DELL Technologies

Solution Guide

Dell EMC PowerScale Powered by Azure Databricks and Faction to accelerate data-driven innovations

Unified data analytics platform: One cloud platform for massive scale data engineering and collaborative data science

Abstract

This paper describes the solution and implementation process of setting up a unified data analytics platform solution, for accelerated data driven innovations powered by Azure Databricks, Faction cloud, and Dell EMC PowerScale.

December 2020

H18628

Revisions

Date	Description
December 2020	Initial release

Acknowledgments

Author: Kirankumar Bhusanurmath, Analytics Solutions Architect, Dell EMC

Support: Anjan Dave, Advisory System Engineer, Dell EMC

Other:

Table of contents

Executive summary

The unified data analytics platform provides a cloud platform solution for massive scale data engineering and collaborative data science workloads for the on-premises data stored on Dell EMC PowerScale data lakes. This solution provides data science workspace collaboration across the full data and machine learning (ML) life cycle, through collaborative notebooks, optimized ML environments, and complete ML life cycles. Solution's unified data services provide high-quality data with great performance, through reliable data lakes, fast and efficient data pipelines and broader business insights. Finally this solution's enterprise cloud service provides a massively scalable and secure multicloud service through platform security, 360-degree administration, elastic scalability, and multicloud management.

To enable this unified data analytics platform, Dell EMC Cloud Storage Services has combined Isilon/PowerScale, the number one scale-out NAS platform powered by OneFS, with the Microsoft Azure public cloud's Databricks service, which offers enterprise-grade Apache Spark compute for operational flexibility. This integration provides a high bandwidth (up to 100 Gbps), low latency (as low as 1.2 milliseconds) connection from Isilon to Azure Databricks using Azure ExpressRoute Local. It also eliminates outbound data traffic costs for data written to Isilon from within Azure. The integration is powered Faction, that provides a fully managed cloud data services platform, along with patented low latency, high throughput connectivity that can deliver ultrahigh performance from PowerScale systems that are next to Azure cloud.

Cloud Storage Services with Azure and Isilon allows for the right combination of compute and storage for data-intensive, high I/O throughput, file-based workloads that require high compute performance on a periodic and/or unpredictable basis. This makes them suitable for a cloud consumption model. Eliminating egress charges enables workloads that require a lot of temporary writes to the Isilon to cost-effectively take advantage of Azure's application services. This is ideal for industries such as Life Sciences and Media and Entertainment, which can require on-demand computing power tied to a massive file system.

For compute, Azure offers the choice of dozens of VMs with a wide variety of CPUs, some optimized for HPC workloads, memory capacity, and network options. For the current solution, we focus only on the Azure Databricks which is a fast, easy, and collaborative Apache Spark™ based analytics service. When combined with Isilon's unmatched performance, reliability and scalability, and a single multi petabyte namespace which supports symmetric data access across its nodes, organizations get a fully managed cloud service that can address the most demanding requirements.

1 Solution overview

Dell Technologies Cloud Storage enables connecting file storage, consumed as a service, directly to the Azure Databricks Apache Spark cluster. This is achieved through native replication from on-premises Dell EMC Isilon storage to a managed service provider location. Dell Technologies has partnered with Faction Inc. to deliver a fully managed, cloud-based service for Dell EMC storage to address various cloud use cases.

Faction, Inc. is a Dell Technologies Gold Cloud Service Provider (CSP) and Tech Connect Select partner founded in 2006 and headquartered in Denver, Colorado. Faction is a multicloud platform-as-a-service provider and VMware partner that offers multicloud-attached storage from various co-locations (Equinix, Coresite, and Digital Reality). Faction has expanded globally to London and Frankfurt. In this hybrid cloud data warehouse solution, we use Factions Cloud Control Volumes (CCVs) storage offerings as storage layer or data lake for Azure cloud.

Native cloud integration

Azure Databricks (a fully managed Databricks service) is used directly with the storage solution, showing interoperability beyond pure mounts to instances; also showing multiprotocol access using **hdfs://** (default CCVs use NFS).

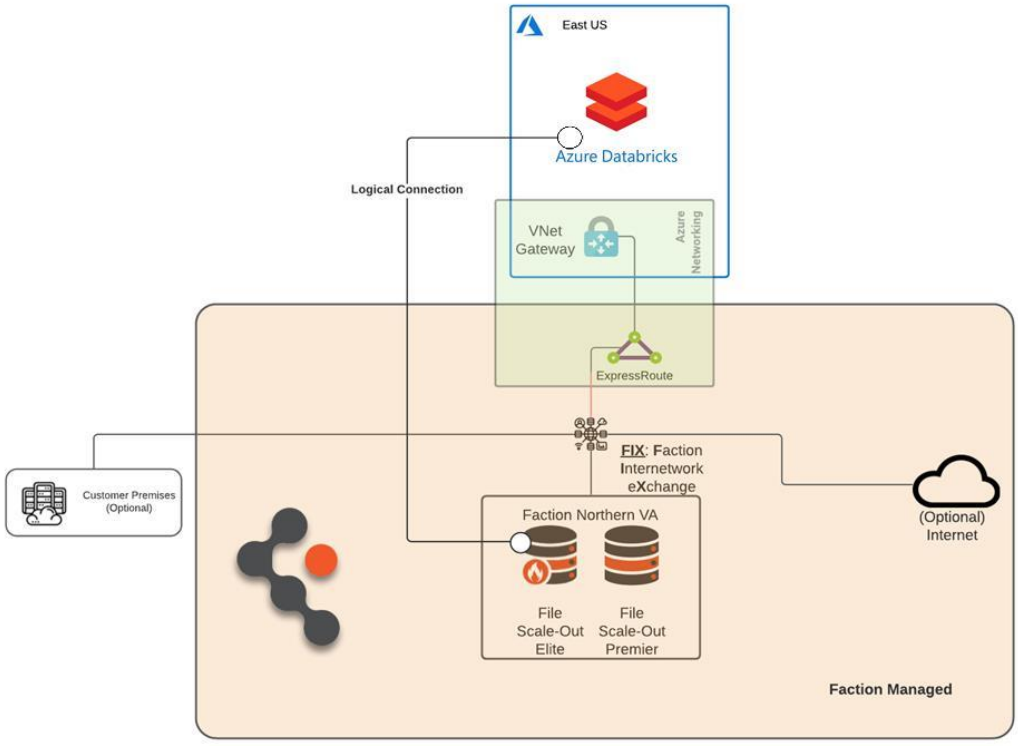


Figure 1 Unified data analytics solution diagram

1.1 Faction Cloud Control Volumes

Cloud Control Volumes (CCVs) provide durable, persistent, cloud-attached, and cloud-adjacent storage directly connected to the Azure cloud.

