

# JOURNEY SCRAPBOOK

Technical Training - Week 6

25-Sep-2023 to 29-Sep-2023 (Days 18 – 22)

YUVA SAHITH VARMA SANGARAJU

BATCH - 3

# Day 18 – PySpark Features

- ▶ Caching vs Persistence
- ▶ Default storage levels for rdd, df and ds
- ▶ Storage types in persistence
- ▶ Aggregation
- ▶ Managed and external tables
- ▶ Unmanaged and hive tables
- ▶ Partitioning by column and saving partitions
- ▶ ForEach loop
- ▶ Azure Databricks and ML – Cluster and compute nodes
- ▶ Overview of Spark

# Day 18 continued...

- ▶ Know your data, Data ingestion and Preparation
- ▶ Cost estimation and parallelism
- ▶ Csv data and its reader modes, options
- ▶ Databricks community edition configuration
- ▶ Retail dataset activity

```
[5]: test_df = [("James","Sales","NY",90000,34,10000),
              ("Michael","Sales","NY",86000,56,20000),
              ("Robert","Sales","CA",81000,30,23000),
              ("Maria","Finance","CA",90000,24,23000),
              ("Raman","Finance","CA",99000,40,24000),
              ("Scott","Finance","NY",83000,36,19000),
              ("Jen","Finance","NY",79000,53,15000),
              ("Jeff","Marketing","CA",80000,25,18000),
              ("Kumar","Marketing","NY",91000,50,21000)
            ]
```

```
[6]: ud_scehma = ["employee_name","department","state","salary","age","bonus"]
```

```
[7]: df = spark.createDataFrame(data=test_df,schema = ud_scehma)
```

```
[8]: df.cache().count()
```

```
[8]: 9
```

```
[12]: from pyspark import StorageLevel
```

```
[17]: test = StorageLevel(useDisk=False,useMemory=True,useOffHeap=False,deserialized=False,replication=2)
```

```
[13]: df.persist(StorageLevel.MEMORY_AND_DISK)
```

```
23/09/25 04:27:59 WARN CacheManager: Asked to cache already cached data.
```

```
[13]: DataFrame[employee_name: string, department: string, state: string, salary: bigint, age: bigint, bonus: bigint]
```

```
[10]: df.show()
```

```
+-----+-----+-----+-----+-----+
|employee_name|department|state|salary|age|bonus|
+-----+-----+-----+-----+-----+
|      James|     Sales|    NY|  90000| 34|10000|
| Michael|     Sales|    NY|  86000| 56|20000|
```

```
[14]: df.unpersist()  
[14]: DataFrame[employee_name: string, department: string, state: string, salary: bigint, age: bigint, bonus: bigint]
```

```
[18]: df.persist(storageLevel = test)  
[18]: DataFrame[employee_name: string, department: string, state: string, salary: bigint, age: bigint, bonus: bigint]
```

```
[19]: import sys  
print(sys.version)  
3.11.4 (main, Jul 5 2023, 14:15:25) [GCC 11.2.0]
```

```
[ ]: # To solve Spark Version Error:  
#import os  
#import sys  
  
#os.environ['PYSPARK_PYTHON'] = sys.executable  
#os.environ['PYSPARK_DRIVER_PYTHON'] = sys.executable
```

```
[21]: df.groupBy('state').max("salary").alias("max_sal_by_state").show()  
+-----+  
|state|max(salary)|  
+-----+  
|   CA|     99000|  
|   NY|     91000|  
+-----+
```

```
[23]: df.groupBy("department", "state").agg(sum("salary").alias("sum_salary"), avg("salary").alias("avg_salary")).where(col("avg_salary")>80000).show()  
+-----+-----+-----+  
|department|sum_salary|      avg_salary|  
+-----+-----+-----+
```

```
[24]: df.createOrReplaceTempView("ida")

[26]: spark.sql("select * FROM ida")

[26]: DataFrame[employee_name: string, department: string, state: string, salary: bigint, age: bigint, bonus: bigint]

[31]: spark.sql("CREATE DATABASE idashell")

[31]: DataFrame[]

[32]: spark.sql("use idashell")

[32]: DataFrame[]

[33]: df.write.saveAsTable("test_tbl")

[34]: test = spark.sql("DESCRIBE test_tbl")

[35]: test.show()
+-----+-----+-----+
| col_name|data_type|comment|
+-----+-----+-----+
| employee_name| string| null|
| department| string| null|
| state| string| null|
| salary| bigint| null|
| age| bigint| null|
| bonus| bigint| null|
+-----+-----+-----+
```

```
[36]: test1 = spark.sql("DESCRIBE EXTENDED test_tbl")
test1.show()
```

col_name	data_type	comment
employee_name	string	null
department	string	null
state	string	null
salary	bigint	null
age	bigint	null
bonus	bigint	null
# Detailed Table ...		
Catalog	spark_catalog	
Database	idashell	
Table	test_tbl	
Created Time	Mon Sep 25 05:29:...	
Last Access	UNKNOWN	
Created By	Spark 3.4.1	
Type	MANAGED	
Provider	parquet	
Location	file:/home/labuse...	

```
[37]: df.write.option("path","/home/labuser/Documents/table_data").saveAsTable("e_emptbl")
```

col_name	data_type	comment
employee_name	string	null
department	string	null
state	string	null
salary	bigint	null
age	bigint	null
bonus	bigint	null
# Detailed Table ...		
Catalog	spark_catalog	
Database	idashell	
Table	e_emptbl	
Created Time	Mon Sep 25 05:32:...	
Last Access	UNKNOWN	
Created By	Spark 3.4.1	
Type	EXTERNAL	
Provider	parquet	
Location	file:///home/labu...	

```
[39]: spark.sql("drop table default.emp_tbl")
```

```
[39]: DataFrame[]
```

```
[41]: spark.sql("drop table idashell.e_emp_tbl")
```

```
***
```

```
[42]: df.write.csv("/home/labuser/testcontainer/")
```

```
[Stage 14:> (0 + 2) / 2]
```

```
[44]: df.write.partitionBy("state","department").csv("/home/labuser/test1/")
```

```
[46]: def f(df):
    print(df.state)
```

```
[47]: test = df.foreach(f)
```

```
NY
```

```
NY
```

```
CA
```

```
CA
```

```
CA
```

```
NY
```

```
NY
```

```
CA
```

```
NY
```

```
: json_data = [
    {"name": "Alice", "age": 25},
    {"name": "Bob", "age": 30},
    {"name": "Charlie", "age": 35}
]
```

```
: from pyspark.sql.types import *

schema = StructType([
    StructField("name", StringType(), True),
    StructField("age", IntegerType(), True)
])
```

```
: df = spark.read.schema(schema).json(sc.parallelize(json_data))
```

```
: json_data = [
    {"name": "Alice", "age": 25, "address": {"city": "New York", "state": "NY"}},
    {"name": "Bob", "age": 30, "address": {"city": "San Francisco", "state": "CA"}},
    {"name": "Charlie", "age": 35, "address": {"city": "Los Angeles", "state": "CA"}}
]
```

```
: schemanest = StructType([
    StructField("name", StringType(), True),
    StructField("age", IntegerType(), True),
    StructField("address", StructType([
        StructField("city", StringType(), True),
        StructField("state", StringType(), True)
    ]), True)
])
```

```
: df = spark.read.schema(schemanest).json(sc.parallelize(json_data))
```

```
In [1]: import os
import sys

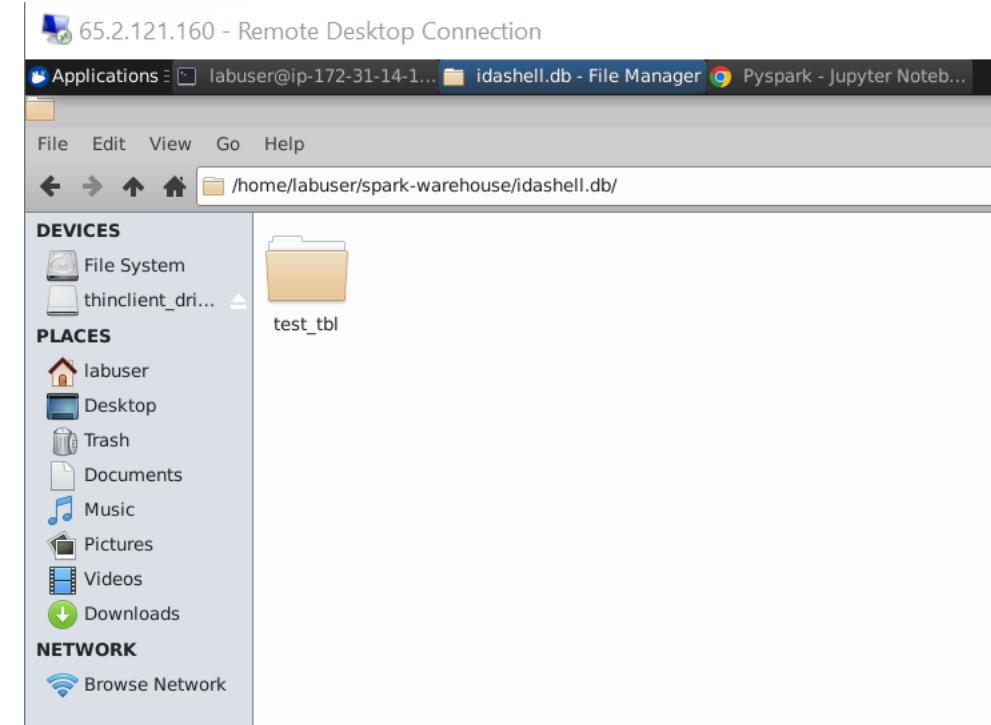
os.environ['PYSPARK_PYTHON'] = sys.executable
os.environ['PYSPARK_DRIVER_PYTHON'] = sys.executable

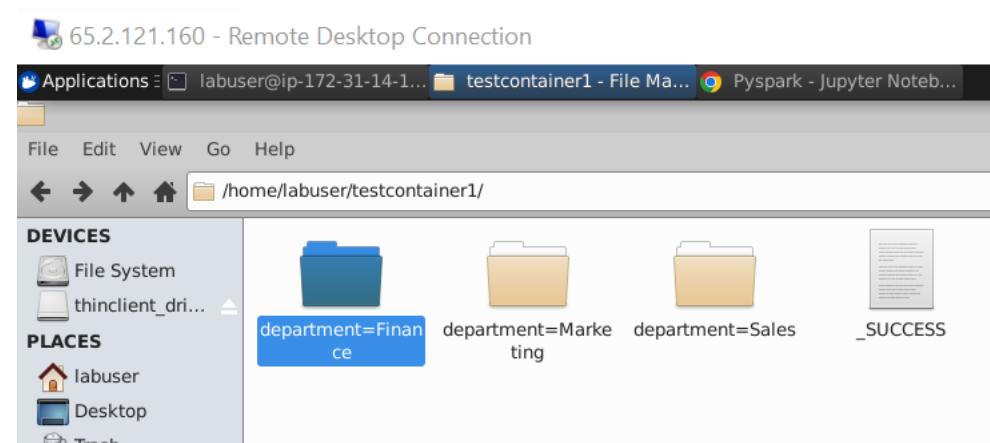
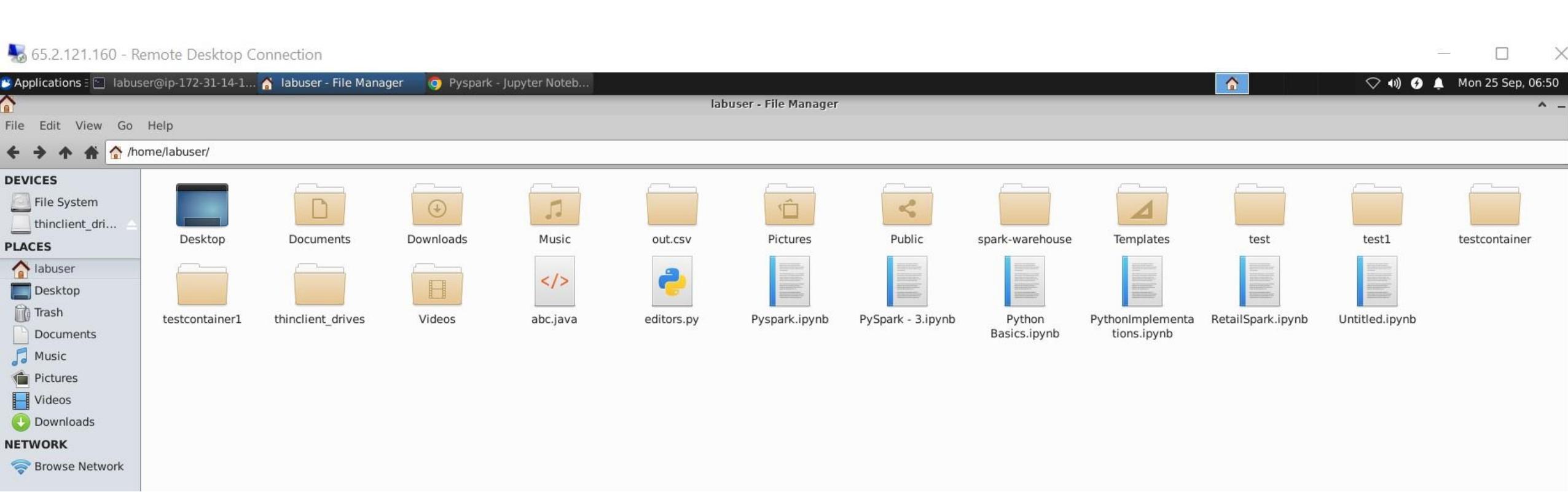
In [2]: from pyspark.conf import SparkConf
from pyspark.sql.session import SparkSession
from pyspark.context import SparkContext
from pyspark.sql.functions import *
from pyspark.sql.types import *
from pyspark import StorageLevel

In [3]: import sys
print(sys.version)

3.11.4 (main, Jul 5 2023, 14:15:25) [GCC 11.2.0]

In [4]: conf = SparkConf().setAppName("NewApp").setMaster("local")
sc = SparkContext(conf = conf)
spark = SparkSession.builder.getOrCreate()
```





