

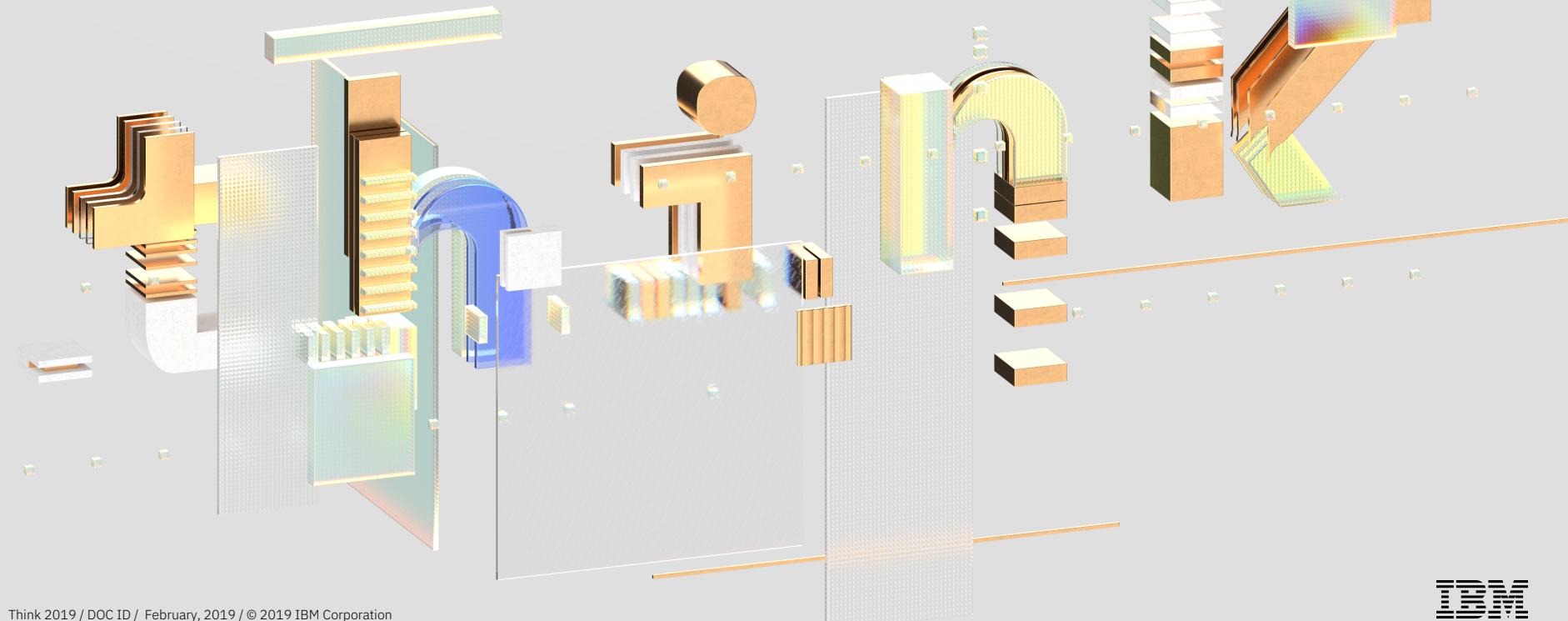
Data Integration and Governance in a Multicloud World