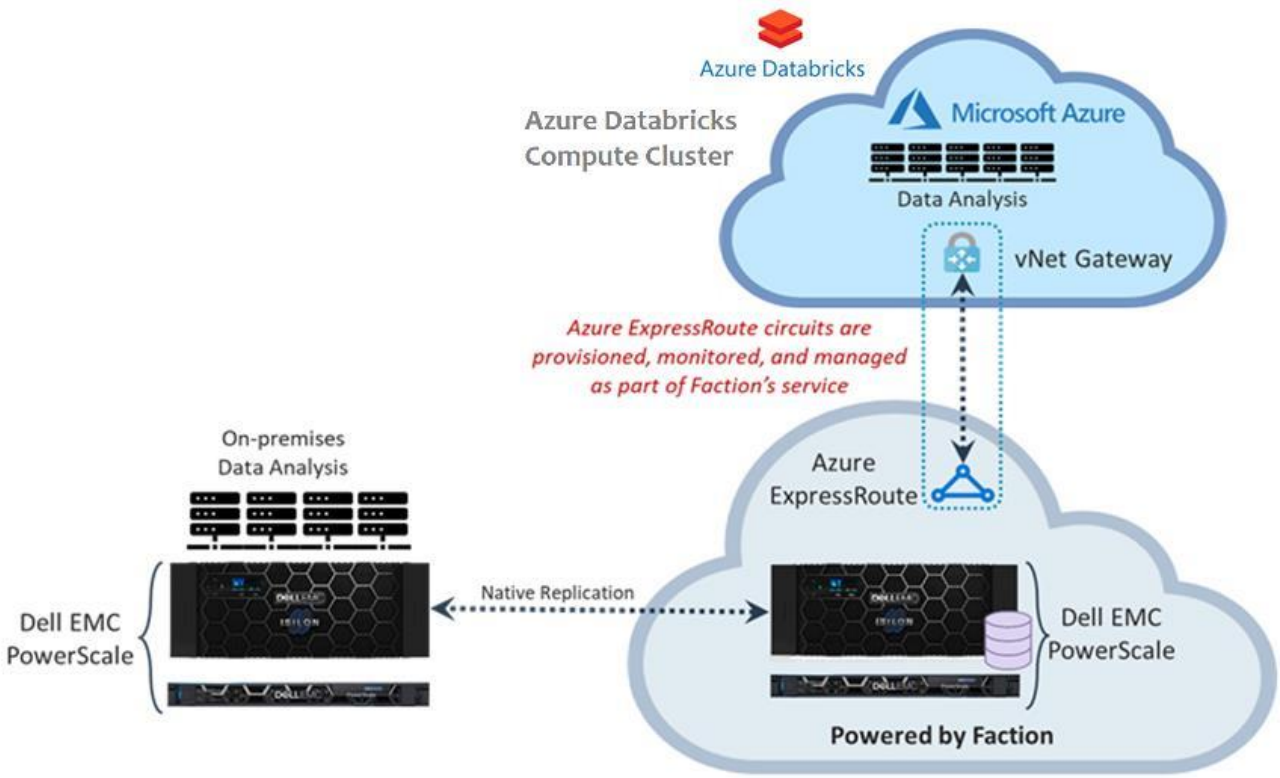


Figure 2 Array-based replication of volumes to Faction directly attached as CCVs across one or more clouds through NFS

Use cases for CCVs could be transient in nature, such as performing data analytics on a large or complex data footprint. A verity of tiers of CCV storages is available in Faction data center. Storage tier specifics are ultimately determined by the Dell EMC arrays and use cases as shown in the below figure.

Archive (Small)	Archive	Standard	Premier	Elite (Small)	Elite
Base Network Connectivity					
10 Gb/s	10 Gb/s	20 Gb/s	40 Gb/s	20 Gb/s	80 Gb/s
Model					
A200	**A2000**	**H5600**	**H500**	**F600**	**F800**
Storage Scaling					
Base includes 162 TB	Base includes 648 TB	Base includes 540 TB	Base includes 162 TB	Base includes 28 TB	Base includes 130 TB
Scale in 90 TB increments	Scale in 360 TB increments	Scale in 300 TB increments	Scale in 90 TB increments	Scale in 12 TB increments	Scale in 77 TB increments

Workloads					
Write-Once-Read-Never/Retention Data Long-term Healthcare Records Retention Long-term Legal Records Retention Web content Management Video Retention Tiering from Standard/Premier/Elite	Write-Once-Read-Never/Retention Data Long-term Healthcare Records Retention Long-term Legal Records Retention Web content Management Video Retention Tiering from Standard/Premier/Elite	Write-Once-Read-Never/Retention Data Long-term Healthcare Records Retention Long-term Legal Records Retention Web content Management Video Retention Tiering from Standard/Premier/Elite	Video Streaming Media Processing Rendering Replace on-premise file servers Test/Dev Big data uses (for example, Genomics, Machine Learning, and so on) Cloud User-level Windows File Sharing	Real-time inference (machine learning) Critical Streaming Analytics Rendering Time-sensitive Data Warehouse workloads Small footprint flash workloads	Real-time inference (machine learning) Critical Streaming Analytics Rendering Time-sensitive Data Warehouse workloads

Figure 3 File scale-out CCV details

For big data analytics, organizations need to migrate volume data from an on-premises data center to a Faction data center. Array-based replication is configured between on-premises Isilon storage and a similar Isilon storage array owned and managed by Faction in the Faction data center.

It is the customer's responsibility to manage the network between their on-premises data center and the Faction data center. A dedicated circuit: should be opted for a dedicated connection for replication traffic between their facility and Faction. Customers may also use a VPN as redundancy to a dedicated link. Faction can source and manage the dedicated link, or the client can work with their carrier directly.

CCVs are presented in close proximity to Azure cloud provider while leveraging redundant connectivity with multiple 10 Gb Ethernet connections and redundant switches to provide highly available connections. Link Aggregation Groups (LAGs) are used to scale to higher levels of bandwidth into the Azure cloud.

2 Solution components

2.1 Azure Databricks

Fast, easy and collaborative Apache Spark™ based analytics service, for big data analytics and AI with optimized Apache Spark. To unlock insights from all your data and build artificial intelligence (AI) solution. With Azure Databricks setup your Apache Spark environment in minutes, autoscale and collaborate on the shared projects in an interactive workspace. Azure Databricks supports Python, Scala, R, Java and SQL, as well as data science frameworks and libraries including TensorFlow, PyTorch and scikit-learn.

See here for more information about the Azure Databricks.

