Red Hat

Microsoft

DATAOPS AND DEVOPS

MAXIMIZE THE INTEGRATION DEVELOPMENT AND DATA

JAN CORDTZ, SNR. CLOUD SOLUTION ARCHITECT, MICROSOFT DENMARK

MICHAEL BANG, SNR. SOLUTION ARCHITECT, RED HAT

INTRODUCTION: THE PROBLEM

- Application Teams are applying modern DevOps principles to deliver business value quicker
- Data teams are being left behind
- The modern data platform is just as sophisticated as the application platform
- They don't need to be different, the same DevOps principles we've been using for the application platform can be used for the data platform
- Data teams can, and should, be 1-to-1 working with the application teams

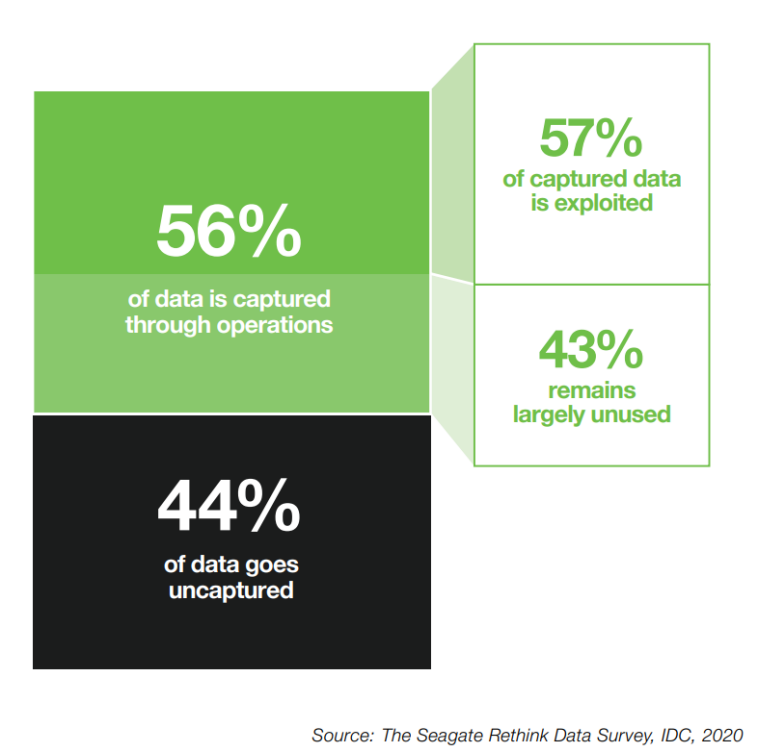
APPLICATION TEAMS ARE MODERNIZING WITH KUBERNETES

Application Developer

- The Data Engineers aren't necessarily going on the same journey

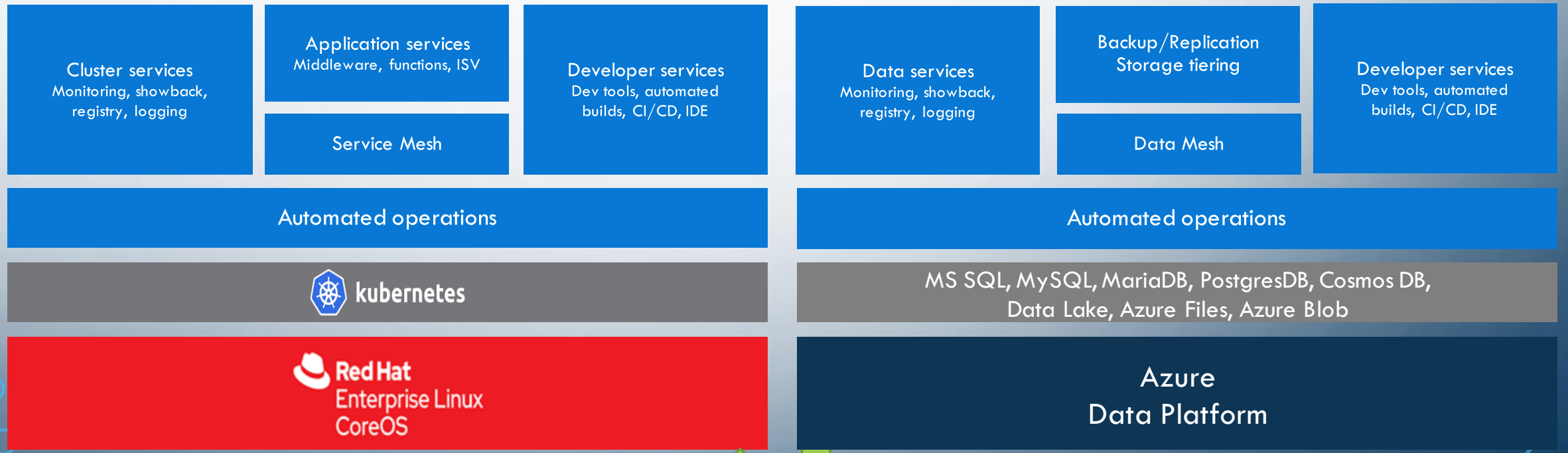
APPLICATIONS AND DATA

- Applications are the main driver for capturing data
- Data needs to be used – ”to be data driven”
- New disciplines occurs like IoT, AI and ML
- So, the application landscape will continue to change
- And Data Platforms needs to be modernized too



- Application and data development must go “hand in hand”

THE IDEAL SOLUTION COULD BE; AZURE RED HAT OPENSHIFT AND DATA PLATFORM



Let Red Hat and Microsoft …

Manage all your clusters/databases

Secure your nodes

Monitor and operate your VMs

Manage environment patches

Backup/recovery

Defend your environment

TOPICS WE WILL DISCUSS

Topic 1 : Dynamic provisioning on-demand

The capability of being able to provision containers on demand and providing the dataset needed for the container to work.

Topic 2: Unsupervised scaling – up and down

Giving the application infrastructure full control of scaling up and down (Machine Set scaling) and ensure that the underlying data infrastructure scales accordingly.
This means not scaling like “1 new application instance = 1 new database instance”, but more something like “20 new application instances = 1 new database instance”

