

Maximizing the Value of Data: Strategy for Modernization of Data & AI Platform

—

Kitman Cheung
CTO, IBM Data & AI – Asia Pacific

Overly Simplistic Theory of The Value of Data...

$$~~V_D = D~~$$

V_D = Value of Data

D = Data

Overly Simplistic Theory of The Value of Data...

$$\cancel{V_D = D}$$

$$D = I + N$$

D = Data

N = Noise

I = Information

Overly Simplistic Theory of The Value of Data...

$$V_D \propto I$$

$$V_D \propto \frac{1}{T}$$

V_D = Value of Data

T = Time to process data

Overly Simplistic Theory of The Value of Data...

$$V_D \propto \frac{I}{T}$$

$$T = \frac{D}{r_p * r_{hw}}$$

r_p = Rate of work
by people

r_{hw} = Rate of work
by hardware

Overly Simplistic Theory of The Value of Data...

$$V_D \propto \frac{I}{D} * r_p * r_{hw}$$

Overly Simplistic Theory of The Value of Data...

$$V_D \propto \frac{I}{D} * r_p * r_{hw} * A$$

V_D = Value of Data

r_p = Rate of Work by People

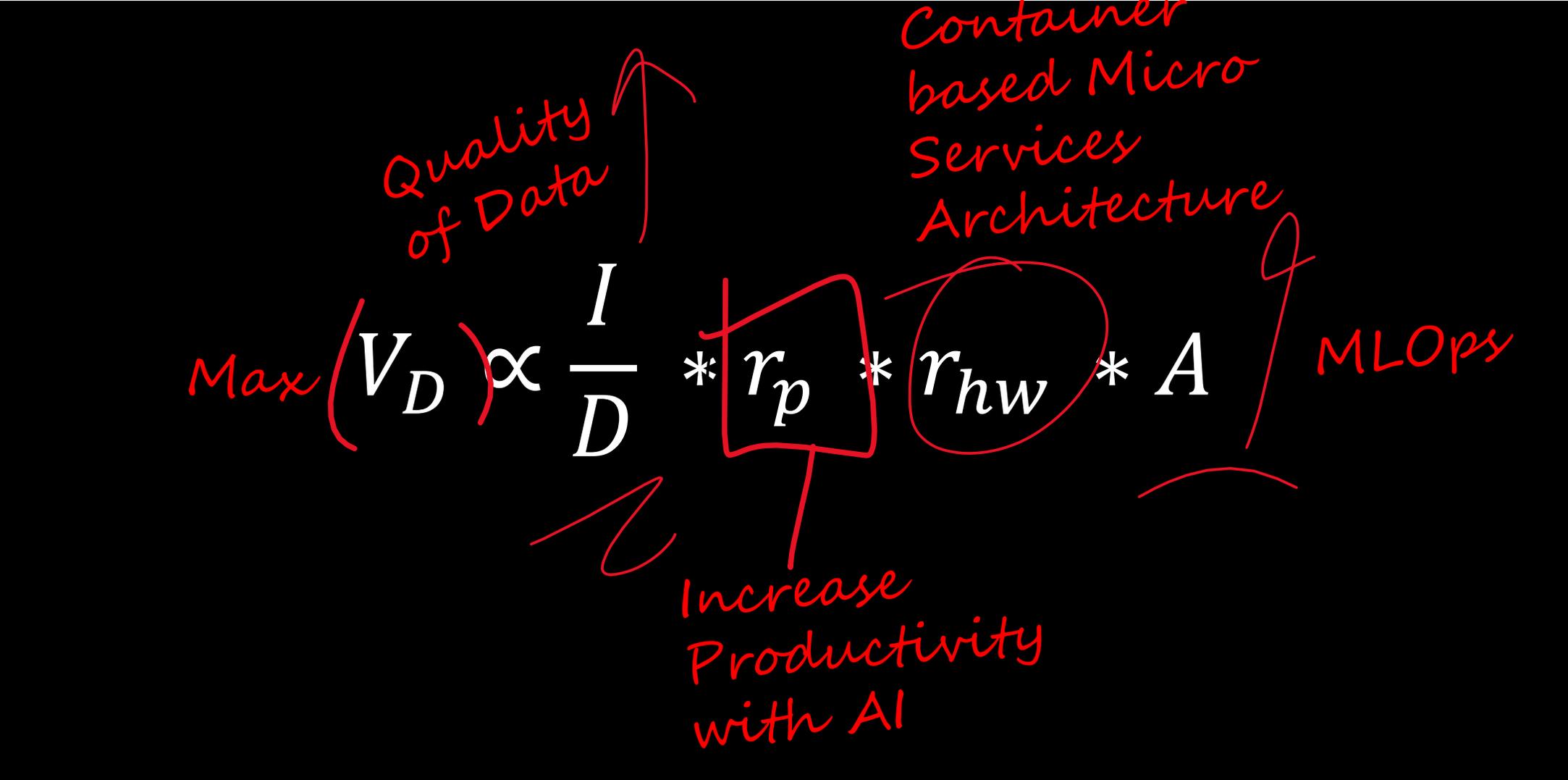
I = Information

r_{hw} = Rate of Work by HW

D = Data = (I + N)

A = Action taken on Insight

Overly Simplistic Theory of The Value of Data...



Evolving to address the challenges

Data Warehouse

Proliferation of data silos and need for enterprise insight, led to data warehouse.
Design principle: focus on KPI, known matrix.

Hadoop Data Lake

Growing digital footprint and web scale analytics, demanded a low cost, scalable infrastructure.
Design principle: focus on ad-hoc analysis and batch processing for large scale data.

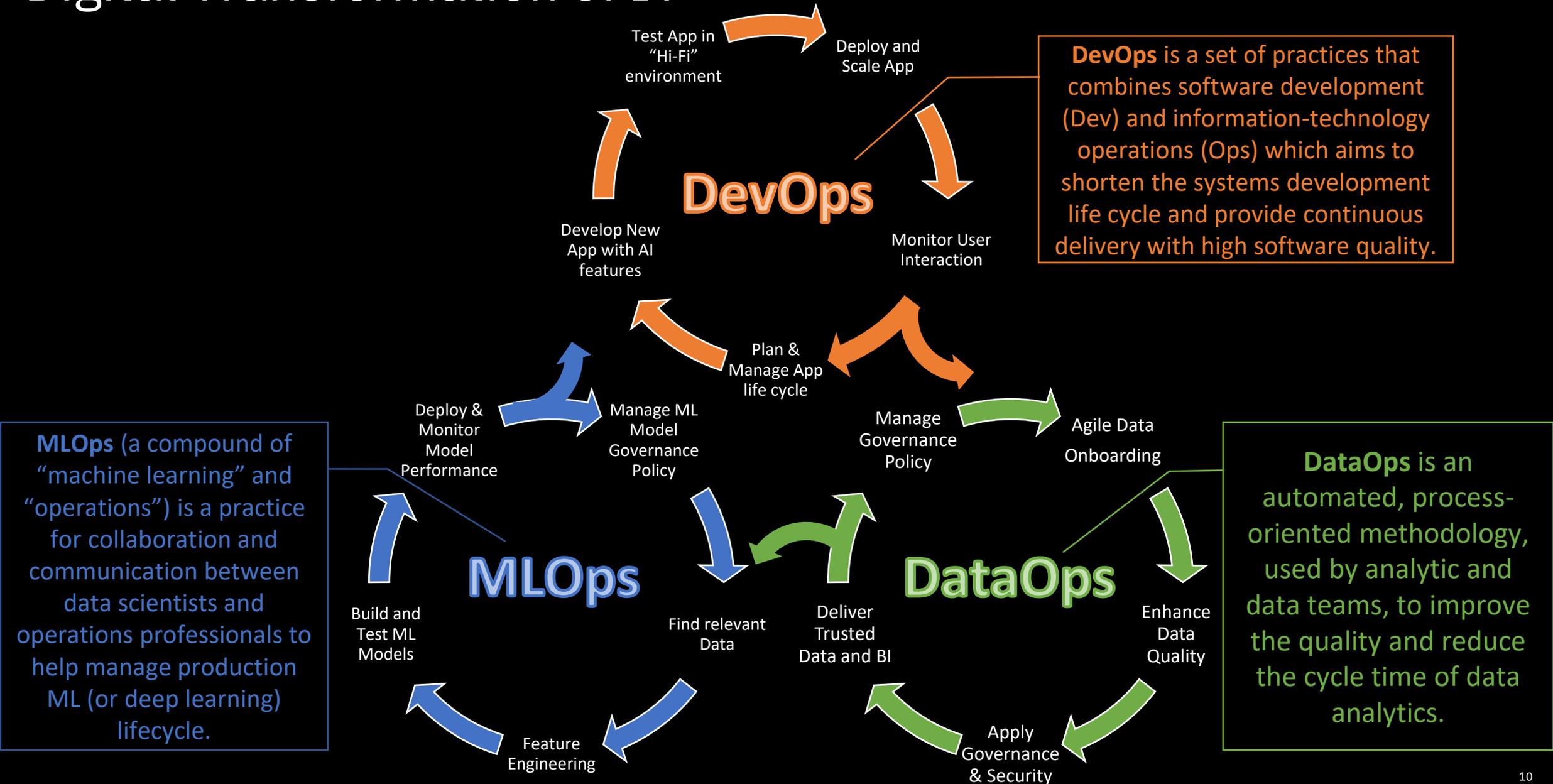
Cloud Data Lake (Object Storage + Spark)

