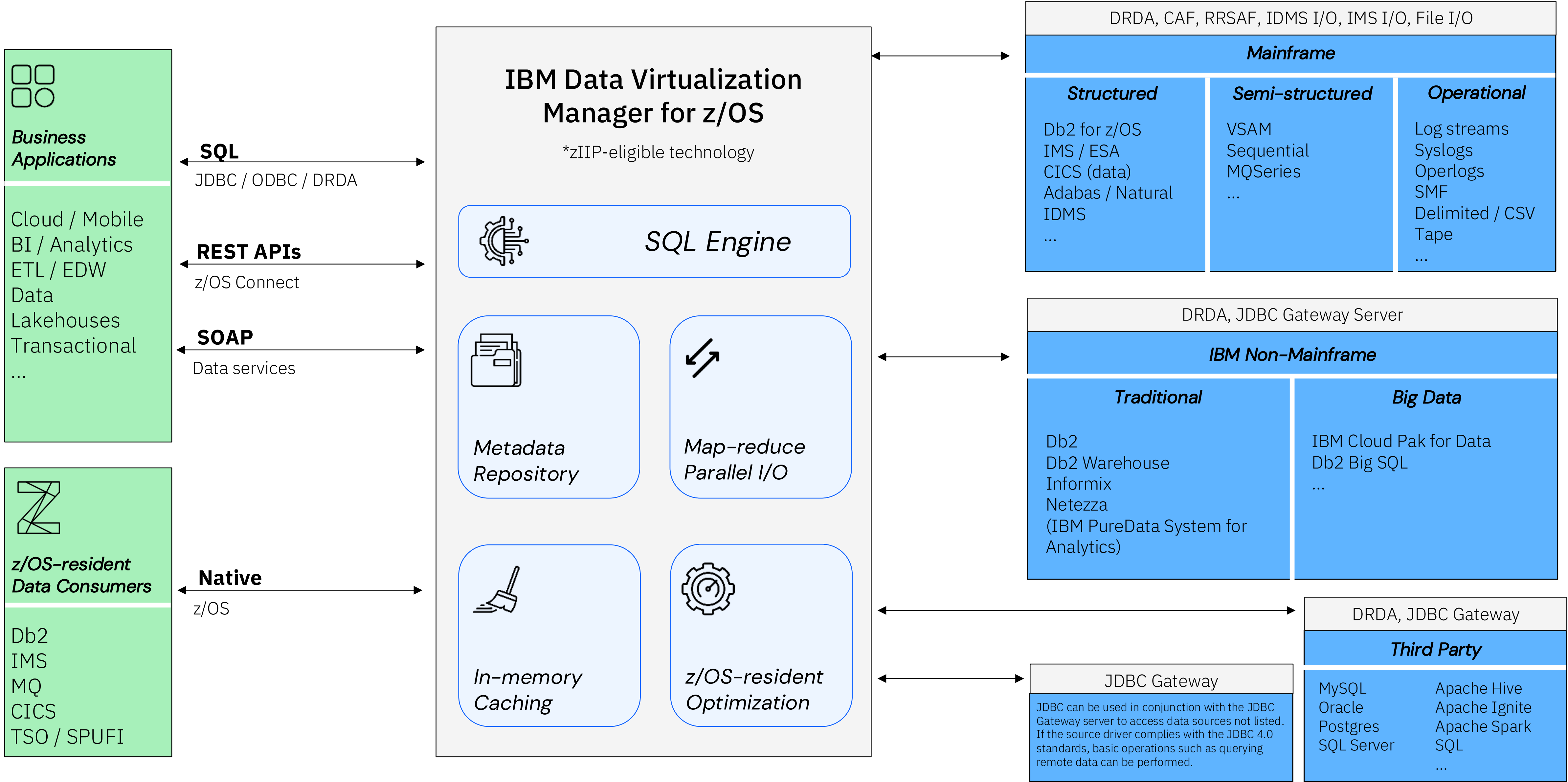
IBM Data Virtualization Manager for z/OS

- A data services layer to access data and content on demand and integrate disparate data sources in real time
- An abstraction layer to shield developers from requiring specific skillsets for multiple data sources
- A metadata catalog to keep track of data location and availability
- Support for transactions that write back to the original data sources (except for sequential files).
- Integration with the IBM watsonx.data lakehouse via PrestoDB connector