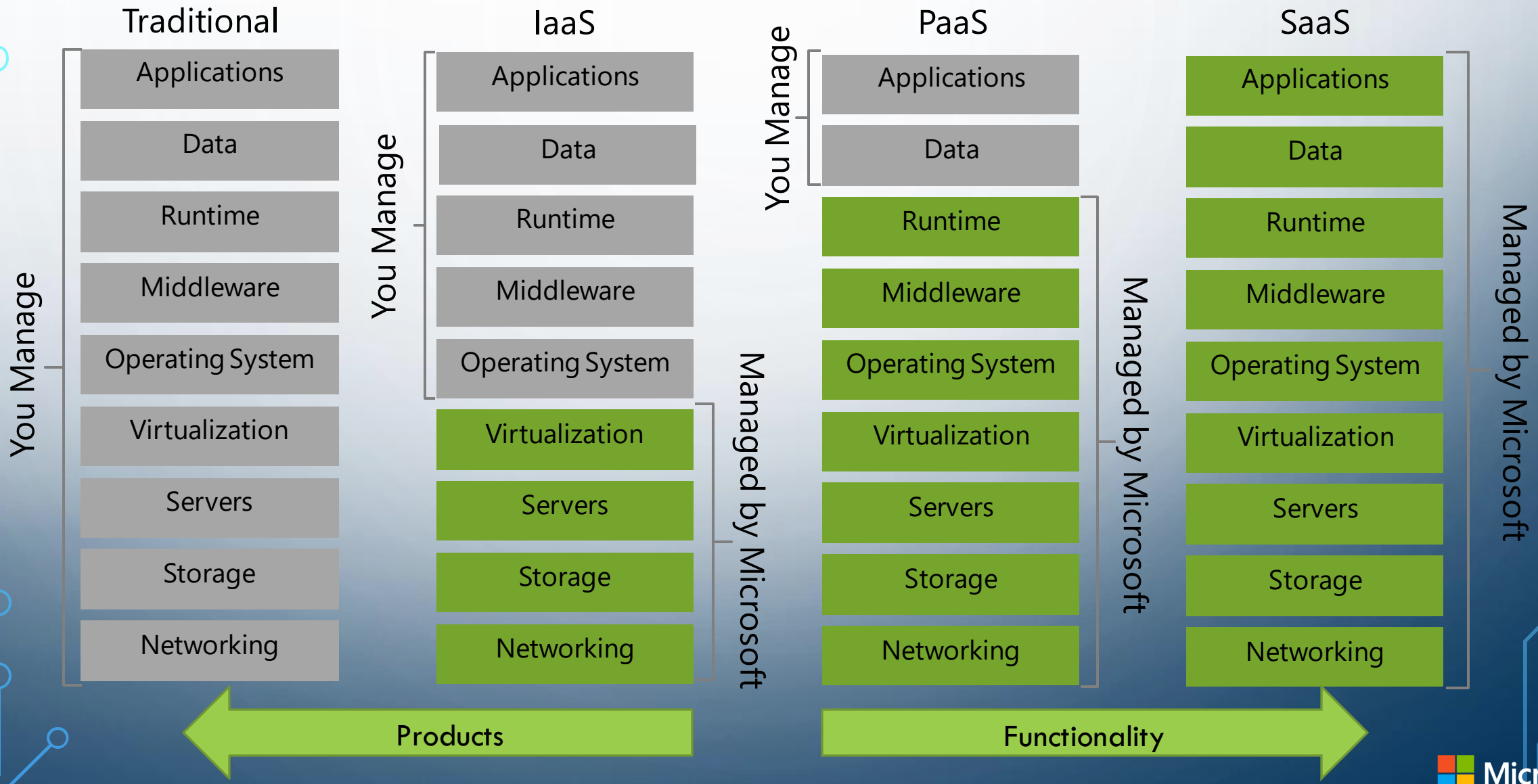
Topic 3: Versioning the application with the database simultaneously

When a new version of the application is created introducing new/changed/removed “fields” a new version of the underlying data infrastructure should be established accordingly.

This to ensure that the “data model” in the database always reflects the “input fields” of the application.

BACKGROUND

CLOUD MODELS



AZURE - A SECURE FOUNDATION AT GLOBAL SCALE

Red Hat

Each *physical datacenter* protected with world-class, multi-layered protection

Secured with cutting-edge *operational security*

- **Restricted access**
- **24x7 monitoring**
- **Global security experts**

***Global cloud infrastructure* with custom hardware and network protection**

Microsoft

Red Hat

Global

Secure

Compliant

Microsoft

Red Hat

Global

Secure

Compliant

Microsoft Azure adheres to the EU Cloud Code of Conduct

May 20, 2021 | Microsoft Corporate Blogs

EU CLOUD COC

Microsoft Azure, a global cloud platform of services, successfully demonstrated its compliance with the EU Cloud Code of Conduct (CoC) through a rigorous, detailed assessment. This accomplishment is the latest example of Microsoft's commitment to meet and exceed data protection requirements in the EU.

Answering Europe's Call: Storing and Processing EU Data in the EU

May 6, 2021 | Brad Smith - President and Chief Legal Officer

Today we are announcing a new pledge for the European Union. If you are a commercial or public sector customer in the EU, we will go beyond our existing data storage commitments and enable you to process and store all your data in the EU. In other words, we will not need to move your data outside the EU. This commitment will apply across all of Microsoft's core cloud services – Azure, Microsoft 365, and Dynamics 365

Microsoft

Red Hat

LIMITLESS

Data estate

Scale

Variety

Cloud & Edge

Microsoft

THE TRENDS

MODERN DATA STRATEGY

Top goals		Focus
Data Monetization, Data Democratization Data-Driven organization	→	Empower the business Citizen Data Analytics, Citizen Data Science
Data Protection & Security & Quality	→	Data Governance, Proper Data owners
Business Value creation Collaboration between IT & Business Time to insight?	→	**Data as a Product** Embrace First Party Cloud Native Solution Delegate to cloud provider to gain more time for business value creation

INNOVATION & TREND

Application development

DevOps

Speed/Stability of development/deployment
Greater collaboration in multiple teams
Improve business

Microservices

Monolithic centralize app to distributed services

Easier to build and maintain Apps
Flexibility, scalability and reusability
Autonomous, Cross-functional Teams