Growth of cloud and AI experimentation led requirement for elastic compute environment.
Design Principle: isolation of compute and storage.

A.I. Data Lake

Automation for Data Delivery(DataOps), focus on ML/AI workload (MLOps), hybrid cloud
Design Principle: Microservices, AI enabled, business outcome.

Digital Transformation of IT



Enterprise Platform for DataOps

Enabling velocity, scalability, and traceability



Data Architects



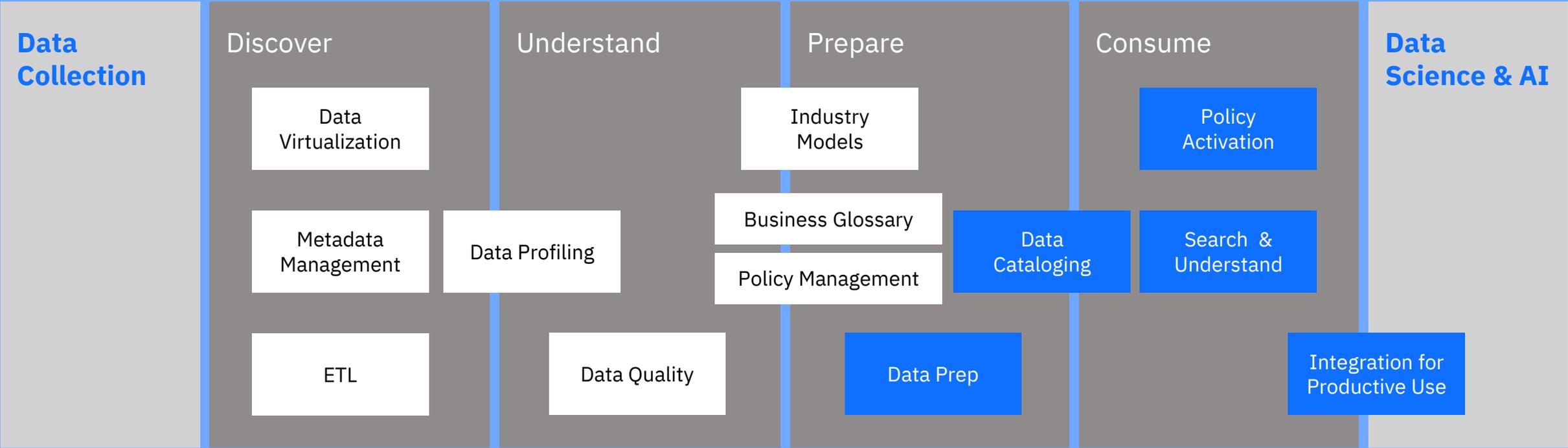
Data Engineers



Data Stewards



Knowledge Worker

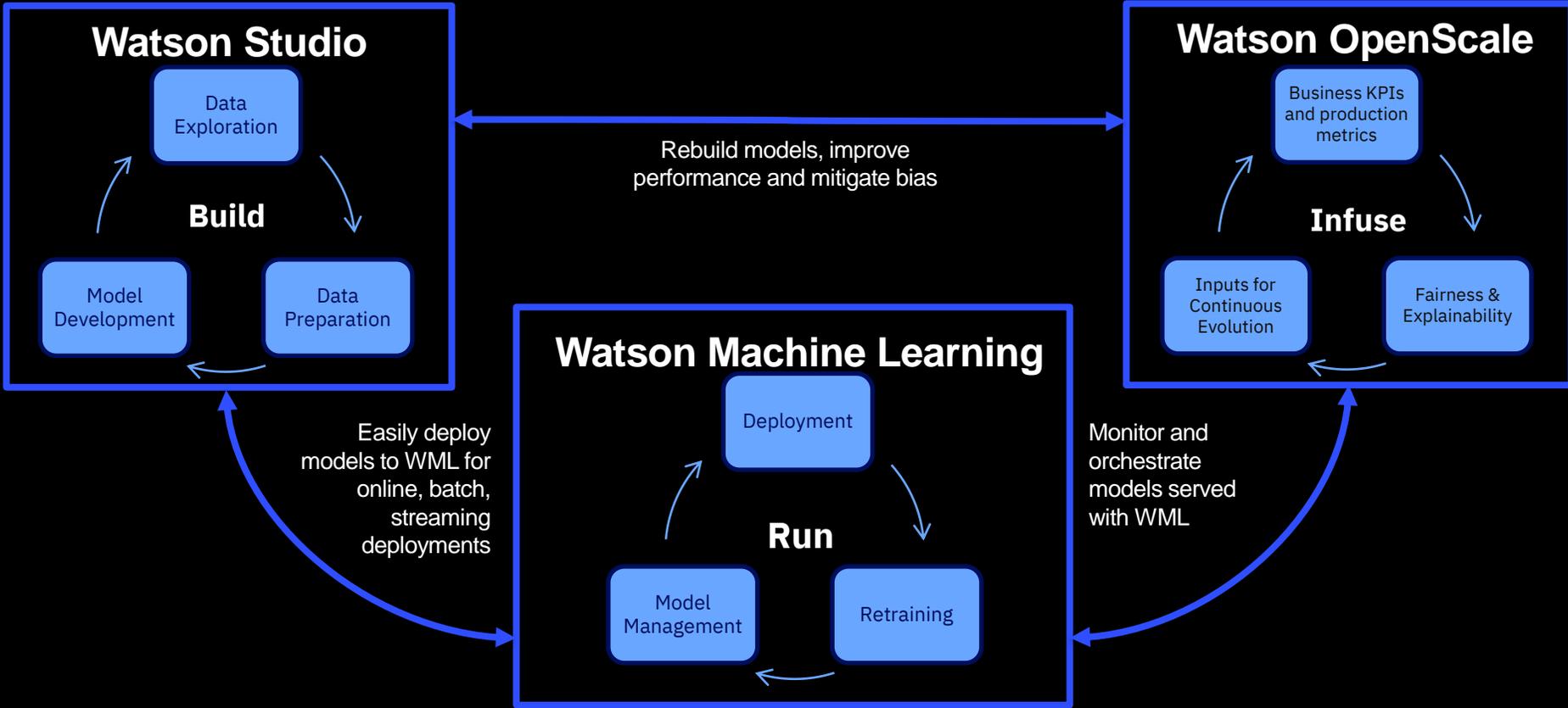


Data Monetization

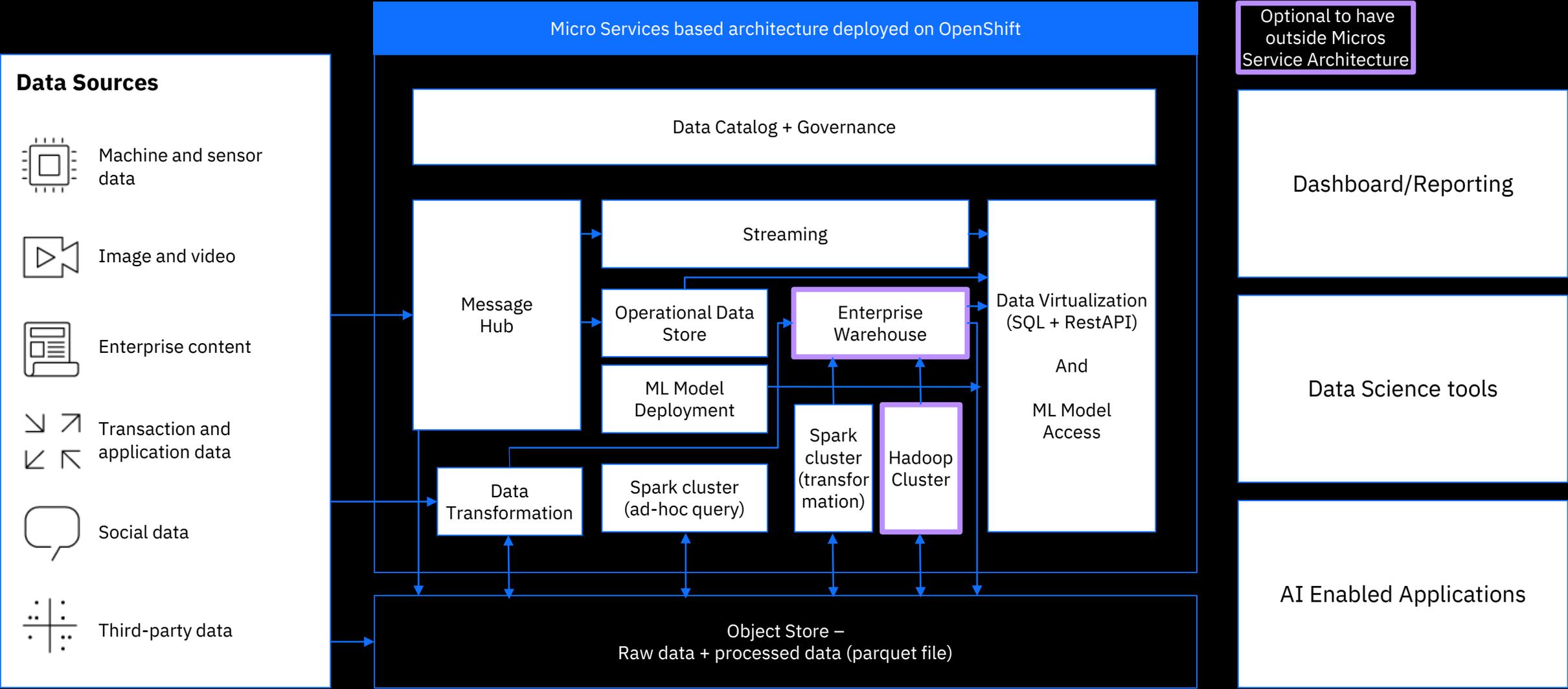


Data Science Platform for MLOps

Delivering Trust, Bias-free ML Models across the Enterprise

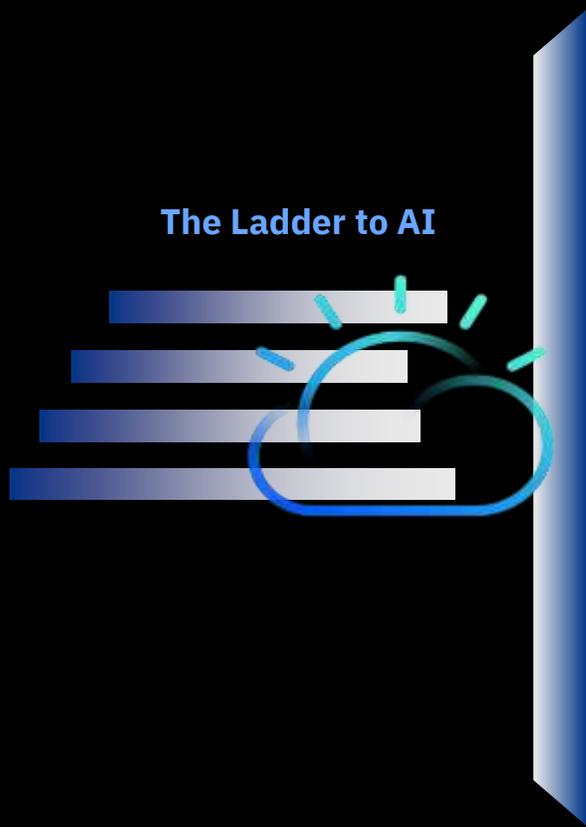


Reference Architecture for A.I. Data Lake



Cloud Pak for Data

Foundational “out of the box” multicloud data & AI services



Collect

- Data virtualization
- Data warehousing
- Databases on-demand
- Data source ingestion
- Distributed processing

Powered by: Db2 and Db2 Warehouse technologies

Organize

- Discovery & search
- Data transformation
- Data cataloging
- Business glossary
- Policies, rules & privacy

Powered by: Infosphere, Data Stage and IGC/WKC

Analyze

- Data visualization
- AutoAI
- Machine learning tools
- Model build & deploy
- Model management
- Dashboards

