

DATABRICKS

Microsoft Azure

Search resources, services, and docs (G+)

Shellunext_1693422264...
UNEXT (NPUNEXT.ONMICROSOFT...)

Home > Azure Databricks >

Create an Azure Databricks workspace ...

Basics Networking Encryption Tags Review + create

Project Details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ npunext-1673505322675

Resource group * ⓘ (New) Databricks
[Create new](#)

Instance Details

Workspace name * priya_databricks ✓

Region * East US

Pricing Tier * ⓘ Trial (Premium - 14-Days Free DBUs)

Managed Resource Group name Enter name for managed resource group

Review + create < Previous Next : Networking >

Microsoft Azure

Search resources, services, and docs (G+)

Shellunext_1693422264...
UNEXT (NPUNEXT.ONMICROSOFT...)

Home >

Databricks_priya_databricks | Overview ...

Deployment

Search << Delete Cancel Redeploy Download Refresh

Overview Inputs Outputs Template

✓ Your deployment is complete

Deployment name : Databricks_priya_databricks Start time : 9/26/2023, 9:43:01 AM
Subscription : npunext-1673505322675 Correlation ID : a06ccb3b-71a7-4eb5-9a57-c2...
Resource group : Databricks

> Deployment details

Next steps

Go to resource

Give feedback
Tell us about your experience with deployment

Cost management

Get notified to stay within your budget and prevent unexpected charges on your bill.
[Set up cost alerts >](#)

Microsoft Defender for Cloud

Secure your apps and infrastructure
[Go to Microsoft Defender for Cloud >](#)

Free Microsoft tutorials

[Start learning today >](#)

Work with an expert

Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
[Find an Azure expert >](#)

Microsoft Azure databricks

Search data, notebooks, recents, and more...CTRL + P

priya_databricks

shellunext_1693422264319@npun...

New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Machine Learning

Experiments

Get started with Databricks

Start the SQL warehouse

A SQL warehouse is a simple compute resource that gives you the power to process data in the cloud. We've created your first warehouse for you. To start it, click the button below.

Start warehouse

Explore sample projects

Don't have your data handy? Learn how to gain insights in just a few steps with these sample projects.

1 SELECT

2 COUNT(distinct(custkey))

3 FROM

4 'samples'..'tpch'..'customer'

5 WHERE

SQL and visualizations

750,000

90,000

KPI dashboard

Data analysis in Python

Bring in your own data

Connect to over 50+ data sources and create tables in Databricks.

Skip onboarding

Microsoft Azure databricks

Search data, notebooks, recents, and more...CTRL + P

priya_databricks

shellunext_1693422264319@npun...

New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Machine Learning

Experiments

Untitled Notebook 2023-09-26 10:05:43 Python

File Edit View Run Help Last edit was 1 minute ago Provide feedback

Run all Connect Schedule Share

Free trial ends in 14 days. Upgrade to Premium in Azure Portal

Cmd 1

Python

1

Shift+Enter to run

Shift+Ctrl+Enter to run selected text

Microsoft Azure

databricks

Search data, notebooks, recent, and more...

CTRL + P

priya_databricks

shellunext_1693422264319@npun...

New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Machine Learning

Experiments

Features

Models

Serving

Compute > New compute > UI preview > Send feedback

Shellunext's Cluster

Policy

Unrestricted

Multi node

Single node

Access mode

Single user access

Single user

shellunext_1693422264319@npun...

Performance

Databricks runtime version

Runtime: 13.3 LTS (Scala 2.12, Spark 3.4.1)

Use Photon Acceleration

Worker type

Standard_DS3_v2

14 GB Memory, 4 Cores

Min workers

1

Max workers

2

Spot instances

Driver type

Same as worker

14 GB Memory, 4 Cores

Enable autoscaling

Terminate after

20

minutes of inactivity

Tags

Create compute

Cancel

Summary

1-2 Workers

14-28 GB Memory

4-8 Cores

1 Driver

14 GB Memory, 4 Cores

Runtime

13.3 x-scala2.12

Standard_DS3_v2

1-3 DBU/h

Microsoft Azure

databricks

Search data, notebooks, recent, and more...