Data & AI

DataOps, MLOps

DevOps for Data & AI

Data Mesh

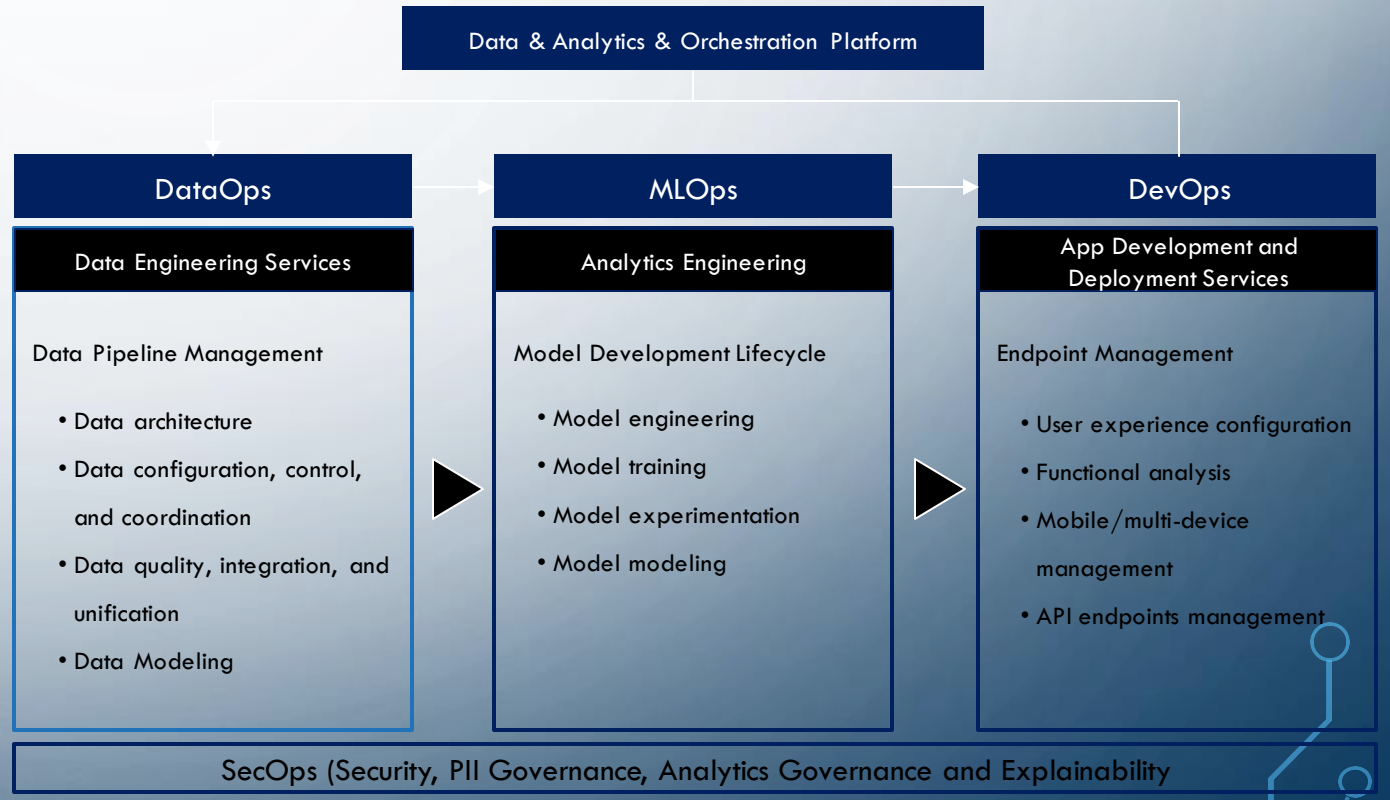
From Monolithic centralize to distributed decentralize

Scalable modern distributed data architecture

END CONSUMERS FOR DATAOPS – MLOPS, BI, ETC.

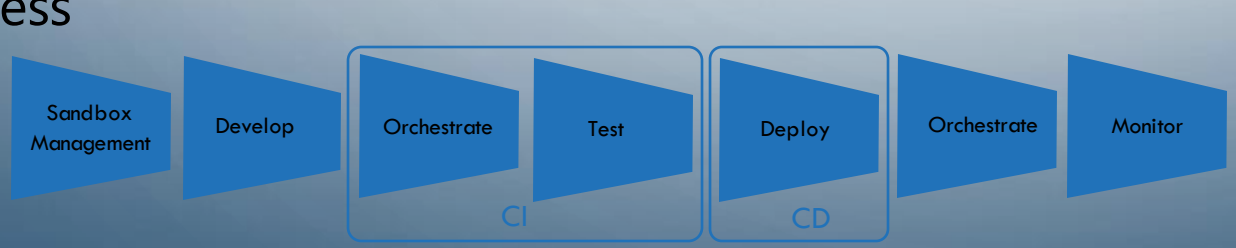
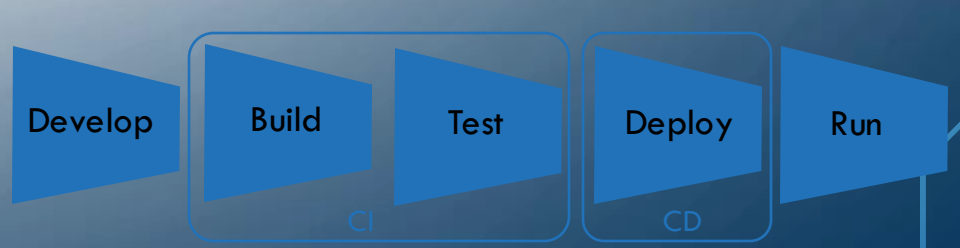
DataOps can be used across different use cases, including unforeseen future use cases. To name a few, end consumers for DataOps represent areas such as MLOps, AIOps, Business Intelligence, and Data Warehouses/Data Lakes

For the most part, the data consumer should be able to use the data coming out of the DataOps pipeline quickly and easily. However, the data consumer should also be able to "peek" into the DataOps pipeline.



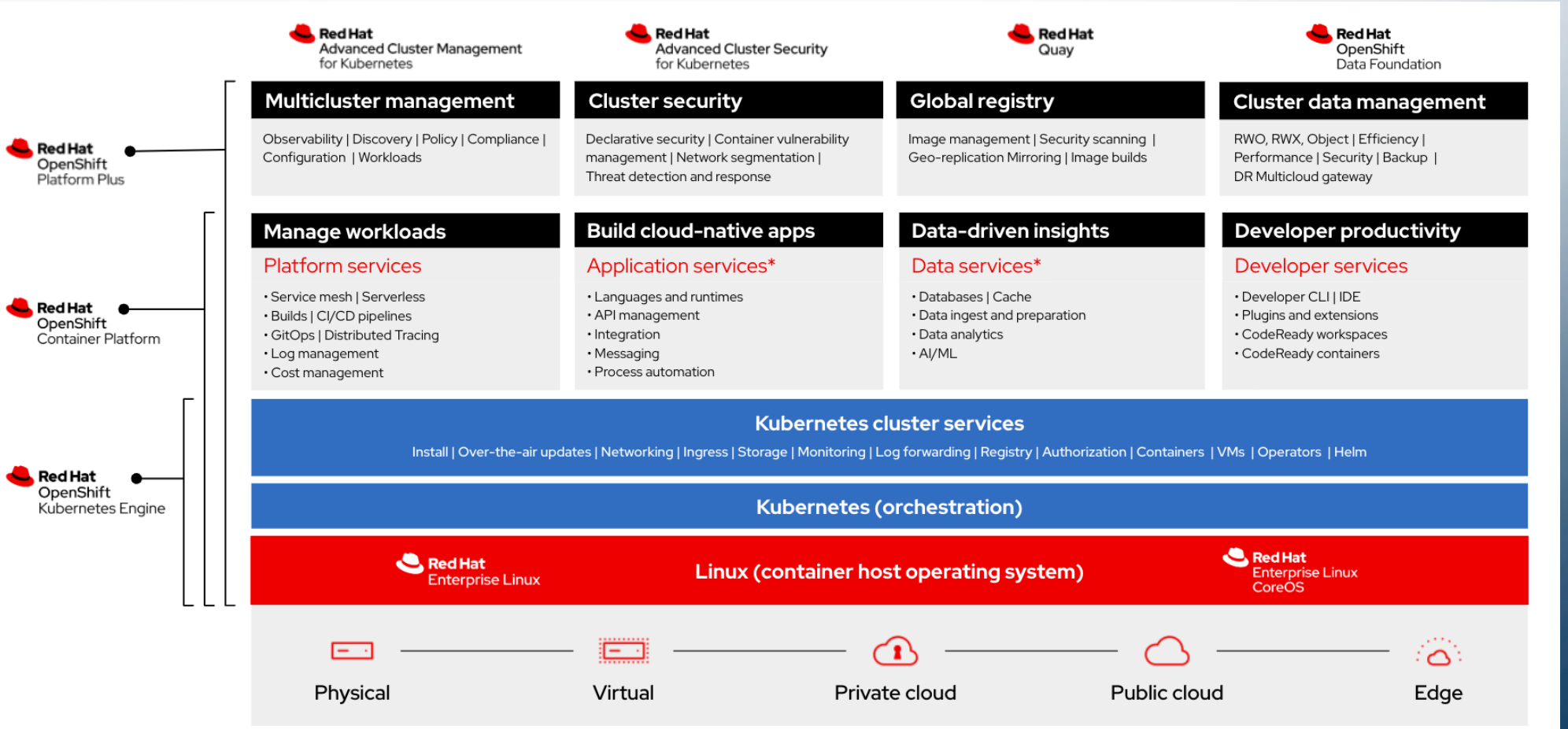
The place of DataOps as one of several AI orchestration platforms per Gartner and the XOps model – via Venturebeat