think 2019

—
Jo Ramos

DE & Director of Development – Data Governance and Integration

jaramos@us.ibm.com



Please note

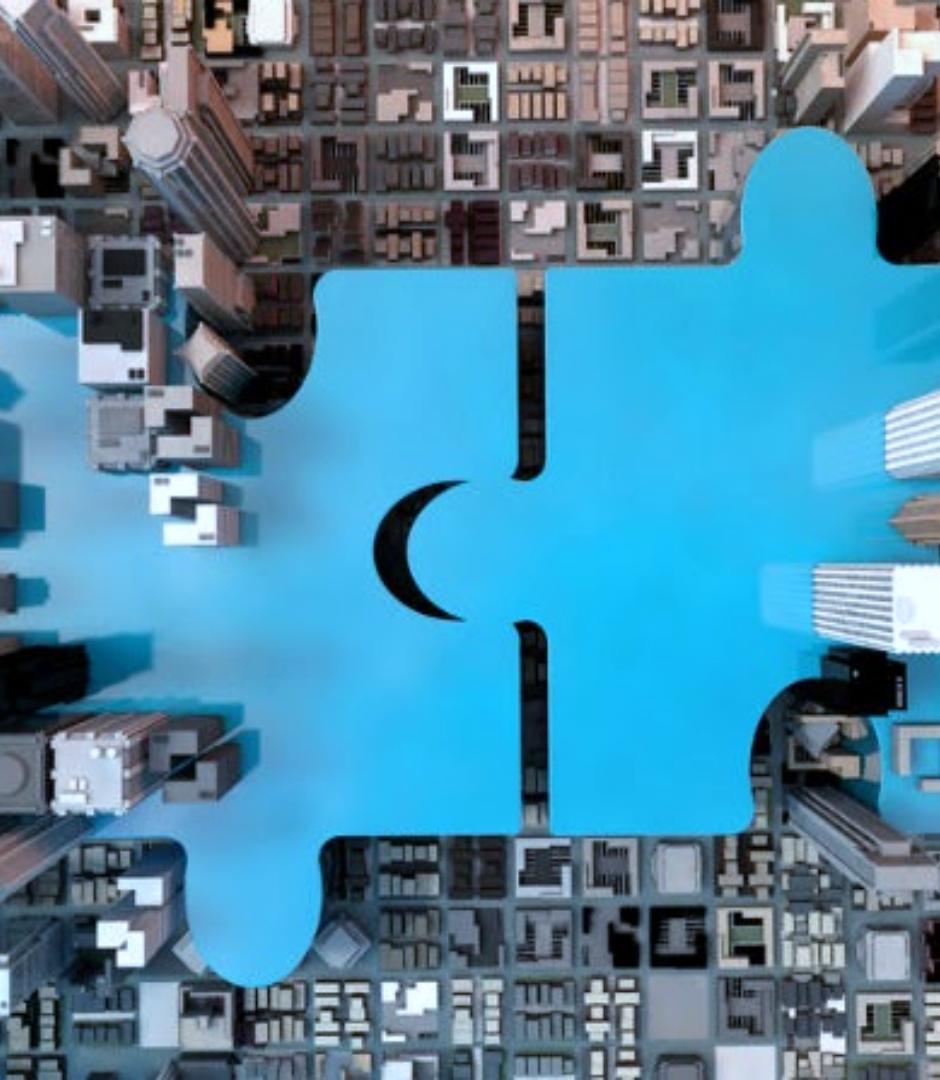
IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice and at IBM's sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.



- Supermarket chain based in Spain
 - Acquired a U.S. grocery chain
 - Wants to run promos based on customer purchase habits
- Mainframe data on-premises
- Private cloud based on data center in EU



Challenges

- Acquisition introduces introducing **another cloud platform and data**
- They have tens of **thousands** of data integration **processes**
- They have constraints
 - **Cost for moving data** out of cloud platforms
 - Processing data as close as possible from their data sources
 - **Regulatory** constraints (GDPR)

Drivers

*More
than just
Big
Data...*

Digital Transformation

89%

Share of organizations have adopted or have plans to adopt a “digital-first” strategy

62%

CEOs/Sr.Execs have a management initiative / transformation program to make business more digital

