KPMG Ignite for OpenShift

Presented by KPMG

February 2022

OpenShift Commons Briefing

With you today

Kevin Martelli
Principal, Cloud Engineering
Email: kevinmartelli@kpmg.com

Hongfei Cao (PhD)
Director, Cloud Engineer
Email: hongfeicao@kpmg.com

What we'll talk about…

1. What is KPMG Ignite?
2. How KPMG Ignite leverages data storage and messaging?
3. Kafka & Postgres
4. PVC and Minio

What is KPMG Ignite?

Ignite runs on a secure OpenShift cluster.

What An artificial intelligence (AI) platform with a modular component architecture with ability to build production AI applications
Who A platform built for data scientists and engineers with a hook for business users.
Why Unlocks the value of unstructured data with surgical precision on complex problems

Use Case Delivery Methodology

Components
are the building blocks of Ignite solutions. Each component represents a single capability or action. Components remove the need to write code for each operation, accelerating service development.

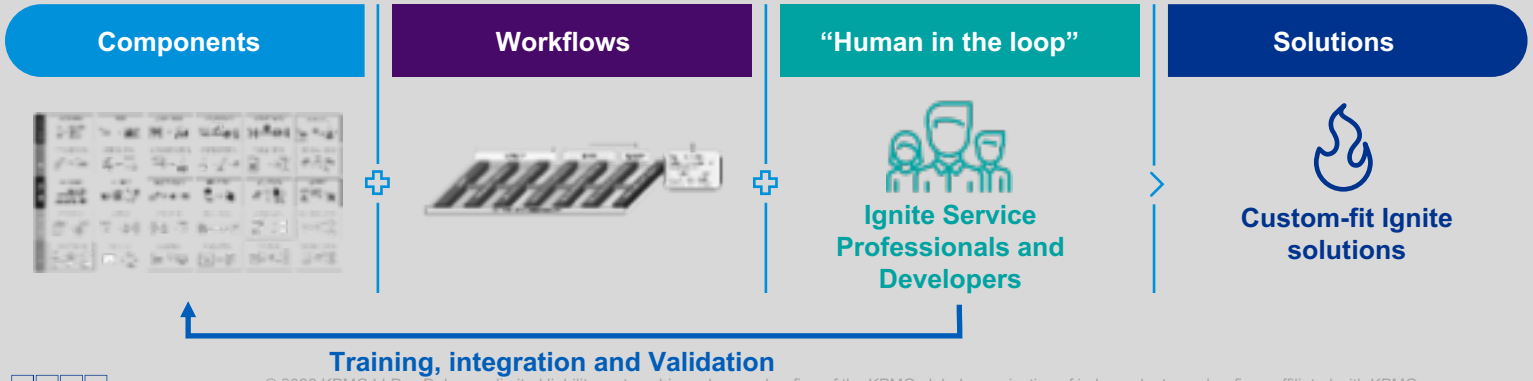
Top quality **components**
(Cognitive Patterns and Scripts)

Workflows
are pipelines of components combined to accomplish AI tasks. Workflows can be sequential or can contain branching paths. Adding components to existing workflows creates new workflows.

Custom **workflows**
(Process pipelines and document classification to accomplish AI tasks)

Solutions
are groups of workflows, models, ontologies, rules, and other configuration that combine to address complex, domain-specific challenges

Underpinned by knowledge, experience and expertise of our people to create **custom-fit Ignite solutions**