2.2 Dell EMC PowerScale

PowerScale is the next evolution of OneFS – the operating system powering the industry's leading scale-out NAS platform. The PowerScale family includes Dell EMC PowerScale platforms and the Dell EMC Isilon platforms configured with the PowerScale OneFS operating system. OneFS provides the intelligence behind the highly scalable, high-performance modular storage solution that can grow with your business. A OneFS powered cluster is composed of a flexible choice of storage platforms including all-flash, hybrid, and archive nodes. These solutions provide the efficiency, flexibility, scalability, security, and protection for you to store massive amounts of unstructured data within a cluster. The new PowerScale all-flash platforms co-exist seamlessly in the same cluster with your existing Isilon nodes to drive your traditional and modern applications

See here for more information about the Dell EMC PowerScale platforms.

3 Solution implementation and validation

Note: The solution is validated functionally, no performance related testing is conducted or presented in this guide.

3.1 Preparing OneFS

Complete the following steps to configure your Isilon OneFS cluster for use with Azure Databricks cluster. Preparing OneFS requires you to configure DNS, SmartConnect, and Access Zones to allow for the Databricks cluster to connect to the Isilon OneFS cluster. If these preparation steps are not successful, the subsequent configuration steps might fail.

Note: For validation purpose, we will skip DNS and SmartConnect configuration. Only setup Access Zone (optional) and use IP address of the Isilon End point from Faction Cloud.

3.1.1 Validate OneFS version and license activation

You must validate your OneFS version, check your licenses, and confirm that they activated.

1. From a node in your Isilon OneFS cluster, confirm the OneFS version using below command.

```
Isi version

Isilon OneFS v9.0.0.0 B_9_0_0_002(RELEASE): 0x900005000000002:Thu Apr 23 13:04:16
PDT 2020    root@sea-build11-
112:/b/mnt/obj/b/mnt/src/amd64.amd64/sys/IQ.amd64.release   FreeBSD clang version
5.0.0 (tags/RELEASE_500/final 312559) (based on LLVM 5.0.0svn).
```

2. Add the license for HDFS and SmartConnect Advanced using the following command:

```
isi license add --evaluation=SMARTCONNECT_ADVANCED,HDFS
```

3. Confirm that licenses for HDFS and SmartConnect Advanced are operational. If these licenses are not active and valid, some commands in this guide might not work.
   Run the following commands to confirm that HDFS and SmartConnect Advanced are installed:

```
isi license licenses list
isi license licenses view HDFS
isi license licenses view "SmartConnect Advanced"
```

4. If your modules are not licensed, obtain a license key from your Dell EMC Isilon sales representative. Type the following command to activate the license:

```
isi license add --path <license file path>
```

3.1.2 Configure OneFS components

After you configure DNS for OneFS, set up and configure the following OneFS components.

- Create an access zone
- (Optional) Create a SmartConnect zone
- Create and configure the HDFS root in the access zone
- (Optional) Create users and groups
- Enable hdfs service

3.1.2.1 Create an access zone

On one of the Isilon nodes, you must define an access zone on the Isilon OneFS cluster and enable the Hadoop(hdfs) node to connect to it.

1. On a node in the Isilon OneFS cluster, create your Hadoop access zone.
```
isi zone zones create --name=hdfs-zone --path=/ifs/hdfs-zone --create-path
```

2. Verify that the access zones are set up correctly.
```
isi zone zones list -verbose
```

   Output similar to the following appears:
```
                Name: System
                Path: /ifs
           Groupnet: groupnet0
      Map Untrusted: -
     Auth Providers: lsa-local-provider:System, lsa-file-provider:System
       NetBIOS Name: -
  User Mapping Rules: -
Home Directory Umask: 0077
  Skeleton Directory: /usr/share/skel
  Cache Entry Expiry: 4H
            Zone ID: 1
-------------------------------------------------------------------------------
                Name: hdfs-zone
                Path: /ifs/hdfs-zone
           Groupnet: groupnet0
      Map Untrusted: -
     Auth Providers: lsa-local-provider:hdfs-zone
       NetBIOS Name: -
  User Mapping Rules: -
Home Directory Umask: 0077
  Skeleton Directory: /usr/share/skel
  Cache Entry Expiry: 4H
            Zone ID: 2
```

3. Create the HDFS root directory within the access zone that you created.
```
mkdir -p /ifs/hdfs-zone
```

4. List the contents of the Hadoop access zone root directory.
```
ls -al /ifs/hdfs-zone
```

3.1.3 Create Network pool and SmartConnect

Note: In this validation, we have not setup a SmartConnect FQDN.

On a node in the Isilon OneFS cluster, add a dynamic IP address pool and associate it with the access zone you created earlier.

1. Modify your existing subnets and specify a service address.
```
isi network subnets modify groupnet0.subnet0 --sc-service-addr=x.x.x.x
```

2. Create network pool for the hdfs access zone.

```
isi network pools create --id=<groupnet>:<subnet>:<name> --ranges=x.x.x.x-x.x.x.x --access-zone=<my-access-zone> --alloc-method=dynamic --ifaces=X-Y: <your interfaces>  --sc-subnet=subnet0 --sc-dns-zone=<my-smartconnectzone-name> --description=hadoop
```

Where:

`--name subnet: <poolname>`—New IP pool in subnet (for example, `subnet0:pool1`).

`--ranges`—IP range that is assigned to the IP pool

`--ifaces`—Node interfaces that are added to the pool

`--access-zone`—Access zone that the pool is assigned to.

`--sc-dns-zone`—SmartConnect zone name

`--sc-subnet`—SmartConnect service subnet that is responsible for this zone

3. View the properties of the existing pool.

```
isi network pools view groupnet0.production.pool-hdfs
                     ID: groupnet0.production.pool-hdfs
               Groupnet: groupnet0
                 Subnet: production
                   Name: pool-hdfs
                  Rules: -
            Access Zone: hdfs-zone
      Allocation Method: static
       Aggregation Mode: lacp
     SC Suspended Nodes: -
            Description:
                 Ifaces: 1:10gige-1, 2:10gige-2, 3:10gige-1, 4:10gige-2
              IP Ranges: 10.1.1.15-10.1.1.18
       Rebalance Policy: auto
SC Auto Unsuspend Delay: 0
      SC Connect Policy: round_robin
                SC Zone:
    SC DNS Zone Aliases: -
     SC Failover Policy: round_robin
              SC Subnet: production
                 SC TTL: 0
          Static Routes: -
```

3.1.3.1 Create and configure the HDFS root in the access zone

On a node in the Isilon OneFS cluster, create new role and configure the backup and restore privileges to the hdfs user.