Hybrid, Multicloud Adoption

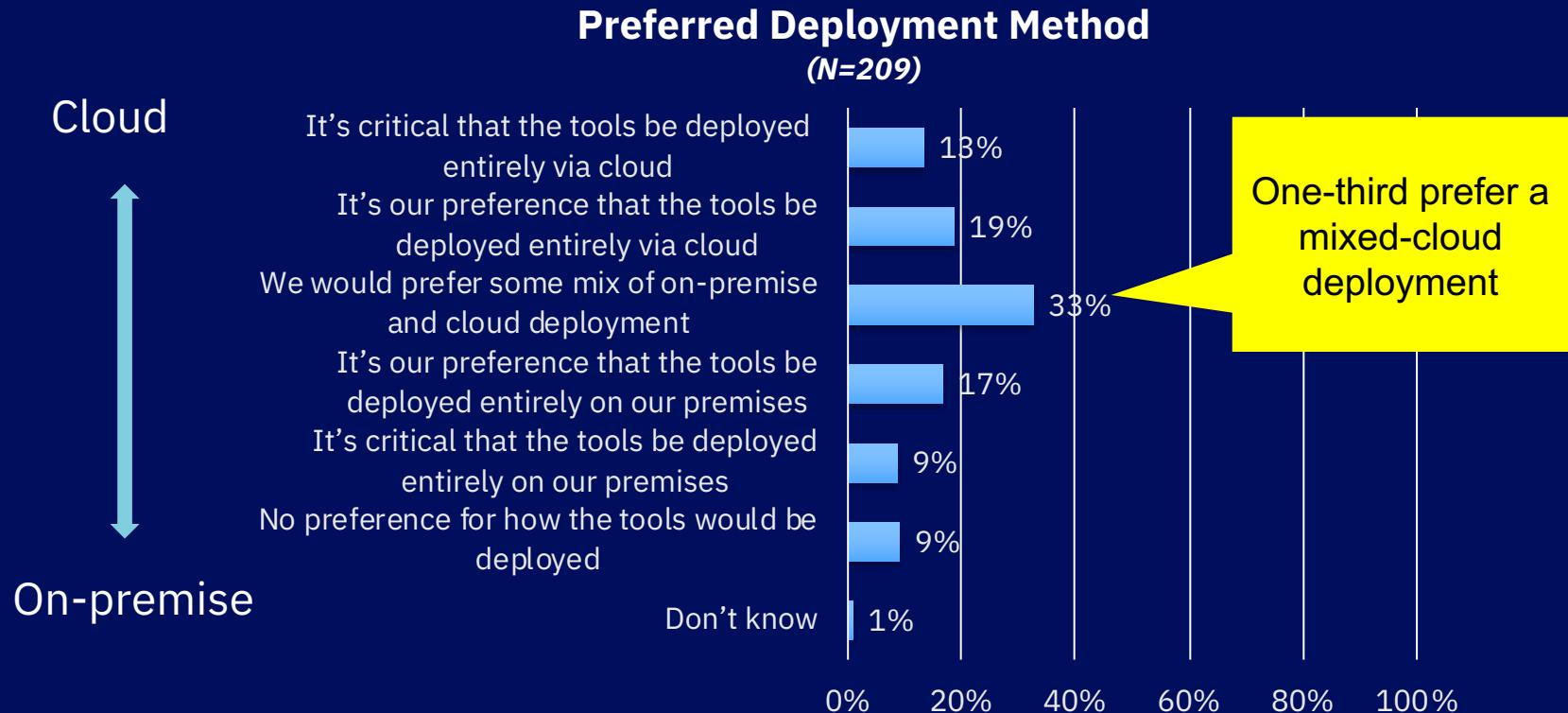
94%

Share of enterprise clients using multiple clouds

67%

Share of enterprise clients using more than one public cloud provider

Preferred Deployment is a mix of on-premise and cloud



Which of the following best describes how your company would prefer that the Data Governance and Integration Platform be deployed?

Problems we hear about...

- **Adapting quickly** to **rapidly changing information landscape** while leveraging existing investment
- **Increasing complexity** of data ecosystems
 - Accessing sources and targets across on-premises and multiple clouds
 - Processing workloads in close proximity to where your data resides
- Modern enterprises demand **real-time insights from multi-structured data**.....from **internal and external sources**
- Opportunity cost of **delayed insights**

How are those being addressed today?

- Hiring third-parties or consultants
- “Stitching” together multiple tools
- Hand-coding
-do nothing!

Resulting in....

- High costs
- Delayed implementations and slow time-to-value
- Inflexibility and sub-optimal performance
- ...competitive disadvantage

What if you had an Integration Service that...