CTRL + P

priya_databricks

shellunext_1693422264319@npun...

New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Machine Learning

Experiments

Compute > UI preview > Send feedback

Shellunext's Cluster

Configuration

Notebooks (0)

Libraries

Event log

Spark UI

Driver logs

Metrics

Apps

Spark compute UI - Master

More

Terminate

Edit

Policy

Unrestricted

Multi node

Single node

Access mode

Single user access

Single user

Shellunext

Performance

Databricks Runtime Version

13.3 LTS (includes Apache Spark 3.4.1, Scala 2.12)

Use Photon Acceleration

Worker type

Standard_DS3_v2

14 GB Memory, 4 Cores

Min workers

1

Max workers

2

Current

1

Spot instances

Driver type

Standard_DS3_v2

14 GB Memory, 4 Cores

Summary

1-2 Workers

14-28 GB Memory

4-8 Cores

1 Driver

14 GB Memory, 4 Cores

Runtime

13.3.x-scala2.12

Standard_DS3_v2

1-3 DBU/h

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

Untitled Notebook 2023-09-26 10:05:43 Python

File Edit View Run Help Last edit was now Provide feedback Run all Shellunext's Cluster Schedule Share

Free trial ends in 14 days. Upgrade to Premium in Azure Portal

Cmd 1

```
1 print("Hello")
```

Hello

Command took 0.96 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 10:23:28 AM on Shellunext's Cluster

Cmd 2

```
1
```

Python

Shift+Enter to run
Shift+Ctrl+Enter to run selected text

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

Untitled Notebook 2023-09-26 10:05:43 Python

File Edit View Run Help Last edit was 1 minute ago Provide feedback Run all Shellunext's Cluster Schedule Share

Cmd 2

```
1 %scala
2 println("hello")
```

hello

Command took 1.54 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 10:27:32 AM on Shellunext's Cluster

Cmd 3

```
1 %sql
2 select "hello hi"
```

(1) Spark Jobs

_sqlidf: pyspark.sql.dataframe.DataFrame = [hello hi: string]

hello hi
1 hello hi

1 row | 4.32 seconds runtime Refreshed now

This result is stored as PySpark data frame `_sqlidf` and in the IPython output cache as `Out[2]`. Learn more

Command took 4.32 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 10:27:47 AM on Shellunext's Cluster

Cmd 4

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

Untitled Notebook 2023-09-26 10:05:43 Python

File Edit View Run Help Last edit was 3 minutes ago Provide feedback Run all Shellunext's Cluster Schedule Share

Command took 0.29 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:09:13 AM on Shellunext's Cluster

Cmd 10

```
1 %sql
2 create database test
```

OK

Command took 1.33 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:09:53 AM on Shellunext's Cluster

Cmd 11

```
1 %sql
2 use test
```

OK

Command took 10.69 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:14:27 AM on Shellunext's Cluster

Cmd 12

```
1 %sql
2 create table test.demo(id int, name string)
```

▶ (4) Spark Jobs

OK

Command took 20.11 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:15:39 AM on Shellunext's Cluster

Cmd 13

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

Dataframe 2023-09-26 11:18:57 Python

File Edit View Run Help Last edit was now Provide feedback Run all Shellunext's Cluster Schedule Share

```
1 users=[(1,'a',30,"Sales"),(2,'b',25,"IT"),(3,'c',28,"Data Science")]
2 schema="id int, name string, age int, dept string"
3 df=spark.createDataFrame(data=users, schema=schema)
```

▼ df: pyspark.sql.dataframe.DataFrame

```
id: integer
name: string
age: integer
dept: string
```

Command took 0.38 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:24:01 AM on Shellunext's Cluster