Powered by: Watson Studio open source and Cognos

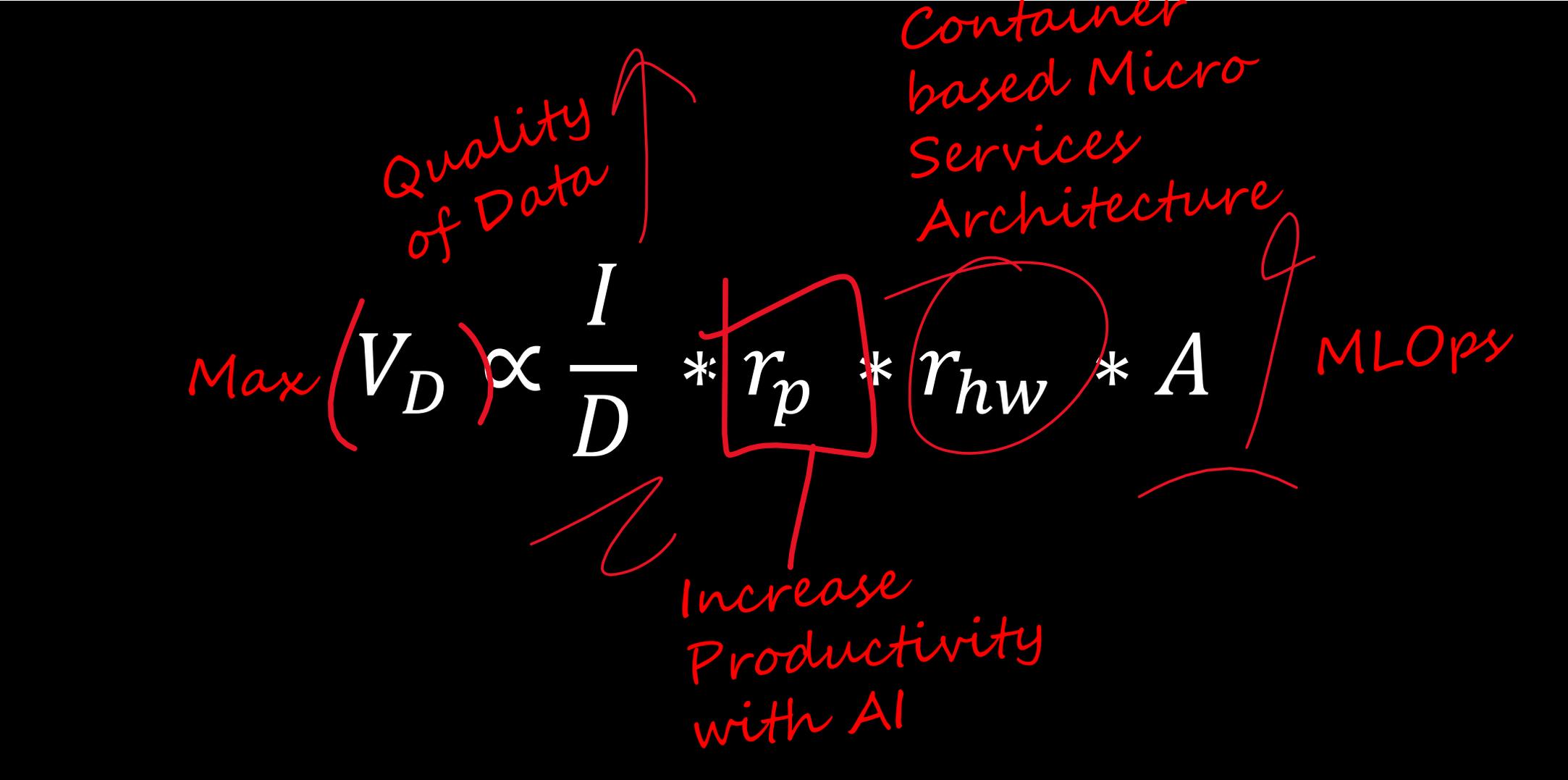
Multicloud Services	<ul style="list-style-type: none">• Logging• Monitoring	<ul style="list-style-type: none">• Metering• Persistent Storage	<ul style="list-style-type: none">• Kubernetes• Security	<ul style="list-style-type: none">• Identity Access Mgmt.• Docker Registry / Helm
----------------------------	--	---	---	--

Cloud Pak for Data



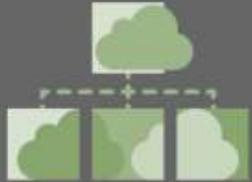
Available as Software on Any Cloud (Private, Public or Hybrid)

Overly Simplistic Theory of The Value of Data...



