

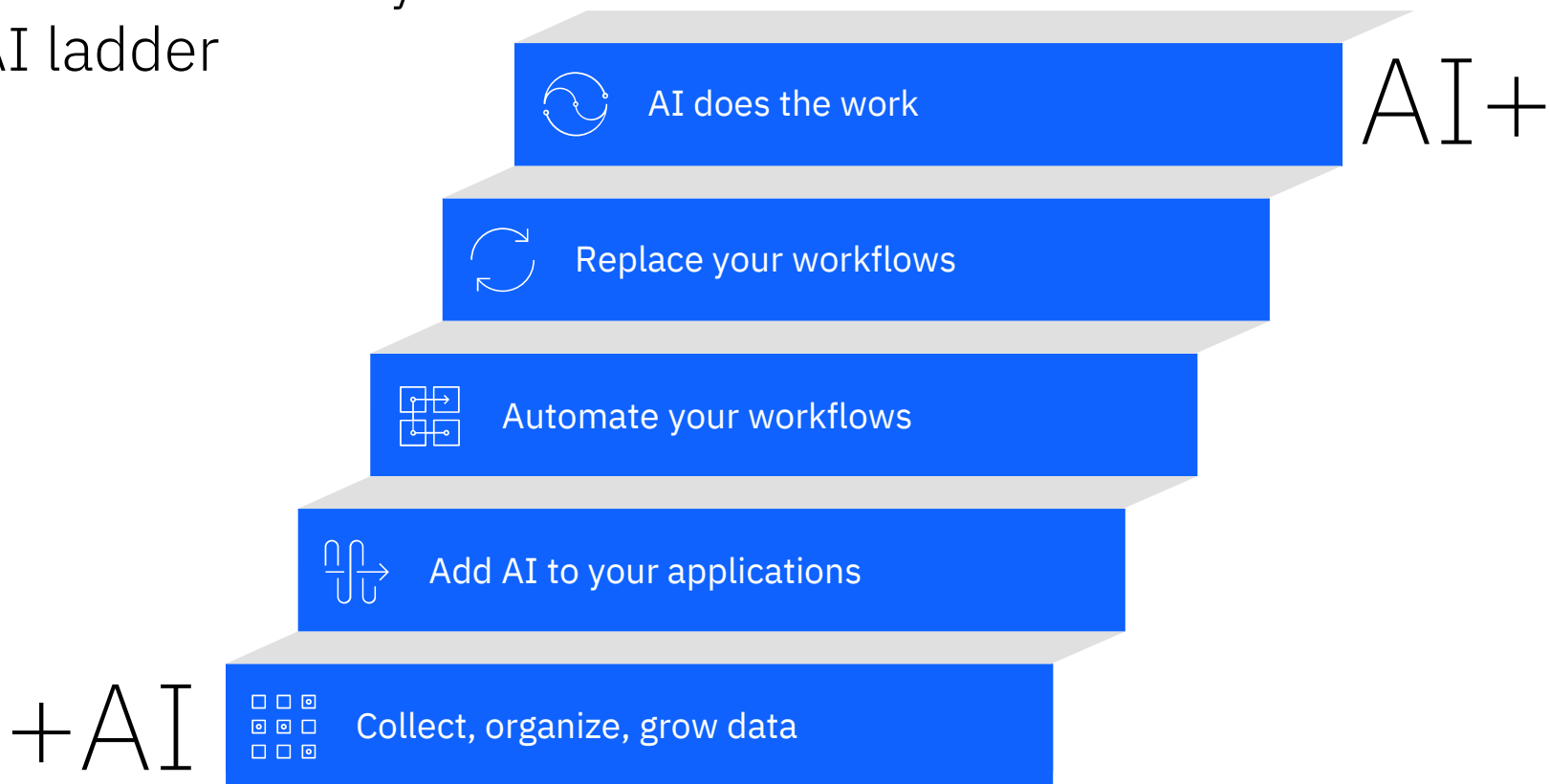
Современный подход к построению корпоративной экосистемы управления данными и аналитикой на базе продуктов IBM

Pavel Maltsev, NCEE Data&AI SME



+ AI → AI +

The modern-day AI ladder



A design concept and architecture for data management
that addresses the challenges of data complexity

Data Fabric



Business benefits of a data fabric

Simplifies data complexity through automating data integration, data governance, and data consumption

Intelligent integration

Unified governance

Knowledge insights

Automates data engineering tasks and augments data integration across hybrid cloud resources

Automates data governance, protection and security tasks; enabled by active metadata

Enables self-service for easy end user data consumption and collaboration

Digital Transformation: 3 Mega Trends

Move to inter-cloud



Hybrid data architecture for BI and AI

Hybrid Cloud Data Fabric

Data Lake & Data Warehouse convergence



Single query engine decoupled from storage to support enterprise BI and AI

Modern Data Warehouse

AI governance



AI lifecycle to support end-to-end Governance, Risk, and Compliance (GRC)

End-to-end governance & privacy

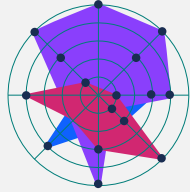
Today, enterprises are faced with multiple data challenges



More data

Exploding data growth

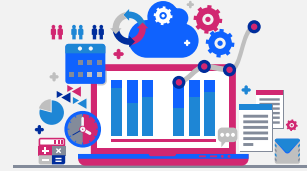
The aggregate volume of data stored is set to grow over **250%** in the next five years