## Create data asset

- Data type
- Data source
- Web URL
- Settings
- Schema
- Review

### Data type

Name  
testDatashell

### Description

--

Type  
tabular

### Data source

Type  
WebURL

### Web URL

Web URL  
[https://azuralexamples.blob.core.windows.net/datasets/credit\\_card/default\\_of\\_credit\\_card\\_clients.csv](https://azuralexamples.blob.core.windows.net/datasets/credit_card/default_of_credit_card_clients.csv)

### Skip data validation

false

### Schema

ID	<i>Integer</i>
LIMIT_BAL	<i>Integer</i>
SEX	<i>Integer</i>
EDUCATION	<i>Integer</i>
MARRIAGE	<i>Integer</i>

(showing 5 of 26 columns)

Unext > testML > Automated ML > Start job

## Create a new Automated ML job

1 Select data asset  
2 Configure job  
3 Select task and settings  
4 Hyperparameter configuration (Computer Vision only)  
5 Validate and test

Data asset  
testDatashell ([View data asset](#))

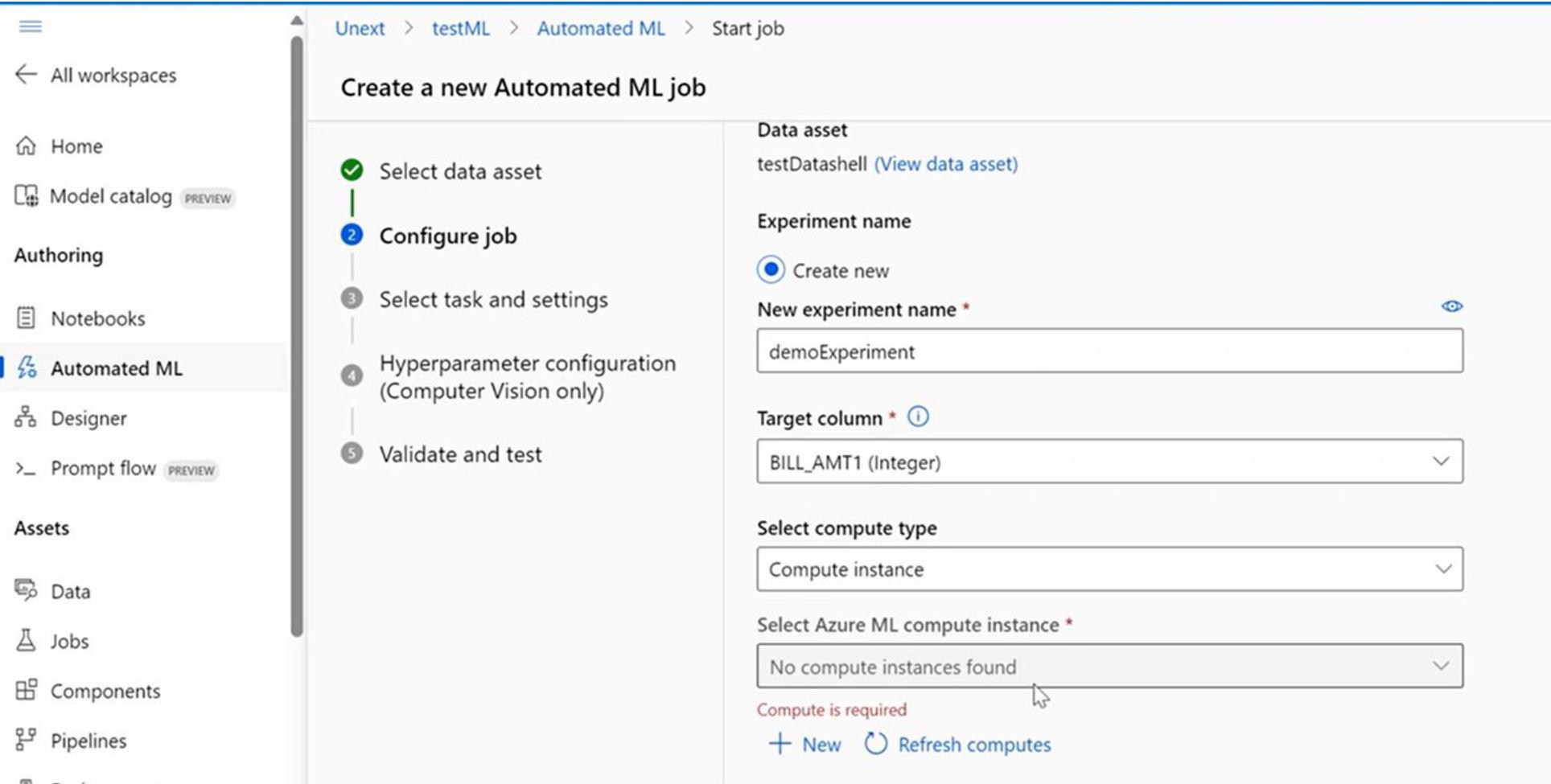
Experiment name  
 Create new  
New experiment name \*

Target column \* [?](#)

Select compute type

Select Azure ML compute instance \*

Compute is required  
+ New Refresh computes



Home >

## Virtual machines

Unext (npunext.onmicrosoft.com)

+ Create ▾ Switch to classic Reservations ▾ Manage view ▾ Refresh Export to CSV Open query Assign tags Start Restart Stop ...

Filter for any field...

Subscription equals all

Type equals all

Resource group equals all

Location equals all

Add filter

Showing 1 to 3 of 3 records.