**... Incorporates
new data
sources in
minutes**

**... Optimizes
data flows for
performance,
cost, SLA –
You decide!**

**... Seamlessly
expands to
adapt new
sources and
targets**

**... Automates
multicloud and
hybrid cloud
execution**

**Autonomous, self learning
systems** that

- plans your data route (flow) to avoid congestion,
- finds the best route and “vehicle” for your “arrival” goal or budget
- Seamlessly adapts to changing technology in either your data ecosystem or for route execution

IBM's Vision

- **Smart and optimized data flows**

- Processing in close proximity to your data
- Distribute processing to multiple clouds or on-prem
- Optimize based on multiple criteria

- **Polyglot Execution Engines**

- Spark, IBM PX, IBM Streams, replication, Hadoop, etc..

- **Smart Data Flow Designer**

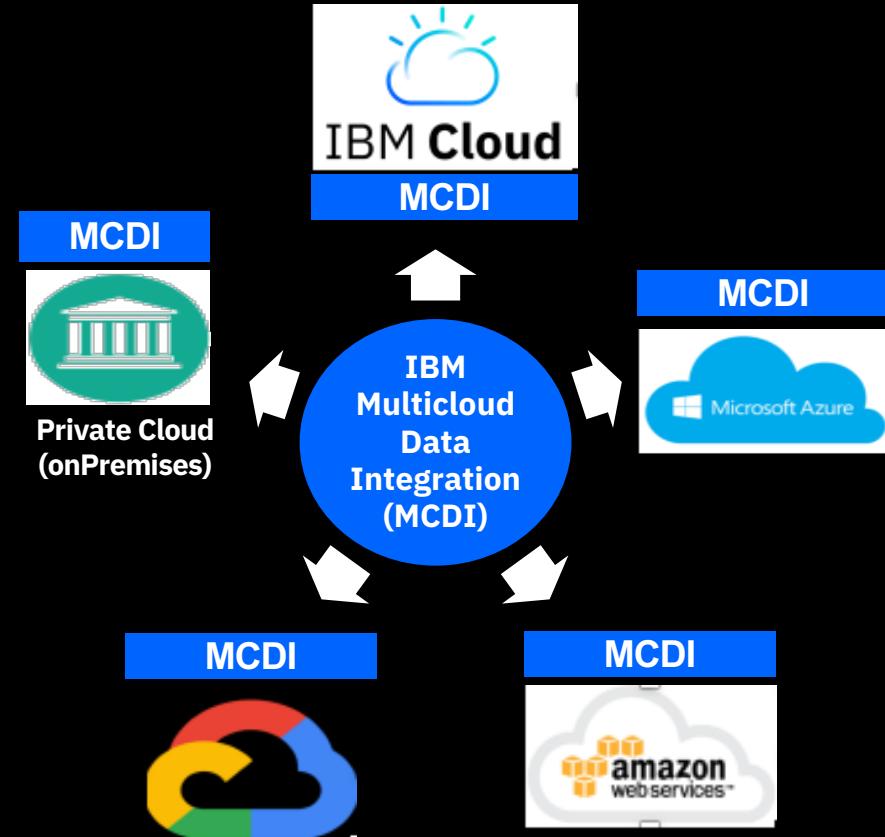
- Intent driven, autonomous and self-learning
- Smart data movement
- Smart data replication
- Smart ETL (transformations)
- Batch, Near Real-Time and Real-Time

- **Governance infused** (data privacy and protection)

- **SaaS and PaaS**

- **Modern Architecture**

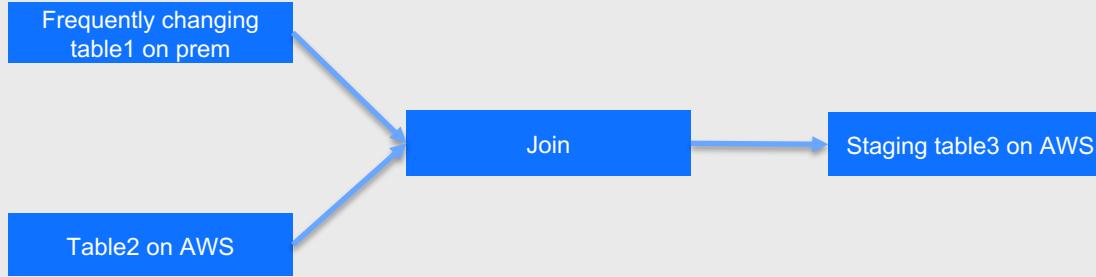
- Cloud native micro-services architecture



Microsoft, Microsoft Azure are trademarks of Microsoft Corporation in the United States, other countries, or both.
"Amazon Web Services, the "Powered by AWS" logo, are trademarks of Amazon.com, Inc. or its affiliates in the United States
and/or other countries."