In multiple locations

Multiple locations, clouds, applications and silos

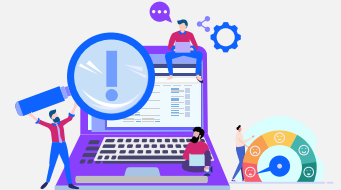
82% of enterprises are inhibited by data silos



In complex forms

Documents, images, video, and more

80% of time spent on data cleaning, integration and preparation

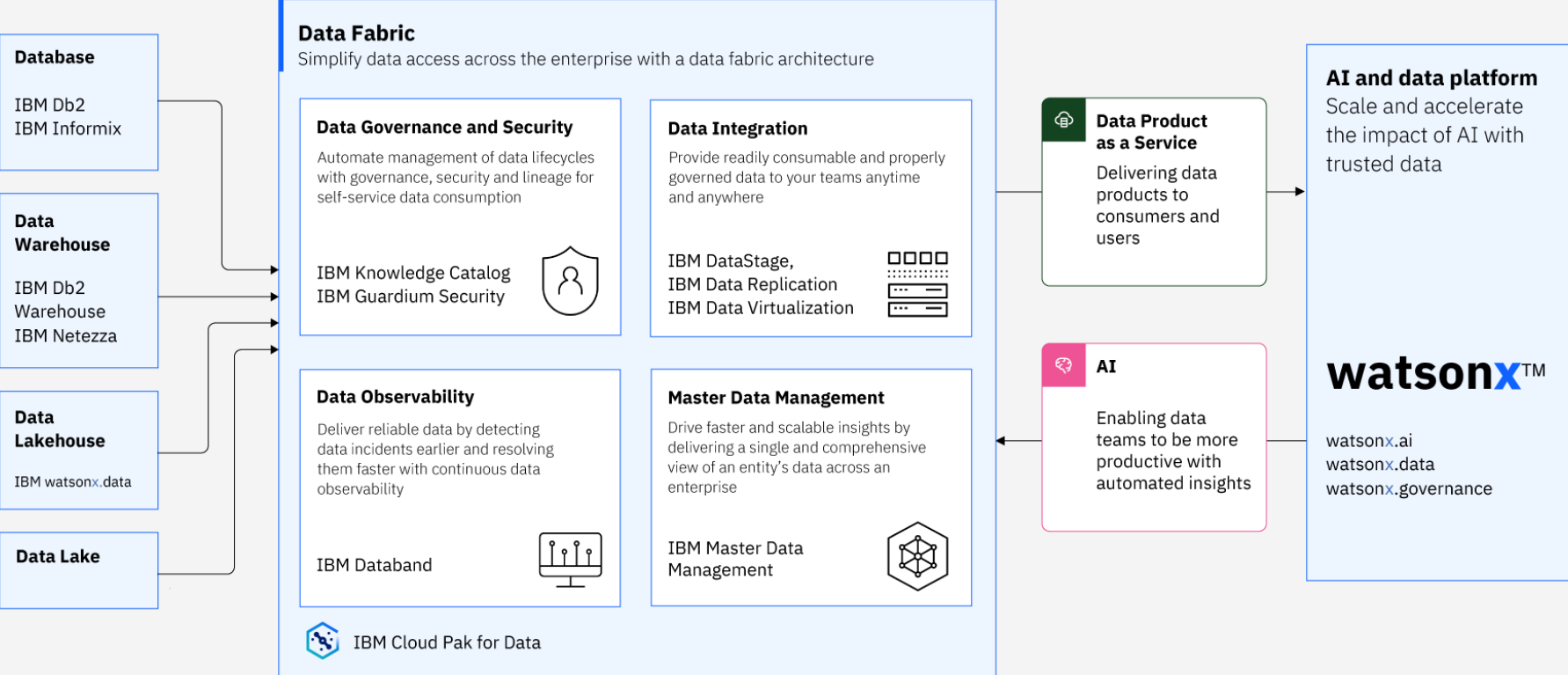


Poor quality

Stale and inconsistent

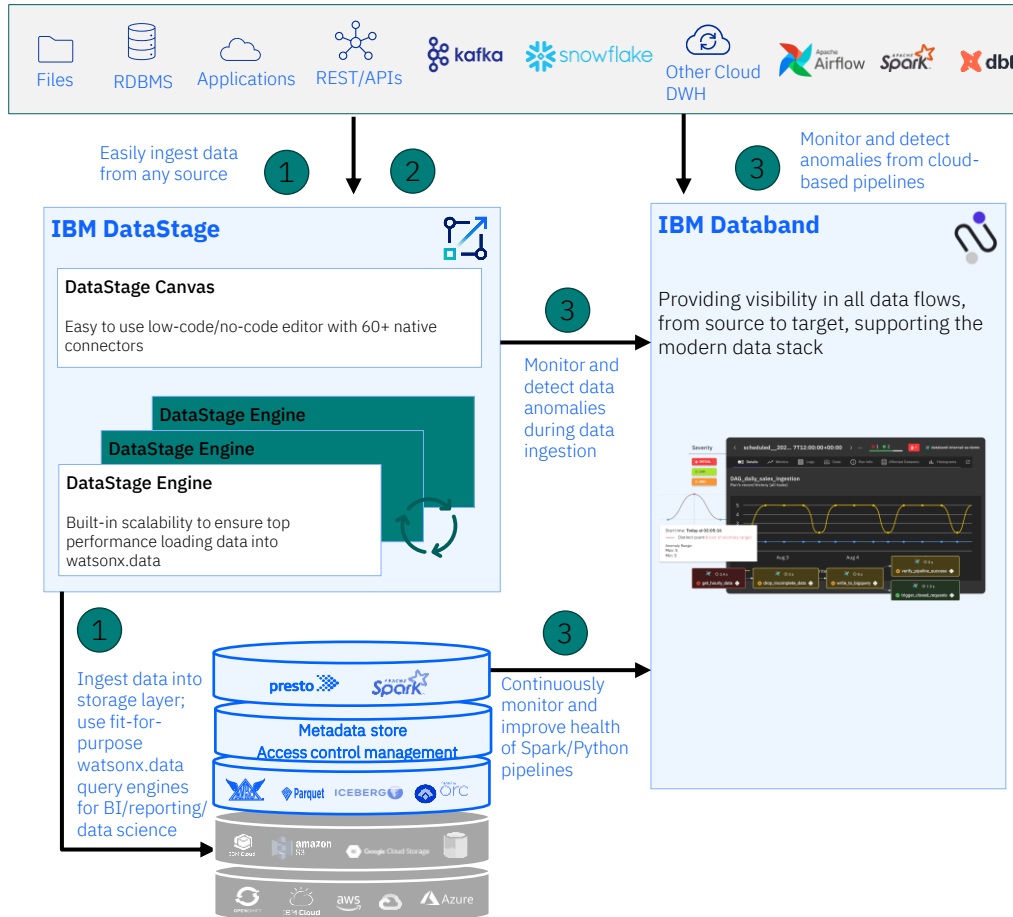
82% of enterprises say data quality is a barrier on their data integration projects