No grouping

List view

<input type="checkbox"/> Name ↑↓	Type ↑↓	Subscription ↑↓	Resource group ↑↓	Location ↑↓	Status ↑↓	Operating system ↑↓	Size ↑↓
<input type="checkbox"/>	0faf6cfcaa7b46d28693c5343f... Virtual machine	npunext-16735054807...	databricks-rg-testData...	East US	Running	Linux	Standard_DS3_v2
<input type="checkbox"/>	a7b9c80de16d4fdebd7fcf6e5... Virtual machine	npunext-16735054807...	databricks-rg-testData...	East US	Running	Linux	Standard_DS3_v2
<input type="checkbox"/>	cd583fa96ac24a2da21bbc6b... Virtual machine	npunext-16735054807...	databricks-rg-testData...	East US	Running	Linux	Standard_DS3_v2

Login ID Password URL    employee

File Edit View

```
empno,ename,sal
111,zzz,8000
111,aaa,8888
121,bbb,8000
NULL,ccc,9000
false,ddd,6000
555,eee,7000
765
666,fff,8890
aaa,,
True
true
a,b,c,d,e
x,NULL,NULL
1000 null null
```

# Day 19 – Azure Databricks

- ▶ Introduction to Databricks and its features
- ▶ Azure databricks workspace
- ▶ Magic command to switch languages
- ▶ Markdown and repos
- ▶ Catalog and delta live tables
- ▶ Compute types
- ▶ Jobs and scheduling
- ▶ Mounting storage to ADB
- ▶ Creating ADLS Gen2 and Key vault service
- ▶ Secret scope and the mount command

# Day 19 continued...

- ▶ Parametrization and widgets
- ▶ Widget functions such as get, remove and text
- ▶ DBFS and dbutils functions
- ▶ Reconfiguring cluster and Utility notebook
- ▶ Registering functions
- ▶ Exporting folder as DBC to keep the backup
- ▶ Structured streaming and Streaming process
- ▶ writeStream modes
- ▶ Checkpoint and droplocation

# 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-1673504988896

Resource group \* ⓘ

(New) rg\_idashell

[Create new](#)

## Instance Details

Workspace name \*

idaws

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 >](#)

# Create a storage account

...

Basics

Advanced

Networking

Data protection

Encryption

Tags

Review

## Basics

Subscription	npunext-1673504988896
Resource Group	rg_idashell
Location	eastus
Storage account name	idaadls
Deployment model	Resource manager
Performance	Standard
Replication	Locally-redundant storage (LRS)

## Advanced

Enable hierarchical namespace	Enabled
Enable network file system v3	Disabled
Allow cross-tenant replication	Disabled

# Create Secret Scope

Cancel

Create

A store for secrets that is identified by a name and backed by a specific store type. [Learn more](#)

Scope Name 

Manage Principal 



Azure Key Vault 

DNS Name

Resource ID



Home &gt; idakeyvault123

# idakeyvault123 | Properties



Key vault



Access policies

Events

## Objects



Keys



Secrets



Certificates

## Settings

Access configuration

Networking

Microsoft Defender for Cloud

Properties

Locks

## Monitoring

Alerts

Metrics



Save



Discard changes



Refresh

Name

idakeyvault123

Sku (Pricing tier)

Standard

Location

eastus

Vault URI

<https://idakeyvault123.vault.azure.net/>

Copied

Resource ID

[/subscriptions/2190116a-14d8-4b0c-8821-f97ab2e43421/resourceGroups/rg\\_idashell/providers/Microsoft.KeyVault/vaults/idakeyvault123](https://api.azure.com/subscriptions/2190116a-14d8-4b0c-8821-f97ab2e43421/resourceGroups/rg_idashell/providers/Microsoft.KeyVault/vaults/idakeyvault123)

Subscription ID

2190116a-14d8-4b0c-8821-f97ab2e43421



Subscription Name

npunext-1673504988896



Directory ID

dce87315-8ffa-4a01-ab40-0de5a7214b2f



Directory Name

Unext



## Soft-delete

**Soft delete has been enabled on this key vault**

Days to retain deleted vaults

90

 Disable purge protection (allow key vault and objects to be purged during retention period) Enable purge protection (enforce a mandatory retention period for deleted vaults and vault objects)

## Shellunext's Cluster

### Policy

Unrestricted

Multi node  Single node

Access mode  Single user access 

Single user

shellunext\_1693422033357@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

2

Max workers

8

Spot instances 

#### Driver type

Same as worker

14 GB Memory, 4 Cores

Enable autoscaling 

Terminate after  minutes of inactivity 

### Tags

Create compute

Cancel

UI | [JSON](#)

### Summary

2-8 Workers 28-112 GB Memory

8-32 Cores

1 Driver 14 GB Memory, 4 Cores

Runtime 13.3.x-scala2.12

Photon

Standard\_DS3\_v2

4-14 DBU/h

# IDA\_Mount\_NB

Python 



File Edit View Run Help Last edit was 9 minutes ago Provide feedback

▶ Run all

● Shellunext's Cluster 

Schedule 

Share 

 Free trial ends in 14 days. [Upgrade to Premium](#) in Azure Portal 

Cmd 1

```
1 dbutils.fs.mount(source = "wasbs://input@idashelladlsgen2.blob.core.windows.net",
2                   mount_point ="/mnt/input",
3                   extra_configs = {"fs.azure.account.key.idashelladlsgen2.blob.core.windows.net":dbutils.secrets.get(scope = "adbScope", key =
"adlsgen2")})
```

True

Command took 12.42 seconds -- by shellunext\_1693422033357@npunext.onmicrosoft.com at 9/26/2023, 11:24:28 AM on Shellunext's Cluster

Cmd 2

Python 

```
1 dbutils.fs.mounts()
```

[MountInfo(mountPoint='/databricks-datasets', source='databricks-datasets', encryptionType=''),  
 MountInfo(mountPoint='/Volumes', source='UnityCatalogVolumes', encryptionType=''),  
 MountInfo(mountPoint='/databricks/mlflow-tracking', source='databricks/mlflow-tracking', encryptionType=''),  
 MountInfo(mountPoint='/databricks-results', source='databricks-results', encryptionType=''),  
 MountInfo(mountPoint='/databricks/mlflow-registry', source='databricks/mlflow-registry', encryptionType=''),  
 MountInfo(mountPoint='/mnt/input', source='wasbs://input@idashelladlsgen2.blob.core.windows.net', encryptionType=''),  
 MountInfo(mountPoint='/Volume', source='DbfsReserved', encryptionType=''),  
 MountInfo(mountPoint='/volumes', source='DbfsReserved', encryptionType=''),  
 MountInfo(mountPoint='/', source='DatabricksRoot', encryptionType=''),  
 MountInfo(mountPoint='/volume', source='DbfsReserved', encryptionType='')]

Command took 0.33 seconds -- by shellunext\_1693422033357@npunext.onmicrosoft.com at 9/26/2023, 11:24:57 AM on Shellunext's Cluster

Cmd 3

test



idashell

[SecretScope(name='adbScope')]

Command took 0.17 seconds -- by shellunext\_1693422046249@npunext.onmicrosoft.com at 9/26/2023, 12:00:55 PM on Shellunext's Cluster

Cmd 6

1 dbutils.widgets.text("test","ida")

Command took 0.65 seconds -- by shellunext\_1693422046249@npunext.onmicrosoft.com at 9/26/2023, 12:33:20 PM on Shellunext's Cluster

Cmd 7

Python ►▼ ▾ - ×

1 a = dbutils.widgets.get("test")  
2 print(a)

idashell

Command took 0.09 seconds -- by shellunext\_1693422046249@npunext.onmicrosoft.com at 9/26/2023, 12:35:35 PM on Shellunext's Cluster

Cmd 8

color dropdown

white

test

idashell

Cmd 9

```
1 dbutils.widgets.remove("ida")
```

Command took 0.07 seconds -- by  
shellunext\_1693422046249@npunext.onmicrosoft.com at  
9/26/2023, 12:36:56 PM on Shellunext's Cluster

Cmd 10

Python ▶️ 🔍 ⏪ ⏴ ⏵ ⏹

```
1 dbutils.widgets.dropdown  
("choose_colors","white",["Red","Black",  
"Blue","white"],"color dropdown")
```

Command took 0.04 seconds -- by  
shellunext\_1693422046249@npunext.onmicrosoft.com at  
9/26/2023, 12:38:37 PM on Shellunext's Cluster

Shift+Enter to run

Shift+Ctrl+Enter to run selected text

color dropdown

ida

input

operation

test

idashell

9   def perform\_operation(value, operation):  
10    if operation == "Double":  
11      return value \* 2  
12    elif operation == "Square":  
13      return value \*\* 2  
14    else:  
15      return "Invalid operation"  
16  
17 # Process the user input  
18 try:  
19    input\_value = float(input\_value)  
20    result = perform\_operation(input\_value, operation\_type)  
21    print(f"Result of {operation\_type} operation on {input\_value}: {result}")  
22 except ValueError:  
23    print("Invalid input. Please enter a numeric value.")  
24  
25 # Sleep for a while to keep the notebook running  
26 time.sleep(5)

Result of Square operation on 5.0: 25.0

Cmd 1

This example notebook closely follows the [Databricks documentation](#) for how to set up Azure Data Lake Store as a data source in Databricks.

Markdown

Cmd 2

## 0 - Setup

To get set up, do these tasks first:

- Get service credentials: Client ID <aaaaaaaa-bbbb-cccc-dddd-eeeeeeeeee> and Client Credential <NzQzY2QzYTAtM2I3Zi00NzFmLWI3MGMtMzc4MzRjZmk=>. Follow the instructions in [Create service principal with portal](#).
- Get directory ID <ffffffff-gggg-hhhh-iiii-jjjjjjjjjj> : This is also referred to as *tenant ID*. Follow the instructions in [Get tenant ID](#).
- If you haven't set up the service app, follow this [tutorial](#). Set access at the root directory or desired folder level to the service or everyone.

Cmd 3

There are two options to read and write Azure Data Lake data from Azure Databricks:

1. DBFS mount points
2. Spark configs

Cmd 4

## 1 - DBFS mount points

Cmd 1

```
1 df = spark.read.option("header",True).option("inferSchema",True).csv("/mnt/input/zipcodes.csv")
```

Cmd 2

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

Cmd 3

```
1 df.write.parquet("/mnt/input/dest")
```

Cmd 4

```
1 df.write.saveAsTable("zip_tbl")
```

Cmd 5

```
1 df.write.option("path","/mnt/input/table").saveAsTable("ext_zip_tbl")
```

Cmd 6

```
1 test = spark.sql("select * from zip_tbl")
2 test.display()
```

confi X <>

Save Discard Refresh Give feedback

**Settings**

**Configuration**

The cost of your storage account depends on the usage and the options you choose below. [Learn more about storage pricing](#)

Account kind  
StorageV2 (general purpose v2)

Performance (i)  
 Standard  Premium

(i) This setting cannot be changed after the storage account is created.

Secure transfer required (i)  
 Disabled  Enabled

Allow Blob anonymous access (i)  
 Disabled  Enabled

(i) Some blobs may become anonymously readable.