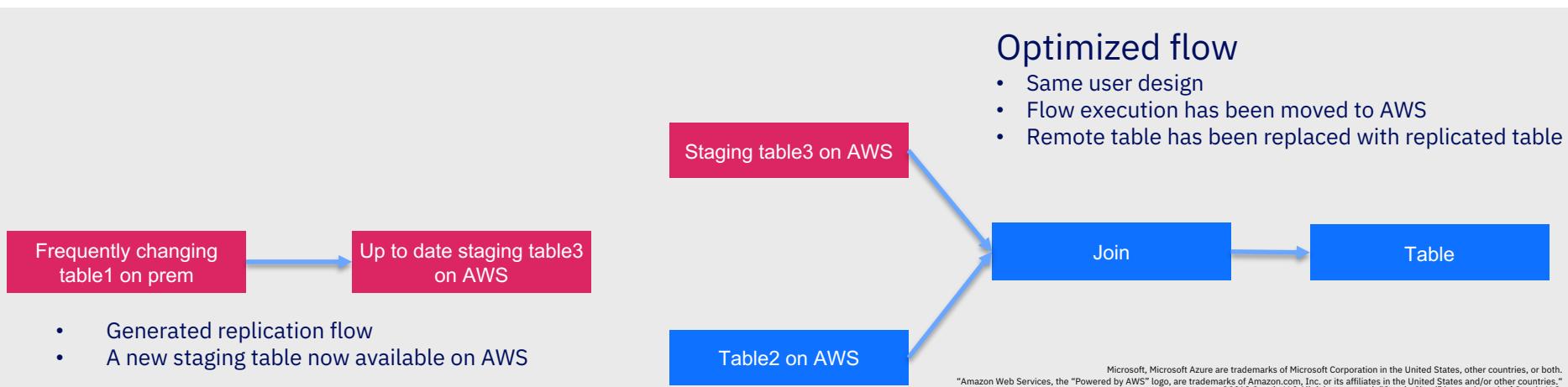
©2018 Google LLC All rights reserved. "Google Cloud" is a trademark of Google LLC.

