Databricks Cluster Configurations

Cluster Types

Compute

All-Purpose Clusters | Job Clusters

+ Create Cluster

All Purpose Cluster

- Need to created manually

- It is Persistent

- Suitable for Interactive workloads

- Shared by Multiple User

- It is Expensive

Job Cluster

- Automatically Created By Job

- Automatically Terminated at End of Job

- Suitable for Automated workloads

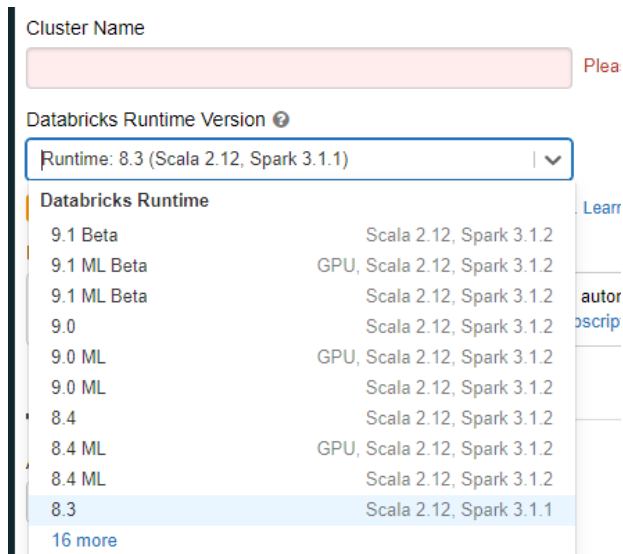- Isolated just for Job

- It is Cheaper

Cluster Modes

Cluster Name		
Test	Test	test

Cluster Mode	Cluster Mode	Cluster Mode
Standard	High Concurrency	Single Node

Standard	High Concurrency	Single Node
Single User Usage	Multiple Users	Single User
No Process Isolation	Provides Process Isolation	No Process Isolation
No Task Preemption	Provides Task Preemption	No Task Preemption
Support for All DSL	Doesn't Support Scala	Support for All DSL
For Production Workload, Ad-hoc Development	For Interactive Analysis, Ad-hoc Development	Light Weight workload for ML, Data Analysis

Runtime

Cluster Name

Please

Databricks Runtime Version ❓

Runtime: 8.3 (Scala 2.12, Spark 3.1.1) ⌄

Databricks Runtime

9.1 Beta	Scala 2.12, Spark 3.1.2	
9.1 ML Beta	GPU, Scala 2.12, Spark 3.1.2	
9.1 ML Beta	Scala 2.12, Spark 3.1.2	
9.0	Scala 2.12, Spark 3.1.2	
9.0 ML	GPU, Scala 2.12, Spark 3.1.2	
9.0 ML	Scala 2.12, Spark 3.1.2	
8.4	Scala 2.12, Spark 3.1.2	
8.4 ML	GPU, Scala 2.12, Spark 3.1.2	
8.4 ML	Scala 2.12, Spark 3.1.2	
8.3	Scala 2.12, Spark 3.1.1	

16 more

Databricks Runtime:

- **Operating System**: Ubuntu 20.04.2 LTS
- **Java**: Zulu 8.54.0.21-CA-linux64
- **Scala**: 2.12.10
- **Python**: 3.8.10
- **R**: 4.1.0 (2021-05-18)
- **Delta Lake** 1.0.0:

Databricks Runtime Photon:

- Faster Delta and Parquet tables

Databricks Runtime Machine Learning (ML):

- Popular ML Libraries
  (**PyTorch**, Keras, TensarFlow, XGBoost etc)

Databricks Runtime Light:

- Runtime option for only jobs
- not requiring advanced features

Auto Termination

- Terminates the cluster after X minutes of inactivity
- Default value for Single Node and Standard Clusters is 120 Minutes
- High Concurrency cluster do not have a default auto termination set
- Users can specify a value between 10 and 10000 mins as the duration

Auto Scaling

- User specify the min and max work nodes
- Auto Scales between min and max based on the workload
- Not recommended for steaming workloads

VM Type/Size

Memory Optimized, Compute Optimized, Storage Optimized, General Purpose, GPU Accelerated

Pool

Clusters   ?   👤

Clusters   **Pools**

[**+ Create Pool**]   ⟳ Refresh

Name ▲	Instance Type	Min Idle	Max Capacity	Idle Instances	Used Instances	Actions
Demo Pool	30.5 GB Memory, 4 Cores	2	10	2	2	✖

1 - 1 of 1   ‹   ›   [20] / Page Go to [1]

Ready to use instances, reduce cluster start and auto scaling times. You can specify difference pool for driver node and worker nodes, or use same pool for both.