Allow storage account key access (i)  
 Disabled  Enabled

Allow recommended upper limit for shared access signature (SAS) expiry interval (i)  
 Disabled  Enabled

Default to Azure Active Directory authorization in the Azure portal (i)  
 Disabled  Enabled

Minimum TLS version (i)

Permitted scope for copy operations (preview) (i)

Blob access tier (default) (i)  
 Cool  Hot

utility\_NB Python ☆

File Edit View Run Help Last edit was now Provide feedback

▶ Run all ⚙ Shellunext's Cluster Schedule Share

```
1 spark.udf.register("test_fun",test_fun,StringType())  
<function __main__.test_fun(a)>  
Command took 0.20 seconds -- by shellunext_1693422033357@npunext.onmicrosoft.com at 9/26/2023, 3:07:44 PM on Shellunext's Cluster
```

Cmd 8

```
1 def concatenate_columns_01(first_name, last_name):  
2     return first_name + " " + last_name  
  
Command took 0.12 seconds -- by shellunext_1693422033357@npunext.onmicrosoft.com at 9/26/2023, 3:33:39 PM on Shellunext's Cluster
```

Cmd 9

```
1 spark.udf.register("concatenate_columns",concatenate_columns_01,StringType())  
Cancel *** Uploading command  
<function __main__.concatenate_columns(df, column1, column2, delimiter=' ')>
```

Cmd 10

```
1
```

```
8
9
10      "value",
        "convictions"
    )
```

▼ ⏷ trimmedDF: pyspark.sql.dataframe.DataFrame

borough: string  
year: string  
month: string  
convictions: string

Command took 0.07 seconds -- by shellunext\_1693422033357@npunext.onmicrosoft.com at 9/26/2023, 3:52:02 PM on Shellunext's Cluster

Cmd 5

Python ► ▶ ▾ - ×

```
1 query = trimmedDF.writeStream\
2             .outputMode("append")\
3             .format("csv") \
4             .option("path", "/mnt/input/destination") \
5             .option("checkpointLocation", "/mnt/input/checkpoint") \
6             .start() I
7             .awaitTermination()
```

Cancel ••• Running command...

Shift+Enter to run

Shift+Ctrl+Enter to run selected text

Home >



...

X

Search

«

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Overview

Diagnose and solve problems

Access Control (IAM)

## Settings

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

**Authentication method:** Access key ([Switch to Azure AD User Account](#))

**Location:** input

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> _\$azuretmpfolder\$				
<input type="checkbox"/> checkpoint				
<input type="checkbox"/> dest				
<input type="checkbox"/> destination				
<input type="checkbox"/> droploc				
<input type="checkbox"/> table				
<input type="checkbox"/> Products.csv	9/26/2023, 11:26:14 ...	Hot (Inferred)		Block blob

Upload Add Directory Refresh | Rename Delete Change tier Acquire lease Break lease Give feedback

**Authentication method:** Access key ([Switch to Azure AD User Account](#))

**Location:** [input](#) / destination

Search blobs by prefix (case-sensitive)



Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
------	----------	-------------	----------------	-----------

<input type="checkbox"/> [..]				
<input type="checkbox"/> _spark_metadata				
<input type="checkbox"/> part-00000-32bc878b-3bf6-4ec9-9cb4-f1f9a1a9ac...	9/26/2023, 4:28:24 PM	Hot (Inferred)		Block blob



Upload Add Directory Refresh | Rename Delete Change tier Acquire lease Break lease Give feedback

**Authentication method:** Access key ([Switch to Azure AD User Account](#))

**Location:** [input](#) / checkpoint

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> [..]				
<input type="checkbox"/> _tmp_path_dir				
<input type="checkbox"/> commits				
<input type="checkbox"/> offsets				
<input type="checkbox"/> sources				
<input type="checkbox"/> metadata	9/26/2023, 4:27:55 PM	Hot (Inferred)		Block blob



↑ Upload + Add Directory ⏪ Refresh | ⏷ Rename 🗑 Delete ⏵ Change tier 🔒 Acquire lease 🔒 Break lease 🤝 Give feedback

**Authentication method:** Access key ([Switch to Azure AD User Account](#))

**Location:** [input](#) / droploc

Search blobs by prefix (case-sensitive)



Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
------	----------	-------------	----------------	-----------

<input type="checkbox"/>	[..]			
<input type="checkbox"/>	Products.csv	9/26/2023, 4:28:20 PM	Hot (Inferred)	Block blob



Upload Add Directory Refresh | Rename Delete Change tier Acquire lease Break lease Give feedback

**Authentication method:** Access key ([Switch to Azure AD User Account](#))

**Location:** input / table

Search blobs by prefix (case-sensitive)



Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> [..]				
<input type="checkbox"/> _delta_log				
<input type="checkbox"/> part-00000-b9f8a24a-fa17-48a7-b229-216a950df8...	9/26/2023, 3:26:07 PM	Hot (Inferred)		Block blob



## Catalog Explorer [Provide Feedback](#)

[+ Add](#)

Shellunext's Cluster

14 GB, 4 Cores



hive\_metastore

default

ext\_prod\_tbl

prod\_tbl

samples

default &gt;

### default.prod\_tbl [A](#)

Owner: Not set [Edit](#) Size: 1.3KiB, 1 file Last Updated: UnknownComment: [Add comment](#)[Columns](#)[Sample Data](#)[Details](#)[Permissions](#)[History](#) Filter columns...

Column	Type	Comment
pid	int	<a href="#">+</a>
pname	string	<a href="#">+</a>
orderqty	int	<a href="#">+</a>
unitprice	int	<a href="#">+</a>

# Catalog Explorer

[Provide Feedback](#)[+ Add](#)

Shellunext's Cluster

14 GB, 4 Cores

[hive\\_metastore](#)[default](#)[ext\\_prod\\_tbl](#)[prod\\_tbl](#)[samples](#)

default &gt;

## default.default.ext\_prod\_tbl

[Create](#)Owner: Not set [Edit](#) Size: 1.3KiB, 1 file Last Updated: UnknownComment: [Add comment](#)[Columns](#)[Sample Data](#)[Details](#)[Permissions](#)[History](#)

Created At	Tue Sep 26 09:56:10 UTC 2023
Last Access	UNKNOWN
Created By	Spark 3.4.1
Type	EXTERNAL
Location	dbfs:/mnt/input/table
Properties	delta.minReaderVersion=1 delta.minWriterVersion=2

# Compute

Free trial ends in **14** days. [Upgrade to Premium in Azure Portal](#)

All-purpose compute Job compute SQL warehouses Pools Policies [?](#)

 Filter compute you have...

Created by [▼](#)

Create with Personal Compute [▼](#)

[Create compute](#)

State <a href="#">☰</a>	Name	Policy	Runtime	Active m...	Active c...	Active D...	Source	Creator	Noteboo...	
 	Shellunext's Cluster	-	13.3	14 GB	4 cores	1.5	UI	shellunext_169...	5	 <a href="#">⋮</a>



IDA\_Shell

Folder

Shellunext

10:48 AM



IDA\_Shell (1)

Folder

Shellunext

4:37 PM



Import from Azure Data Lake Store ... Notebook

Shellunext

Open in new tab

Create &gt;

Clone

Import

Export &gt;

Copy &gt;

Rename

Share (Permissions)

Move

Add to favorites

URL

Path

Move to Trash

# Day 20 – Azure Databricks continued...

- ▶ Table creation through UI and through Notebook
- ▶ Delta tables and their features
- ▶ Parquet files and the delta log folder
- ▶ Time travelling and CRUD operations
- ▶ Vacuum and optimize functions
- ▶ Table history and version control
- ▶ Medallion/Multihop architecture
- ▶ Bronze, silver and gold data layers
- ▶ Restore ... as version of ... command
- ▶ Workflow and jobs

# Day 20 continued...

- ▶ Job cluster creation and scheduling trigger
- ▶ JDBC connection for Azure SQL
- ▶ Unity catalog and its features
- ▶ Creating metastore and sharing catalog
- ▶ Access control and role assignment



+ New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

NEW

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Machine Learning

Experiments

## 2023-09-27 - DBFS Example

Python



File Edit View Run Help Last edit was now Provide feedback

Run all

Connect

Schedule

Share

⚠️ Free trial ends in 14 days. [Upgrade to Premium](#) in Azure Portal X

Cmd 1

```
%md ## Overview This notebook will show you how to create and query a table or ...
```

Cmd 2

```
1 # File location and type
2 file_location = "/FileStore/tables/zipcodes.csv"
3 file_type = "csv"
4
5 # CSV options
6 infer_schema = "false"
7 first_row_is_header = "false"
8 delimiter = ","
9
10 # The applied options are for CSV files. For other file types, these will be ignored.
11 df = spark.read.format(file_type) \
12     .option("inferSchema", infer_schema) \
13     .option("header", first_row_is_header) \
14     .option("sep", delimiter) \
15     .load(file_location)
16
17 display(df)
```

## DBFS

[Preview Table](#)

## Specify Table Attributes

Specify the Table Name, Database and Schema to add this to the data UI for other users to access

Table Name [?](#)

zipcodes\_1\_csv

Table Preview

Create in Database [?](#)

default

File Type [?](#)

CSV

Column Delimiter [?](#) First row is header [?](#) Infer schema [?](#) Multi-line [?](#)[Create Table](#) Create Table in  
Notebook



Type to filter

- ▼ hive\_metastore
- ▼ default
  - zipcodes\_1\_csv
- > samples

default &gt;

**default.zipcodes\_1\_csv** 

Create

Owner: Not set Size: 6.3KiB, 1 file Last Updated: Unknown

Comment: [Add comment](#)[Columns](#)[Sample Data](#)[Details](#)[Permissions](#)[History](#)

Created At	Wed Sep 27 04:17:53 UTC 2023
Last Access	UNKNOWN
Created By	Spark 3.4.1
Type	MANAGED
Location	dbfs:/user/hive/warehouse/zipcodes_1_csv
Is Managed Location	true
Properties	delta.minReaderVersion=1 delta.minWriterVersion=2

```
1 %sql  
2 describe history zipcodes_1_csv
```

▶ (1) Spark Jobs

▶ \_sqldf: pyspark.sql.dataframe.DataFrame = [version: long, timestamp: timestamp ... 13 more fields]

Table +

	version	timestamp	userId	userName
1	1	2023-09-27T04:17:58.000+0000	878179360858262	shellunext_1693422046249@npunext.onmicrosoft.com
2	0	2023-09-27T04:17:44.000+0000	878179360858262	shellunext_1693422046249@npunext.onmicrosoft.com

2 rows | 1.58 seconds runtime

Refreshed 22 minutes ago

This result is stored as PySpark data frame `_sqldf` and in the IPython output cache as `Out[4]`. [Learn more](#)

Command took 1.58 seconds -- by shellunext\_1693422046249@npunext.onmicrosoft.com at 9/27/2023, 10:08:35 AM on Shellunext's Cluster

Cmd 4

```
1 %sql  
2 select * from zipcodes_1_csv
```

▶ (1) Spark Jobs

▶ \_sqldf: pyspark.sql.dataframe.DataFrame = [RecordNumber: integer, Zipcode: integer ... 18 more fields]

```
1 dbutils.help()
```

This module provides various utilities for users to interact with the rest of Databricks.

