

CLUSTERA DATAFLOW FOR THE PUBLIC CLOUD

WHY CLUSTERA DATAFLOW FOR THE PUBLIC CLOUD?

- **Deploy flows quickly** - Leverage the deployment wizard to deploy NiFi flows faster and in a managed manner.
- **Monitor flow deployments** - Visualize pre-defined KPIs, system metrics, and alerts of all your flow deployments across multiple clusters in a centralized dashboard.
- **Scale flows up/down automatically** - Easily scale up high-performant flows based on pre-defined metrics such as CPU utilization. Also, scale down lighter flows automatically that are not consuming as many resources.
- **Embrace hybrid cloud model** - With the ability to deploy flows into CDF for Public Cloud from both on-premises NiFi and Flow Management for Data Hub, you can truly adopt a hybrid cloud model.
- **Use ReadyFlows to boost productivity** - With a pre-built set of flows for some of the common use cases to move data from a data source like Kafka into other common targets such as S3, you can now quickly configure and deploy such flows into production.

Cloudera DataFlow (CDF) for the Public Cloud is a new experience on the Cloudera Data Platform (CDP) that addresses operational challenges with deployment, scaling, and infrastructure sizing for Apache NiFi data flows. CDF for Public Cloud offers comprehensive data flow monitoring to help administrators identify and avert performance bottlenecks before they impact the business.

Operational Challenges for NiFi Administrators

Resource contention

Multiple high-performing NiFi flows deployed on the same cluster can compete for compute and infrastructure resources, impacting the performance of other flows running on the same cluster.

Overestimating cluster sizes

IT administrators often err on the side of caution when it comes to estimating NiFi cluster sizes. In order to plan for any peak surges, they overestimate the cluster size leading to a lot of unused nodes as well as high infrastructure costs.

Scaling clusters on-demand

Manually scaling clusters on-demand based on peak loads is not an easy task for administrators. While Kubernetes can help in this regard, it still requires special skills for such administrators.

Troubleshooting errors and bottlenecks

Administrators are required to open up the NiFi user interface to understand the health and metrics of any flow running in production. Not all administrators are trained in NiFi and so, this can be challenging. Moreover, monitoring NiFi flows across multiple clusters can be even more cumbersome.

CLUSTERA DATAFLOW FOR THE PUBLIC CLOUD

Cloudera DataFlow for Public Cloud takes away the operational and monitoring challenges by providing cloud-native flow management capabilities powered by Apache NiFi. It is a purpose-built framework to modernize the data flow user experience so that the NiFi developer and administrator can be prepared to easily handle sophisticated flows in production as well as boost operational efficiencies with such flows using comprehensive monitoring and management.

CDF for the public cloud offers a Control Plane on CDP. The control plane has three key capabilities

- **Catalog** - A central store for flow definitions and allow users to manage the flow definition lifecycle from import to versioning and deletion. The catalog also allows users to initiate new deployments.
- **Dashboard** - A central monitoring component showing all flow deployments across multiple environments and clusters at a glance. Helps define important KPIs to manage the data flows and get alerts on critical data flow metrics such as throughput, disk usage, or CPU utilization.
- **Environment Management** - Enables Cloudera DataFlow for any AWS environment registered with Cloudera Data Platform (CDP). The process creates the Kubernetes infrastructure required by Cloudera DataFlow, and each environment maps to one Kubernetes cluster.

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at: cloudera.com

CLOUDERA DATAFLOW FOR PUBLIC CLOUD



Cloud-native architecture

Cloudera DataFlow for Public Cloud follows a two-tier architecture where product capabilities like the Dashboard, Catalog and Environment management are hosted on the CDP Control Plane while the flow deployments processing your data are provisioned in a CDP environment which represents infrastructure in your cloud provider account.

When you enable DataFlow for one of your registered CDP environments, DataFlow creates and configures the required infrastructure including a Kubernetes cluster, Kubernetes Operators and the DataFlow workload application in your cloud account. After DataFlow has been successfully enabled for an environment, users can deploy Flow Definitions into this environment. Deploying a Flow Definition creates a dedicated NiFi cluster on Kubernetes allowing you to treat NiFi flows as isolated flow deployments.

Key Benefits

- Boost your operational efficiencies by deploying flows in a streamlined manner and by defining key metrics to measure their performance.
- Optimize your infrastructure setup by allowing CDF for Public Cloud to auto-scale your flows based on needs. This ensures that you are not over-sizing your infrastructure unnecessarily.
- Prevent resource contention on crowded clusters by isolating your flows to their own individual cloud-native clusters.
- Speed up the execution of your flows with this cloud-native way of execution of your standard NiFi flows.
- Monitor all your flow deployments across multiple cloud clusters from a single dashboard. This enables seamless troubleshooting from an administrator's perspective. In-built alerting capabilities also make administration a lot easier.
- Boost productivity by getting a headstart on common flow use cases by leveraging pre-built flows from a gallery of ReadyFlows