Forrester Study: The Projected Total Economic Impact™ Of IBM Cloud Pak For Data

Key Benefits



Reduced infrastructure management effort:
65% to 85%



Reduced ETL requests:
25% to 65%



Data science, ML, and AI benefits:
\$1.2 million to \$3.4 million

“Quantified projected benefits. The following benefits reflect the financial analysis associated with the composite organization.

› **Containers and container management efficiencies totaling \$12.5 million to \$14.4 million.** With Cloud Pak for Data, companies can improve their readiness for cloud migration, improve licensing flexibility with IBM, and reduce both hardware purchases and infrastructure management efforts.

› **Data virtualization benefits totaling \$932,569 to \$2.4 million.** Data virtualization “democratizes” data accessibility across the organization, improves data governance and security, and can allow companies to avoid costly data migration projects.

› **Data science, ML, and AI benefits totaling \$1.2 million to \$3.4 million.** Cloud Pak for Data helps data scientists improve productivity through faster model development and deployment. Additionally, due to Cloud Pak for Data’s integrated platform, companies avoided costs associated with legacy analytics tools or otherwise building comparable integrated data management and science solutions internally.”

<https://www.ibm.com/downloads/cas/V5GNQKGE>



INDONESIA NATIONAL STRATEGY FRAMEWORK FOR AI 2020-2045

MISSIONS

Creating an ethical Artificial Intelligence in accordance with the values of Pancasila.

Preparing Artificial Intelligence Talents that are competitive and have character.

Creating a data ecosystem and infrastructure that supports the contribution of Artificial Intelligence for the benefit of the country.

Cultivate an ecosystem of research collaboration and artificial intelligence innovations to accelerate bureaucratic reform and a superior national industry.

AI FOR INDONESIA'S GRAND 2045 VISION

5 PRIORITY AREAS



HEALTHCARE



BUREAUCRATIC REFORM



EDUCATION & RESEARCH



FOOD SECURITY



MOBILITY/ SMART CITY



RESEARCH & INDUSTRIAL INNOVATION



DATA & INFRASTRUCTURE



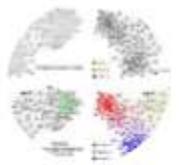
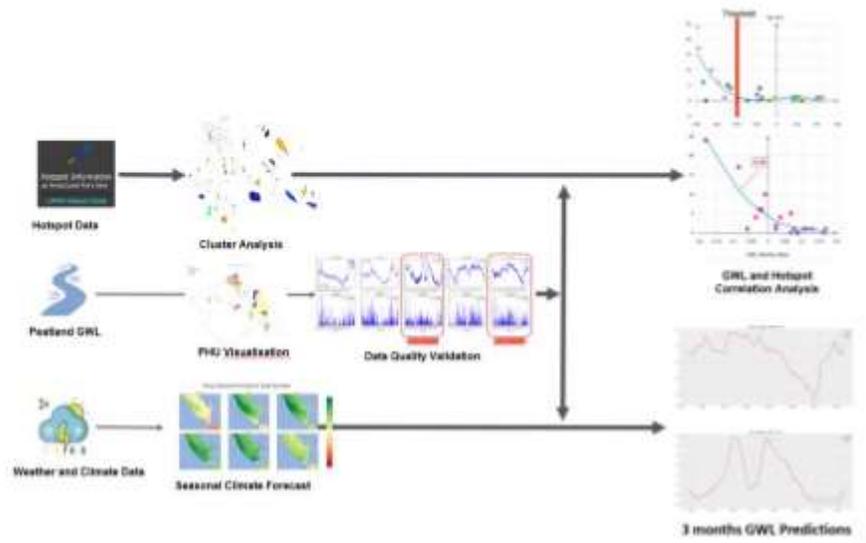
AI TALENT DEVELOPMENT



ETHICS & POLICY

AI INNOVATION FOR FOREST FIRES PREVENTION

BPPT PROGRAM FOR AI AND BIG DATA



Clustering result can give the weather modification team a mapping of the potential hotspot area