Multi-cloud Data Integration Topology

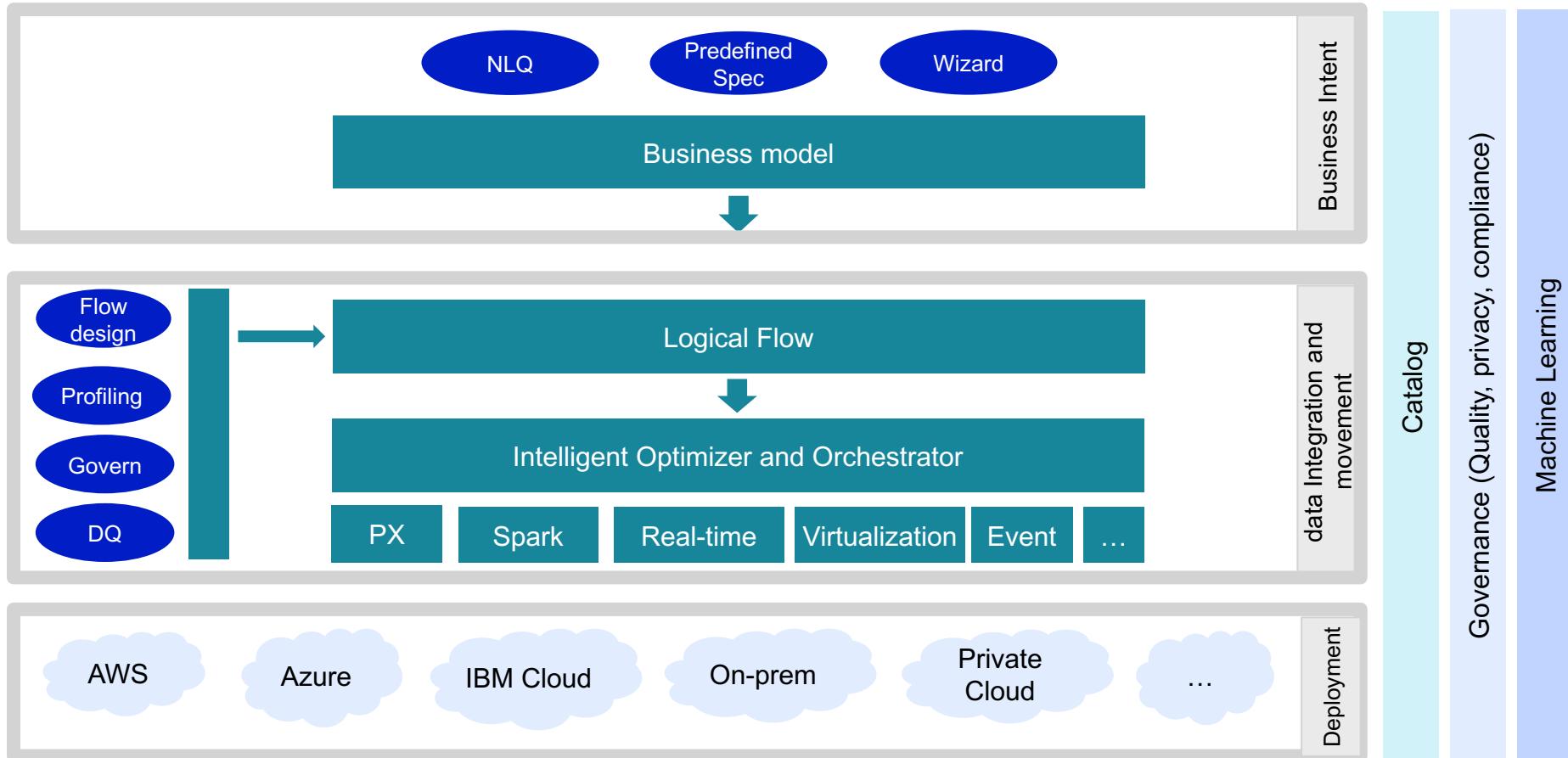


Today

- Runs on-prem
- Move data out of AWS for each job execution



Multi-cloud Data Integration Architecture



Intelligent Optimization

Optimize for

Cost and/or SLA

- Maximize the transformation close to the data
- Minimize the amount of data extracted from a cloud platform
- Replicate the data to a staging area when multiple jobs access the same remote data
- Optimize resource utilization based on SLA vs cost

