Cloudera Data Engineering 1.25.0

Managing Cloud Costs in Cloudera Data Engineering

Date published: 2020-07-30
Date modified: 2025-11-20

CLOUDERA

Legal Notice

Contents

Cloudera Data Engineering Spot Instances

Cloudera Data Engineering supports spot instances to facilitate cloud cost savings for workloads that are not SLA-bound. If you have enabled spot instances for your Cloudera Data Engineering service, when you create a virtual cluster, you can specify whether drivers and executors run on spot instances or on-demand instances. The default setting uses on-demand instances for drivers, and spot instances for executors.

> **Note:** Spot Instances are not currently supported for Cloudera Data Engineering on Microsoft Azure.

Spot instances are discounted instances that can be reclaimed at any time by the cloud provider. For fault-tolerant workloads such as Apache Spark, utilizing spot instances can provide significant cost savings. Because spot instances can disappear, impacting job performance, Cloudera recommends only using them for workloads without strict SLA requirements.

For development, testing, and other workloads that have more flexible requirements, Cloudera recommends using on-demand instances for drivers, and spot instances for executors. This reduces the impact of losing a spot instance, because executor failure recovery is more efficient than recovery from a driver failure.

To use spot instances, you must enable them on the Cloudera Data Engineering service. For instructions, see Enabling Cloudera Data Engineering. You cannot enable spot instances on an existing Cloudera Data Engineering service. To use spot instances, create a new service.

After you have enabled spot instances for the Cloudera Data Engineering service, you can configure individual virtual clusters to use on-demand instances only, spot instances only, or the default hybrid model (on-demand for drivers, spot for executors).

For more information on spot instances, see the cloud provider documentation:

Amazon AWS

    Amazon EC2 Spot Instances