Cmd 2

```
1 display(df)
```

▶ (2) Spark Jobs

	id	name	age	dept
1	1	a	30	Sales
2	2	b	25	IT
3	3	c	28	Data Science

3 rows | 2.76 seconds runtime Refreshed now

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

Dataframe 2023-09-26 11:18:57 Python

File Edit View Run Help Last edit was 3 minutes ago Provide feedback Run all Shellunext's Cluster Schedule Share

Command took 0.04 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:26:02 AM on Shellunext's Cluster

Cmd 4

```
1 df1=df.withColumnRenamed("id","emp_id").withColumn("current_date",current_date())
```

df1: pyspark.sql.dataframe.DataFrame = [emp_id: integer, name: string ... 3 more fields]

Command took 0.09 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:27:38 AM on Shellunext's Cluster

Cmd 5

```
1 display(df1)
```

(2) Spark Jobs

Table +

	emp_id	name	age	dept	current_date
1	1	a	30	Sales	2023-09-26
2	2	b	25	IT	2023-09-26
3	3	c	28	Data Science	2023-09-26

3 rows | 0.72 seconds runtime Refreshed now

Command took 0.72 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:27:52 AM on Shellunext's Cluster

Cmd 6

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

Dataframe 2023-09-26 11:18:57 Python

File Edit View Run Help Last edit was 3 minutes ago Provide feedback Run all Shellunext's Cluster Schedule Share

3 rows | 0.72 seconds runtime Refreshed 8 minutes ago

Command took 0.72 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:27:52 AM on Shellunext's Cluster

Cmd 6

```
1 df1.write.mode("overwrite").saveAsTable("test.emp_table")
```

(6) Spark Jobs

Command took 15.03 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:34:34 AM on Shellunext's Cluster

Cmd 7

```
1 %sql
2 use test;
3 select * from emp_table
```

(3) Spark Jobs

_sqldf: pyspark.sql.dataframe.DataFrame = [emp_id: integer, name: string ... 3 more fields]

Table +

	emp_id	name	age	dept	current_date
1	3	c	28	Data Science	2023-09-26
2	1	a	30	Sales	2023-09-26
3	2	b	25	IT	2023-09-26

Microsoft Azure databricks Search data, notebooks, recent, and more... CTRL + P priya_databricks shellunext.1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

Dataframe 2023-09-26 11:18:57 Python File Edit View Run Help Last edit was now Provide feedback Run all Shellunext's Cluster Schedule Share

```

1 %sql
2 use test;
3 select * from emp_table where age=25

```

(1) Spark Jobs

_sqlidf: pyspark.sql.dataframe.DataFrame = [emp_id: integer, name: string ... 3 more fields]

	emp_id	name	age	dept	current_date
1	2	b	25	IT	2023-09-26

1 row | 1.28 seconds runtime Refreshed now

This result is stored as PySpark data frame _sqlidf and in the IPython output cache as Out[12]. Learn more

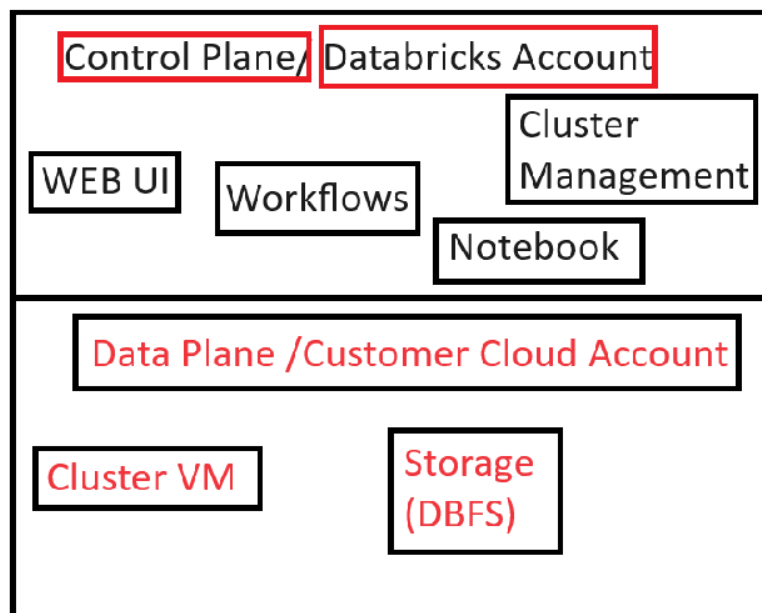
Command took 1.28 seconds -- by shellunext.1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:39:29 AM on Shellunext's Cluster