Data volume

- Use virtualization when more efficient than moving data

Entitlement

- Optimize which runtime to use based on your entitlement

Location

- Make optimization decisions based on the location of the data

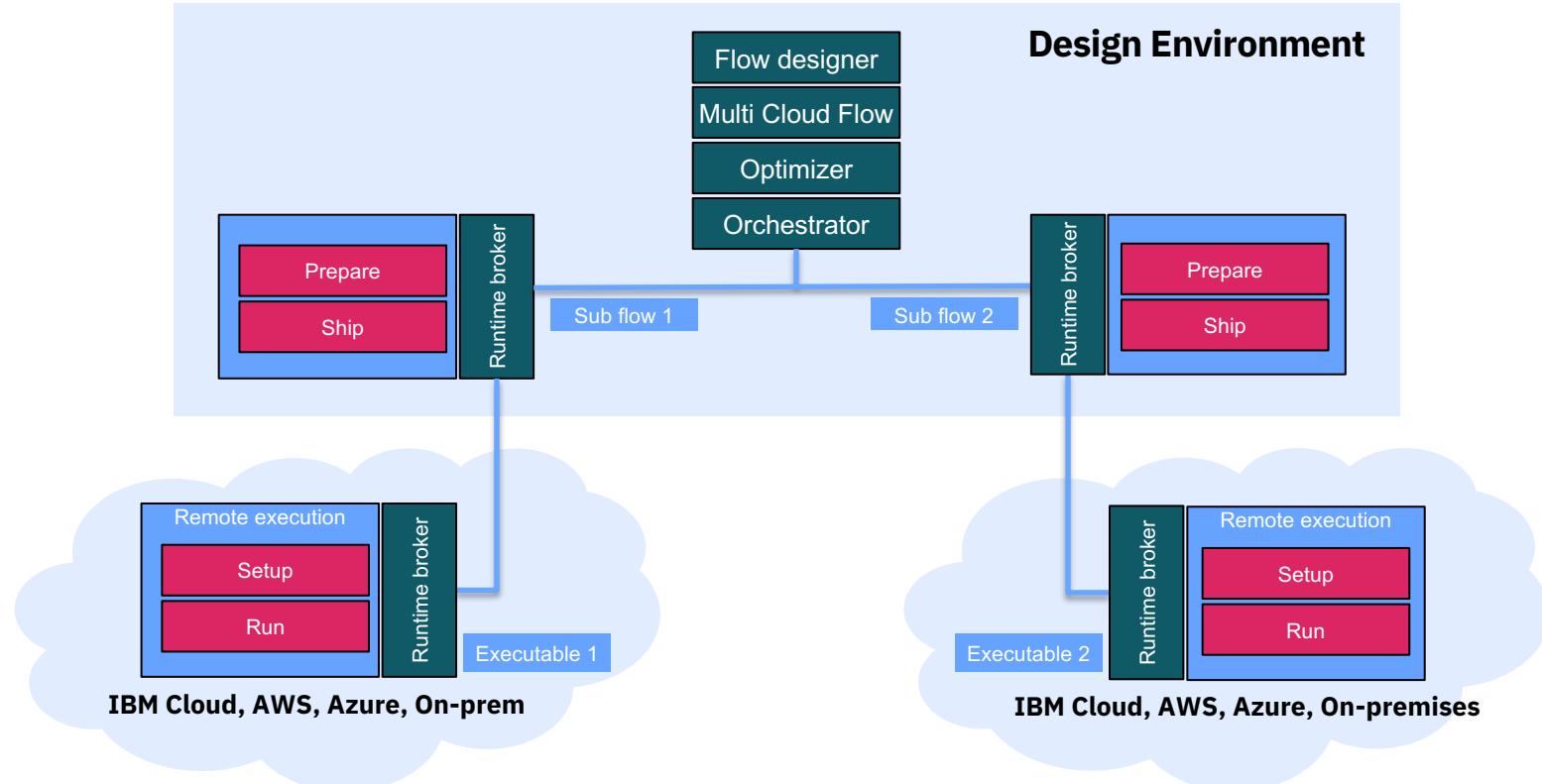
Time and frequency

- If allowed, optimize job execution schedule to optimize idle resources
- Optimize for a collection of jobs based on collective data sources access