DATAOPS IS NOT JUST DEVOPS FOR DATA

	DataOps	DevOps
Key focus	Business-ready trusted, high-quality data available for use fast.	Application and software development
Users & Tools	Data Scientists, Engineers, and Analysts who want to just analyze data and build models.	Software Engineers, comfortable with coding and complexity of multiple languages, tools, and hardware/software.
Process		

THE ARCHITECTURE

OPENSHIFT ARCHITECTURE



* Red Hat OpenShift® includes supported runtimes for popular languages/frameworks/databases. Additional capabilities listed are from the Red Hat Application Services and Red Hat Data Services portfolios.

** Disaster recovery, volume and multicloud encryption, key management service, and support for multiple clusters and off-cluster workloads requires OpenShift Data Foundation Advanced

DATA : CAPABILITIES

Capability	Description
Consistency	A consistent solution design to ensure simple operation and further development of solutions. Replacing resources must be flexible.
Modulated	Solutions in the data infrastructure are developed modularly with a well-defined interface, so it is easy to replace subcomponents.
Technology independency	The architecture must be independent of the technology being used. This means that no matter what technology is being used, the different processes, functionalities and layers are the same.
Scalability	Scalability (up/down) is part of the solution design from the beginning, so that implementation and operations are not affected by bottlenecks, downtime or a possible mandatory purchase of unforeseen licenses.
Agility	The focus is on MVP (minimum viable product) and ongoing feedback to previous steps in the data flow. It is OK to fail, and solution designs must be tested as soon as possible during the development process.
Security	Security is incorporated into the general architecture as well as in the concrete solution designs, both in terms of information security and privacy.
Data encapsulation	Data in the data infrastructure is accessed through an interface that enables control of who has access to what. The interface allows you to make changes to the data infrastructure without affecting external systems.
Reuse	Solutions are being developed for reuse. The architecture will consist of templates for solution designs that shorten time-to-market and ensure standardization.
Feedback	The architecture is continuously adapted and improved based on feedback from the different users of data.

DATA : LAYERS

New systems

Legacy systems

Ingest

Characteristics
- Consume raw data
- High degree of integration points
- Very technical
- No transformation

Transform

Characteristics
- Technical useable data
- Getting data from ingest layer - only
- Very technical
- Place for technical transformations
- Adhere to company chosen standards

Publish

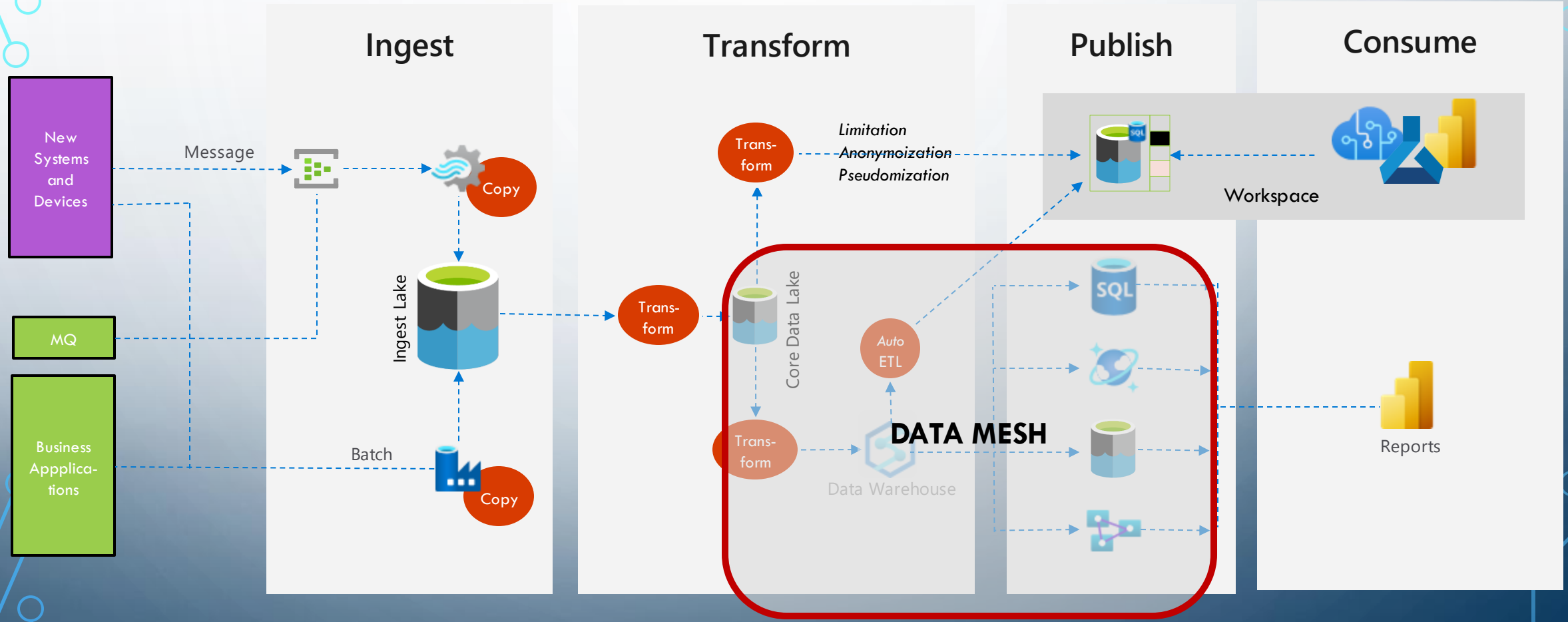
Characteristics
- Business useable data
- Data placed in whatever suitable technology for the end usage
- “Lives” as long as being used
- Can be recreated (by self-service)

Consume

Characteristics
- End user tools
- Business Intelligence
- Reporting
- Machine Learning
- Artificial Intelligence