Investments in a trusted data foundation will accelerate and scale AI



Effortlessly populate watsonx.data with trusted data

Leverage best in class Data Ingestion and Observability



Data Pipelines with IBM DataStage

Easily build EL(T) pipelines with an intuitive visual design

1 Ingest data from any source

Leverage over 60+ native connectors to ingest data into watsonx.data from any type of source, ensuring top performance with built-in engine scalability

2 Reduce cost by offloading data from cloud data warehouses

Offload data from cloud data warehouses to enable shifting workloads like BI, reporting, or data science to fit-for-purpose query engines

Data Observability with IBM Databand

Continuously detect and resolve data quality incidents

3 Monitor, detect, and resolve data quality incidents

Monitor and improve the health of DataStage, Spark, or Python pipeline workloads running on watsonx.data

Detect data anomalies and accelerate issue resolution

IBM industry leadership

Forrester Wave™ and Forrester New Wave™ Leaders

IBM Watson Studio:
Multimodal Predictive Analytics and Machine Learning Solutions

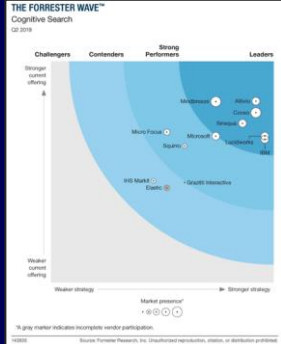
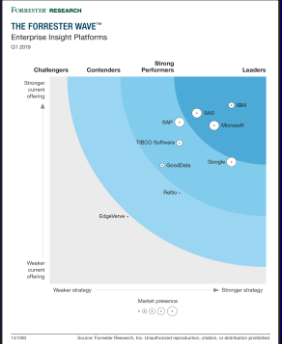
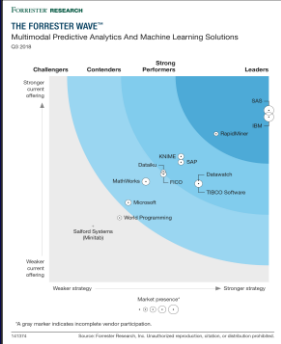
IBM Watson Knowledge Catalog:
Machine Learning Data Catalogs

IBM Cloud Pak for Data:
Enterprise Insight Platforms

IBM Watson Assistant:
Conversational Computing Platforms

IBM Watson Discovery:
Cloud AI Developer Services

IBM Watson Discovery:
AI-Based Text Analytics Platforms



IBM #1 in AI Market Share



Industry Awards



A'DESIGN AWARD

Full source information for all analyst reports on this page is contained in the speaker notes.

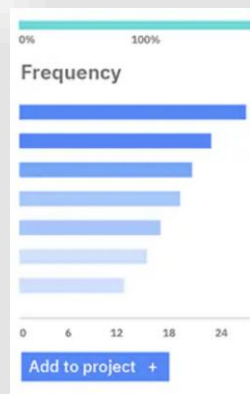
The Forrester Wave™ and Forrester New Wave™ are copyrighted by Forrester Research, Inc. Forrester, Forrester Wave™, and Forrester New Wave™ are trademarks of Forrester Research, Inc. The Forrester Wave™ and Forrester New Wave™ are graphical representations of Forrester's call on a market. Forrester Wave™ is plotted using a detailed spreadsheet with exposed scores, weightings, and comments. Forrester does not endorse any vendor, product, or service depicted in the Forrester Wave™ or the Forrester New Wave™. Information is based on best available resources. Opinions reflect judgment at the time and are subject to change.

Cloud Pak for Data 4.7 – The Platform

Cloud Pak for Data platform delivers the must have foundational capabilities for a robust enterprise grade system which is table stakes for top Tier-1 companies.

It's all about having the right tools to access the right data at the right time. Cloud Pak for Data platform strives to bring this notion to reality with its 40+ services for accessing and leveraging data efficiently, ensuring security and connectivity.

Cloud Pak for Data 4.7 continues to focus on Enterprise Readiness, Security & Connectivity for production deployments. Additional focus on quality fixes and security vulnerability patches in monthly releases.