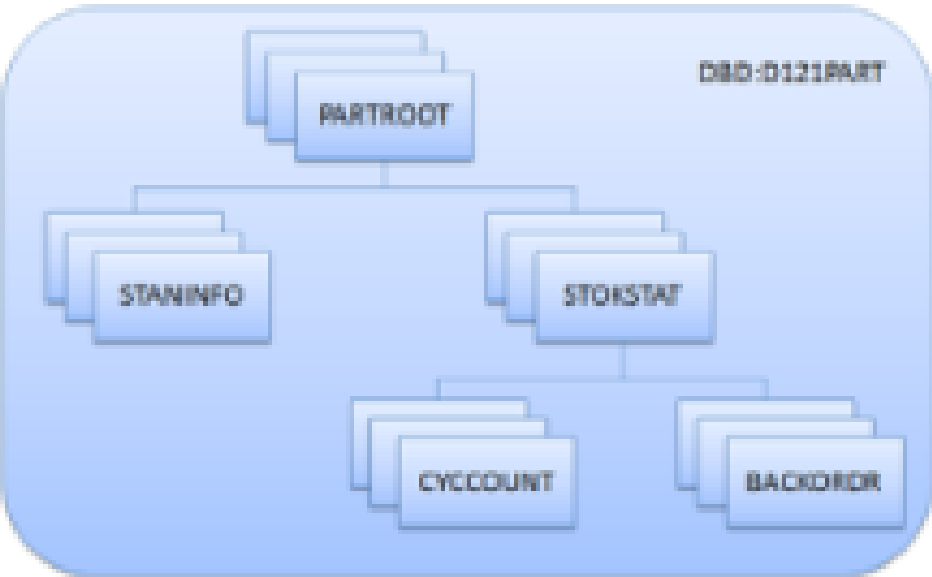
DVM technical overview

A look under the hood and how we integrate with the broader virtualized ecosystem



DVM - IMS into Relational

Secure, read/write access to non-relational data Adabas, IMS, IDMS, VSAM, SMF....



```
Database Definition (DBD)
DBD NAME=DI21PART,ACCESS=(VSAM VSAM)
DATASET DD1=DI21PART,DEVICE=(3390,ONLINE)DD2=PARTD,
SIZE=(2048,2048),RECORD=(578,478)
SEGN NAME=PARTROOT, PARENT=, BYTES=64, FREQ=250
FIELD NAME=PARTKEY, SEQ, TYPE=C, BYTES=17, START=1
SEGN NAME=STANINFO, PARENT=PARTROOT, BYTES=85, FREQ=1
FIELD NAME=(STANKEY, SEQ, TYPE=C, BYTES=2, START=1
SEGN NAME=STOKSTAT, PARENT=PARTROOT, BYTES=160, FREQ=2
FIELD NAME=(STOCKKEY, SEQ, TYPE=C, BYTES=15, START=1
SEGN NAME=CYCCOUNT, PARENT=STOKSTAT, BYTES=25, FREQ=1
FIELD NAME=(CYCLKEY, SEQ, TYPE=C, BYTES=2, START=1
SEGN NAME=BACKORDR, PARENT=STOKSTAT, BYTES=75, FREQ=6
FIELD NAME=(BACKKEY, SEQ, TYPE=C, BYTES=16, START=1
DROGEN
FINISH
END
```

```
Program Specification Block (PSB)
DBPCB01 PCB TYPE=DB,DBDNAME=DI21PART,PROCOPT=GOT,KEYLEN=43
SENSEG NAME=PARTROOT
SENSEG NAME=STANINFO,PARENT=PARTROOT
SENSEG NAME=STOKSTAT,PARENT=PARTROOT
SENSEG NAME=CYCCOUNT,PARENT=STOKSTAT
SENSEG NAME=BACKORDR,PARENT=STOKSTAT
PSBGEN LANG=COBOL,PSBNAME=DFSSAM07
END
```

Mainframe non-relational data structure

Relational Tables and Views with Column Names				
Table-name PARTROOT View-name IMS_PARTREC Column-names USERKEY PARTKEY DESCRIPTION Child	Table-name STANINFO View-name IMS_STANINFO Column-names STANKEY PROCUREMENT_CODE MAKE_DEPT Parent	Table-name STOKSTAT View-name IMS_STOKSTAT Column-names STOCKKEY SS_AREA SS_DEPT Parent Child	Table-name CYCCOUNT View-name IMS_CYCLREC Column-names CYCLEKEY PHYSICAL_COUNT TOTAL_STOCK Parent	Table-name BACKORDR View-name IMS_BACKREC Column-names BACKKEY WORK-ORDER WO_QTY Parent