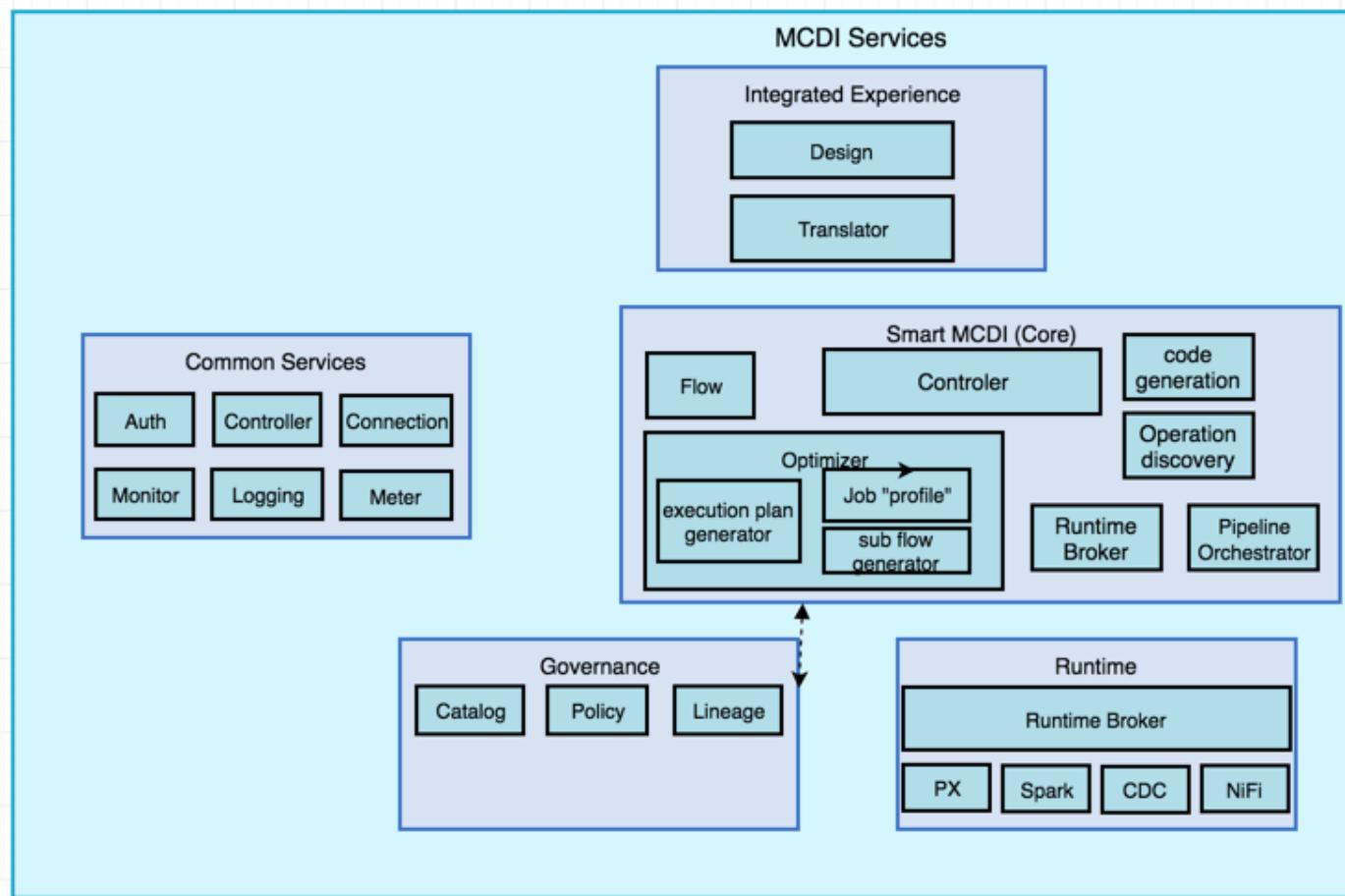
Infuse governance

- Keep data where it is if we can't move it (e.g. GDPR)
- Enforce governance policies on data (automatic enforcement)

Multi-cloud Data Integration Topology



Multi-cloud Data Integration Topology



Key differentiators

Embedded Governance



- Enforce data quality and policies

Optimize for your needs



- Cost, SLA, etc.
- Distribute across multiple clouds and runtimes
- Embedded machine learning

Operational Controls



- Single view of operational metadata- logs, monitoring, performance

With intelligent multicloud data integration, IBM will...

Equip **Data engineers and Knowledge workers** with **smart design and data auto pilot** that lets **users focus on their data goals** rather than spending time to determine technology or how data sources are accessed allowing Data-driven enterprise to **derive insights faster**

Notices and disclaimers

© 2018 International Business Machines Corporation. No part of this document may be reproduced or transmitted in any form without written permission from IBM.

U.S. Government Users Restricted Rights – use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed “as is” without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.

IBM products are manufactured from new parts or new and used parts. In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply.”

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer’s responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer’s business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

Notices and disclaimers continued

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products.

Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

IBM, the IBM logo, ibm.com and [names of other referenced IBM products and services used in the presentation] are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.