Overview
Deployment spaces
After you create deployment spaces, you'll see them here.
New deployment space +

Experiment completed
8 PIPELINES GENERATED

Built for Multi-cloud
Avoid vendor lock-in & get started on your cloud journey today

PICK YOUR ADD-ON
Containerized Services
DaaS, DaaS ES, Mongo, Cognos, DSP, ISV, DaaS, Custom

DATA & AI PLATFORM
#1 Ranked by Forrester
IBM Cloud Pak for Data
Collect, Organize, Analyze, Infuse

KUBERNETES BASED
Containerized, easy to manage
RED HAT OPENShift Container Platform

PICK YOUR CLOUD
Private or Public
On-Premise, IBM Cloud, Azure, AWS, openstack

My virtualized data		
<input type="checkbox"/> Table	Schema	Created on
<input type="checkbox"/> RANKING_RESULTS	EMPLOYEE	Sep 27, 2021
<input type="checkbox"/> TRAINING_DETAILS	EMPLOYEE	Sep 27, 2021
<input type="checkbox"/> EMPLOYEE	EMPLOYEE	Sep 27, 2021
<input type="checkbox"/> EXPENSE_DETAIL	EMPLOYEE	Sep 27, 2021

STORAGE
15.78 KB used | File system



IBM Cloud Pak for Data

High-level architecture

Extended Services

Ecosystem of open-source, partner, and IBM services

Base Data Services

Integrated self-service data analytics tools

Cloud Pak Control Plane

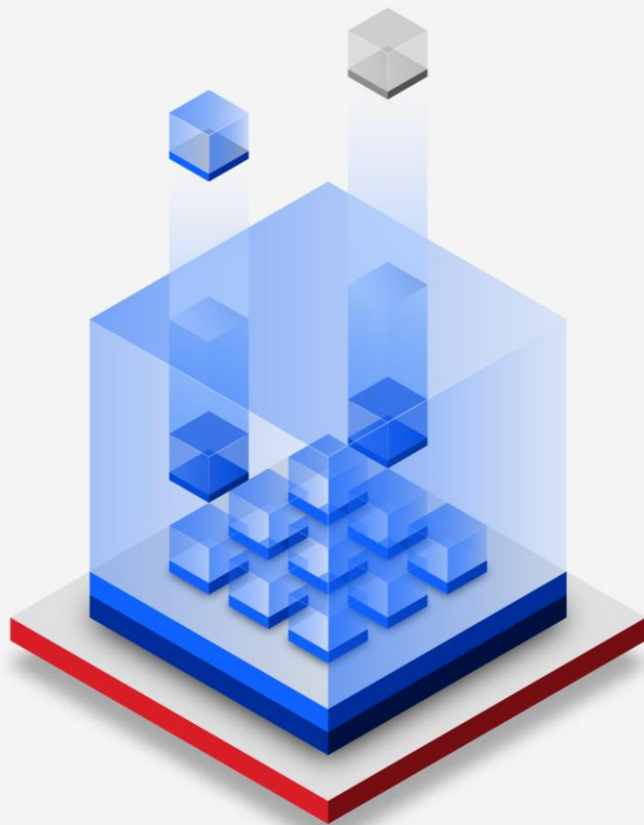
Administration tools, services catalog, and the central user experience

Red Hat OpenShift

Container-based software deployment platform that runs CP4D services on underlying infrastructure

Deployment flexibility

Deployable on any vendor's cloud platform or on-premises hardware; available as managed services