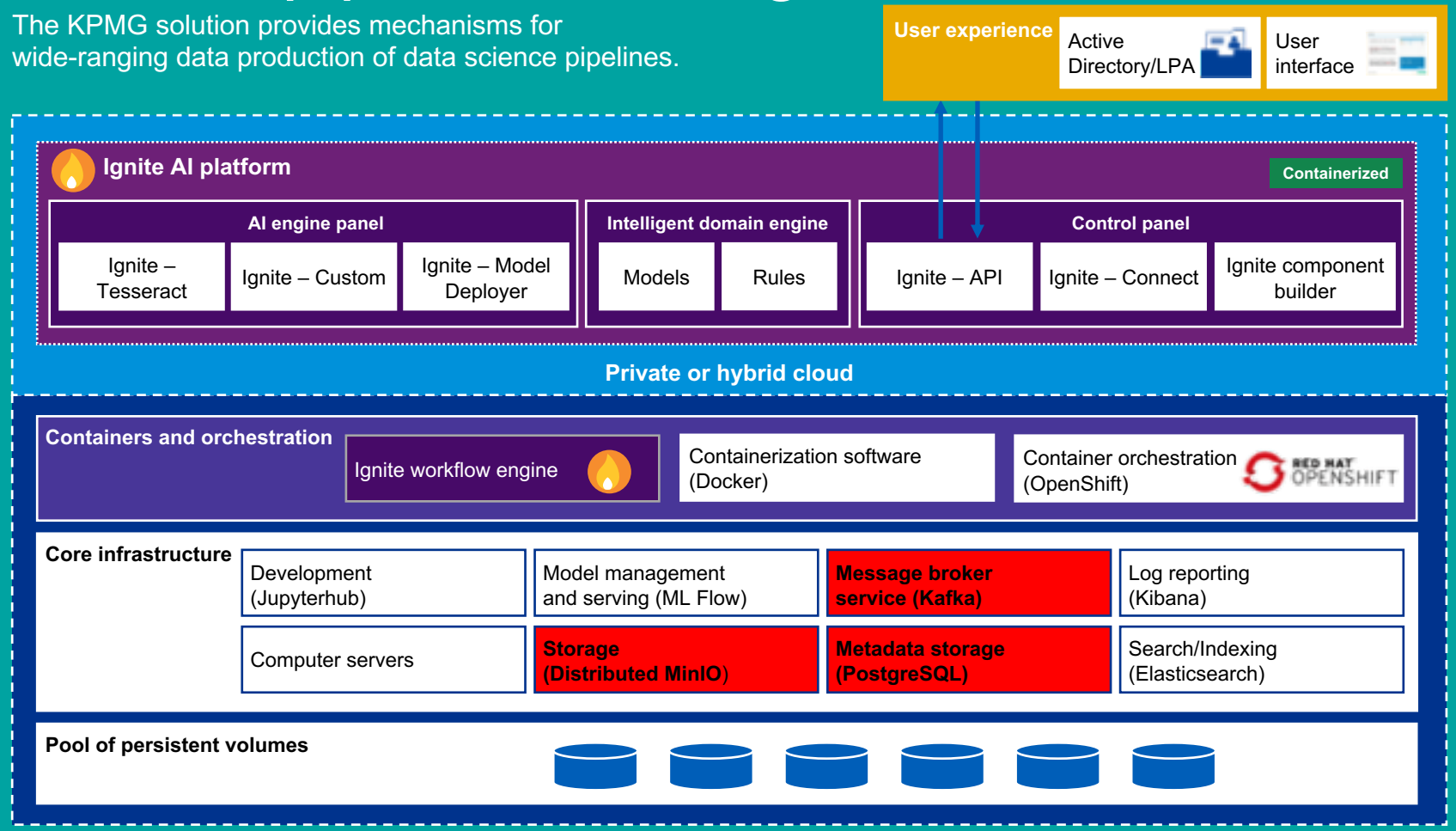
Platform Key Features:

CI/CD, Logging and Monitoring, Pipeline Orchestration, AI/Model Management/Deployment, Multi-Tenant, scalability, pre-built connectors, development

Build, orchestrate, and deploy data science pipelines with Ignite

Ignite runs on a secure OpenShift cluster.

The KPMG solution provides mechanisms for wide-ranging data production of data science pipelines.