1. Create new role for the Hadoop access zone

```
isi auth roles create --name=<role_name> --description=<role_description> --zone=<access_zone>
```

For example:

```
isi auth roles create --name=RunAsRoot --description="Bypass FS permissions" --zone=hdfs-zone
```

2. Add restore privileges to the new “RunAsRoot” role

```
isi auth roles modify <role_name> --add-priv=ISI_PRIV_IFS_RESTORE --
zone=<access_zone>
```

For example:

```
isi auth roles modify RunAsRoot --add-priv=ISI_PRIV_IFS_RESTORE --zone=hdfs-zone
```

3. Add backup privileges to the new “RunAsRoot” role

```
isi auth roles modify <role_name> --add-priv=ISI_PRIV_IFS_BACKUP --
zone=<access_zone>
```

For example:

```
isi auth roles modify RunAsRoot --add-priv=ISI_PRIV_IFS_BACKUP --zone=hdfs-zone
```

4. Add user hdfs to the new “RunAsRoot” role

```
isi auth roles modify <role_name> --add-user=hdfs --zone=<access_zone>
```

For example:

```
isi auth roles modify RunAsRoot --add-user=hdfs --zone=hdfs-zone
```

5. Verify the role setup, backup / restore privileges and hdfs user setup.

```
isi auth roles view <role_name> --zone=<access_zone>
```

For example:

```
isi auth roles view RunAsRoot --zone=hdfs-zone
       Name: RunAsRoot
Description: Bypass FS permissions
    Members: - hdfs
Privileges
             ID: ISI_PRIV_IFS_BACKUP
      Read Only: True

             ID: ISI_PRIV_IFS_RESTORE
      Read Only: True
```

6. (Optional) Flush auth mapping and auth cache to make hdfs user take immediate effect as “RunAsRoot” role created above.

```
isi_for_array "isi auth mapping flush --all"
isi_for_array "isi auth cache flush --all"
```

7. Alternate way is to add hdfs user to the ZoneAdmin role as below.

```
isi auth users view --user=hdfsuser --zone=hdfs-zone
                    Name: hdfsuser
                      DN: CN=hdfsuser,CN=Users,DC=DC15-ISI04
              DNS Domain: -
                  Domain: DC15-ISI04
                Provider: lsa-local-provider:hdfs-zone
        Sam Account Name: hdfsuser
                     UID: 2000
                     SID: S-1-5-21-1437622239-4266375620-3563931565-1000
                 Enabled: Yes
                 Expired: No
                  Expiry: -
```

```
                  Locked: No
                   Email: -
                   GECOS: -
            Generated GID: No
            Generated UID: No
            Generated UPN: Yes
            Primary Group
                          ID: GID:1800
                        Name: Isilon Users
          Home Directory: /ifs/hdfs-zone/home/hdfsuser
        Max Password Age: 4W
        Password Expired: No
         Password Expiry: 2020-12-22T15:50:52
       Password Last Set: 2020-12-22T12:43:31
        Password Expires: No
                   Shell: /bin/zsh
                     UPN: hdfsuser@DC15-ISI04
User Can Change Password: Yes
```

3.1.3.2 Enable hdfs service

By default, hdfs and SmartConnect services are disabled in OneFS 9.x, these services needs to be manually enabled to connect to the access zone using hdfs protocol.

```
isi services hdfs enable
```

3.2 Preparing Azure Databricks

Below steps are referred from the Azure Databricks documentation to create a new Azure Databricks workspace and Spark cluster.

1. Login into Azure portal and search Azure Databricks service
   Click on Add to add a new workspace, select resource group, region and pricing tiers, for validation we have Trial (Premium – 14 days Free DBUs).