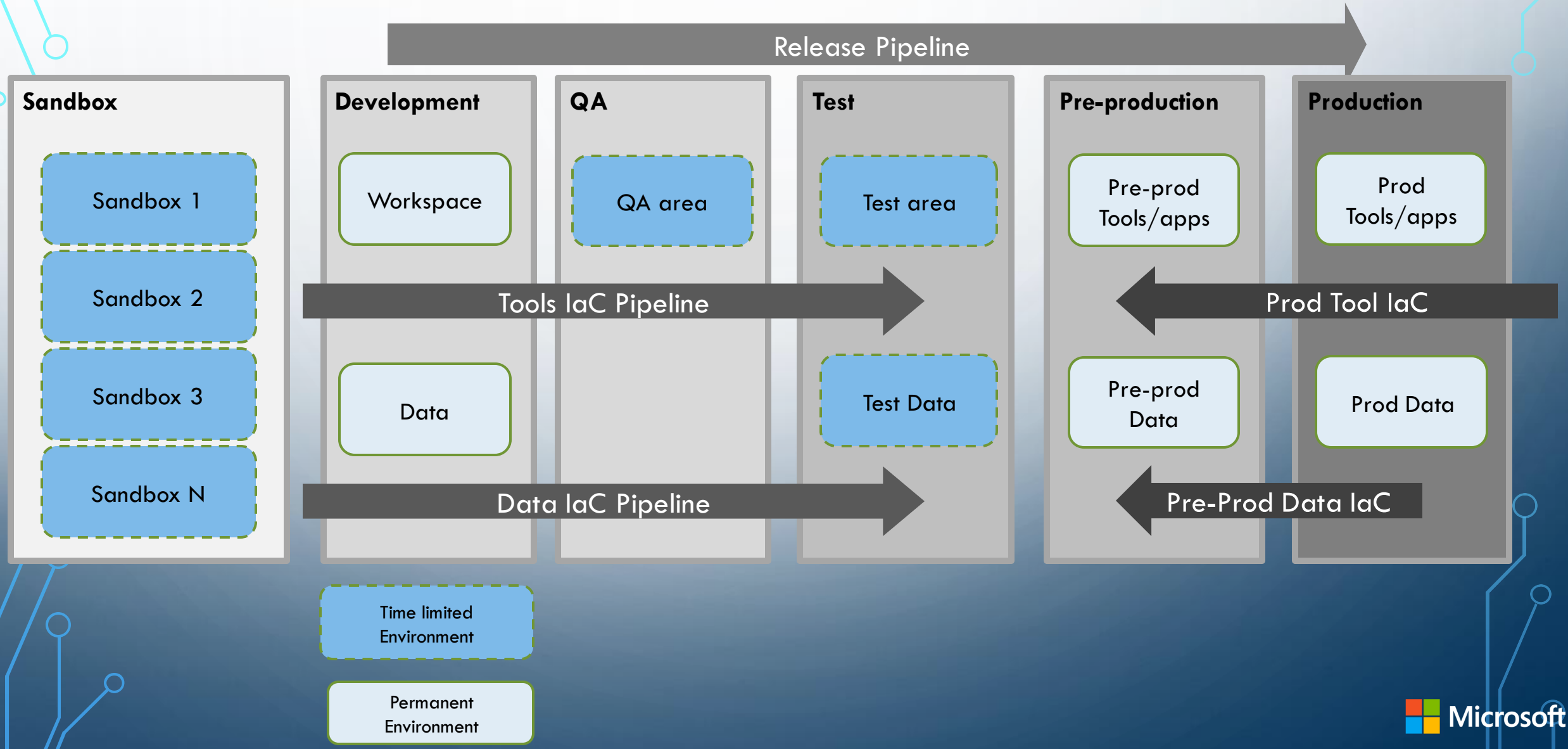
Supporting Data Services Data Catalog | Common Data Model | Data Discovery | Data Lineage | Data Governance | Master Data

DATA : PROCESS



Supporting Data Services Data Catalog | Common Data Model | Data Discovery | Data Lineage | Data Governance | Master Data