Ignite currently has a proprietary data framework, two development entry points, and a user interface for building training data:

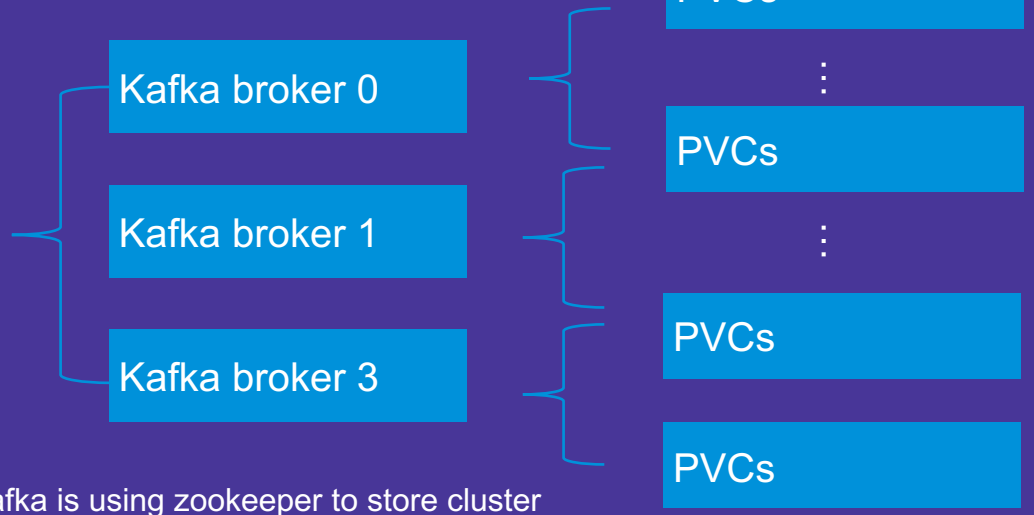
- **LUME:** The common language across Ignite, allowing data and accelerators to be shared across the ecosystem; it is lightweight, flexible, and extendable to structured and unstructured data
- **Data Scientist Notebook:** Developer entry point for creating AI models and interacting with the platform
- **Annotation UI:** User interface for domain and technical resources to annotate documents for NLP use cases
- **Management UI:** User interface for to define high-level model requirements, data access/ workflows, and integrate business processes with Ignite solutions
- **Model Management:** Manage model lifecycle and serving

Kafka on Ignite

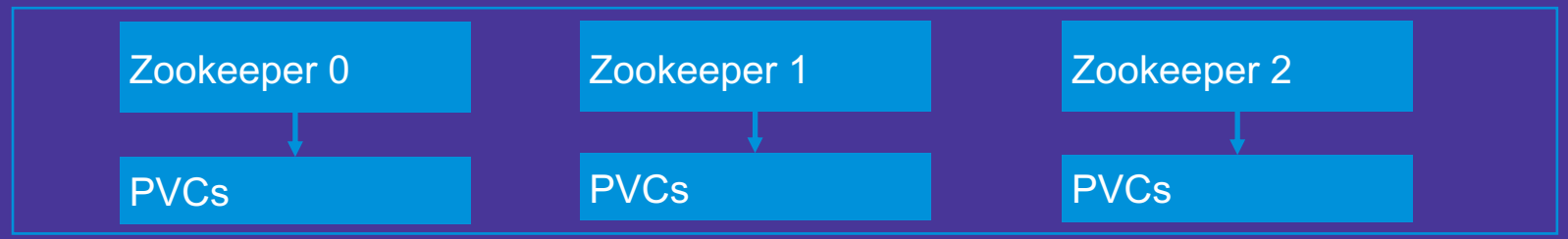
Ignite runs on a secure OpenShift cluster.