Cmd 9 Python

1

Shift+Enter to run
Shift+Ctrl+Enter to run selected text

Databricks Architecture



Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

csv to table 2023-09-26 11:58:34 Python Run all Shellunext's Cluster Schedule Share

Free trial ends in 14 days. Upgrade to Premium in Azure Portal

Cmd 1

```
1 %fs ls
```

	path	name	size	modificationTime
1	dbfs:/Volume/	Volume/	0	0
2	dbfs:/Volumes/	Volumes/	0	0
3	dbfs:/databricks-datasets/	databricks-datasets/	0	0
4	dbfs:/databricks-results/	databricks-results/	0	0
5	dbfs:/user/	user/	0	1695706795000
6	dbfs:/volume/	volume/	0	0
7	dbfs:/unlimac/	unlimac/	0	0

7 rows | 1.53 seconds runtime Refreshed 10 minutes ago

Command took 1.53 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 11:59:14 AM on Shellunext's Cluster

Cmd 2

```
1 %fs ls dbfs:/FileStore/tables/
2
```

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

csv to table 2023-09-26 11:58:34 Python Run all Shellunext's Cluster Schedule Share

```
1 %fs ls dbfs:/FileStore/tables/
2
```

	path	name	size	modificationTime
1	dbfs:/FileStore/tables/FilteredData_2023_09_08.csv	FilteredData_2023_09_08.csv	335	1695710156000

1 row | 1.47 seconds runtime Refreshed 3 minutes ago

Command took 1.47 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 12:06:48 PM on Shellunext's Cluster

Cmd 3

```
1 df=spark.read.option("header",True).option("inferSchema",True).csv("dbfs:/FileStore/tables/FilteredData_2023_09_08.csv")
```

(2) Spark Jobs

df: pyspark.sql.dataframe.DataFrame = [ID: integer, First_Name: string ... 6 more fields]

Command took 1.60 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 12:09:30 PM on Shellunext's Cluster

Cmd 4

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


Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

csv to table 2023-09-26 11:58:34 Python Run all Shellunext's Cluster Schedule Share

File Edit View Run Help Last edit was 1 minute ago Provide feedback

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

(1) Spark Jobs

ID	First_Name	Last_Name	ProductName	ItemsSold	SoldPrice	Country	Region
3	Christy	Ladson	TV	3	1600	USA	North America
4	John	Rivers	Laptop	5	2400	USA	North America
7	Chirag	Patel	Cell Phone	5	1500	India	Asia
9	Petra	Henry	TV	10	5000	France	Europe
10	Rita	Roger	Laptop	7	2100	France	Europe
11	Tamara	Tony	Cell Phone	2	1200	Germany	Europe

Command took 0.45 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 12:09:42 PM on Shellunext's Cluster

Cmd 5 Python

1

Shift+Enter to run
Shift+Ctrl+Enter to run selected text

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

csv to table 2023-09-26 11:58:34 Python Run all Shellunext's Cluster Schedule Share

File Edit View Run Help Last edit was 1 minute ago Provide feedback

```
1 df.write.mode("overwrite").saveAsTable("test.my_table")
```

(6) Spark Jobs

Command took 4.00 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 12:16:27 PM on Shellunext's Cluster

```
1 %sql
2 select * from my_table
```

(2) Spark Jobs

_sqlcmd: pyspark.sql.dataframe.DataFrame = [ID: integer, First_Name: string ... 6 more fields]