Transformed into relational format

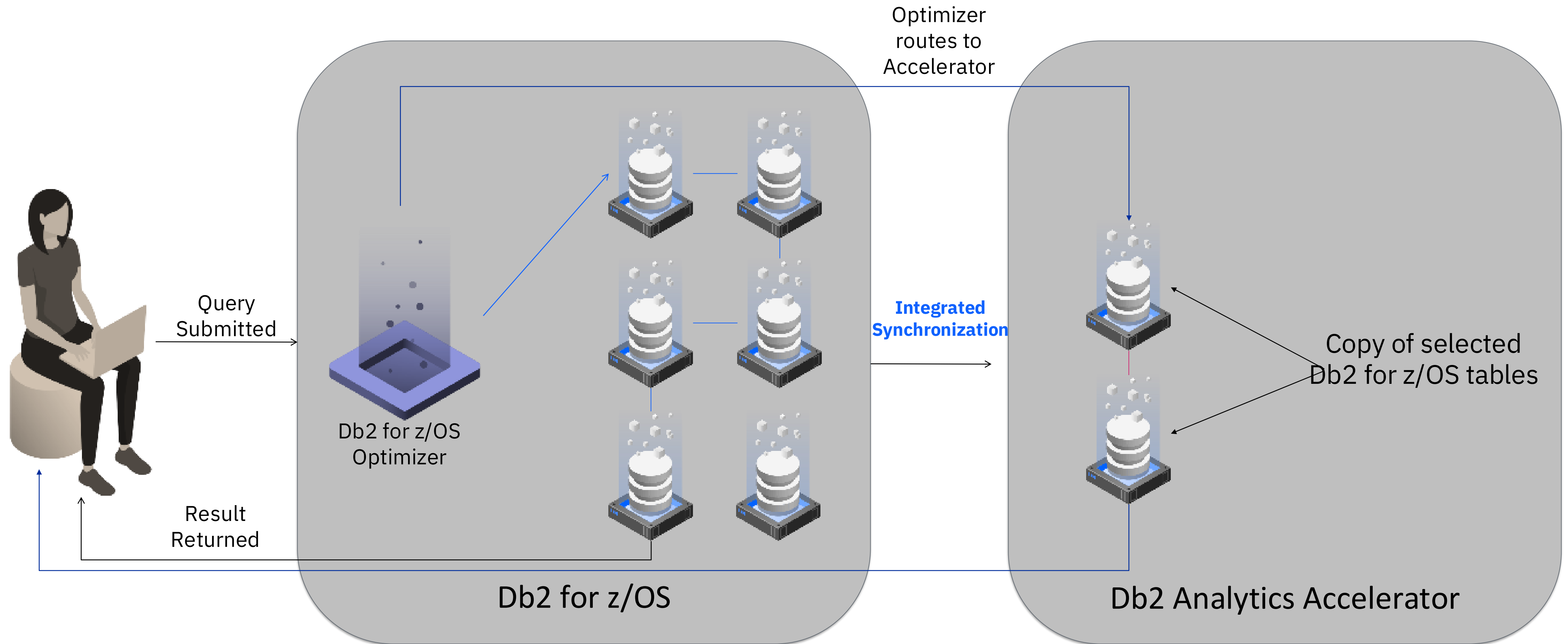
IBM Db2 Analytics Accelerator for z/OS

High speed analysis of enterprise data



- A high-performance component tightly integrated with Db2 for z/OS
- High-speed processing of complex and resource intensive Db2 queries
- Supports business-critical reporting and analytic workloads
- Turns Db2 into a robust transactional and analytics processing systems

How the Accelerator works (simplified)



Two options for copying IMS data to Accelerator

IDAA Loader

PROS

Dual-Load option (Db2 and/or Accelerator)

Limited zIIP eligibility

HA Load (Loading Multiple Accelerators)