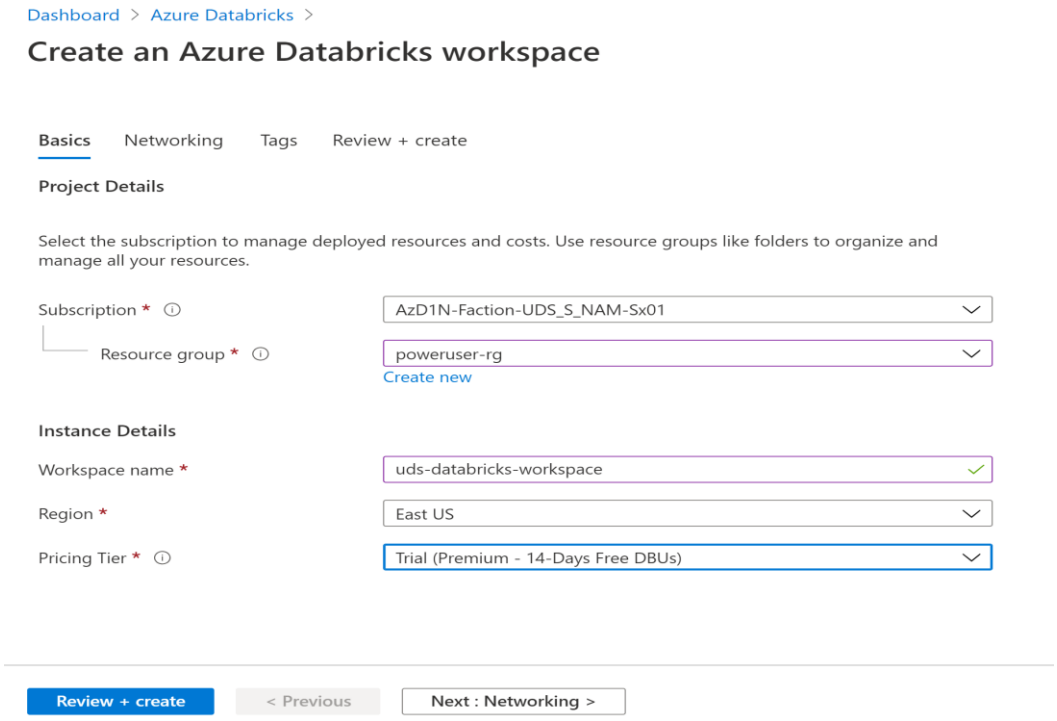


Figure 4 Create an Azure Databricks workspace part 1

2. Networking
   Choose existing Virtual Network and subnet not in use.

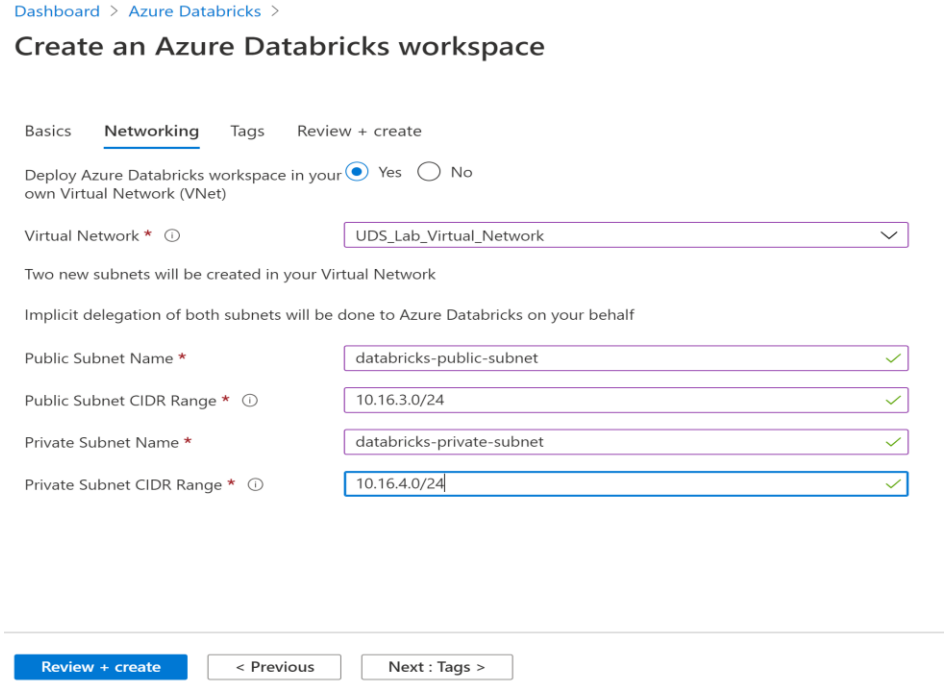


Figure 5 Create an Azure Databricks workspace part 2

3. Tag is optional, click Review/Create

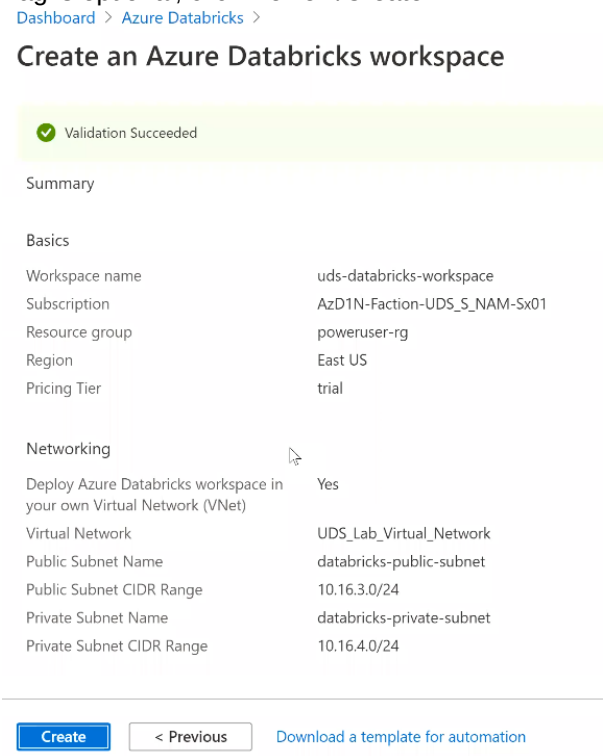


Figure 6 Create an Azure Databricks workspace part 3

4. Give it 10mins to create the workspace, then click on the Workspace, and then click on Launch Workspace:

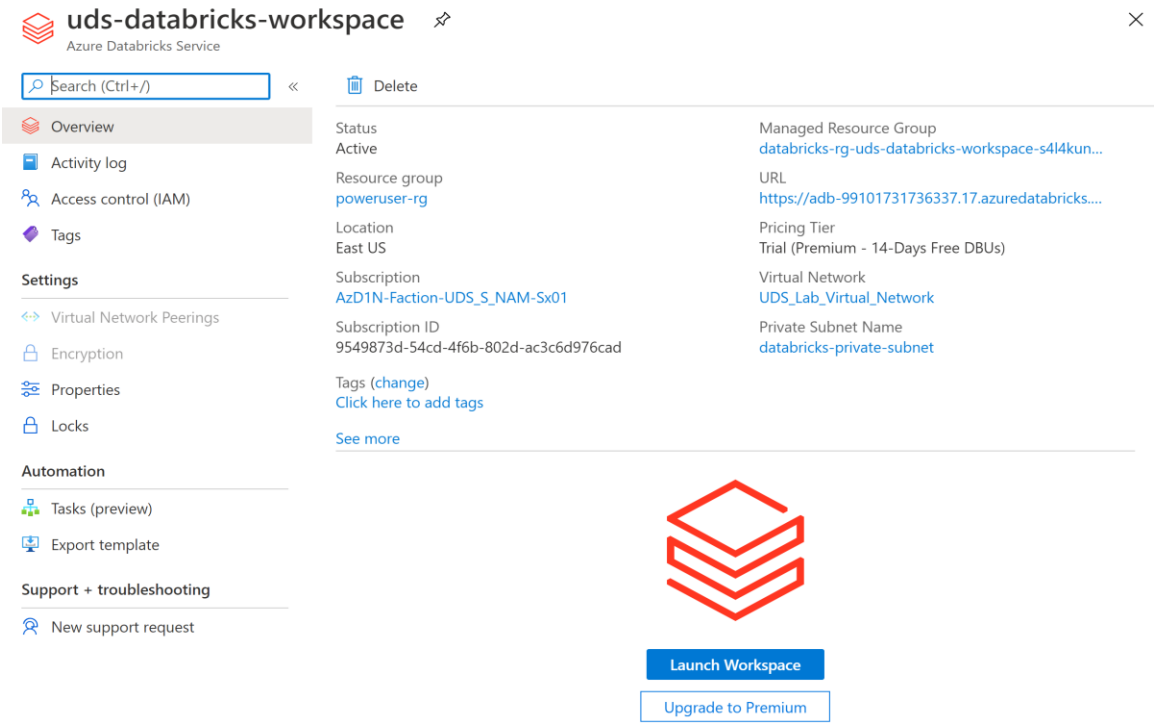


Figure 7 Create an Azure Databricks workspace part 4

5. The new Azure Databricks workspace created with open in a new page as below.