	ID	First_Name	Last_Name	ProductName	ItemsSold	SoldPrice	Country	Region
1	3	Christy	Ladson	TV	3	1600	USA	North America
2	4	John	Rivers	Laptop	5	2400	USA	North America
3	7	Chirag	Patel	Cell Phone	5	1500	India	Asia
4	9	Petra	Henry	TV	10	5000	France	Europe
5	10	Rita	Roger	Laptop	7	2100	France	Europe
6	11	Tamara	Tony	Cell Phone	2	1200	Germany	Europe

Microsoft Azure databricks Search data, notebooks, recent, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace

File Edit View Run Help Last edit was 2 minutes ago Provide feedback Run all Shellunext's Cluster Schedule Share

csv to table 2023-09-26 11:58:34 Python

Table

	path	name	size	modificationTime
1	dbfs:/user/hive/	hive/	0	1695706795000

1 row | 0.37 seconds runtime Refreshed 3 minutes ago

Command took 0.37 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 12:25:02 PM on Shellunext's Cluster

Cmd 9

```
1 %fs ls
2 dbfs:/user/hive/warehouse/test.db/my_table
```

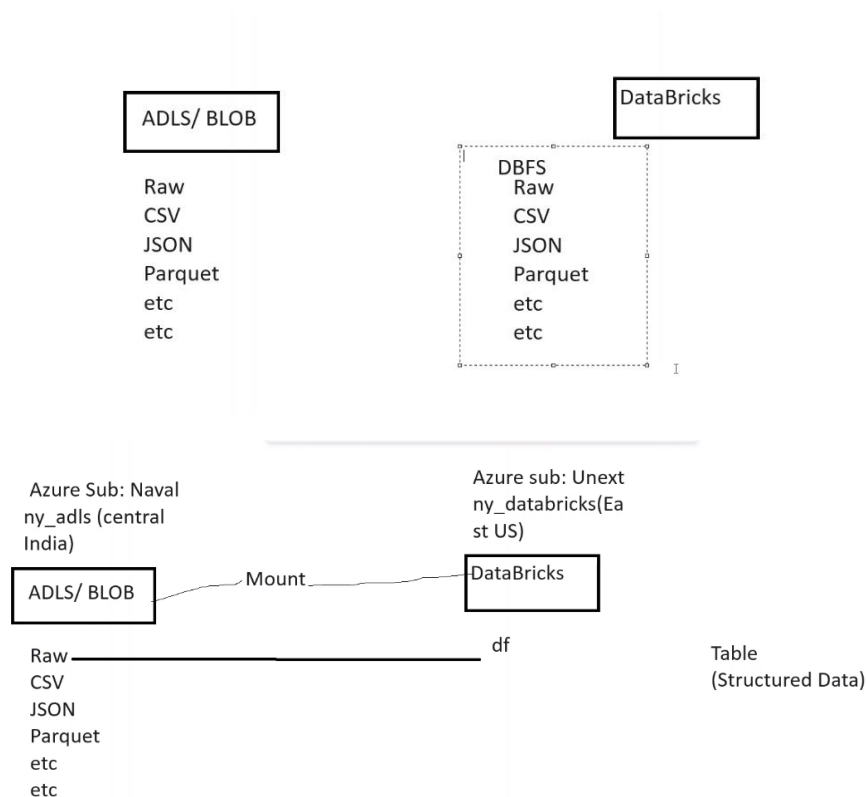
Python

Table

	path	name
1	dbfs:/user/hive/warehouse/test.db/my_table/_delta_log/	_delta_log/
2	dbfs:/user/hive/warehouse/test.db/my_table/part-00000-2d3e8b91-151d-4739-bd9f-a9ef0dd5f319-c000.snappy.parquet	part-00000-2d3e8b9

2 rows | 0.36 seconds runtime Refreshed 2 minutes ago

Command took 0.36 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 12:25:15 PM on Shellunext's Cluster



Mounting

Notebook Authentication

- Access Key (Deprecated)
- SAS (Shared Access Key)
- Service Principal(Recommend)