CONS

Snapshot of the IMS source, Mini Batches

Always full refresh is needed

Change Data Capture

PROS

Near Real Time Replication

Only changes are replicated

Replicate directly to Accelerator

CONS

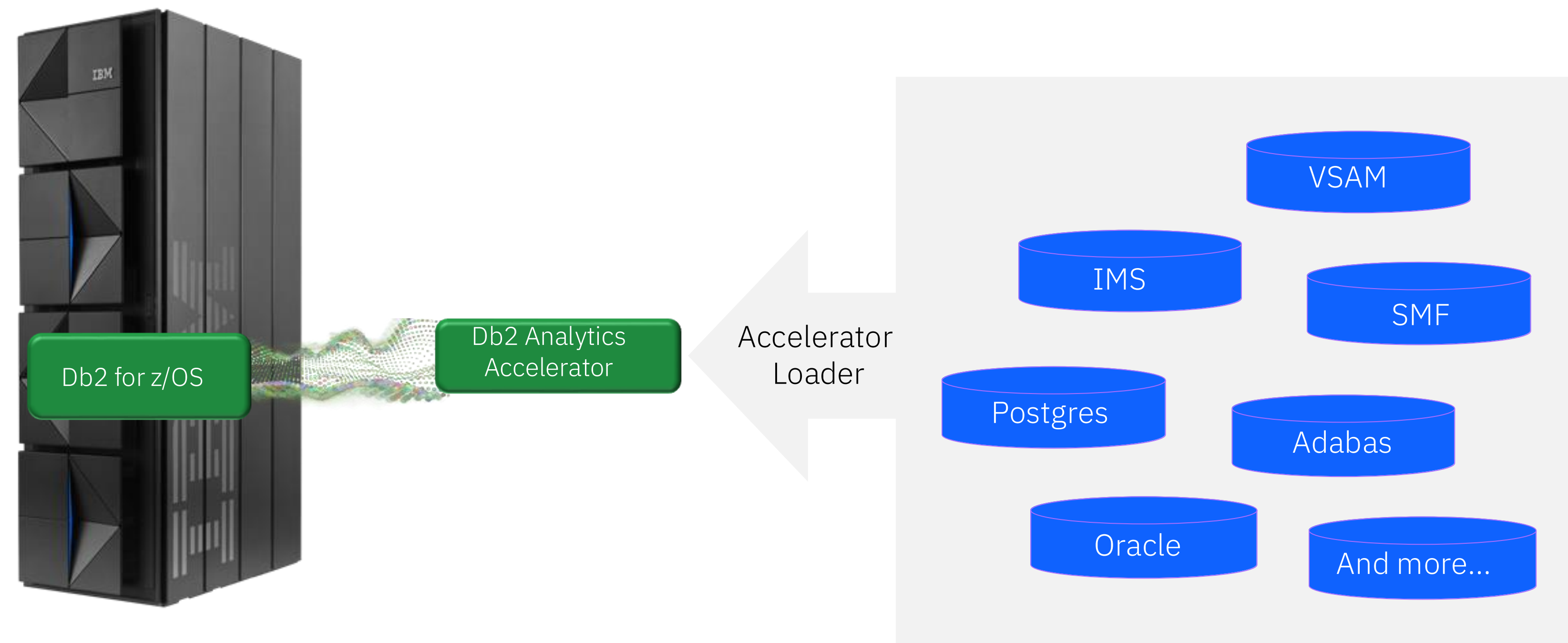
No zIIP Eligibility

IBM Db2 Analytics Accelerator Loader

*A separately priced offering that extends Accelerator benefits to VSAM, **IMS**, Adabas, and SMF*

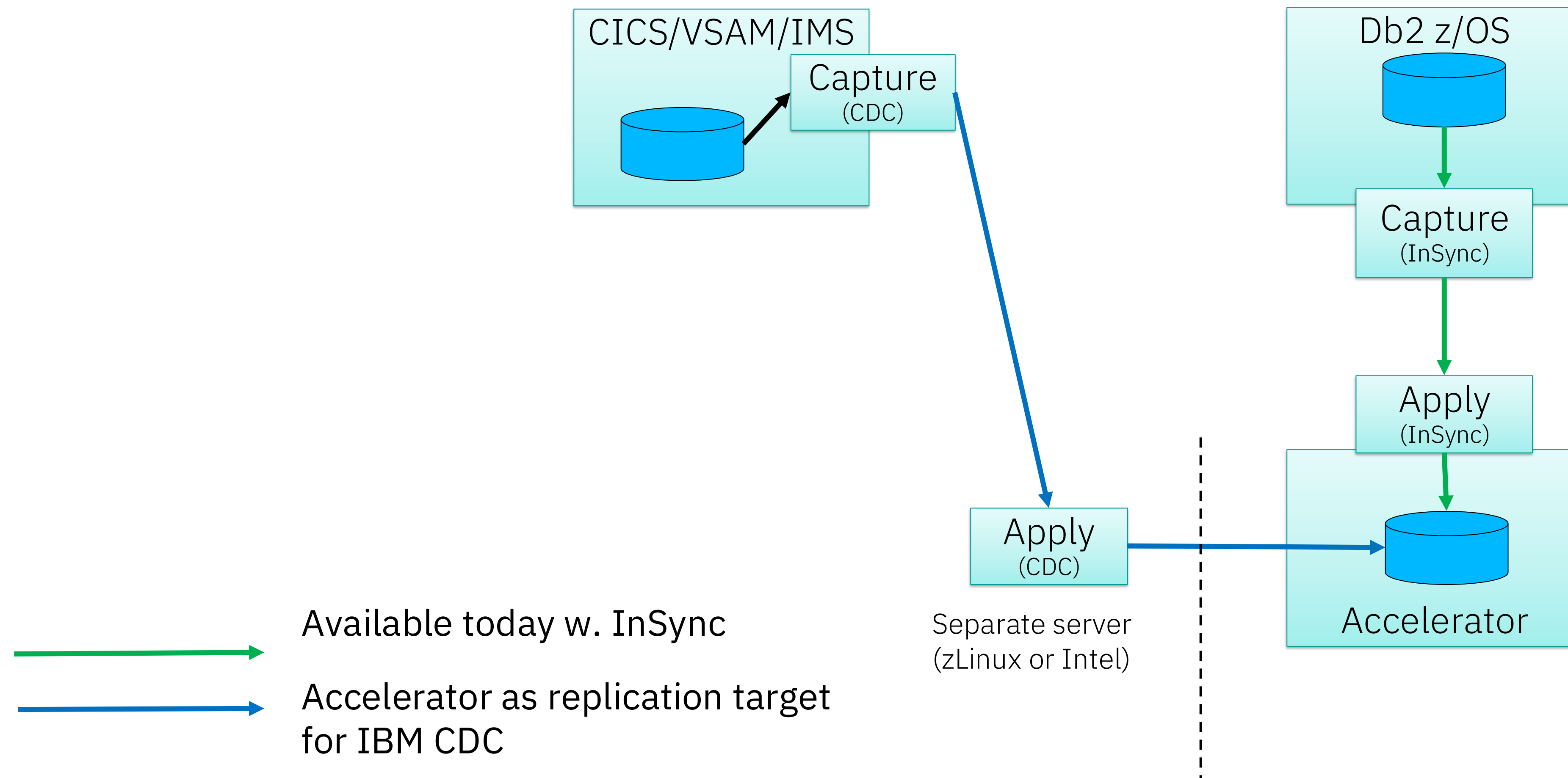
Loader increases analytics capabilities by adding access to data:

- Outside of Db2 for z/OS
- Outside of IBM Z
- HA Load Capabilities



Use-Case: External Replication – Simpler View

- External replication of non-relational mainframe data, such as IMS
- Use-case: Run analytical SQL queries against IMS or VSAM data, or against a combined view of IMS or VSAM data and Db2 for z/OS data.





IBM SQL Data Insights Pro

Uncover and monetize hidden insights in your most important data

- AI-powered extension for IBM Db2 for z/OS that brings vector-based intelligence and semantic understanding directly to data on IBM Z - without moving data off-platform.
- Automatically builds an unsupervised, neural network powered Large Database Model, to answer business questions
- Integrates AI within any Db2 application or tool via common SQL – no data scientist required!
- Can integrate non-Db2 data
- *An example: Based upon historical purchases made by similar buyers, [show me the most likely upsell products](#)*

Bring AI to the data — not data to AI

SQL Data Insights Pro Business Value

SQL Data Insights Pro adds AI models to Db2 that understand data meaning, enabling semantic queries—running securely on IBM Z without moving data.

More Relevant Insights

Get answers based on meaning — not just exact matches

Faster decisions

AI insights directly from Db2
No waiting on data pipelines

Low Cost

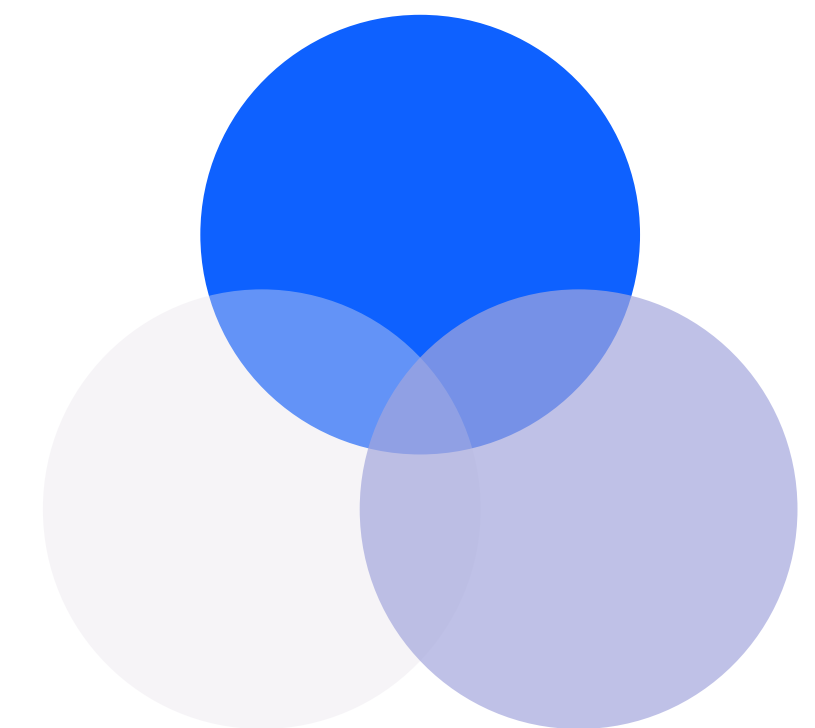
No data movement
No duplication

Reduced Risks

Data stays secure and governed on IBM Z

Broader Access to Insights

Use familiar SQL to explore AI-driven insights



AI is Changing Data Interaction

Combined relational predicate with AI semantic functions

From queries → **Discovery**

From data access → **Insights Generation**

Meaning-based semantic queries in SQL DI Pro

AI_SIMILARITY

- Find claims similar to a claim involving vehicle collision and minorly body injury.

AI_SEMANTIC_CLUSTER

- Find similar insurance policies as a group of policies in “comprehensive auto coverage, collision coverage, roadside assistance”.

AI_COMMONALITY

- Find the most common coverages in the homeowners insurance category.

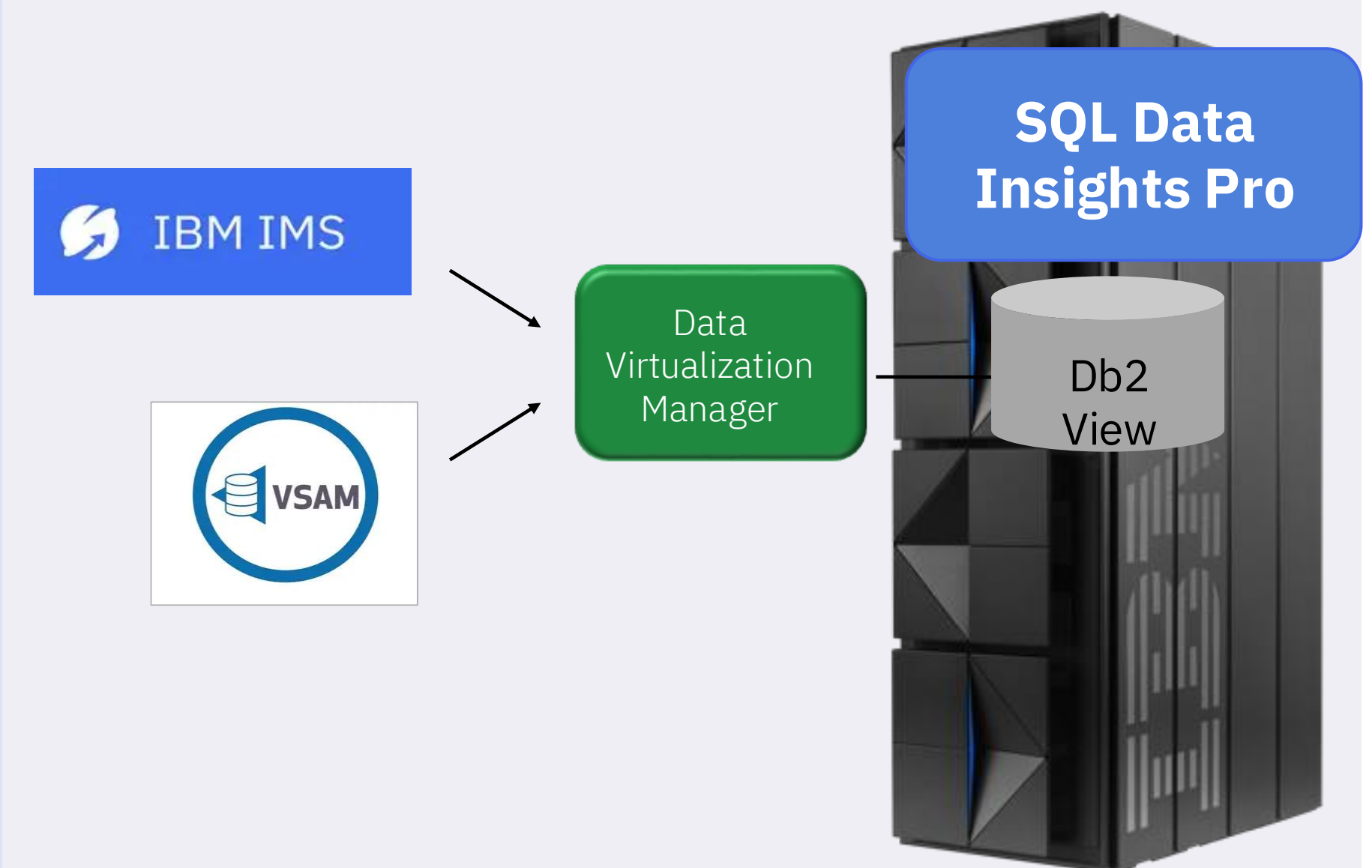
AI_ANALOGY

- Give me a life insurance policy that has a similar risk classification as Level Term Life Insurance.