**credentials:** **DatabricksCredentialUtils** -> Utilities for interacting with credentials within notebooks  
**data:** **DataUtils** -> Utilities for understanding and interacting with datasets (EXPERIMENTAL)  
**fs:** **DbfsUtils** -> Manipulates the Databricks filesystem (DBFS) from the console  
**jobs:** **JobsUtils** -> Utilities for leveraging jobs features  
**library:** **LibraryUtils** -> Utilities for session isolated libraries  
**meta:** **MetaUtils** -> Methods to hook into the compiler (EXPERIMENTAL)  
**notebook:** **NotebookUtils** -> Utilities for the control flow of a notebook (EXPERIMENTAL)  
**preview:** **Preview** -> Utilities under preview category  
**secrets:** **SecretUtils** -> Provides utilities for leveraging secrets within notebooks  
**widgets:** **WidgetsUtils** -> Methods to create and get bound value of input widgets inside notebooks

Command took 0.09 seconds -- by shellunext\_1693422046249@npunext.onmicrosoft.com at 9/27/2023, 10:30:18 AM on Shellunext's Cluster

Cmd 7

Python ▶ ▾ ▷ - ×

```
1 dbutils.fs.mount(source = "wasbs://input@adlsstore321.blob.core.windows.net",
2                   mount_point = "/mnt/input",
3                   extra_configs = {"fs.azure.account.key.adlsstore321.blob.core.windows.net":dbutils.
4                     secrets.get(scope = "adbScope", key = "idasecret")})
```

True

Command took 12.32 seconds -- by shellunext\_1693422046249@npunext.onmicrosoft.com at 9/27/2023, 10:29:42 AM on Shellunext's Cluster

 Upload  Add Directory  Refresh |  Rename  Delete  Change tier  Acquire lease  Break

**Authentication method:** Access key ([Switch to Azure AD User Account](#))

**Location:** [input](#) / [opt](#) / State=AL

Search blobs by prefix (case-sensitive)

Show

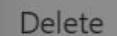
Name	Modified	Access tier	Archive status
<input type="checkbox"/>  [...]			
<input type="checkbox"/>  City=SPRING GARDEN			
<input type="checkbox"/>  City=SPRINGVILLE			
<input type="checkbox"/>  City=SPRUCE PINE			

Schedule 

## Trigger Status

- Active  
 Paused

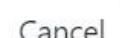
Chennai, Kolkata, Mumbai...



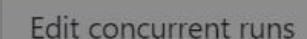
## Trigger type

Scheduled Schedule Every 5 Minutes 

from

4 minutes past the hour  Show cron syntaxSource\* Workspace Maximum concurrent runs 

1

Duration thresholds 

No thresholds defined



Jobs Job runs Delta Live Tables

 Only my job runs

Run status ▾

2023-09-25 14:30:00

→ 2023-09-27 14:30:00



Create job

1

1

0

26 Sep, 12 AM

26 Sep, 12 PM

27 Sep, 12 AM

27 Sep, 12 PM

■ Failed   ■ Skipped   ■ Succeeded

Start time	Job	Run as	Launched	Duration	Status	Run parameters	⋮
Sep 27, 2023, 01:54 PM	Job1	Shellunext	By scheduler	15s	Pending	<span style="color: darkblue;">■</span>	⋮
Sep 27, 2023, 01:52 PM	Job1	Shellunext	By scheduler	1m 44s	<span style="color: green;">✓</span> Succeeded		⋮
Sep 27, 2023, 01:50 PM	Job1	Shellunext	By scheduler	1m 52s	<span style="color: red;">✖</span> Failed		⋮
Sep 27, 2023, 01:48 PM	Job1	Shellunext	By scheduler	1m 53s	<span style="color: red;">✖</span> Failed		⋮
Sep 27, 2023, 01:46 PM	Job1	Shellunext	By scheduler	1m 52s	<span style="color: red;">✖</span> Failed		⋮

# Task1 run

Original (latest) · Running

[Cancel task run](#)[Edit task](#) 

## Output

[Hide code](#)[Export as HTML](#)

```
df = spark.read.option("header",True).option("inferSchema",True).csv("/mnt/input/zipcodes.csv")
```

Running command...

```
df.write.option("path","/mnt/input/table").saveAsTable("ext_zip_tbl")
```

Command submitted to compute for execution

## Task run details

Job ID	<a href="#">662898694622057</a>
Job run ID	1390
Task run ID	1795
Run as	Shellunext
Launched	By scheduler
Started	09/27/2023, 01:46:07 PM
Ended	-
Duration	1m 42s
Status	Running - <a href="#">Cancel</a>

## Notebook

[/Users/shellunext\\_1693422046249@npunext.onmicrosoft.com](#)  
[Scheduling](#)

## Compute

Cmd 1

```
1 jdbc_url = "jdbc:sqlserver://idaservernew.database.windows.net:1433;databaseName=idadb02"
2 connection_properties = {
3     "user": "sqladmin",
4     "password": "IDAshell@123",
5     "driver": "com.microsoft.sqlserver.jdbc.SQLServerDriver"
6 }
```

Command took 0.13 seconds -- by shellunext\_1693422046249@npunext.onmicrosoft.com at 9/27/2023, 2:31:48 PM on Shellunext's Cluster

Cmd 2

```
1 # query = "select * from SalesLT.Customer"
```

Command took 0.16 seconds -- by shellunext\_1693422046249@npunext.onmicrosoft.com at 9/27/2023, 2:25:38 PM on Shellunext's Cluster

Cmd 3

```
1 df = spark.read.jdbc(url=jdbc_url,table="SalesLT.Customer",properties=connection_properties)
```

▶ df: pyspark.sql.dataframe.DataFrame = [CustomerID: integer, NameStyle: boolean ... 13 more fields]

Command took 0.79 seconds -- by shellunext\_1693422046249@npunext.onmicrosoft.com at 9/27/2023, 2:31:51 PM on Shellunext's Cluster

Cmd 4

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



## Get started with Unity Catalog



Unity Catalog is a fine-grained governance solution for data and AI on the Lakehouse

As an account admin you can create and administer a metastore. You can then add specific workspaces to this metastore. A metastore and assigned workspaces must all be in the same region.

Follow the steps below to get started

### As an account admin

- Create a metastore
- Add users and groups

## Workspaces



Filter workspaces



Name	Status	Resource g...	Region	Subscription	Created	Metastore	Actions
idaadbuc	Running	RG_IDAshell	eastus	205450d4-d...	today at 2:5...	-	<a href="#">Open</a>
databhai	Running	rgseptshaile...	eastus	438c9c8e-ce...	today at 2:5...	-	<a href="#">Open</a>
a2590db	Running	a2590usw	westus2	60ff146d-74...	today at 2:3...	-	<a href="#">Open</a>
ws27septsid...	Running	rg27septsid...	eastus	63176cd6-d...	today at 2:3...	-	<a href="#">Open</a>
tandb2222	Running	tanrg27	eastus	324a7224-f0...	today at 2:2...	-	<a href="#">Open</a>
azuredatabri...	Running	27septdatab...	eastus	51ca1e02-af...	today at 2:2...	-	<a href="#">Open</a>
ws27septsu...	Running	rg27septsu...	eastus	54f01b07-e...	today at 2:2...	-	<a href="#">Open</a>
ws27septch...	Running	rg27septcha...	eastus	59b391ce-b...	today at 2:1...	-	<a href="#">Open</a>
databricks27	Running	27asdrg	eastus	33380e94-1...	today at 2:1...	-	<a href="#">Open</a>
draxdb27	Running	rgdrax27	eastus	61877913-2...	today at 2:1...	-	<a href="#">Open</a>
adb27septra...	Running	rg27septrag...	eastus	62ba2fc3-3...	today at 2:1...	-	<a href="#">Open</a>
ws27septadi...	Running	rgsept27adi...	eastus	5ce489e0-fb...	today at 2:1...	-	<a href="#">Open</a>
ws27septsw...	Running	rg27septswa...	eastus	52b86789-fc...	today at 2:1...	-	<a href="#">Open</a>