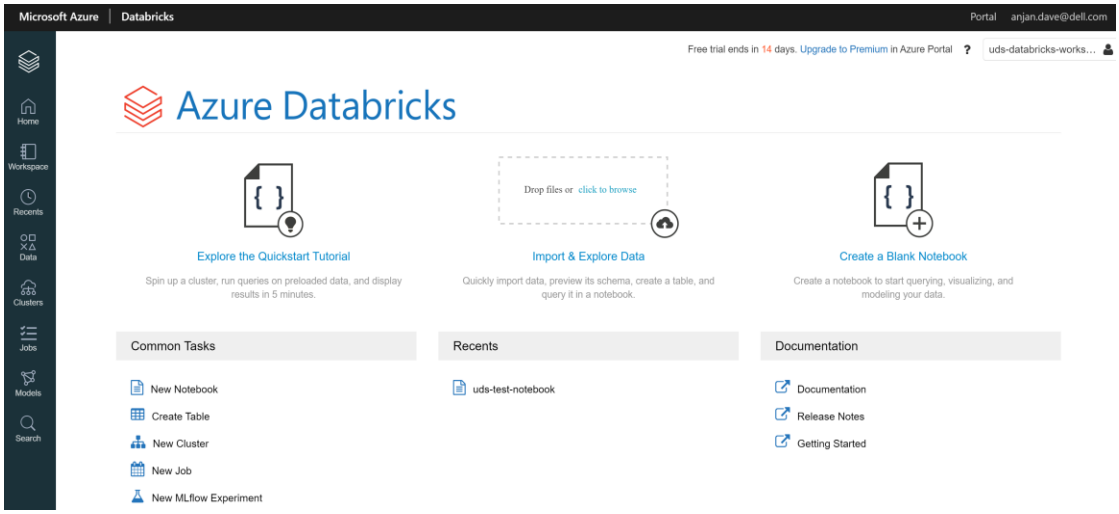


Figure 8 Create an Azure Databricks workspace part 5

6. Click on the New Cluster and create a new Cluster within the workspace.

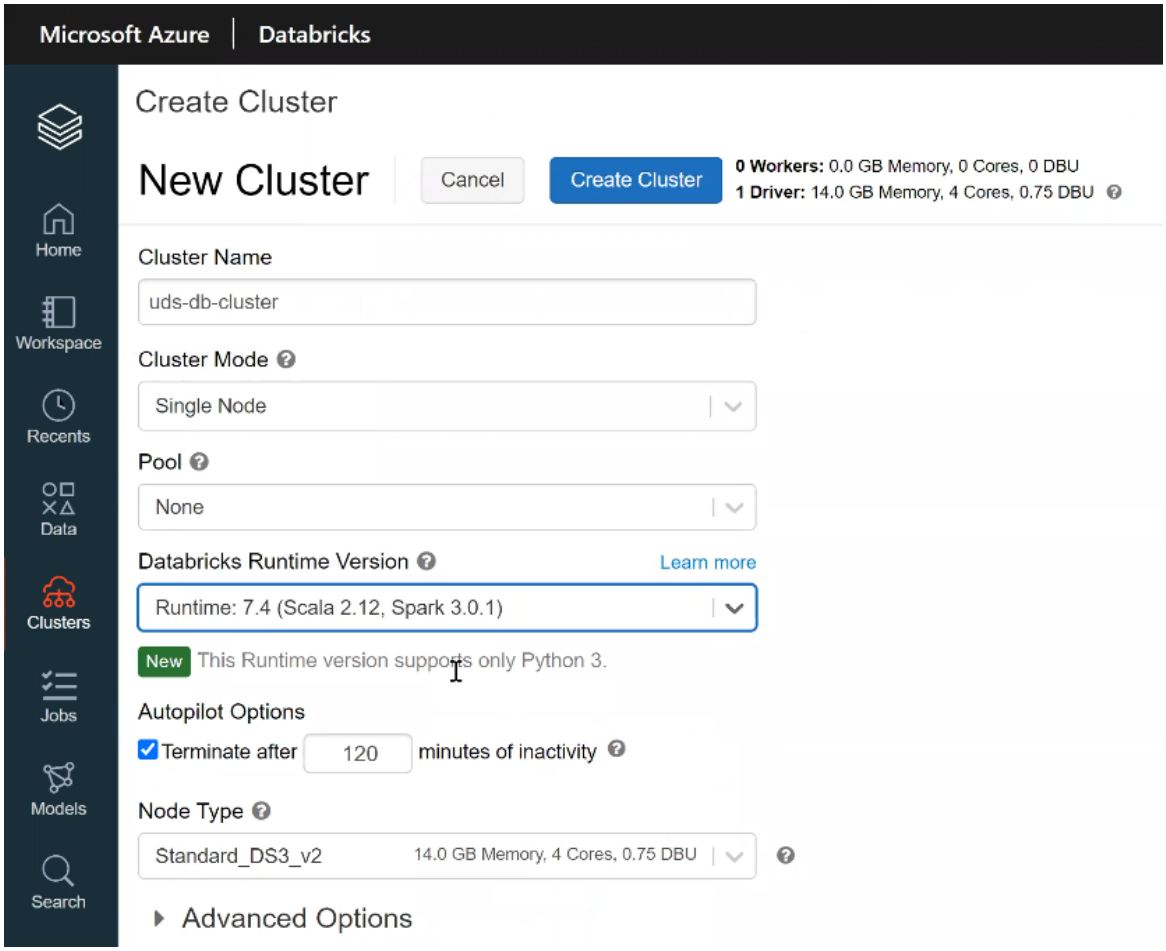


Figure 9 Create an New Spark cluster within Azure Databricks workspace

7. Create new Notebook within the new cluster created.

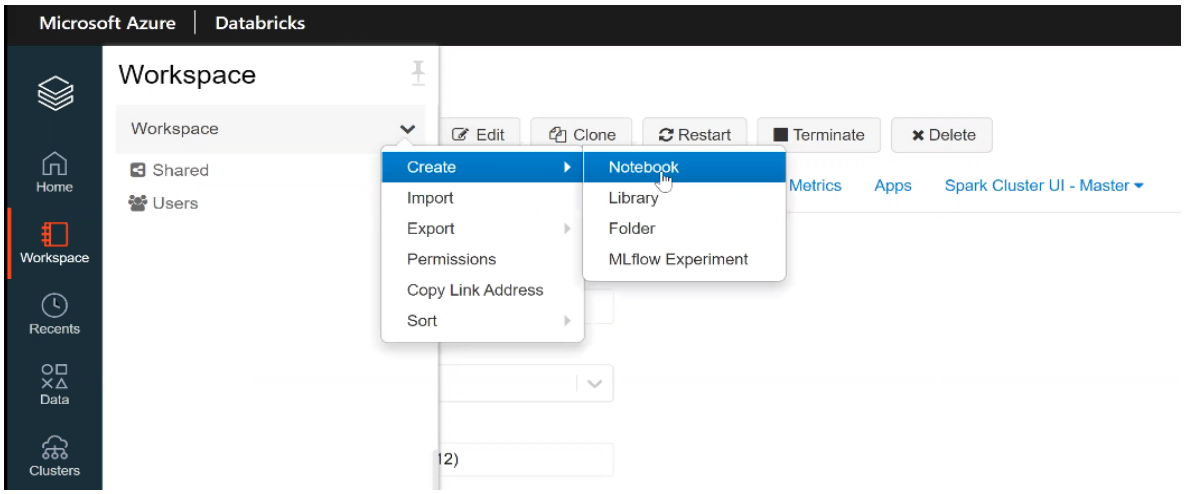


Figure 10 Create an New Notebook within new Spark cluster inside Azure Databricks workspace part1

8. Click on the New Notebook, int his case, Python was selected as the language), and then you can start typing the Spark commands.