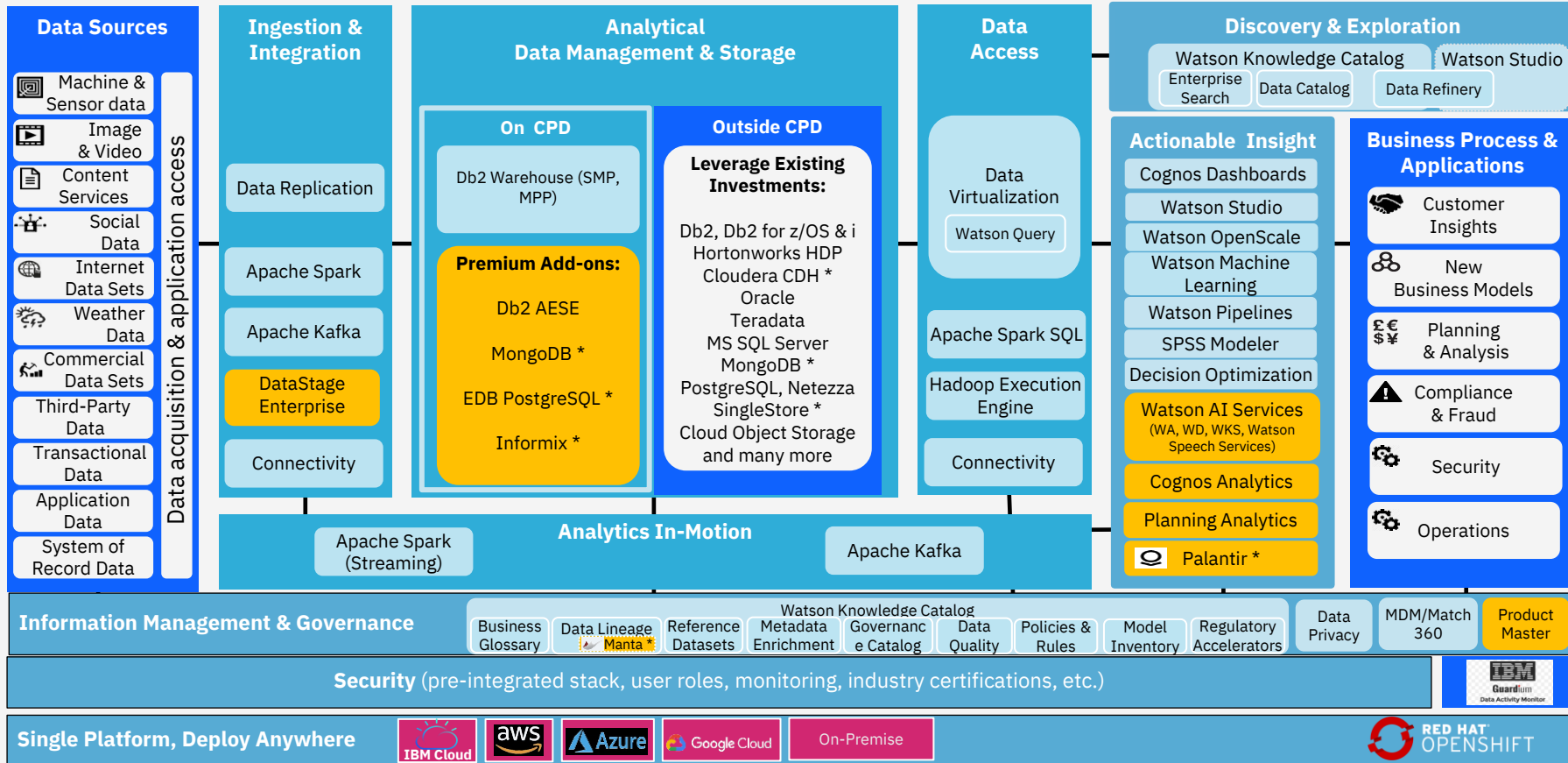


Reference Architecture – Cloud Pak for Data 4.x

IBM Cloud Pak for Data (Base)

IBM Cloud Pak for Data – Premium Add-ons

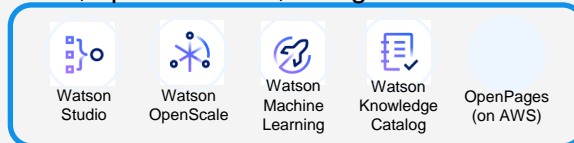
Customer Investments Outside CPD * partner



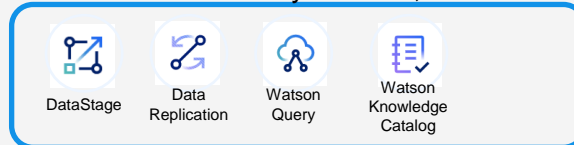
What can I do with these services?

Featured combinations of services to form mission-critical solutions for our clients

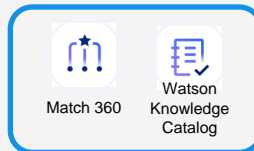
Build, operationalize, and govern AI



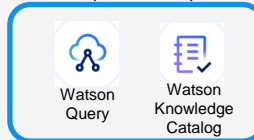
Provide access to all your data, without moving it



Create a 360-view of customer data



Share, enrich, and govern data



Unified orchestration of data preparation, transformation and AI activities in IBM Cloud Pak for Data

The screenshot displays the IBM Cloud Pak for Data interface, showing a workflow titled "Data Transformation Pipeline". The interface includes a top navigation bar with "IBM Cloud Pak for Data", a search bar, and user information. Below the navigation bar, there are several panels and a central workflow diagram.

Left Panel (Find palette nodes):

- Copy
- Copy assets
- Export assets
- Import assets
- Create
 - Create AutoAI experiment
 - Create batch deployment
 - Create data asset
 - Create deployment space
 - Create web service
- Wait
 - Wait for all results
 - Wait for any result
 - Wait for file
- Control

Central Workflow Diagram:

- Data asset operations in project or between projects:** Includes "Wait for Client data ..." and "Place assets in prop... Copy assets".
- Data Refinery for semi-automated data preparation:** Includes "Data Preparation Run Data Refinery flow".
- Classic DataStage Sequence Jobs orchestration:** Includes "Address Validation Run DataStage flow" and "Customer transact... Run DataStage flow".
- Control:** Includes "Both jobs complete Wait for all results", "Execution Fa...", "Success", and "Terminate pipeline Terminate pipeline".
- AutoAI Experiments:** Includes "Run AutoAI experi... Run AutoAI experiment".
- Data Science Notebooks:** Includes "Run notebook Run notebook" and "Post-processing Run DataStage flow".

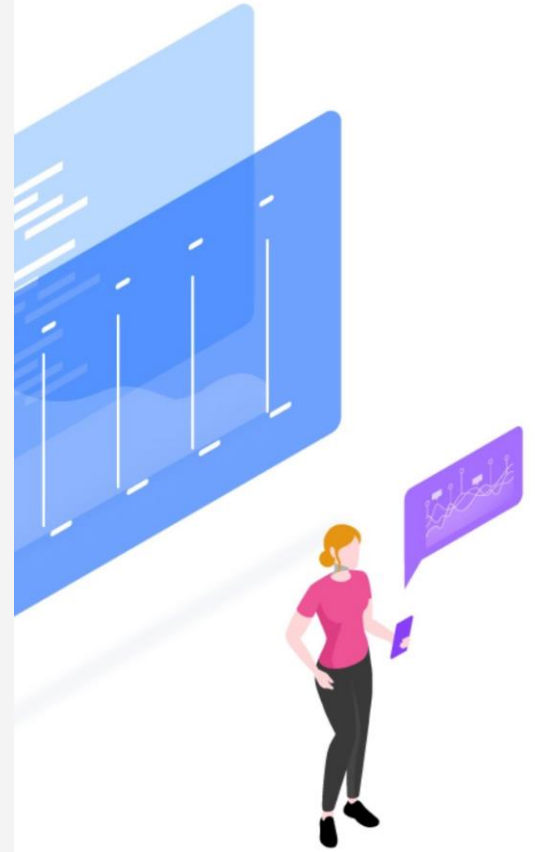
Right Panel (Control and Delete):

