Explore Spark Streaming in Azure Synapse Analytics
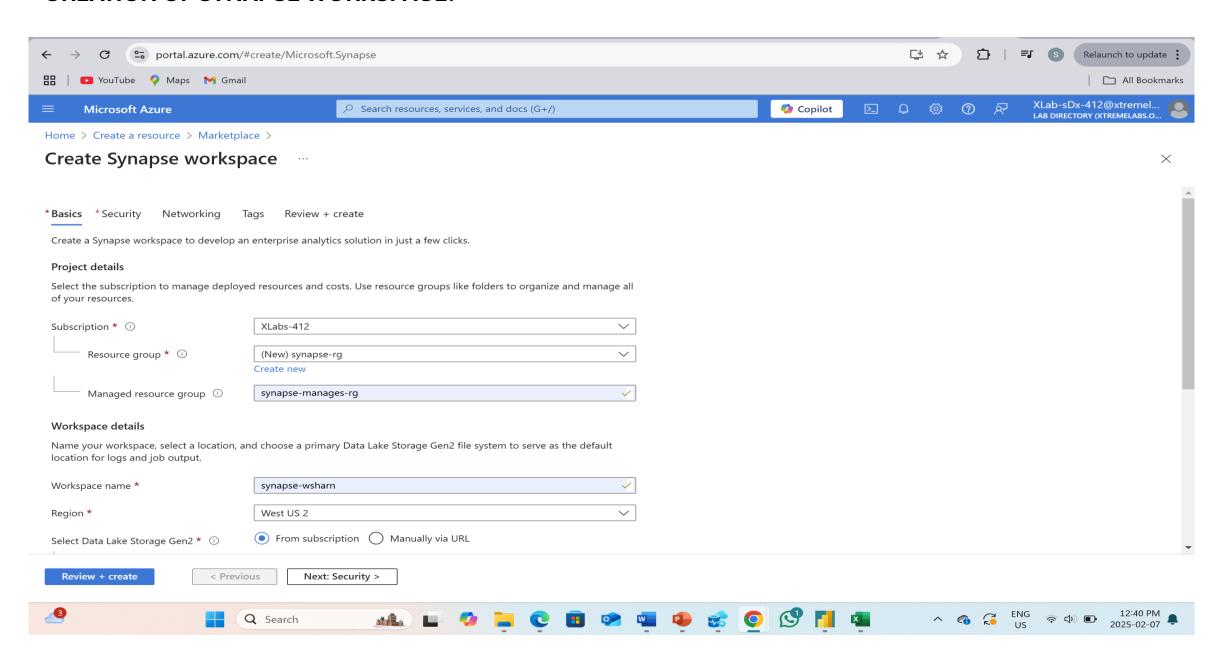
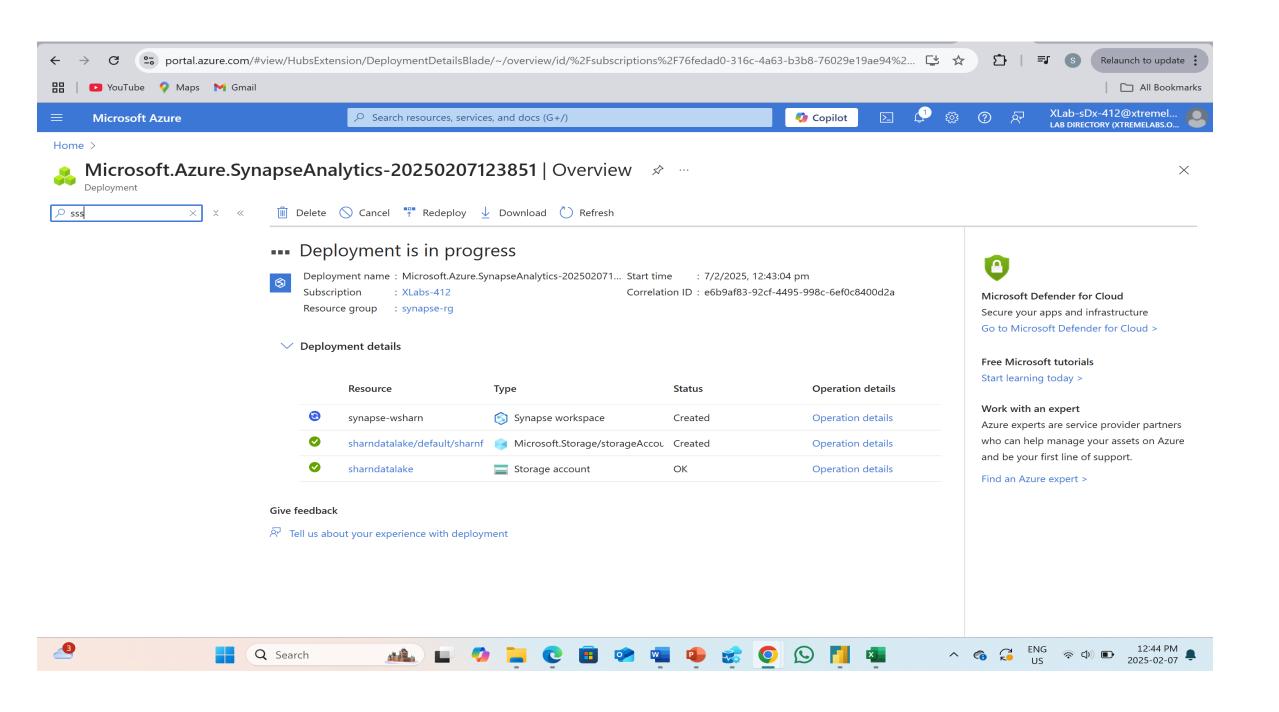**SUBMITTED BY:**