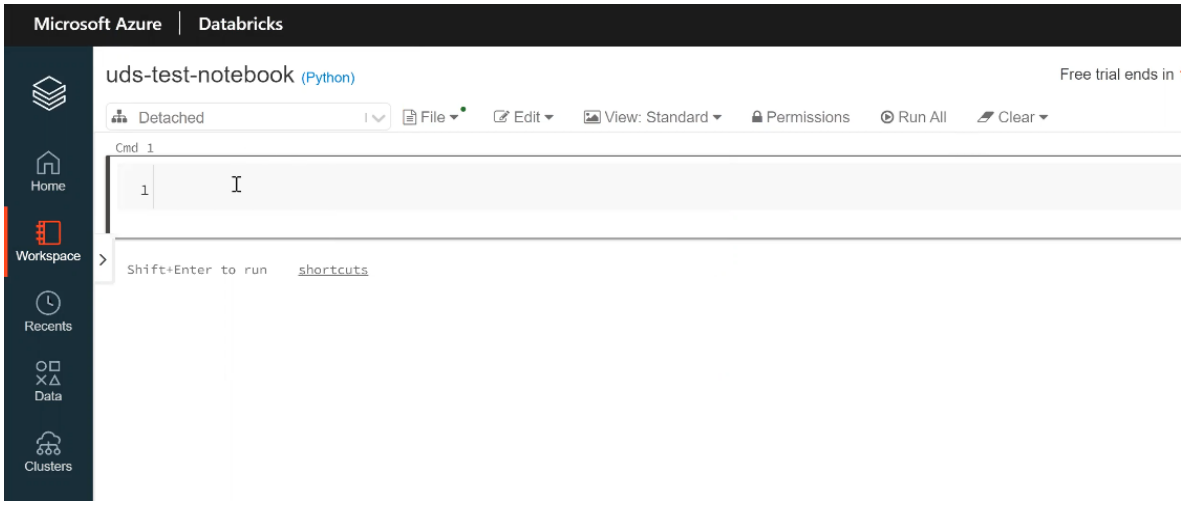


Figure 11 Create an New Notebook within new Spark cluster inside Azure Databricks workspace part2

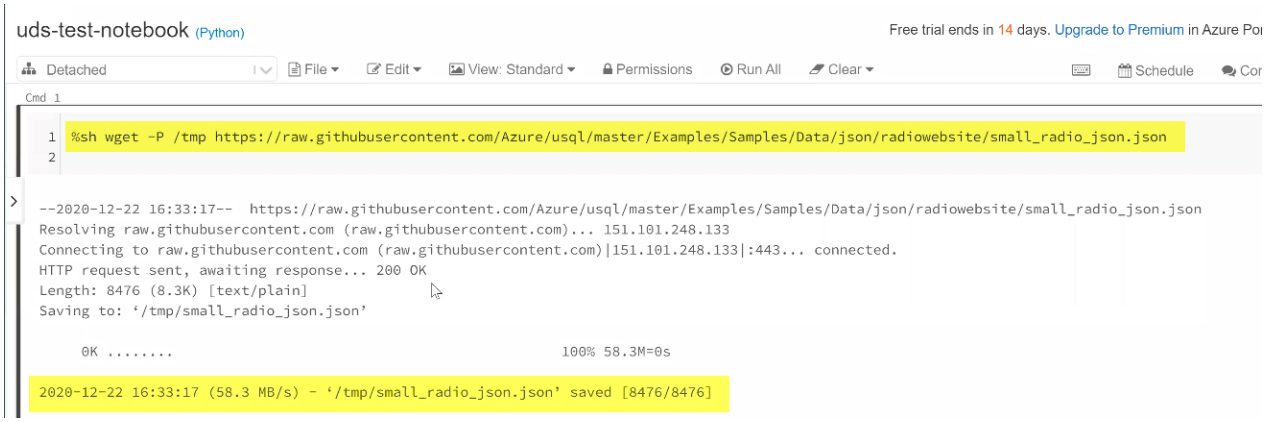
3.3 Solution validation

In this section we will demonstrate the unified data analytics platform solution validation, how the data from On-premises data center is replicated into Isilon in the Faction cloud, and the same is made available to the Azure Databricks cluster on Azure public cloud for in place data analytics.

Note: For simplicity purpose we are using Isilon IP provided by Faction as the data endpoint, Databricks Spark cluster can connect to this endpoint using HDFS protocol and read/write data into Isilon. If the Isilon is configured with a DNS and SmartConnect is enabled, then a Fully Qualified Domain Name (FQDN) can be set to hdfs access zone, and the FQDN can be used instead of IP to read and write data to/from Isilon.

1. Download sample dataset into Databricks File System (DBFS)
   In the new notebook, run a unix shell command wget to download a sample data file into temporary folder of DBFS, as shown below.

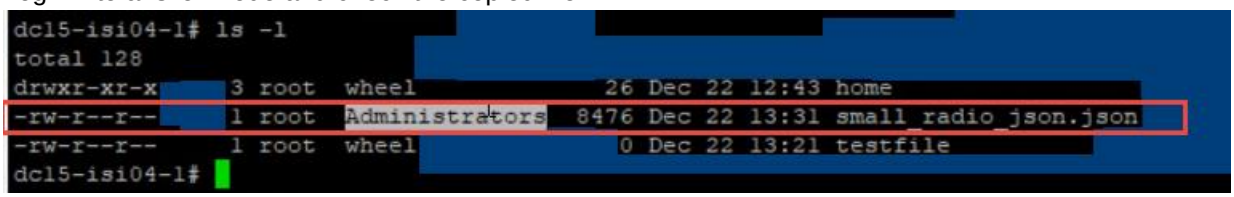
   ```
   %sh wget -P /tmp
   https://raw.githubusercontent.com/Azure/usql/master/Examples/Samples/Data/json/radiowebsite/small_radio_json.json
   ```
   (this downloads the .json file from a github site and stores in it local /tmp dir)



2. Write data into Isilon from Databricks File System.
   Using DBFS copy command copy the sample json data set downloaded into temporary folder to Isilon hdfs access zone.