SQL Data Insights Pro

What about VSAM & **IMS** Data?

- ✓ A view in Db2 for z/OS is a way of representing a subset of data as a “virtual” Db2 table
- ✓ IMS and VSAM data can be virtualized into Db2 views using the DVM product
- ✓ SQL Data Insights Pro can access and analyze that data during model training and query execution

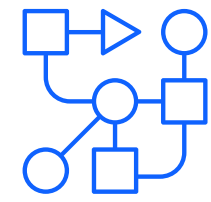


Machine Learning for IBM z/OS - Enterprise Edition

Accelerated inferencing at scale for Transactional AI workloads on z/OS

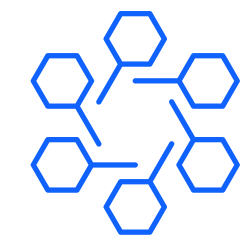
A full-featured ML platform on z/OS that brings AI into mission-critical applications by deploying models close to data and transactions—enabling high performance, low latency, and real-time insights.

Benefits



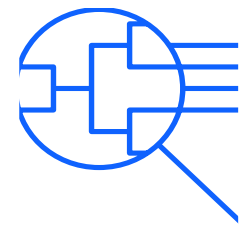
End-end platform

GUI and script driven features for end-end AI model lifecycle management.



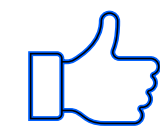
Scalable AI

Leverage Telum I and Telum II accelerator to score transactions natively in CICS, IMS and batch applications at scale.



Easy import and deploy

Leverage predictive models and Encoder LLMs trained anywhere or with IBM watsonx.ai and deploy with MLz.