The screenshot shows a Databricks notebook interface. The top bar includes the Microsoft Azure logo, the Databricks logo, a search bar, and the user's session information. The left sidebar contains navigation options like New, Workspace, Recents, Catalog, Workflows, Compute, SQL, SQL Editor, Queries, Dashboards, Alerts, Query History, SQL Warehouses, Data Engineering, Job Runs, Data Ingestion, Delta Live Tables, Machine Learning, and Experiments. The main area displays the notebook content for 'Access Key 2023-09-26 13:53:23'. It shows three commands: Cmd 1 mounts the storage account using dbutils.fs.mount with an Access Key. Cmd 2 reads a CSV file from the mounted path. Cmd 3 displays the data. The output of Cmd 2 shows the DataFrame schema and the command execution time.

```
1 dbutils.fs.mount(  
2   source = "wasbs://raw@sanly.blob.core.windows.net",  
3   mount_point = "/mnt/sanly/raw",  
4   extra_configs = {"fs.azure.account.key.sanly.blob.core.windows.net": "wZyMjdwqIETizCNMc/uvE0AJQ/2  
+fIGVKVfx4um7LsU00EPZjLx3efLhF901hdkaV1TBwq77j+AStSZRQ1Q=="})  
  
True  
  
Command took 17.56 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 2:12:07 PM on Shellunext's Cluster
```

```
1 df=spark.read.option("header",True).option("inferSchema",True).csv("dbfs:/mnt/sanly/raw/Baby_Names.csv")  
  
(2) Spark Jobs  
df: pyspark.sql.dataframe.DataFrame = [Year: integer, First Name: string ... 3 more fields]  
Command took 13.97 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 2:17:50 PM on Shellunext's Cluster
```

```
1 df.display()
```

(1) Spark Jobs

The screenshot shows the same Databricks notebook interface, but now displaying the output of the data display command. The output shows a table with 7 rows and 6 columns: Year, First Name, County, Sex, and Count. The data is truncated, showing only the first 7 rows. The command execution time is 0.93 seconds.

	Year	First Name	County	Sex	Count
1	2007	ZOEY	KINGS	F	11
2	2007	ZOEY	SUFFOLK	F	6
3	2007	ZOEY	MONROE	F	6
4	2007	ZOEY	ERIE	F	9
5	2007	ZOE	ULSTER	F	5
6	2007	ZOE	WESTCHESTER	F	24
7	2007	ZOE	BRONX	F	12

10,000 rows | Truncated data | 0.93 seconds runtime | Refreshed now

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

Access Key 2023-09-26 13:53:23 Python Last edit was 2 minutes ago Provide feedback Run all Shellunext's Cluster Schedule Share

File Edit View Run Help 10,000 rows | Truncated data | 0.93 seconds runtime Refreshed 4 minutes ago

Command took 0.93 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 2:18:16 PM on Shellunext's Cluster

Cmd 4

```
1 output="dbfs:/mnt/sanly/raw/output"
```

Command took 0.02 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 2:20:35 PM on Shellunext's Cluster

Cmd 5

```
1 df.write.mode("overwrite").parquet(f"{output}/Priyadarsini/babyname")
```

(1) Spark Jobs

Command took 7.05 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 2:20:51 PM on Shellunext's Cluster

Cmd 6

Python

Shift+Enter to run
Shift+Ctrl+Enter to run selected text

Streaming

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

Streaming 2023-09-26 15:36:33 Python Last edit was now Provide feedback Interrupt Shellunext's Cluster Schedule Share

Free trial ends in 14 days. Upgrade to Premium in Azure Portal

Cmd 1

```
1 dbutils.fs.mount(
2   source = "wasbs://inputfiles@saunext.blob.core.windows.net",
3   mount_point = "/mnt/saunext/inputfiles",
4   extra_configs = {"fs.azure.account.key.saunext.blob.core.windows.net":"UUDMjJk8YIiTwHhYh8WCs3BShkFIL//HM/cUrbOrRmUH+HaoR/35bM9MLWTYefbkqNo/bQzgsIM+ASTEn3dkA="})
```

