ROBIN

Big Data-as-a-Service with Kubernetes

Automate your big data infrastructure using cloud-native architecture. Improve the agility and efficiency of your Data Scientists, Data Engineers, and Developers.

Top 5 Challenges for Big Data Management

Big data has transformed how we store and process data. However, following challenges keep organizations from unlocking the full potential of big data and maximizing ROI:

Provisioning agility for ephemeral workloads: Certain workloads, such as ad-hoc analysis, require significant compute resources for a short period of time. Developers need the ability to quickly provision and decommission compute-only clusters for such workloads.

Separation of compute and storage: Big data needs converged nodes with both compute and storage for data locality. However, compute is significantly more expensive than storage, and with ever increasing data volumes, infrastructure costs are rising.

Dynamic scaling to meet sudden demands: If critical services such as the NameNode run out of resources, it is not easy to scale-up nodes on the fly to add more memory or CPU.

Cluster sprawl and hardware underutilization: Due to lack of reliable multi-tenancy and performance isolation, Hadoop Admins often deploy separate clusters for critical workloads, resulting in cluster sprawl and poor utilization of server resources.

Cloud migration: There is no easy way to migrate big data clusters to public clouds, or leverage public cloud compute and storage as needed for on-prem clusters.

Highlights

- Decouple compute and storage and scale independently to achieve public cloud flexibility

- Migrate big data clusters to public cloud or leverage public cloud to off-load compute

- Provision/Decommission compute-only clusters within minutes for ephemeral workloads

- Provide self-service experience to improve developer and data scientist productivity

- Eliminate planning delays, start small and dynamically scale-up/out nodes to meet demand

- Consolidate multiple workloads on shared infrastructure to reduce hardware footprint

- Trade resources among big data clusters to manage surges & periodic compute requirements

Less than an hour:
You can provision a 500 node Cloudera cluster with end-to-end security in 45 minutes or less

Solution Benefits and Business Impact

Robin brings the agility, scalability, and portability of cloud-native architecture to all of your applications.

Deliver Insights Faster

Self-service experience
Robin provides self-service provisioning and management capabilities to developers, data engineers, and data scientists, significantly improving their productivity. It saves valuable time at each stage of application lifecycle.

Provision clusters in minutes
Robin has automated the end-to-end cluster provisioning process for Hortonworks, Cloudera, Spark, Kafka, and custom stacks. The entire provisioning process takes only a few minutes.

Provision compute-only clusters
You can create and decommission compute-only clusters for Hortonworks, Cloudera, and your custom big data stacks. Perfect for ephemeral workloads, these clusters simply point to existing data lake cluster in your organization, do the required processing, and store the data in the target systems.

Eliminate "right-size" planning delays
DevOps and IT teams can start with small deployments, and as applications grow, they can add more resources. Robin runs on commodity hardware, making it easy to scale-out by adding commodity servers to existing deployments.

Scale on-demand during surges
No need to create IT tickets wait for days to scale-up NameNodes, or to add more DataNodes. Cut the response time to few minutes with one-click scale-up and scale-out.

Reduce Costs

Decouple compute and storage
Enjoy cloud-like cost efficiencies by disaggregating compute and storage. Scale independently so you can store massive data on inexpensive hardware, while only utilizing compute when needed. For high performance, turn on data locality with one click.

Improve hardware utilization
Consolidate multiple big data workloads and improve hardware utilization without compromising SLAs or QoS. The Robin platform provides multi-tenancy with RBAC which enables sharing resources between clusters.

Simplify lifecycle operations
Native integration between Kubernetes, storage, network, and application management layer enables one-click operations to scale, snapshot, clone, backup, migrate applications, reducing the administrative cost of your big data infrastructure.

Trade resources among clusters
Reduce your hardware cost by sharing the compute between clusters. If a cluster runs majority of its batch jobs during the night-time, it can borrow resource from an adjacent application cluster with day-time peaks, and vice versa.

Future-Proof Your Enterprise

Migrate or extend to public cloud
Robin provides one-click lift-and-shift for big data clusters. Simply clone your entire cluster and migrate to the public cloud of your choice. You can also scale-out your clusters to public cloud from on-prem to create hybrid cloud environment.

Standardize on Kubernetes
Modernize your data infrastructure using cloud-native technologies such as Kubernetes and Docker. Robin solves the storage and network persistency challenges in Kubernetes to enable its use in the provisioning, management, high availability and fault tolerance of mission critical Hadoop deployments.

No vendor lock-in
A Kubernetes-based architecture gives you complete control of your infrastructure. With the freedom to move your workloads across private and public clouds, you avoid vendor lock-in.

The Robin Platform Enables an "as a Service" Experience

The Robin application automation platform enables enterprises to deliver complex data- and network-centric application pipelines as a service. Only Robin is architected with application and infrastructure topology awareness, automating the provisioning and life cycle management of your critical big data applications.

Developers

Data Scientists

Data Engineers

Self-Service Provisioning and Management

Hortonworks   cloudera   Apache Spark™   kafka   APACHE HBASE   + **Custom Apps and Pipelines**

Robin Workload Management

Kubernetes + Built-In Networking

Compute-Only Nodes

Converged Nodes

Public Cloud Compute

AWS
EC2

Azure
VM

Google
COMPUTE

Robin Application-Aware Scale-Out Storage

Storage-Only Nodes

SSD   SSD   SSD   HD   HD   HD

SSD   SSD   HD   HD

Plublic Cloud Storage

EBS   S3   BLOCK   OBJECT

ROBIN