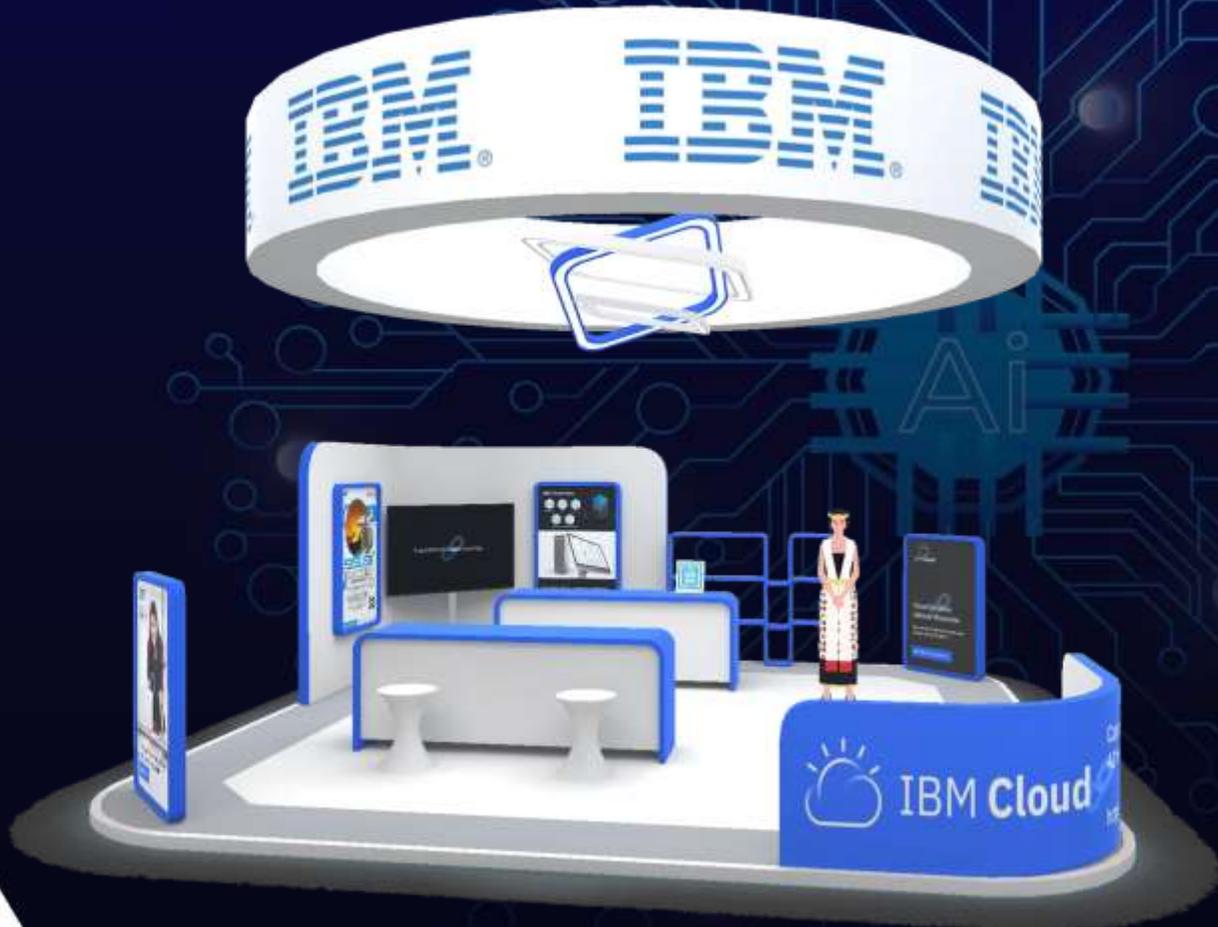


Result from Ground Water Level (GWL) and Hotspot Correlation Analysis can give the weather modification team alert by giving them number of minimum GWL in an area that will trigger a hotspot



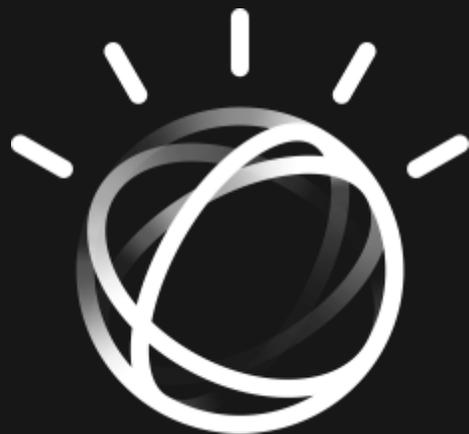
GWL prediction can help the weather modification team to plan an effective way to prevent forest fires





Updated daily!

- Products information
- Live chat
- FREE Training and certification
- Startups Program



Maximizing the Value of Data: Strategy for Modernization of Data & AI Platform

—
BPPT AI Summit 2020 – 10th-13th November 2020

Kitman Cheung
CTO, IBM Data & AI – Asia Pacific