

## Databricks configuration

The screenshot displays the Databricks console interface for configuring a cluster named 'Rahul Kaileasa's Cluster'. The left sidebar contains navigation options like 'New', 'Workspace', 'Recent', 'Catalog', 'Workflows', 'Compute', 'SQL', 'SQL Editor', 'Demos', 'Dashboards', 'Jobs', 'Alerts', 'Query History', 'SQL Worksheets', 'Data Engineering', 'Job Run', 'Data Ingestion', 'Data Line Tables', 'Machine Learning', 'Playground', 'Experiments', 'Features', 'Models', and 'Serving'.

The main panel shows the 'Configuration' tab for the cluster. Key settings include:

- Policy:** Unrestricted
- Access mode:** Single user access
- Single user:** Rahul Kaileasa
- Performance:** Databricks Runtime Version 16.4 LTS (includes Apache Spark 3.5.0, Scala 2.12)
- Use Photon Acceleration:** Enabled
- Worker type:** m5d.xlarge (32 GB Memory, 4 Cores)
- Driver type:** m5d.xlarge (32 GB Memory, 4 Cores)
- Instance profile:** None
- Tags:** No custom tags
- Advanced options:** On-demand/Spot composition (1 On-demand first, followed by 8 Spot; Spot fall back to On-demand)
- Instance:** m5d.xlarge, Logging: text, format: JSON, size: 10GB
- Availability zone:** us-east-1
- Max spot price:** 100% of on-demand instance price

A 'Summary' panel on the right provides a quick overview: 2-8 Workers, 64-256 GB Memory, 8-32 Cores, 1 Driver, 32 GB Memory, 4 Cores, Runtime 16.4-latest-LTS, and a total cost of 6-18.08h.

## S3 Configuration

The screenshot shows the AWS S3 console interface for the 'Data\_2020\_2024/' bucket. The left sidebar lists navigation options: 'Amazon S3', 'General purpose buckets', 'Directory buckets', 'Table buckets', 'Access Grants', 'Access Points', 'Object Lambda Access Points', 'Multi-Region Access Points', 'Batch Operations', 'IAM Access Analyzer for S3', 'Block Public Access settings for this account', 'Storage Lens', 'Dashboards', 'Storage Lens groups', 'AWS Organizations settings', 'Feature spotlight', and 'AWS Marketplace for S3'.

The main panel displays the 'Data\_2020\_2024/' bucket with the 'Objects' tab selected. It shows a list of 8 objects, all of type 'parquet' and stored in the 'Standard' storage class. The objects are:

Name	Type	Last modified	Size	Storage class
spot_15m.parquet	parquet	December 2, 2024, 21:10:46 (UTC-08:00)	2.5 GB	Standard
spot_1h.parquet	parquet	December 2, 2024, 21:11:36 (UTC-08:00)	648.0 MB	Standard
spot_30m.parquet	parquet	December 2, 2024, 21:12:51 (UTC-08:00)	1.3 GB	Standard
spot_5m.parquet	parquet	December 2, 2024, 21:20:20 (UTC-08:00)	7.0 GB	Standard
um_15m.parquet	parquet	December 2, 2024, 21:22:05 (UTC-08:00)	1.2 GB	Standard
um_1h.parquet	parquet	December 2, 2024, 21:22:31 (UTC-08:00)	313.9 MB	Standard
um_30m.parquet	parquet	December 2, 2024, 21:23:08 (UTC-08:00)	638.8 MB	Standard
um_5m.parquet	parquet	December 2, 2024, 21:26:06 (UTC-08:00)	5.4 GB	Standard

## EC2 Configuration

Dashboard

EC2 Global View

Events

Instances

Instance Types

Launch Templates

Spot Requests

Savings Plans

Reserved Instances

Dedicated Hosts

Capacity Reservations

Images

AMI Catalog

Elastic Block Store

Volumes

Snapshots

Lifecycle Manager

Network & Security

Security Groups

Elastic IPs

Placement Groups

Key Pairs

Network Interfaces

Load Balancing

Load Balancers

Target Groups

Trust Stores

Auto Scaling

Auto Scaling Groups

Settings

EC2 > Instances > i-0d1df4daa7ac71235

Instance summary for i-0d1df4daa7ac71235 (CSV-Processing-Instance)

Updated less than a minute ago

Instance ID

i-0d1df4daa7ac71235

IPv4 address

-

IPv6 address

-

Hostname type

IP name: ip-172-31-52-129-us-east-2.compute.internal

Answer private resource DNS name

IPv4 (A)

Auto-assigned IP address

3.16.215.242 [Public IP]

IAM Role

ec2-S3-Access-Role

IMDSv2

Required

Operator

-

Public IPv4 address

3.16.215.242 [open address]

Instance state

Running

Private IP DNS name (IPv4 only)

ip-172-31-52-129-us-east-2.compute.internal

Instance type

r5.2xlarge

VPC ID

vpc-0f5e84bfc014899ad

Subnet ID

subnet-0b0345e1b7d936ae9

Instance ARN

arn:aws:ec2:us-east-2:396608785330:instance/i-0d1df4daa7ac71235

Connect

Instance state

Actions

Private IPv4 addresses

172.31.32.129

Public IPv4 DNS

ec2-3-16-215-242-us-east-2.compute.amazonaws.com [open address]

Elastic IP addresses

-

AWS Compute Optimizer finding

Opt-in to AWS Compute Optimizer for recommendations. [Learn more]

Auto Scaling Group name

-

Managed

false

Details

Status and alarms

Monitoring

Security

Networking

Storage

Tags

Networking details

Public IPv4 address

3.16.215.242 [open address]

Public IPv4 DNS

ec2-3-16-215-242-us-east-2.compute.amazonaws.com [open address]

Subnet ID

subnet-0b0345e1b7d936ae9

Availability zone

us-east-2c

Use RBN as guest OS hostname

Disabled

Private IPv4 addresses

172.31.32.129

Private IP DNS name (IPv4 only)

ip-172-31-52-129-us-east-2.compute.internal

IPv6 addresses

-

Carrier IP addresses (ephemeral)

-

Answer RBN DNS hostname IPv4

Enabled

VPC ID

vpc-0f5e84bfc014899ad

Secondary private IPv4 addresses

-

Output ID

-