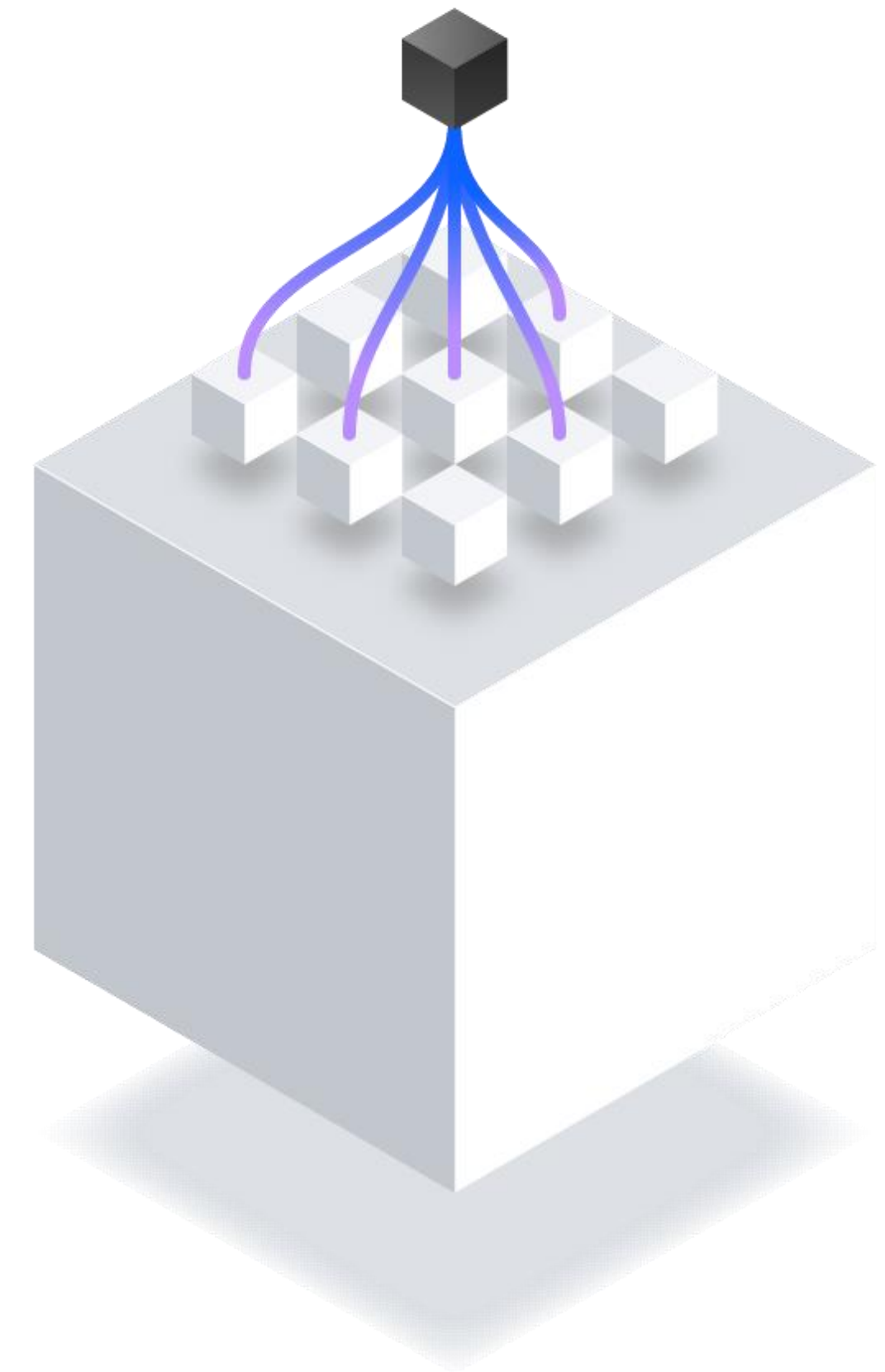


Trustworthy features

Leverage built in explainability and drift detection features or enterprise AI governance with IBM watsonx.governance.

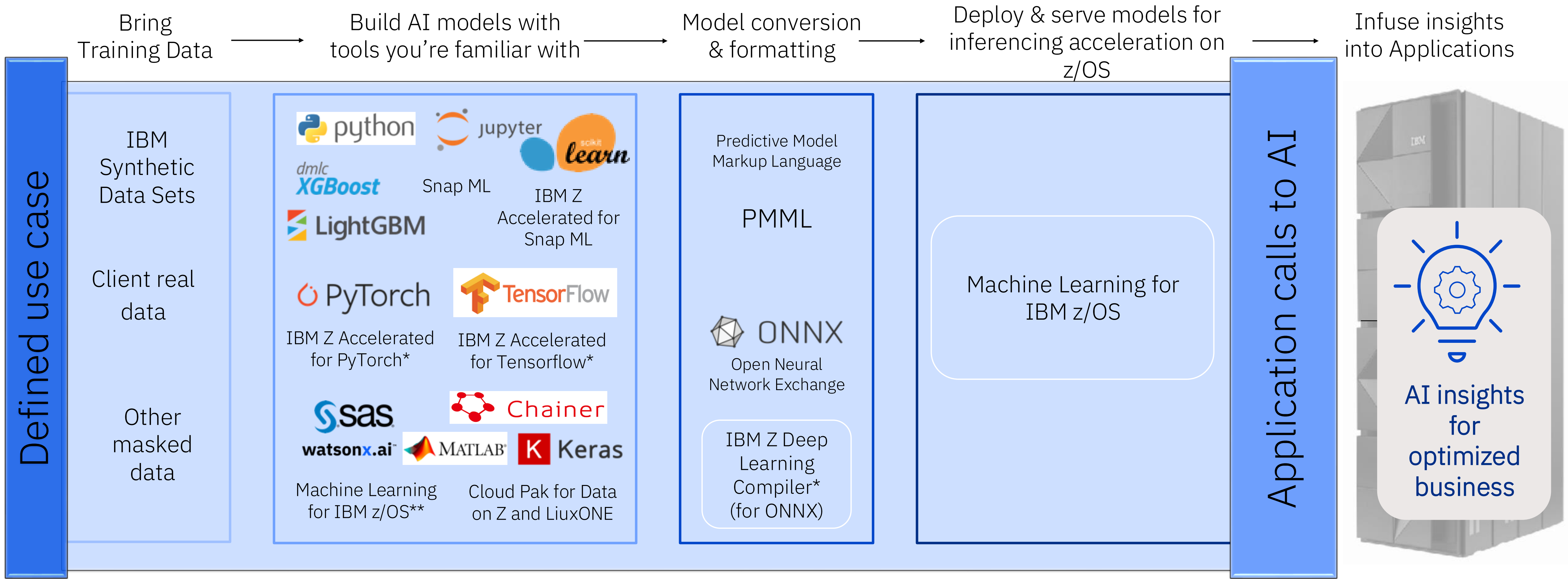
Transactional AI

Infuse AI into every transaction



Train anywhere, deploy on IBM Z

Leverage & extend AI investments with MLz



Train anywhere... but execute where transactions happen.

