

Chronic Conditions Warehouse

Your source for national CMS Medicare and Medicaid research data



Chronic Conditions Warehouse Virtual Research Data Center

Databricks Credit Upgrade and Purchase Policy

JULY 2023 | VERSION 1.0

This page intentionally left blank.

Revision Log

Date	Changed by	Revisions	Version
July 2023	CMS OEDA D. Taylor	Created initial document	1.0

Table of Contents

1.0 Policy Statement	1
2.0 Databricks Credit Purchases	1
3.0 Databricks Cluster Upgrades	1
4.0 Databricks Package Requests	2
Appendix A — List of Acronyms	3

List of Tables

Table 1. Databricks cluster configuration options	1
---	---

1.0 Policy Statement

This document provides Centers for Medicare & Medicaid Services' (CMS) Databricks use policies within the Chronic Conditions Warehouse (CCW) Virtual Research Data Center (VRDC) environment and specifically identifies the policies related to Databrick Unit (DBU) credit purchases, Databricks cluster upgrades, and Databricks package requests.

2.0 Databricks Credit Purchases

A Databricks credit, also known as a DBU, is a normalized unit of processing power on the Databricks platform used for measurement and pricing purposes. Processing metrics drive the number of DBUs a workload consumes, which may include the compute resources used and the amount of data processed.

CCW VRDC seatholders may purchase DBUs up to two times during their project year — at the time of the DUA's project renewal and one additional purchase during the project year. DBUs must be purchased through the DUA Amendment Request process. Contact the Research Data Assistance Center (ResDAC) at resdac@umn.edu or 1-888-973-7322 for assistance with submitting a DUA amendment request.

3.0 Databricks Cluster Upgrades

A Databricks cluster is a set of computational resources users can execute within their Databricks notebook. Users need to attach a notebook to a cluster before executing their code. As stated in the [CCW VRDC Databricks User Guide](#), CCW VRDC users consume Databricks units at one DBU per hour per node. The default Databricks cluster in the CCW VRDC running all five nodes consumes five DBUs per hour with a node type of i3.xlarge.

CCW VRDC project teams are required to specify their Databricks cluster size when initially setting up their DUA. In addition to the default cluster, two additional cluster configuration options, described in [Table 1](#), are available. If an upgraded cluster is not requested at the time of your initial DUA request, the CCW team will assign the DUA to the default cluster.

Project teams may subsequently request changes to your Databricks cluster size annually during the project fee renewal process. These requests cannot be made during your project year. There are no additional fees associated with cluster changes. Contact ResDAC for assistance with submitting a DUA amendment request to renew your CCW VRDC project fee.

Table 1. Databricks cluster configuration options

Cluster type	Node details	DBU consumption
Standard data analysis cluster (default)	<ul style="list-style-type: none"> Driver Node — one i3.xlarge (4 vCPU, 30.5 GB memory) Worker Nodes — up to 4 i3.xlarge 	<ul style="list-style-type: none"> This cluster consumes 1 DBU per hour per node A cluster running all five nodes consumes 5 DBUs per hour
Enhanced data analysis cluster	<ul style="list-style-type: none"> Driver Node — one i3.8xlarge (32 vCPU, 244 GB memory) Worker Nodes — up to 4 i3.8xlarge 	<ul style="list-style-type: none"> This cluster consumes 8 DBUs per hour per node A cluster running all five nodes consumes 40 DBUs per hour
Artificial intelligence (AI)/Machine Learning cluster	<ul style="list-style-type: none"> Driver Node — one c5d.9xlarge (36 vCPU, 72 GB memory) Worker Nodes — up to 8 c5d.9xlarge 	<ul style="list-style-type: none"> This cluster consumes 5.46 DBUs per hour per node A cluster running all five nodes consumes 49.14 DBUs per hour

4.0 Databricks Package Requests

CCW VRDC Databricks users must follow the CCW software request process and adhere to the [CCW Software Request End-User Guidelines](#). Refer to the [CCW VRDC Databricks User Guide](#) for additional information on requesting Databricks packages.

Appendix A — List of Acronyms

Acronym	Definition
AI	Artificial Intelligence
CARS	CCW Access Request System
CCW	Chronic Conditions Warehouse
CMS	Centers for Medicare & Medicaid Services
DART	Data Access Request Tracking
DBU	Databricks Unit
DUA	Data Use Agreement
HI	Health Informatics
ML	Machine Learning
OY	Option Year