- Control:**
 - Loop in parallel
 - Loop in sequence
 - Set user variables
 - Terminate pipeline
- Update:**
 - Update AutoAI experiment
 - Update batch deployment
 - Update deployment space
 - Update web service
- Delete:**
 - Delete AutoAI experiment
 - Delete batch deployment
 - Delete deployment space
 - Delete web service
- Run:**
 - Run AutoAI experiment
 - Run Bash script
 - Run batch deployment
 - Run Data Refinery flow
 - Run DataStage flow
 - Run notebook

Introducing...

watsonx

Market evolution

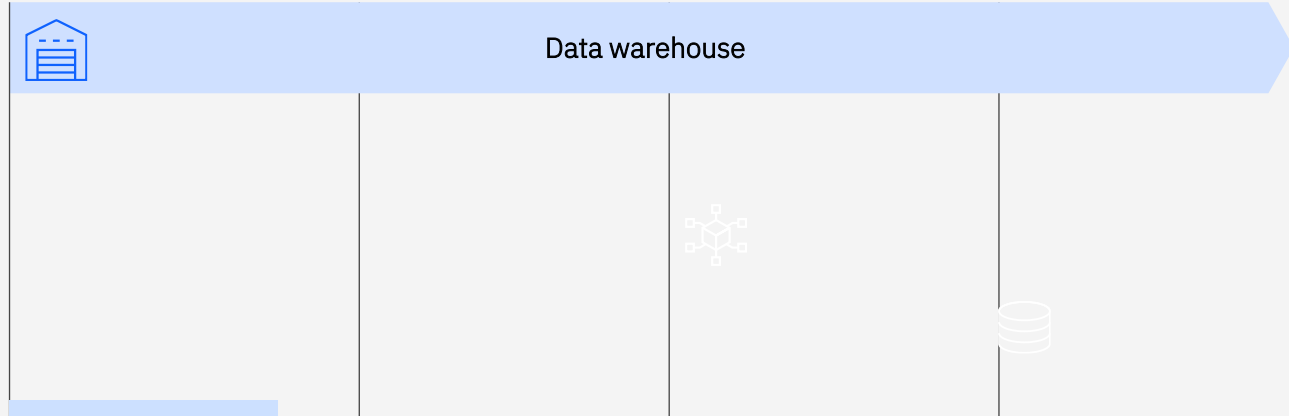


The data warehouse remains the center of analytics at most organizations

Late 1990s

Early 2000s

Present



High up-front costs

Structured data only

ETL required

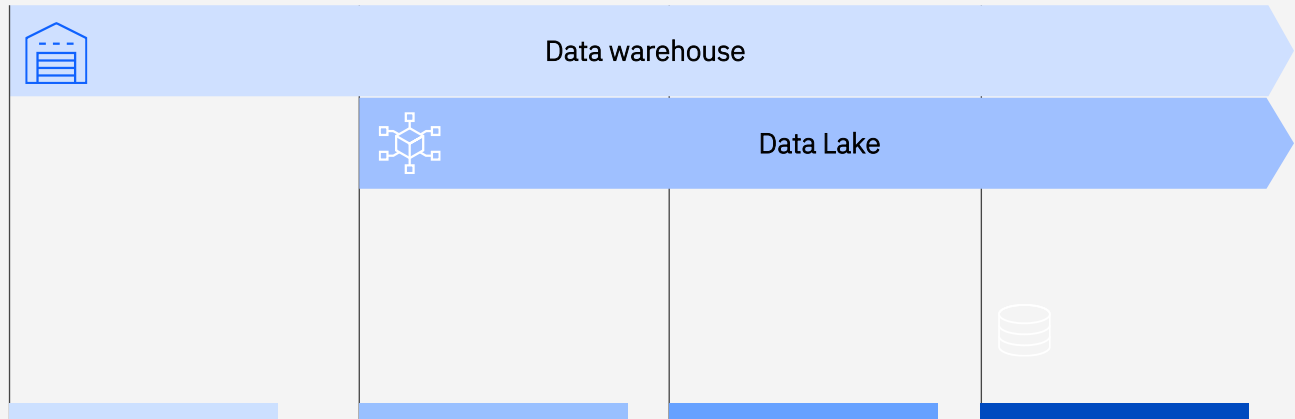
Vendor lock-in

Limited scalability

- Data warehouses emerged as the dominant method to analyze data
- Normalized and trusted data made it easy to analyze, however it's an expensive choice
- Warehouses technology has evolved continuously to improve... from appliance form factors to in-memory technologies