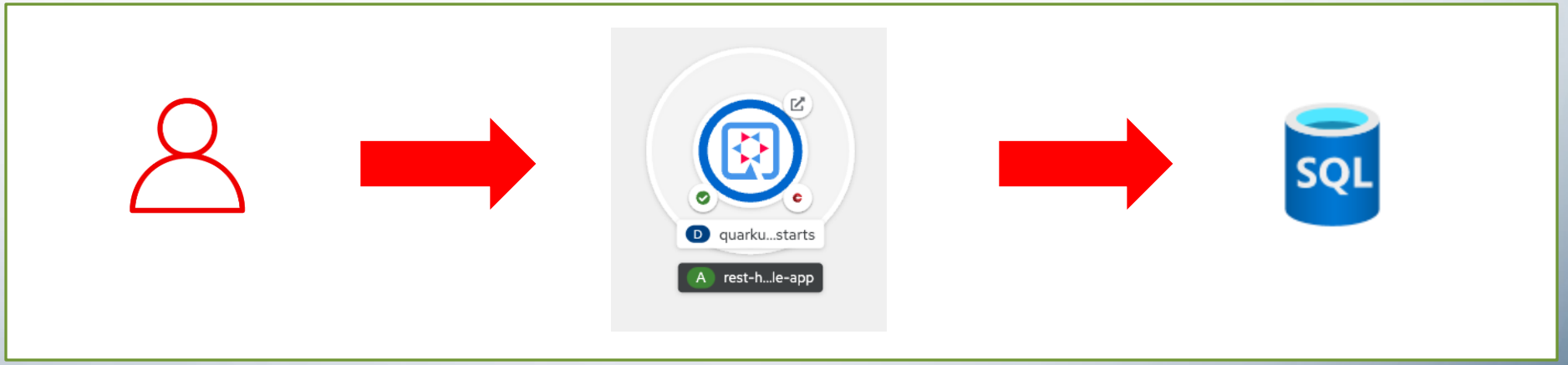
DATA : ENVIRONMENTS / PIPELINES



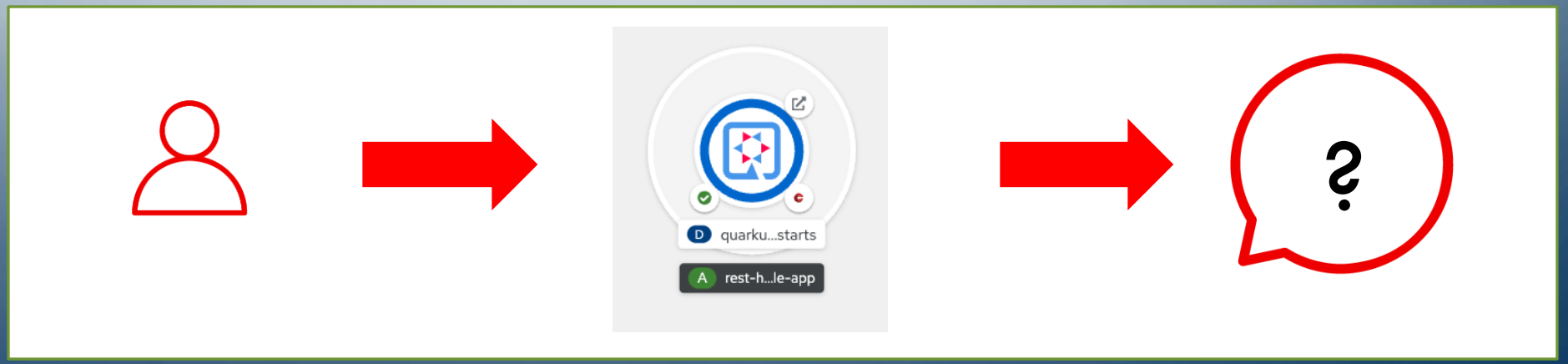
THE TOPICS

TOPIC 1: DATA BEING PART OF THE PROCESS

Having an active development branch being tested



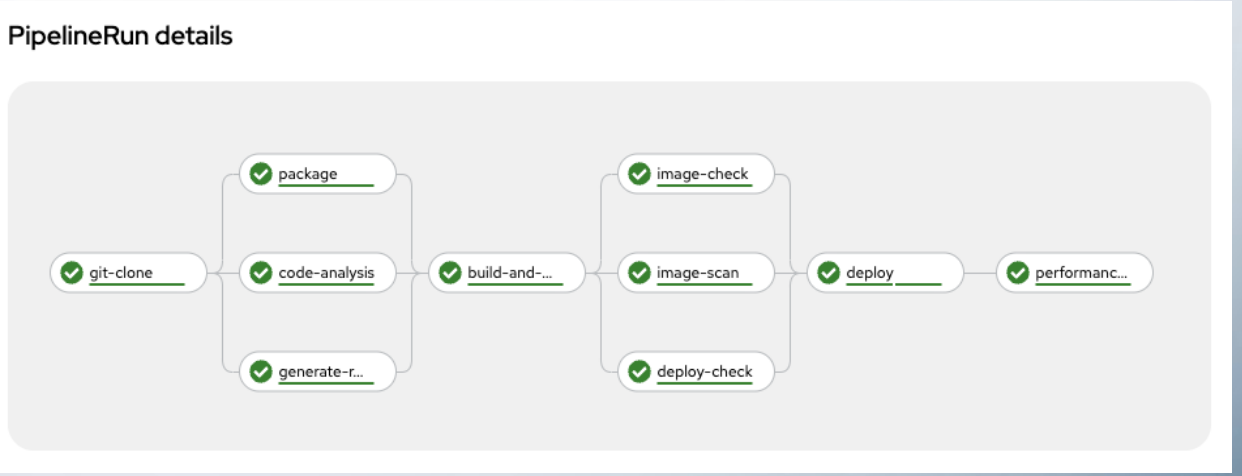
Created a new feature branch to test a new feature



TOPIC 1: DATA BEING PART OF THE PROCESS

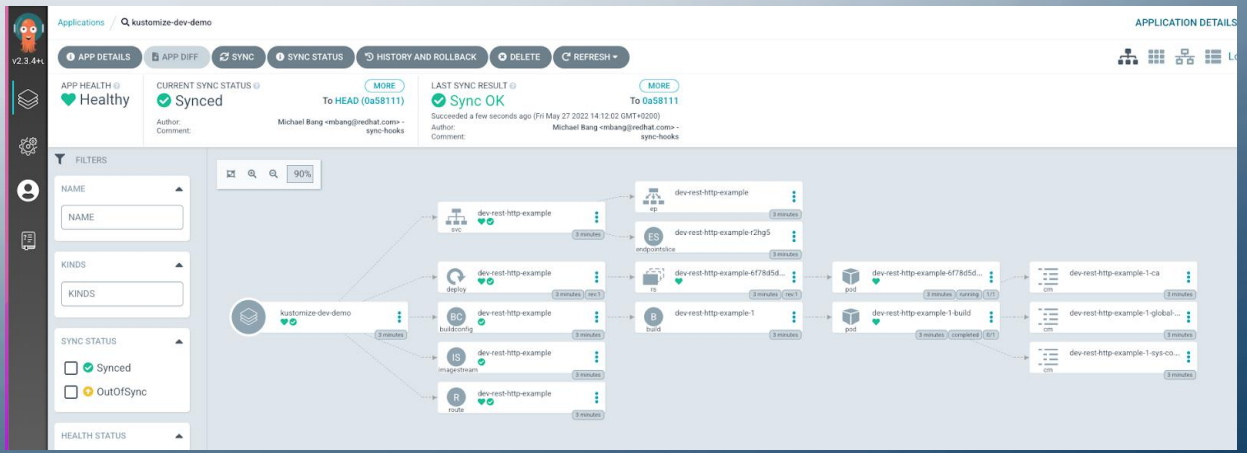
OpenShift Pipelines:

- Create a new project as part of the pipeline

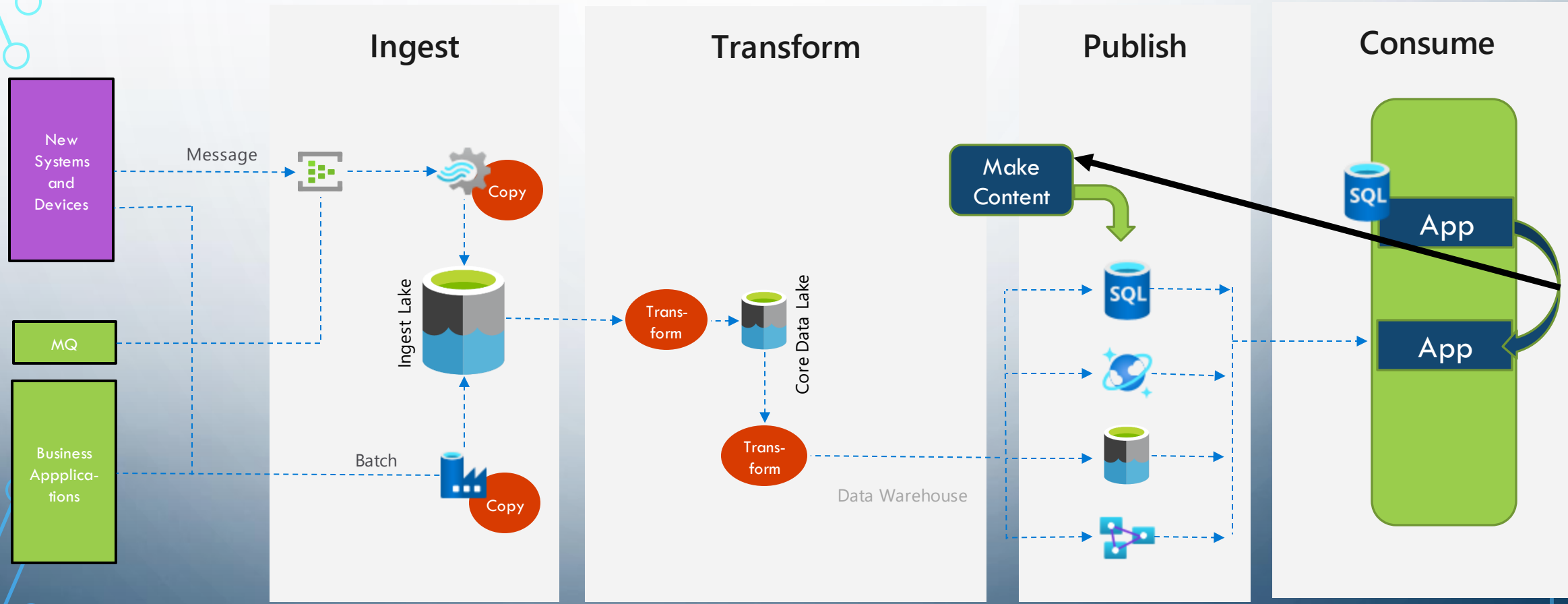


OpenShift GitOps:

- Or use OpenShift GitOps with kustomize to add a new project with the new feature



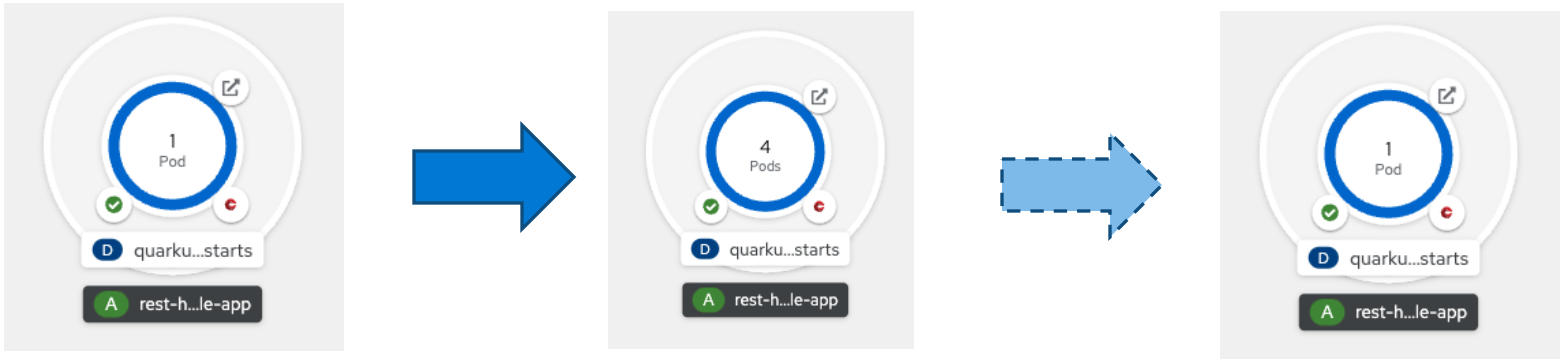
TOPIC 1: DATA BEING PART OF THE PROCESS



TOPIC 2 - UNSUPERVISED SCALING

How do we support an application that fluctuates in usage over time ?

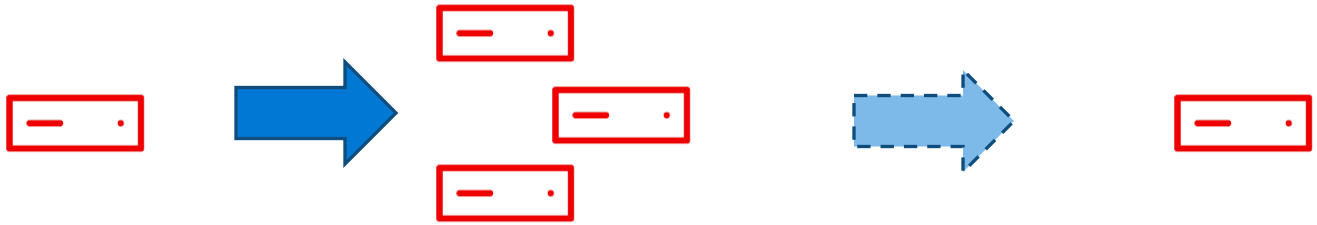
Horizontal pod autoscaler can be used to scale the deployment



TOPIC 2 - UNSUPERVISED SCALING

How do we support an application that fluctuates in usage over time ?

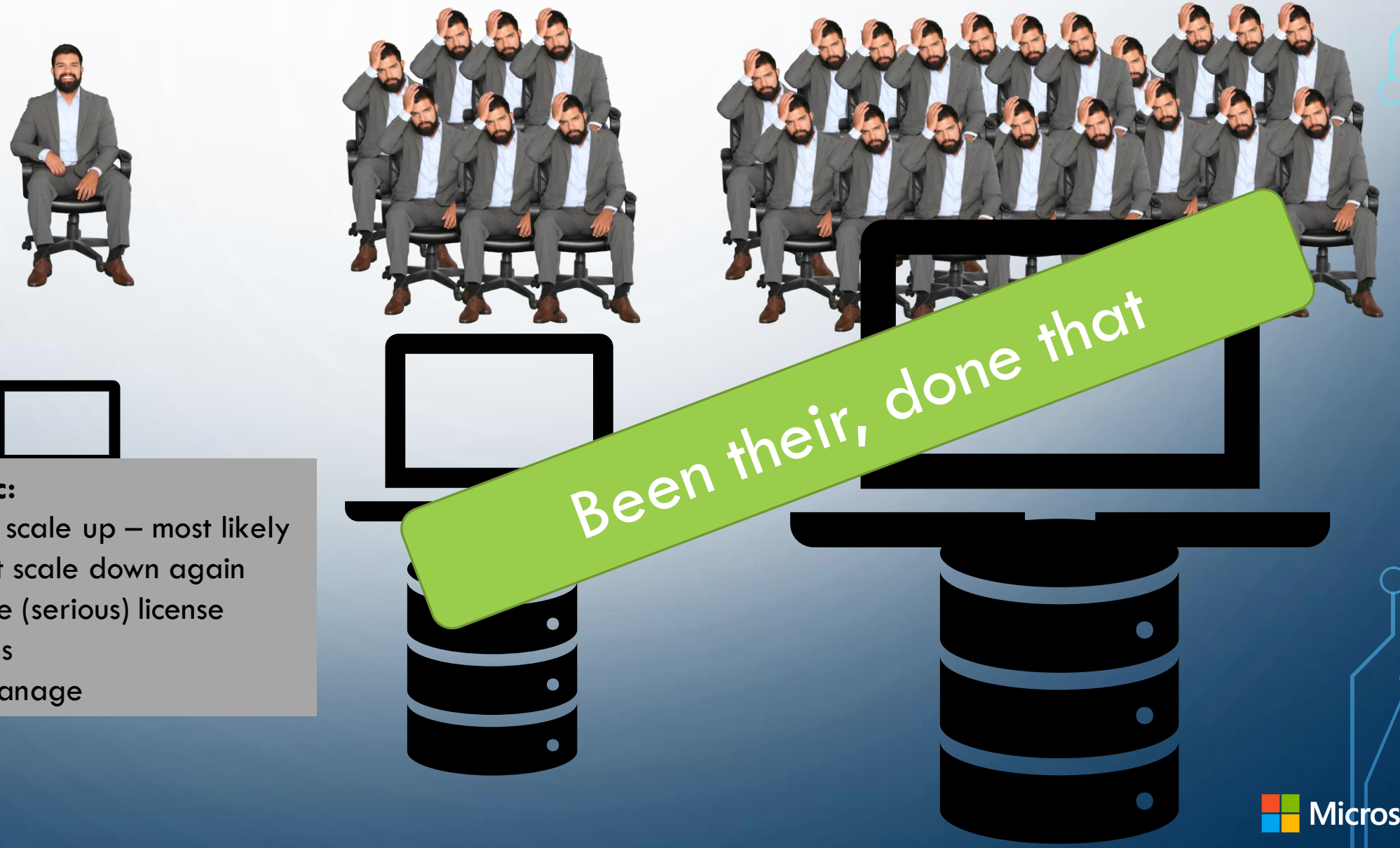
Cluster autoscaler can be used to scale the machinesets in the cluster