< Previous

Next >

# Day 21 – Docker and Kubernetes

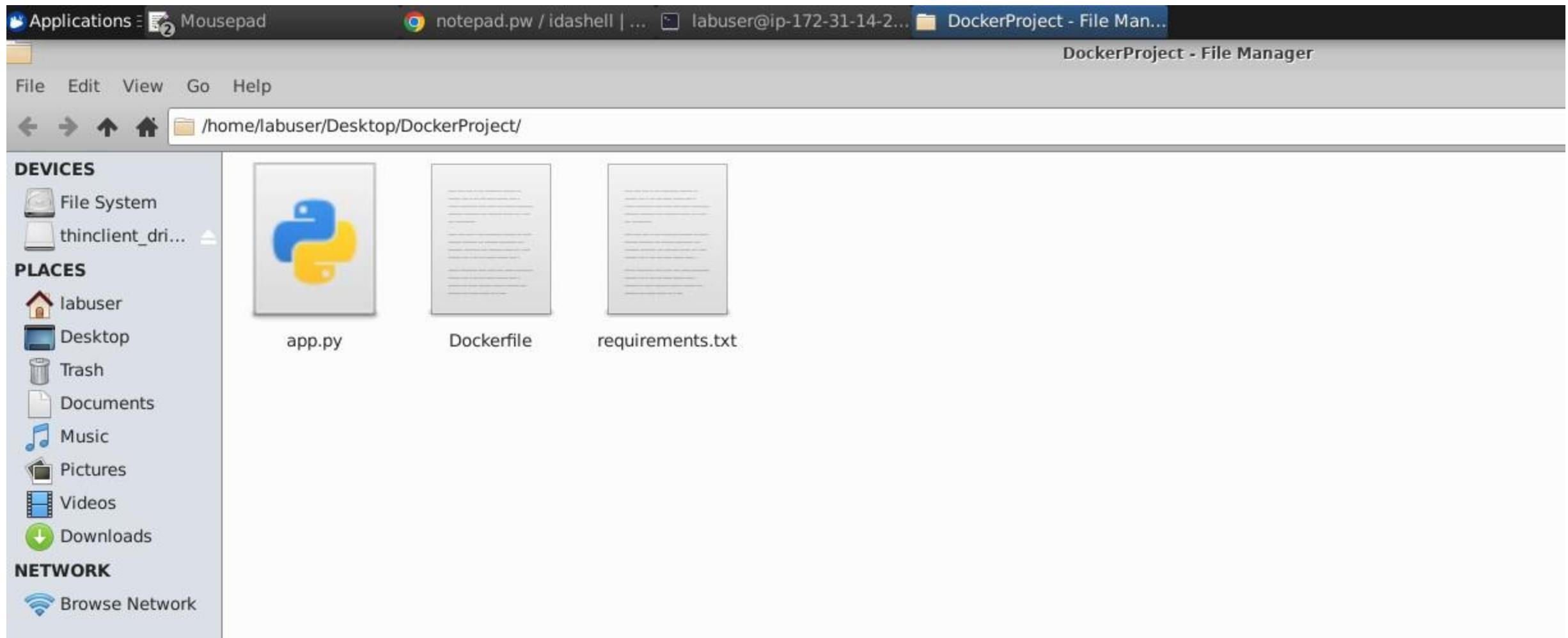
- ▶ Docker introduction
- ▶ Start docker and create docker project
- ▶ App.py, Dockerfile and requirements.txt files
- ▶ Build docker image and run the app
- ▶ Azure Container Registry (ACR) and Azure CLI
- ▶ Azure login, Resource group creation from terminal
- ▶ Creating ACR and ACR login
- ▶ Tag docker image and push it to ACR
- ▶ Kubernetes cluster configuration
- ▶ Yaml files for deployment and service
- ▶ Attach Azure Kubernetes Service (AKS) with ACR

13.233.136.23 - Remote Desktop Connection

Applications: Mousepad, notepad.pw / idashell | ..., labuser@ip-172-31-14-2..., DockerProject - File Man...      labuser@ip-172-31-14-209: ~

File Edit View Search Terminal Help

```
labuser@ip-172-31-14-209:~$ sudo systemctl start docker
labuser@ip-172-31-14-209:~$ docker ps
Got permission denied while trying to connect to the Docker daemon socket at unix:///var/run/docker.sock: Get http://var/run/docker.sock/v1.24/containers/json: dial unix /var/run/docker.sock: connect: permission denied
labuser@ip-172-31-14-209:~$ sudo docker ps
CONTAINER ID        IMAGE               COMMAND       CREATED          STATUS           PORTS          NAMES
labuser@ip-172-31-14-209:~$
```



```
    "name": "SHELLCLOUD_1095422040249@shellcloud.0001ctossl.com",
    "type": "user"
  }
}
]
labuser@ip-172-31-14-209:~$ az group create --name RG_IDAshell --location eastus
{
  "id": "/subscriptions/2190116a-14d8-4b0c-8821-f97ab2e43421/resourceGroups/RG_IDAshell",
  "location": "eastus",
  "managedBy": null,
  "name": "RG_IDAshell",
  "properties": {
    "provisioningState": "Succeeded"
  },
  "tags": null,
  "type": "Microsoft.Resources/resourceGroups"
```



Search



Move Delete

JSON View

Overview

Activity log

Access control (IAM)

Tags

Quick start

Events

## Settings

Access keys

Encryption

Identity

Networking

Microsoft Defender for Cloud

Properties

Locks

## Services

### ^ Essentials

Resource group ([move](#))

[RG\\_IDAshell](#)

Location

East US

Subscription ([move](#))

[npunext-1673504988896](#)

Subscription ID

2190116a-14d8-4b0c-8821-f97ab2e43421

Soft delete (Preview)

[Disabled](#)

Tags ([edit](#))

[Add tags](#)

Login server

idaacr01.azurecr.io

Creation date

9/28/2023, 11:34 AM GMT+5:30

Provisioning state

Succeeded

Pricing plan

Standard

Get started

Monitoring

Capabilities (9)

Tutorials

## Simplify container lifecycle management

Container registry allows you to build, store, and manage container images and artifacts in a private registry for all types of container deployments. [Learn more](#)

# idaacr01 | Access keys

Container registry



Registry name

idaacr01



Overview

Activity log

Access control (IAM)

Tags

Quick start

Events

Settings

Access keys

Encryption

Identity

Networking

Microsoft Defender for Cloud

Properties

Login server

idaacr01.azurecr.io



Admin user (i)

Username

idaacr01



Name

Password

Regenerate

password

auqXru70x1qyVv+D7r/yxxFiAjdcTTeiPh0JmyHvj5+ACRDXM...



password2

eh66VtlcPCr7HCRZOF8xVmUs8ksn7wv+uILrRcW7GC+ACR...



# python-app:v1

...

sha256:8b4f0d42b5aeae17a4495b927f5afd8636ebdfca53971f912ee4a0939fc4b6b3



## ^ Essentials

[JSON View](#)

Repository	:	python-app	Digest	:	sha256:8b4f0d42b5aeae17a4495b927f5afd8636ebdfca53971f9...
Tag	:	v1	Manifest creation date	:	9/28/2023, 12:01 PM GMT+5:30
Tag creation date	:	9/28/2023, 12:01 PM GMT+5:30	Platform	:	linux / amd64
Tag last updated date	:	9/28/2023, 12:01 PM GMT+5:30	Media type	:	application/vnd.docker.distribution.manifest.v2+json

## Manifest Referrers

Docker pull command

docker pull idaacr01.azurecr.io/python-app:v1



## Manifest

```
{  
    "schemaVersion": 2,  
    "mediaType": "application/vnd.docker.distribution.manifest.v2+json",  
    "config": {  
        "mediaType": "application/vnd.docker.container.image.v1+json",  
        "size": 9137,  
        "digest": "sha256:bd09ebc9a2e8d63dc4b2420e226600c2dc8e10558cf1de4960e71c74ba8c19d7"  
    },  
    "layers": [  
        {  
            "mediaType": "application/vnd.docker.image.rootfs.diff.tar.gzip",  
            "size": 27138450,  
            "digest": "sha256:8b91b88d557765cd8c6802668755a3f6dc4337b6ce15a17e4857139e5fc964f3"  
        },  
        {  
            "mediaType": "application/vnd.docker.image.rootfs.diff.tar.gzip",  
            "size": 2781342,  
            "digest": "sha256:8b4f0d42b5aeae17a4495b927f5afd8636ebdfca53971f912ee4a0939fc4b6b3"  
        }  
    ]  
}
```

## Add a node pool

...

kubcluster

Node pool name \* ⓘ

idademo



Mode \* ⓘ

User

System

OS type ⓘ

OS type \*

Linux

Windows

Availability zones ⓘ

None



Enable Azure Spot instances ⓘ



Node size \* ⓘ

Choose a size

Scale method ⓘ

Manual

Autoscale - **Recommended**

This option is recommended so that the cluster is automatically sized correctly for the current running workloads.

Minimum node count \* ⓘ

1

Add

Cancel

52.66.64.204 - Remote Desktop Connection

Applications Mousepad notepad.pw / idashell | ... [labuser@ip-172-31-14-... Kubernetes - File Manager