Pods

- Kafka is used for Ignite components to exchange messages
- Kafka HA



- Kafka is using zookeeper to store cluster metadata
- Zookeeper HA



Storage

Benefits of deploying Kafka on OpenShift

- Cloud platform agnostic
- Orchestration and failover
- Automated workflow with CI/CD
- Easy scale up and down

Postgres components

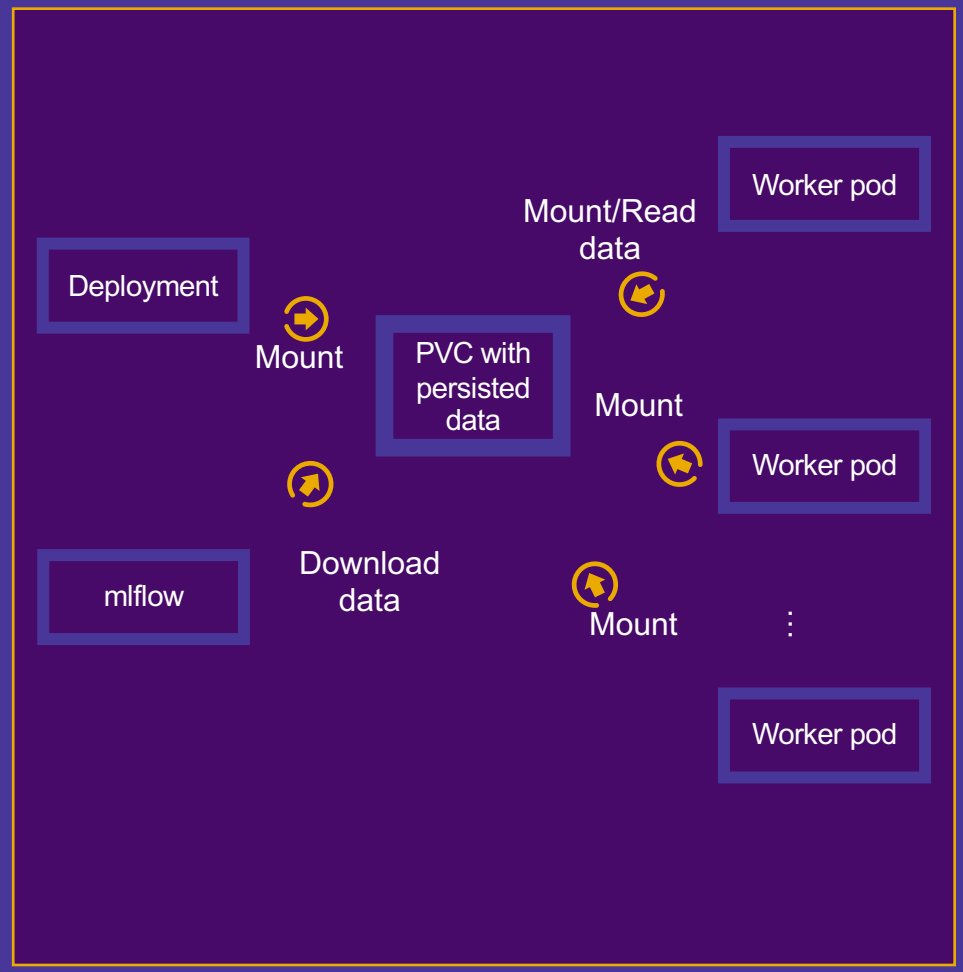
Ignite runs on a secure OpenShift cluster.

1. Used to store Ignite workflow metadata
2. HA deployment using PostgreSQL Operator
3. Leverage customized pod with PostgreSQL tool for data backup
4. — Solution:
   - Spin up a separate deployment, where it connects to Postgres via psql
   - pg.dump the entire database and put file to MinIO API
   - From MinIO API, back up to other environments
   - Schedule this deployment to run every day, etc.
5. **Benefits of deploying Postgres on OpenShift:**
   - Agile deployment with operator and integrated storage support
   - Enterprise grade, High availability out of the box
   - Dev, test, production parity
   - Cloud Platform agnostic (onprem vs. Azure/AWS/GCP)
   - Enable DevOps, CI/CD integration
   - Security support with image vulnerability scanning, built-in security policies for hardened container platform, RBAC integrated with enterprise directory enables principle of least privilege

Stand-alone PVC as storage for AI Models

Ignite runs on a secure OpenShift cluster.