How do we ensure that we have the needed throughput to the data ?

- Resource:
  - Pod scaling: https://docs.openshift.com/container-platform/4.10/nodes/pods/nodes-pods-autoscaling.html
  - Cluster scaling:https://docs.openshift.com/container-platform/4.10/machine_management/applying-autoscaling.html

TOPIC 2: UNSUPERVISED SCALING – UP AND DOWN



Characteristic:

- Difficult to scale up – most likely we will not scale down again
- Could have (serious) license implications
- Hard to manage

TOPIC 2: UNSUPERVISED SCALING – UP AND DOWN

Characteristics:

- **Extremely user friendly**
- **Extremely expensive**
- **Unable to manage**

TOPIC 2: UNSUPERVISED SCALING – UP AND DOWN

Characteristics:

- **Extremely user friendly**
- **Automatic**
- **Cost effective**
- **Requires automatic "data sync"**
- **What is the "formular"?**

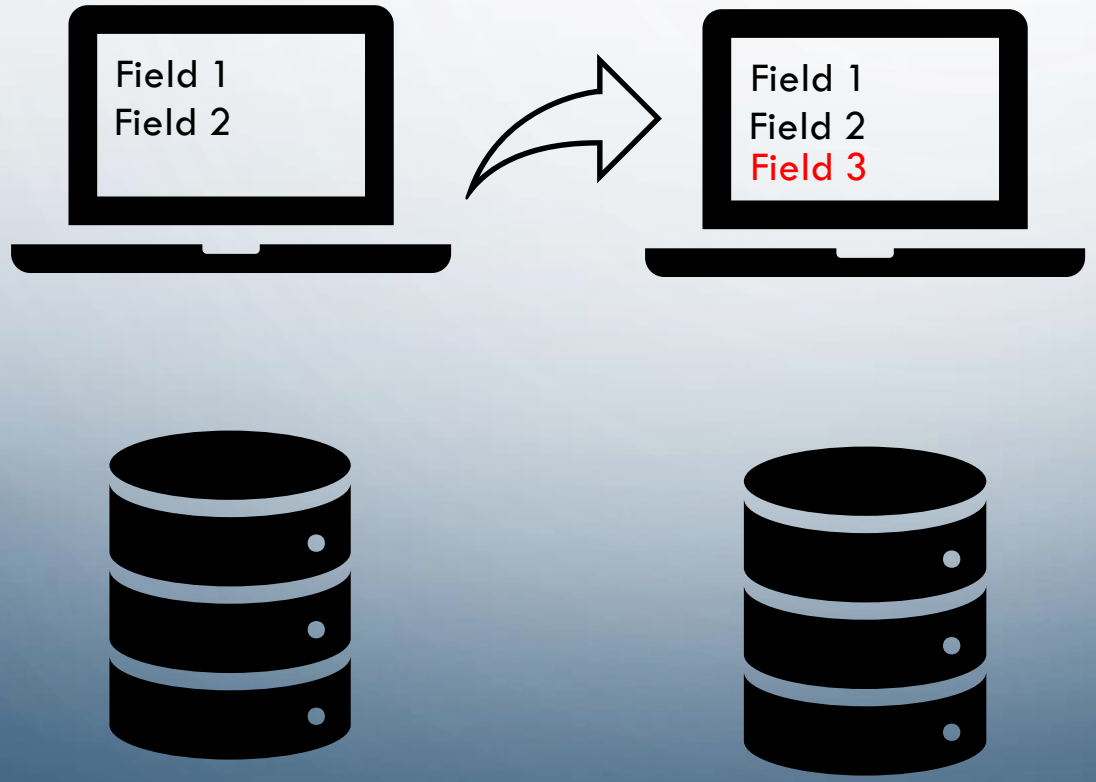
The new way

When x <y and
instance > 1then
instance -1

TOPIC 3 - VERSIONING THE APPLICATION WITH THE DATABASE SIMULTANEOUSLY

- OpenShift GitOps has made it easy to update an application
  - Even with changes to kubernetes manifests
- With new versions of an application often also include changes to the database
  - New fields, changes to fields, new tables etc.
- How do ensure that the database is updated with the application ?

TOPIC 3: VERSIONING THE APPLICATION WITH THE DATABASE SIMULTANEOUSLY



A way of doing this:

- Field 3 is registered in Microsoft Purview
- On database creation/update time this info is extracted
- The change script is created and executed

SUMMARY

- With Cloud technologies it is easier to think Application development and Data operations in to one process.
- “You build it, you run it” works much better
- Scalability and security by design
- Governance/Compliance part of the process

Red Hat

100%

of the **Fortune Global 500** companies
in these industries rely on Red Hat

Airlines

Telecommunications

Healthcare

Commercial Banking

Microsoft

>95%

of Fortune 500 use Microsoft Azure

Microsoft

RED HAT + MICROSOFT PARTNERSHIP

Red Hat

Red Hat Enterprise Linux in Azure

Cost savings and operational efficiency gained from using consistent/standard OS platforms across your hybrid infrastructures

Integrated support for RHEL in the Azure Marketplace

Red Hat subscription flexibility/portability

Red Hat OpenShift Container Platform in Azure

Easily build, deploy, and manage modern container-based apps

Technology that enables digital transformation and application modernization

Consistent application platform for hybrid cloud infrastructures.

Fully managed Red Hat OpenShift service

SQL Server on Red Hat Enterprise Linux

Industry-leading, most-secure data platform on a leading OS and cloud platform

Optimize with a modern data platform

Red Hat Enterprise Linux for SAP Solutions in Azure

Most-powerful and scalable cloud for SAP HANA

Deep partnership among SAP, Microsoft and Red Hat

First-class hybrid support experience for Red Hat on Azure

Integrated management portal experience

Hybrid application framework — Hybrid cloud storage — Hybrid cloud management

Microsoft

CALL TO ACTION

Red Hat

Microsoft

Red Hat

THE END

Microsoft