/home/labuser/Desktop/DockerProject/Kubernetes/deployment.yaml - Mousepad

File Edit Search View Document Help

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: my-app
spec:
  replicas: 2
  selector:
    matchLabels:
      app: my-app
  template:
    metadata:
      labels:
        app: my-app
    spec:
      containers:
        - name: my-app
          image: idaacr0707.azurecr.io/my-python-app:v1
        ports:
          - containerPort: 5000
```



# kubcluster



Kubernetes service

[Create](#)[Connect](#)[Start](#)[Stop](#)[Delete](#)[Refresh](#)[Open in mobile](#)[Give feedback](#)

## Overview

[Activity log](#)[Access control \(IAM\)](#)[Tags](#)[Diagnose and solve problems](#)[Microsoft Defender for Cloud](#)

## Kubernetes resources

[Namespaces](#)[Workloads](#)[Services and ingresses](#)[Storage](#)[Configuration](#)[Custom resources](#)[Run command](#)

## Settings

### Essentials

Resource group

[RG\\_IDAshell](#)

Status

Succeeded (Running)

Location

East US

Subscription

[npunext-1673504988896](#)

Subscription ID

2190116a-14d8-4b0c-8821-f97ab2e43421

[Tags \(edit\)](#)[Add tags](#)[Get started](#)[Properties](#)[Monitoring](#)[Capabilities \(4\)](#)[Recommendations](#)[Tutorials](#)

#### Kubernetes services

Encryption type

Encryption at-rest with a platform-managed key

Virtual node pools

Not enabled



#### Networking

API server address

kubcluster-dns-rlf7top9.hcp.eastus.azmk8s.io

Network type (plugin)

Azure CNI

Pod CIDR

Service CIDR

10.0.0.0/16



#### Node pools

52.66.64.204 - Remote Desktop Connection

labuser@ip-172-31-14-209:~/Desktop/DockerProject/Kubernetes\$ az aks update -n kubcluster -g RCI  
IDAshell --attach-acr idaacr01  
AAD role propagation done[#####] 100.0000% - Starting  
AAD role propagation done[#####] 100.0000%{  
    "aadProfile": null,  
    "addonProfiles": {  
        "azureKeyvaultSecretsProvider": {  
            "config": null,  
            "enabled": false,  
            "identity": null  
        },  
        "azurepolicy": {  
            "config": null,  
            "enabled": true,  
            "identity": {  
                "clientId": "5afc9d8a-88dd-4481-9fbc-ac621f5fa542",  
                "objectId": "1bc76767-c173-47a0-9f83-048986963929",  
                "resourceId": "/subscriptions/2190116a-14d8-4b0c-8821-f97ab2e43421/resourcegroups/MC\_F\_IDAshell\_kubcluster\_eastus/providers/Microsoft.ManagedIdentity/userAssignedIdentities/azurepolicy-kubcluster"  
            }  
        },  
        "omsAgent": {  
            "config": {  
                "logAnalyticsWorkspaceResourceID": "/subscriptions/2190116a-14d8-4b0c-8821-f97ab2e43421/resourceGroups/DefaultResourceGroup-EUS/providers/Microsoft.OperationalInsights/workspaces/DefaultWorkspace\_2190116a\_14d8\_4b0c\_8821\_f97ab2e43421\_EUS"  
            }  
        }  
    }  
}

# Create a starter application

...

Deploy a basic web or single-image application to get started with Kubernetes. No prior Kubernetes experience required.

If you already have a YAML file or want to create a more complex deployment, [add with YAML](#) instead.



## Create a basic web application

Learn how to create and deploy a basic web app with just a few clicks.

[Create>](#)



## Create a single-image application

Learn how to create and deploy a single-image from a container registry with a quick tutorial.

[Create>](#)

# Create a single-image application

...

✓ Get image

✓ Application details

✓ Review YAML

4 Deploy

## ✓ The application has been deployed

You can see the deployment you created in the list below and click on the deployment name to see more details, including the individual pods that were created as part of the deployment.

### 1. Connecting your Azure Container Registry

Connecting your Azure Container Registry to your cluster gives your cluster the permissions it needs to pull images from that registry.

✓ idaacr01

### 2. Deploying resources

Select a deployment name to see more details, including the individual pods that were created as part of the deployment.

Resource	Type	Status
✓ default-1695897647119	Namespace	Success
✓ python-app	Deployment	Success

Previous

Close

# Azure Voting App

---

Cats

Dogs

Reset

---

Cats - 2 | Dogs - 6

---

# Day 22 – Azure DevOps

- ▶ DevOps introduction and CI/CD
- ▶ Collaboration (main) and feature branches
- ▶ Azure DevOps project creation and setup
- ▶ Azure Boards and Work items
- ▶ Epic, Feature, User story and Task
- ▶ Sprint and typical sprint duration
- ▶ Creating Azure boards for a real time scenario
- ▶ Billing configuration in Organization settings
- ▶ Dashboards and Wiki
- ▶ Waterfall and Agile models

# Day 22 continued...

- ▶ Work item creation and linking them with proper hierarchy
- ▶ Repos overview and implementation
- ▶ Creating and cloning repos
- ▶ Pull requests to merge feature branch with the main branch
- ▶ Version control through revert
- ▶ Creating pipelines using a sample GitHub repository
- ▶ Configure with starter or python package
- ▶ Run the pipelines after generating a yaml file
- ▶ Case study discussion

## idashell



Overview

Summary

Dashboards

Wiki

Boards

Repos

Pipelines

Test Plans

Artifacts

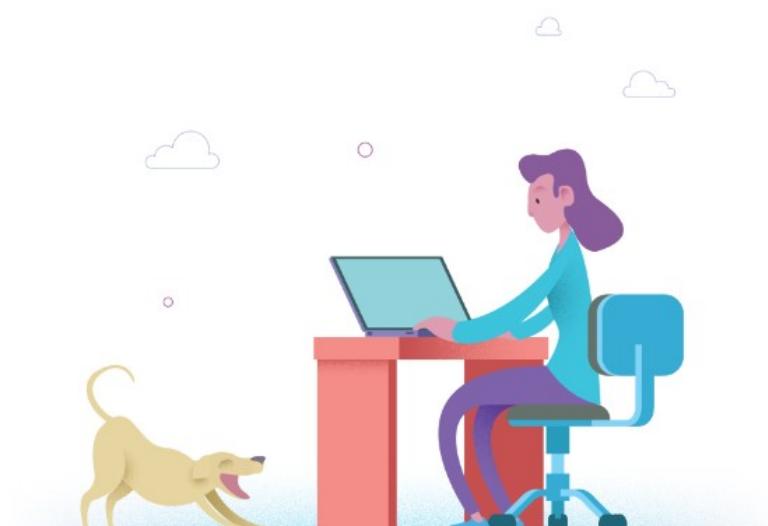
Project settings



## idashell

Private

Invite



## Welcome to the project!

What service would you like to start with?

Boards

Repos

Pipelines

Test Plans



## **Set up billing**

Select an Azure subscription for billing your organization's Azure DevOps charges

X

Select an Azure subscription

- npunext-1673504988896  
2190116a-14d8-4b0c-8821-f97ab2e43421

  Linking subscription to Organization

+ New Azure subscription

Cancel

Save

2190116a-14d8-4b0c-8821-f97ab2e43421

Change billing

Configure user billing

Pipelines for private  
projects

Free

Paid parallel jobs

MS Hosted CI/CD [↗](#)

1

Self-Hosted CI/CD [↗](#) 1

0

Visit [parallel jobs](#) for full details on free pipelines and public concurrency

## idashell



## Overview

## Boards

## Work items

## Boards

## Backlogs

## Sprints

## Queries

## Delivery Plans

## Analytics views

## Queries &gt; My Queries &gt; dev epic with workitem ▾ ★

Query returned no results.

Results Editor Charts

Run query

New ▾

Save query

Save as...

Rename

...



Type of query Flat list of work items

Query across projects 

Filters for top level work items

And/Or	Field*	Operator	Value
<input type="checkbox"/>	Work Item Type	=	Epic
<input type="checkbox"/> And	Tags	Contains	dev
Add new clause			

I

**idashell**

## About this project

Help others to get on board!

Describe your project and make it easier for others to get involved.

+ Add Project Description

## Project stats

Boards



1

Work items created

Members 1

## About this project

X

### Description

idashell

### Tags

Add tags

### About

Readme file  Wiki

### Preview

This IDA shell project is to train all the new resources in Data Engineering task.

Cancel

Save

## idashell



## Work items

Recently updated

+ New Work Item

Open in Queries

Column Options



Filter by keyword

Types

Assigned to

States

Area

Tags



ID	Title	Assigned To	State	Area
174	Design code so that user lands on the home page on successful login	Unassigned	New	idashell
173	Design the web page	Unassigned	New	idashell
172	Check user credentials	Unassigned	New	idashell
171	Planning the web page layout	Unassigned	New	idashell
170	User Authentication	Unassigned	New	idashell
169	Design the application	Unassigned	New	idashell
168	Online Ordering System	Unassigned	New	idashell
1	Test	Unassigned	To Do	idashell



## 177 US - Create a login page for webapp

su Shellunext une...

0 comments

Save &amp; Close

Follow



Add tag

State

● Active

Area

idashell

Reason

🔒 Implementati...

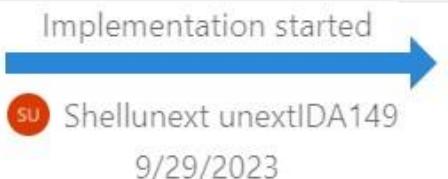
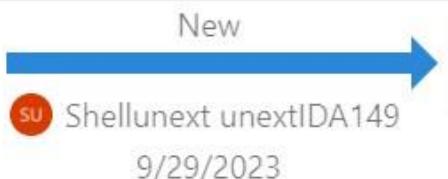
Iteration

idashell\Sprint 1

Details



(2)



## History

▼ Today	SU	Shellunext unextIDA... Fri 9/29/2023 6:11 AM	Changed state from ● New to ● Active, ...
	SU	Shellunext unextIDA... 6:11 AM	Activated By Shellunext unextIDA149
	SU	Shellunext unextIDA... 6:11 AM	Reason Implementation started New

State	<span>● Active</span>	Area	idashell
Reason	<span>🔒 Implementation</span>	Iteration	idashell\Sprint 1

Details



🔗 (2)



## Acceptance Criteria

As a developer, I should develop the webapp with below criteria:

1. proper header
2. company color palette should match
3. alignment

## Planning

Story Points

## Deployment

To track releases associated with this



## Create a repository

Repository type



Git



Repository name \*

adfrepo

 Add a READMEAdd a .gitignore: **None**Your repository will be initialized with a **main** branch.

Cancel

Create



## Commit

adfrepo

M README.md

config.json

Comment

Added config.json

Branch name

main

Work items to link

Search work items by ID or title



Cancel

Commit

config.json

README.md

## config.json

 Edit

Contents History Compare

## Create a branch

Name \*

sahith

Based on

main

Work items to link

Search work items by ID or title

Cancel

Create

# New pull request

from sahith into main

Overview Files 1 Commits 1

## Title

Updated config.json

## Description

Updated config.json

19/4000

① [Markdown supported.](#) Drag & drop, paste, or select files to insert.

① [Link work items.](#)

@ # ⌂ Ø A ↴ B I ↵ ⊞ ⊞ ⊞

Updated config.json

## Reviewers

Add required reviewers

② Search users and groups to add as reviewers

## Work items to link

# Updated config.json

Approve



Complete



Active

!1

su

Shellunext unextIDA149 proposes to merge sahith into main

su

Overview Files Updates Commits



No merge conflicts

Last checked Just now

Description



Updated config.json

Reviewers

Add

Required

No required reviewers

Optional



su

Shellunext unextIDA149

No review yet

Tags





g sahith ▾



## Commits



On

Commit ID



Commit

**Updated config.json**

e1d62bc6 Shellunext unextIDA149 Today at 11:53 AM

**Added config.json**

35d44cf3 Shellunext unextIDA149 Today at 11:49 AM

**Added README.md**

22a8d634 Shellunext unextIDA149 Today at 11:45 AM

I

sahith ▾

+

Pushes

≡

9/29/2023 (2 updates)

> Updated to e1d62bc6: Updated config.json  
e1d62bc6 Shellunext unextIDA149 Today at 11:53 AM

> Created at 35d44cf3: Added config.json  
35d44cf3 Shellunext unextIDA149 Today at 11:53 AM

# Branches

New branch

Mine All Stale

≡ Search branch name

Branch

Behind | Ahead

main

Default

Compare



sahith

1 | 0

The screenshot shows a GitHub repository interface for a repository named "adfrepo".

**Left Sidebar:**

- A red icon with a white letter "I" is at the top.
- A plus sign (+) icon is below it.
- Icons for Issues, Pull Requests, Commits, and Other branches are listed vertically.

**Repository Overview:**

- The repository name "adfrepo" is displayed with a small orange diamond icon.
- A dropdown menu is open next to the repository name, showing options like "main" and "config.json".
- File list:
  - "config.json" is highlighted with a blue background.
  - "README.md" is shown below it.

**config.json File View:**

- The file name "config.json" is displayed prominently.
- Actions: "Edit" (with a pencil icon) and three-dot menu.
- Navigation links: "Contents", "History", "Compare", and "Blame".
- A message box indicates "1 changed json input".

## Added hello.py

Completed

!2

su

Shellunext unextIDA149 proposes to merge sahith

Overview Files Updates Commits



Shellunext unextIDA149 completed this pull request 3m ago

Cherry-pick

Merged PR 2: Added hello.py

b3b304d0 su Shellunext unextIDA149 Today at 12:14 PM

Show details



No merge conflicts  
Last checked 3m ago

Description

Added hello.py

Show everything

## Revert pull request 2



Revert the changes made in this pull request into a target branch. A topic branch will be created with the reverted changes, and then you will be prompted to create a pull request to the target branch.

Target branch

main ▾

Topic branch name required

sahith-revert-from-main

Cancel

Revert

 idashell +

 Overview

 Boards

 Repos

 Files

 Commits

 Pushes

 Branches

 Tags

 Pull requests

 Advanced Security

 Pipelines

 Test Plans

 Project settings <

## New pull request

(i) The revert operation successfully created the new branch  sahith-revert-from-main. X

 sahith-revert-from-main  into  main  

[Overview](#) [Files 1](#) [Commits 1](#)

**Title**  
Revert 'Added hello.py'

**Description**  
Added hello.py  
  
Reverts !2

26/4000

(i) [Markdown supported.](#) Drag & drop, paste, or select files to insert. (i) [Link work items.](#)



Added hello.py

Reverts  PR 2: Added hello.py

**Reviewers** Add required reviewers

## Revert 'Added hello.py'

Active

!3

su

Shellunext unextIDA149 proposes to merge sahith-revert-from-main into main

Overview Files Updates Commits



No merge conflicts

Last checked Just now

### Description

Added hello.py

Reverts [PR 2: Added hello.py](#)

Show everything



su

Add a comment...

✓

su

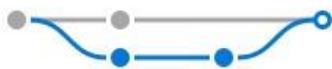
Shellunext unextIDA149 approved the pull request

## Complete pull request



Merge type

Merge (no fast forward)



Post-completion options



Complete associated work items after merging



Delete sahith-revert-from-main after merging



Customize merge commit message

Cancel

Complete merge

## Updated config.json

Active

!3 su Shellunext unextlDA155 proposes to merge idashell into main 0/1 comments resolved

Overview Files Updates Commits



No merge conflicts

Last checked Just now

### Description



Updated config.json to change name to clara and gender to female.

Show everything (2) ▾



SU

Add a comment...



SU

Shellunext unextlDA155 Just now

↪ Active ▾

change the code based on the latest discussion we had through teams.

SU

Shellunext unextlDA155 Just now





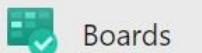
idashell

+

⚠ You selected a public repository, but this is not a public project. Go to project settings to change the visibility of the project. [Learn more](#)



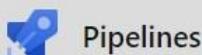
Overview



Boards



Repos



Pipelines



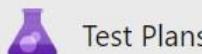
Pipelines



Environments



Library



Test Plans



Artifacts



Project settings

&lt;

✓ Connect

✓ Select

Configure

Review

New pipeline

## Configure your pipeline



Docker

Build a Docker image



Docker

Build and push an image to Azure Container Registry



Deploy to Azure Kubernetes Service

Build and push image to Azure Container Registry; Deploy to Azure Kubernetes Service



Python package

Create and test a Python package on multiple Python versions.



Python to Linux Web App on Azure

Build your Python project and deploy it to Azure as a Linux Web App.



Starter pipeline

Start with a minimal pipeline that you can customize to build and deploy your code.



Existing Azure Pipelines YAML file

Select an Azure Pipelines YAML file in any branch of the repository.

⚠ You selected a public repository, but this is not a public project. Go to [project settings](#) to change the visibility of the project. [Learn more](#) X

✓ Connect

✓ Select

✓ Configure

Review

New pipeline

## Review your pipeline YAML

Variables

Save and run



Sahith03/python-sample-vscode-flask-tutorial / azure-pipelines.yml \* □

↶ Show assistant