True

Command took 17.83 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 3:50:03 PM on Shellunext's Cluster

Cmd 2

```
1 users_sch="timestamp timestamp, event_type string, user_id string, page_id string"
```

Command took 0.10 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 3:55:54 PM on Shellunext's Cluster

Cmd 3

```
1 df=spark.readStream.schema(users_sch).json("dbfs:/mnt/saunext/inputfiles/inputstream/")
```

df: pyspark.sql.dataframe.DataFrame = [timestamp: timestamp, event_type: string ... 2 more fields]

Command took 1.84 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 3:56:09 PM on Shellunext's Cluster

Cmd 4

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

Streaming 2023-09-26 15:36:33 Python Interrupt Shellunext's Cluster Schedule Share

File Edit View Run Help Last edit was 1 minute ago Provide feedback

Cmd 4

```
1 df.display()
```

Cancel **

(1) Spark Jobs

display_query_1 (id: bf268c19-493b-429b-be19-d5579a099e26) Last updated: 5 seconds ago

	timestamp	event_type	user_id	page_id
1	2023-07-01T10:15:00.000+0000	login	user1	null
2	2023-07-02T10:20:00.000+0000	logout	user2	null
3	2023-07-03T10:25:00.000+0000	page_view	user1	page123
4	2023-07-04T10:30:00.000+0000	login	user3	null
5	2023-07-05T10:35:00.000+0000	page_view	user2	page456
6	2023-07-06T10:40:00.000+0000	logout	user3	null
7	2023-07-07T10:45:00.000+0000	page_view	user1	page789

30 rows

Cmd 5

```
1
```

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P priya_databricks shellunext_1693422264319@npun...

New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning Experiments

Streaming 2023-09-26 15:36:33 Python Interrupt Shellunext's Cluster Schedule Share

File Edit View Run Help Last edit was now Provide feedback

```
1 outputstream="/mnt/saunext/inputfiles/outputstream"
```

Command took 0.02 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 4:05:46 PM on Shellunext's Cluster

Cmd 6

```
1 df.writeStream\
2 .option("checkpointlocation",f"{outputstream}/nava1/checkpoint")\
3 .option("path",f"{outputstream}/Priyadarsini/output")\
4 .table("test.jsonsample")
```

Cancel **

(1) Spark Jobs

fb91293e-1330-4410-ad62-8cf89b0906ba Last updated: 5 seconds ago

<pyspark.sql.streaming.query.StreamingQuery at 0x7f7d70581480>

Cmd 7

```
1 %sql
2 select * from test.jsonsample
```


(1) Spark Jobs

_sqldf: pyspark.sql.dataframe.DataFrame = [timestamp: timestamp, event_type: string ... 2 more fields]

Query returned no results

This result is stored as PySpark data frame _sqldf and in the IPython output cache as Out[8]. Learn more


Microsoft Azure


 databricks

Search data, notebooks, recents, and more...

CTRL + P

priya_databricks



 shellunext_1693422264319@npun...

New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Machine Learning

Experiments

Streaming 2023-09-26 15:36:33 Python

File Edit View Run Help Last edit was now Provide feedback

Run all Shellunext's Cluster Schedule Share

```
1 for stream in spark.streams.active:
2     stream.stop()
```

Command took 0.18 seconds -- by shellunext_1693422264319@npunext.onmicrosoft.com at 9/26/2023, 4:24:04 PM on Shellunext's Cluster

Cmd 9

```
1 (spark
2 .readStream
3 .schema(users_sch)
4 .json("dbfs:/mnt/saunext/inputfiles/inputstream/")
5 .writeStream
6 .option("checkpointlocation",f"{outputstream}/naval/checkpoint")
7 .option("path",f"{outputstream}/naval/output")
8 .trigger(once=True)
9 .table("test.jsonsample")
10 )
```

Stream stopped...

<pyspark.sql.streaming.query.StreamingQuery at 0x7f7d705e99f0>

Command complete

Cmd 10

```
1
```