- Designed for specific use case that:
  - Shares large-size object storage across worker pods
  - Requires minimal data download time
  - Data is relatively static
- Use a separate deployment to load objects in one-time
- Requires a process in place for data updates



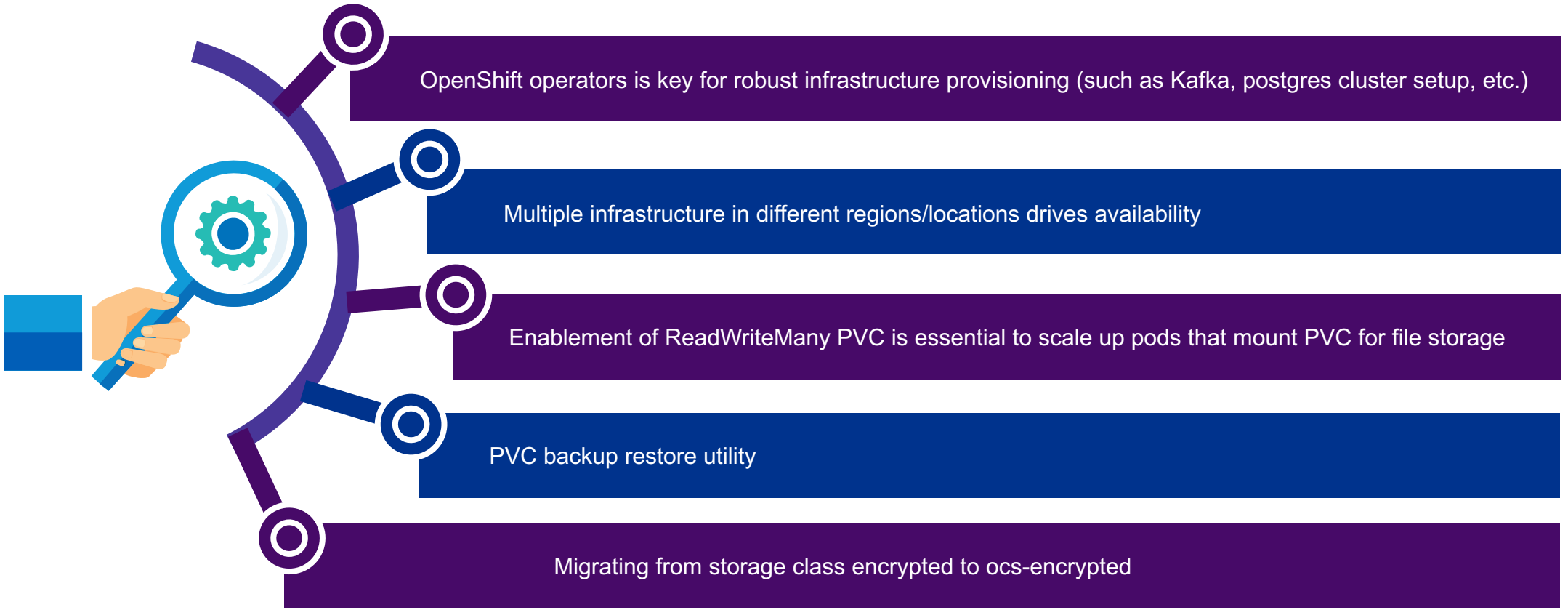
MinIO on Ignite

Ignite runs on a secure OpenShift cluster.

- Readwritemany ocs-encrypted storage class for PVC
- MinIO access key stored as secrets
- Usage shared across use cases by connecting to same MinIO API
- Data organized within MinIO within buckets
- Easily access data from multiple environments through MinIO API

Other observations

Ignite runs on a secure OpenShift cluster.



- OpenShift operators is key for robust infrastructure provisioning (such as Kafka, postgres cluster setup, etc.)
- Multiple infrastructure in different regions/locations drives availability
- Enablement of ReadWriteMany PVC is essential to scale up pods that mount PVC for file storage
- PVC backup restore utility
- Migrating from storage class encrypted to ocs-encrypted

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

kpmg.com/socialmedia