```
1 # Starter pipeline
2 # Start with a minimal pipeline that you can customize to build and deploy your code.
3 # Add steps that build, run tests, deploy, and more:
4 # https://aka.ms/yaml
5
6 trigger:
7 - main
8
9 pool:
10   vmImage: ubuntu-latest
11
12 steps:
13 - script: echo Hello, world!
14   displayName: 'Run a one-line script'
15
16 - script: |
17   echo Add other tasks to build, test, and deploy your project.
18   echo See https://aka.ms/yaml
19   displayName: 'Run a multi-line script'
```



## #20230929.1 • Set up CI with Azure Pipelines

Sahith03.python-sample-vscode-flask-tutorial

Run new

:

This run is being retained as one of 3 recent runs by pipeline.

[View retention leases](#)

### Summary

Triggered by Sahith03

[View change](#)

Repository and version Sahith03/python-sample-vscode-flask-tutorial  
 main 79281dbd

Time started and elapsed Just now  
 16s

Related 0 work items  
 0 artifacts

Tests and coverage [Get started](#)

### Jobs

Status

Duration

## Summary Tests

- ✖ Job Python27 • Use Python 2.7
- ✖ Version spec 2.7 for architecture x64 did not match any version in Agent.ToolsDirectory. Versions in /opt/hostedtoolcache: 3.10.13 (x64) 3.11.5 (x64) 3.7.17 (x64) 3.8.18 (x64) 3.9...  
Job Python27 • Use Python 2.7
- ✖ Failed to download Python from the Github Actions python registry (<https://github.com/actions/python-versions>). Error: Error: Could not find Python matching spec 3.5 (x64) i...  
Job Python35 • Use Python 3.5
- ✖ Version spec 3.5 for architecture x64 did not match any version in Agent.ToolsDirectory. Versions in /opt/hostedtoolcache: 3.10.13 (x64) 3.11.5 (x64) 3.7.17 (x64) 3.8.18 (x64) 3.9...  
Job Python35 • Use Python 3.5
- ✖ Failed to download Python from the Github Actions python registry (<https://github.com/actions/python-versions>). Error: Error: Could not find Python matching spec 3.6 (x64) i...  
Job Python36 • Use Python 3.6
- ✖ Version spec 3.6 for architecture x64 did not match any version in Agent.ToolsDirectory. Versions in /opt/hostedtoolcache: 3.10.13 (x64) 3.11.5 (x64) 3.7.17 (x64) 3.8.18 (x64) 3.9...  
Job Python36 • Use Python 3.6



[View documentation for troubleshooting failed runs](#)

## Jobs

Name	Status	Duration
<span style="color: red;">✖</span> Job Python27	Failed	⌚ 5s
<span style="color: red;">✖</span> Job Python35	Failed	⌚ 7s
<span style="color: red;">✖</span> Job Python36	Failed	⌚ 6s
<span style="color: green;">✓</span> Job Python37	Success	⌚ 24s

## ← Jobs in run #20230...

Sahith03.python-sample-vscode-flask-tutorial

### Jobs

	Job	6s
✓	Initialize job	<1s
✓	Checkout Sahith...	2s
✓	Run a one-line ...	<1s
✓	Run a multi-lin...	<1s
✓	Post-job: Check...	<1s
✓	Finalize Job	<1s

## ✓ Job



View raw log



```
1 Pool: Azure Pipelines
2 Image: ubuntu-latest
3 Agent: Hosted Agent
4 Started: Just now
5 Duration: 6s
6
7 ▼ Job preparation parameters
8 ContinueOnError: False
9 TimeoutInMinutes: 60
10 CancelTimeoutInMinutes: 5
11 Expand:
12 MaxConcurrency: 0
13 ##### System Pipeline Decorator(s) #####
14
15 Begin evaluating template 'system-pre-steps.yml'
16 Evaluating: eq('true', variables['system.debugContext'])
17 Expanded: eq('true', Null)
18 Result: False
19 Evaluating: resources['repositories']['self']
20 Expanded: object
21 Result: True
22 Evaluating: not(containsValue(job['steps'][*]['task']['id'], '6d15af64-176c-496d-b583-fd2ae21d4df4'))
23 Expanded: not(containsValue(Object, '6d15af64-176c-496d-b583-fd2ae21d4df4'))
24 Result: True
25 Evaluating: resources['repositories']['self']['checkoutOptions']
26 Result: object
```



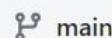
Sahith03 / python-sample-vscode-flask-tutorial

Type  to search[Code](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

## Commit

✓ Set up CI with Azure Pipelines

[skip ci]

[Browse files](#)

main



Sahith03 committed 3 minutes ago

1 parent e792293 commit 79281db

Showing 1 changed file with 19 additions and 0 deletions.

[Split](#) [Unified](#)

▼ 19 ██████ azure-pipelines.yml

...

```
...     ... @@ -0,0 +1,19 @@
1      + # Starter pipeline
2      + # Start with a minimal pipeline that you can customize to build and deploy your code.
3      + # Add steps that build, run tests, deploy, and more:
4      + # https://aka.ms/yaml
5      +
6      + trigger:
7      + - main
8      +
9      + pool:
10     +   vmImage: ubuntu-latest
```

I idashell +

---

 Overview

 Boards

 Repos

 Pipelines

 Pipelines

 Environments

 Library

 Test Plans

 Artifacts

## Pipelines

New pipeline

Recent All Runs

Filter pipelines

### Recently run pipelines

Pipeline

Last run



Sahith03.python-sample-vscode-flask-tut...

#20230929.1 • Set up CI with Azure Pipelines

⌚ Individual CI for  main

🕒 4m ago

⌚ 16s