The emergence of the data lake

Late 1990s Early 2000s Present

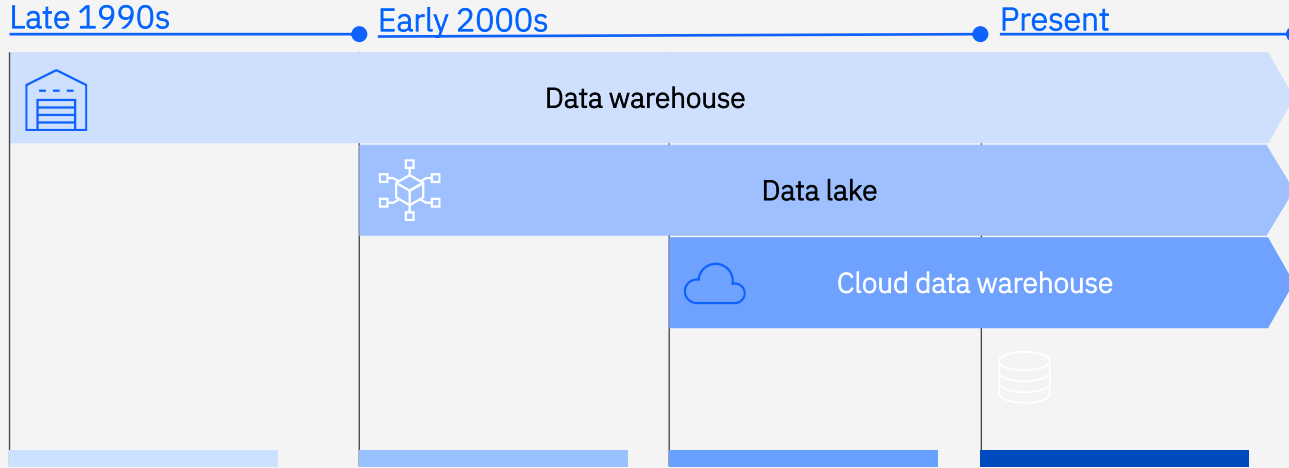


High up-front costs
Structured data only
ETL required
Vendor lock-in
Limited scalability

High complexity
Poor data quality
Limited performance
Expensive to maintain

- As volume, velocity, and variety of data grew, data lakes emerged as the new technology to replace data warehouses
- Data stored in raw and unstructured format = lower cost for large volumes of data
- Highly flexible and scalable
- Difficult to use & complex to maintain, and required a data scientist
- Ultimately, most data lakes failed and required a two-tier architecture

Cloud data warehouses evolved to address specific challenges of data warehouses



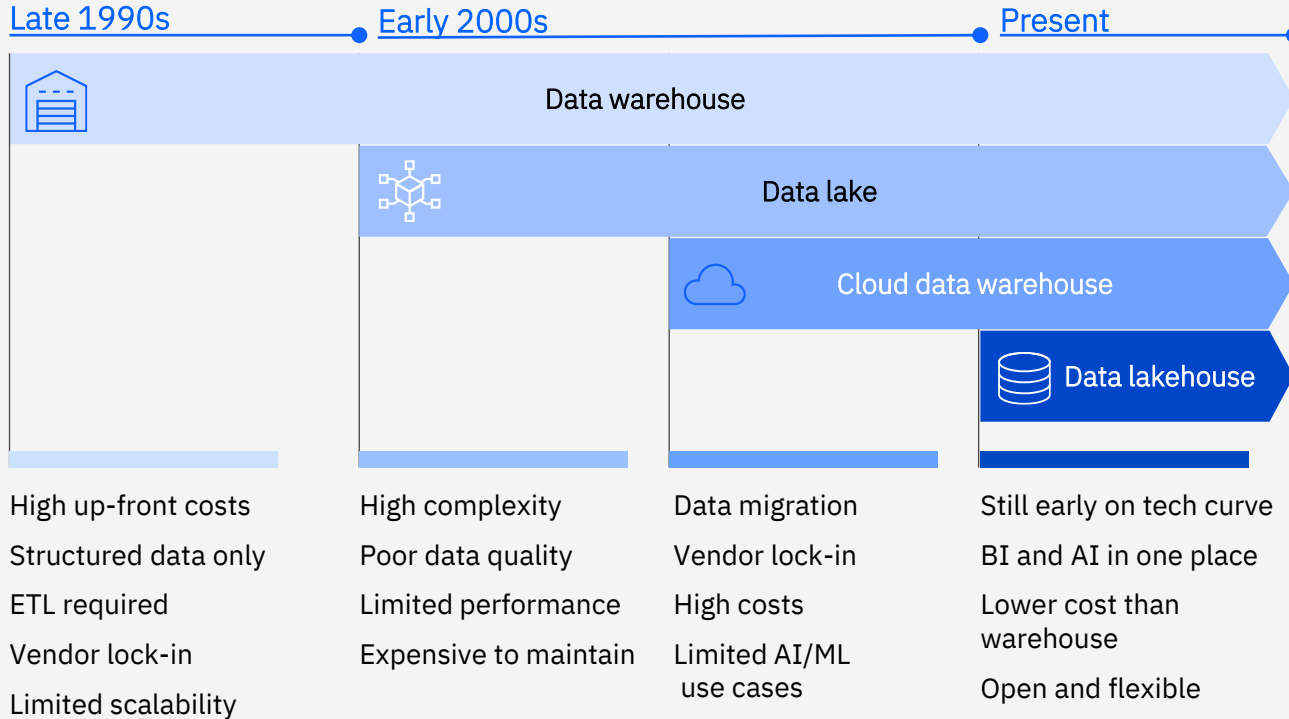
High up-front costs
Structured data only
ETL required
Vendor lock-in
Limited scalability