   ```
   dbutils.fs.cp("file:///tmp/small_radio_json.json", "hdfs://10.1.1.15/")
   ```
   (cp the file over HDFS protocol to the PowerScale cluster – here a nodeIP is used instead of smartconnect)

   Login into a Isilon node and check the copied file.



3. Create a new Spark Dataframe pointing to the sample dataset on Isilon using hdfs protocol (Spark Action).

```
df = spark.read.json("hdfs://10.1.1.15/small_radio_json.json")
```

```
1 df = spark.read.json("hdfs://10.1.1.15/small_radio_json.json")
2
```

```
▸ (1) Spark Jobs
▾ df: pyspark.sql.dataframe.DataFrame
      artist: string
      auth: string
      firstName: string
      gender: string
      itemInSession: long
      lastName: string
      length: double
      level: string
      location: string
      method: string
      page: string
      registration: long
      sessionId: long
      song: string
      status: long
      ts: long
      userId: string
```

4. Read the data from Isilon through hdfs protocol from Databricks read (spark Transform)
   `df.show()`

```
1 df.show()
```

```
▸ (1) Spark Jobs
+--------------------+---------+---------+------+-------------+----------+---------+-----+--------------------+------+--------+-------------+---------+--------------------+-----
-+-------------+------+
|              artist|     auth|firstName|gender|itemInSession|  lastName|   length|level|            location|method|    page| registration|sessionId|                song|statu
s|           ts|userId|
+--------------------+---------+---------+------+-------------+----------+---------+-----+--------------------+------+--------+-------------+---------+--------------------+-----
-+-------------+------+
|         El Arrebato|Logged In| Annalyse|     F|            2|Montgomery|234.57914| free|  Killeen-Temple, TX|   PUT|NextSong|1384448062332|     1879|Quiero Quererte Q...|   20
0|1409318650332|   309|
|Creedence Clearwa...|Logged In|   Dylann|     M|            9|    Thomas|340.87138| paid|       Anchorage, AK|   PUT|NextSong|1400723739332|       10|        Born To Move|   20
0|1409318653332|    11|
|           Gorillaz|Logged In|     Liam|     M|           11|     Watts|246.17751| paid|New York-Newark-J...|   PUT|NextSong|1406279422332|     2047|                DARE|   20
0|1409318685332|   201|
|                null|Logged In|     Tess|     F|            0|  Townsend|     null| free|Nashville-Davidso...|   GET|    Home|1406970190332|     2136|                null|   20
0|1409318686332|   779|
|       Otis Redding|Logged In|  Margaux|     F|            2|     Smith|135.57506| free|Atlanta-Sandy Spr...|   PUT|NextSong|1406191211332|      400| Send Me Some Lovin'|   20
0|1409318697332|   401|
|    Slightly Stoopid|Logged In|     Alan|     M|           39|     Morse|198.53016| paid|Chicago-Napervill...|   PUT|NextSong|1401760632332|      520|         Mellow Mood|   20
0|1409318714332|   521|
|                NOFX|Logged In| Gabriella|     F|            1|   Shelton| 130.2722| free|San Jose-Sunnyval...|   PUT|NextSong|1389460542332|     2261|            Linoleum|   20
0|1409318743332|   244|
|             Nirvana|Logged In|   Elijah|     M|            0|  Williams|260.98893| paid|Detroit-Warren-De...|   PUT|NextSong|1388691347332|      968|The Man Who Sold ...|   20
```

5. Verify the user from Databricks and POSIX user on the Isilon.
   Check the service user id and posix user on the Isilon, the Authentication and authorization can be handled from the Azure cloud and Faction.

   `%sh id`
   (who's the user that is running all these commands in databricks? That is root)

```
1 %sh id
2
```

```
uid=0(root) gid=0(root) groups=0(root)
```

On Isilon the user is root and Administrator group.

```
dc15-isi04-1# ls -l
total 128
drwxr-xr-x    3 root  wheel                26 Dec 22 12:43 home
-rw-r--r--    1 root  Administrators   8476 Dec 22 13:31 small_radio_json.json
-rw-r--r--    1 root  wheel                 0 Dec 22 13:21 testfile
dc15-isi04-1#
```

3.4 Validation summary

The Dell EMC PowerScale powered by Azure Databricks and Faction cloud was able to provide a unified data analytics platform as a solution for advanced analytics with accelerated data-driven innovation. This solution could demonstrate:

1. High-speed data movement into and out of the Azure Databricks cluster
2. Simplified connectivity process between Azure Databricks cluster and PowerScale storage
3. In place data analytics on the on-premises large scale data stored on PowerScale

4 Conclusion

The unified data analytics platform solution for enterprises must effectively address the data deployment challenges and costs associated with the storage and consumption of data for insights. The solution presented herein combines the strengths of Dell EMC PowerScale powered by Faction multicloud Platform-as-a-Service and Azure Databricks to offer enterprises both multicloud flexibility, leading to deployment freedom, superior performance and industry-leading costs. With this reference architecture, enterprises can share and leverage data across public clouds in both an agile and secure fashion, more efficiently use cloud compute, eliminate cloud lock-in and reduce cloud egress costs.

In addition to deployment flexibility and superior price-performance, the unified data analytics platform solution for enterprises must also meet a number of tactical demands. First, it must effectively process data by using solutions like Databricks Spark for data discovery and AI/ML. Second, it must effectively support real-time streaming with the ability to scale to high message rates and large datasets. Third, it must satisfy the concurrency requirements resulting from today's business intelligence solutions.

While enterprise customers demand for more data with faster access to insights will continue to grow, traditional architectural approaches relying on commodity solutions built on virtualized instances will fall short of these stringent demands and will come at a premium price. As such, enterprise customers are best served by leveraging optimized instances and purpose-built analytics solutions like this to achieve the flexibility and best price performance for today's multicloud challenges.

A Technical support and resources

Dell.com/support is focused on meeting customer needs with proven services and support.

Storage technical documents and videos provide expertise that helps to ensure customer success on Dell EMC storage platforms.

A.1 Related resources

Provide a list of documents and other assets that are referenced in the paper; include other resources that may be helpful.

- Dell Technologies Cloud Storage Hybrid Disaster Recovery as a Service
- Dell technologies cloud storage for multi-cloud – powered by Faction
- Dell EMC PowerScale
- Azure Databricks