Accelerate The Data Lifecycle From Data Ingest To Insight

CLOUDERA DATA ENGINEERING

Cloudera Data Engineering is purpose-built for data engineers and enables them to:

DELIVER QUALITY DATASETS

Datasets are the backbone of any advanced analytics initiative. CDE lets you deliver curated, quality datasets while also ensuring security and governance compliance — all with open industry standard tools like Apache Airflow.

DEPLOY WITH EASE & CONTROL COSTS

CDE is a containerized, scalable, and portable service, so it's easy to deploy in cloud environments. It also features on-demand automatic resource scaling, so you only pay for what you use and maintain efficient resource utilization.

MANAGE RESOURCES & USERS EFFECTIVELY

Platform administrators can control costs, security, and governance from a centralized control plane. CDE empowers users to deliver agile, self-service data engineering for governed data while also regulating provisions and ensuring isolation across business stakeholders.

MONITOR JOBS AND TROUBLESHOOT WITH CLARITY

Data engineers can use CDE to efficiently deploy and monitor the lifecycle for every job and then use visual troubleshooting tools to quickly solve issues. Through a single pane of glass, users can make adjustments and review visualized analytics, which helps deliver and continuously maintain production-ready data pipelines.

With exponentially growing amounts of business data everywhere, delivering quality curated datasets securely is essential to enterprise data and analytics success. Building data pipelines is resource-intensive, requiring a focus on data integration, modeling, optimization, quality, governance, security and reusability. The ability to prepare data for analysis and production use-cases across the data lifecycle is critical for transforming data into business value.

Data Engineering For Enterprise Scale

Cloudera Data Engineering (CDE) for Cloudera Cloud Platform (CDP) empowers data engineers to leverage integrated and fully managed Spark to process large data volumes and streamlines data pipeline delivery to analytics and machine learning teams across the business. CDE provides comprehensive tooling to orchestrate, deploy, monitor, manage, and operationalize data pipelines at any scale.

CDE is an integrated platform delivered through a central interface that simplifies the job management life cycle to alleviate many of the challenges with running Spark jobs in production. CDE allows data engineers and platform admins to control costs and efficiently utilize resources while meeting key business requirements and meeting SLAs.

Deliver Quality Datasets Everywhere

Datasets are the backbone of any advanced analytics initiative. CDE lets you deliver curated, quality datasets while also ensuring security and governance compliance — all with open industry standard tools built in.

• **Containerized, Managed Spark Service**—CDE allows you to seamlessly manage and deploy different versions of spark without time-intensive preparation. By removing the need for version uniformity within and across teams, data engineers using different versions of Spark can seamlessly collaborate

• **Apache Airflow Scheduling**—With native Apache Airflow built in, you can automate the orchestration of complex data pipeline scenarios. This gives you the flexibility to use whatever language you prefer for a given task. Choose from Scala, Java, Python, and others.

• **Real Time Pipeline Monitoring**—Always have an up to date view of your data pipelines and proactively set alerts to quickly catch and fix issues before they become problems for your analytics and machine learning teams.

• **Robust Visual Troubleshooting & Tuning**—Data engineers can efficiently identify and resolve problems for every job with visual troubleshooting tools. Through a single pane of glass, users can make adjustments and review visualized analytics, which helps deliver and continuously maintain production-ready data pipelines.

• **Simplified Job Management**—Platform administrators can use CDE as a centralized way to control costs, security, and governance. CDE empowers users to deliver agile, self-service data engineering for governed data while also regulating provisions and ensuring isolation across business stakeholders.

• **Powerful APIs**—Use APIs to automate the lifecycle management of clusters, applications, and more. This allows you to integrate with partner services for additional authoring, scheduling, and monitoring functionality.

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

Modern Data Engineering Built For Cloudera Data Platform

CDE is built on Cloudera Data Platform (CDP) and is fully integrated with your enterprise data cloud to enable full lifecycle management and seamless collaboration between IT, Data Engineers, and analytic teams. CDE on CDP is available on a cloud native hybrid data architecture enabling best-in-class data engineering on any cloud or data center.
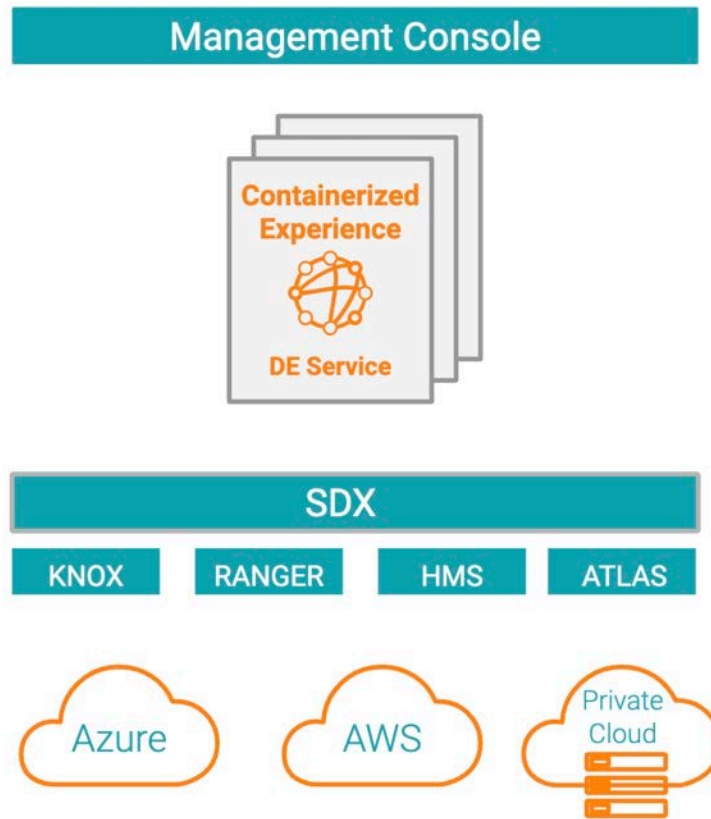

Figure 1. CDE on CDP enables full lifecycle management and analytic service integration on any public or on premises private cloud.

Governance Security & Compliance with Cloudera SDX

Cloudera Shared Data Experience (SDX) ensures full data security, governance and interoperability from the edge to AI — delivering unified data management with policies set once and consistently enforced throughout CDP. Cloudera's unique context layer enables safe and compliant self-service access to data and analytics, offering deeper insights from data with increased agility, at lower cost and with reduced risk. Users can set data and metadata security and governance policies once, and SDX will automatically enforce them in CDE and across all analytics and workloads in CDP.