High complexity
Poor data quality
Limited performance
Expensive to maintain

Data migration
Vendor lock-in
High costs
Limited AI/ML use cases

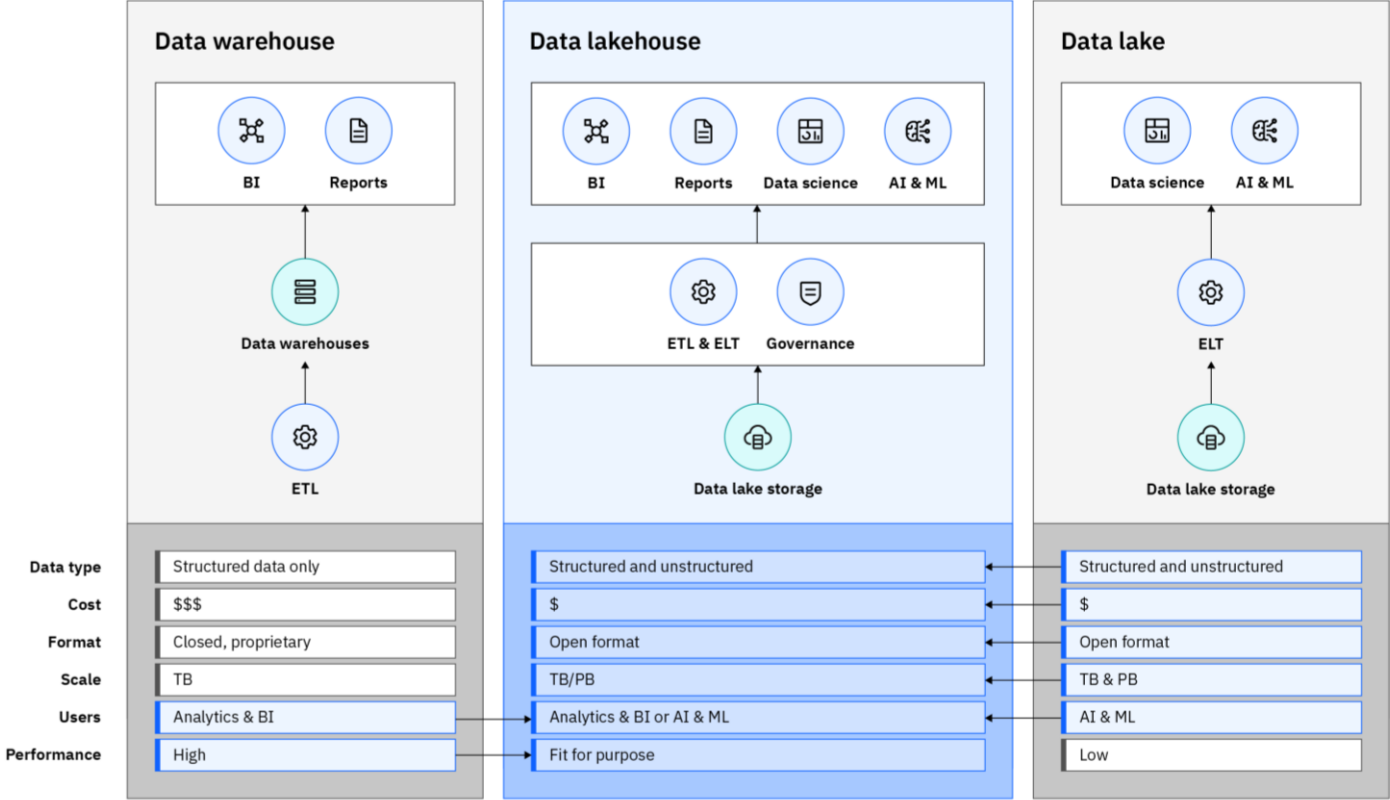
- Specifically, cloud data warehouses introduced the **separation of compute and storage**
- Addressed the scalability challenge with traditional warehouses – no data redistribution
- Ability to add more compute resources to the same data to solve the problem
- Easier to manage, however, much more expensive vs. on-prem warehouses

Data lakehouses are emerging technologies that solve for a new age of analytics



- Most enterprises today require two-tier architectures – both a data lake and multiple warehouses
- Data is moved and replicated from lake to warehouse, and the warehouse is still the access layer for key data
- Data lakehouses combine the best of warehouses and data lakes
- Data warehouse engine plus data lake storage

Lakehouses are meant to be a new class of data store that combines the best of data warehouses and data lakes



First generation lakehouses are still limited by their ability to address cost and complexity challenges:

- Single query engines set up to support limited workloads ... typically just BI or ML
- Typically deployed on cloud only with no support for multi-/hybrid-cloud deployments
- Minimal governance and metadata capabilities to deploy across the entire ecosystem

The platform
for AI and data

watsonx

Scale and
accelerate the
impact of AI with
trusted data.

watsonx.ai

Train, validate, tune
and deploy AI models

A next generation enterprise studio for AI builders to train, validate, tune, and deploy both traditional machine learning and new generative AI capabilities powered by foundation models. It enables you to build AI applications in a fraction of the time with a fraction of the data.

watsonx.data

Scale AI workloads, for
all your data, anywhere

Fit-for-purpose data store optimized for governed data and AI workloads, supported by querying, governance and open data formats to access and share data.

watsonx.governance

Enable responsible,
transparent and explainable
data and AI workflows

End-to-end toolkit encompassing both data and AI governance to enable responsible, transparent, and explainable AI workflows.

Scaling AI adoption with Ecosystem Partners



Embedding Watson Assistant and Watson Discovery into SAP applications to aid business users.



Embedding Watson Discovery and NLP libraries for Adobe Acrobat to help better process PDFs for subscribers.



With Watson Orders, McDonalds uses AI to automate drive thru order taking, enabling employees to increase focus on food delivery and customer service.



Watson Code Assistant and Red Hat Ansible bring the power of AI to IT Automation to transform the developer experience.



Embedding the Red Hat stack in support of GM's in-car applications. Enabling future embedded intelligent applications based on enhancements to Red Hat OpenShift for AI workloads.

+ AI → AI +

IBM