SHARNDEEP KAUR

https://www.linkedin.com/in/sharndeep-kaur-b866441a6/

LEARNING OUTCOMES:
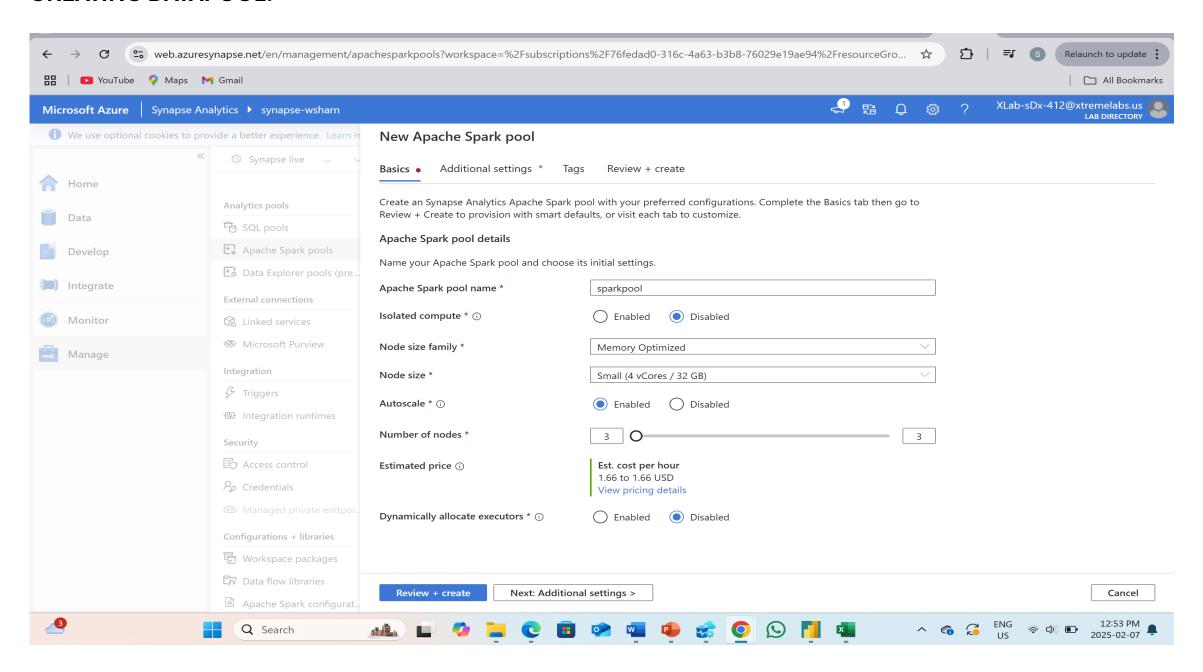
❑ **Provisioning a Synapse Workspace** – Learn how to set up an Azure Synapse Analytics workspace, including creating necessary resources like a Data Lake storage account and Spark pool.

❑ **Setting Up and Using Spark Pools** – Understand how to create and configure an Apache Spark pool within Synapse Analytics for processing big data workloads.

❑ **Exploring Stream Processing** – Gain hands-on experience with Spark Structured Streaming and Delta Tables to process streaming data efficiently.

❑ **Managing Azure Resources** – Learn how to clean up resources in Azure to optimize costs by deleting unnecessary resource groups after completing tasks.
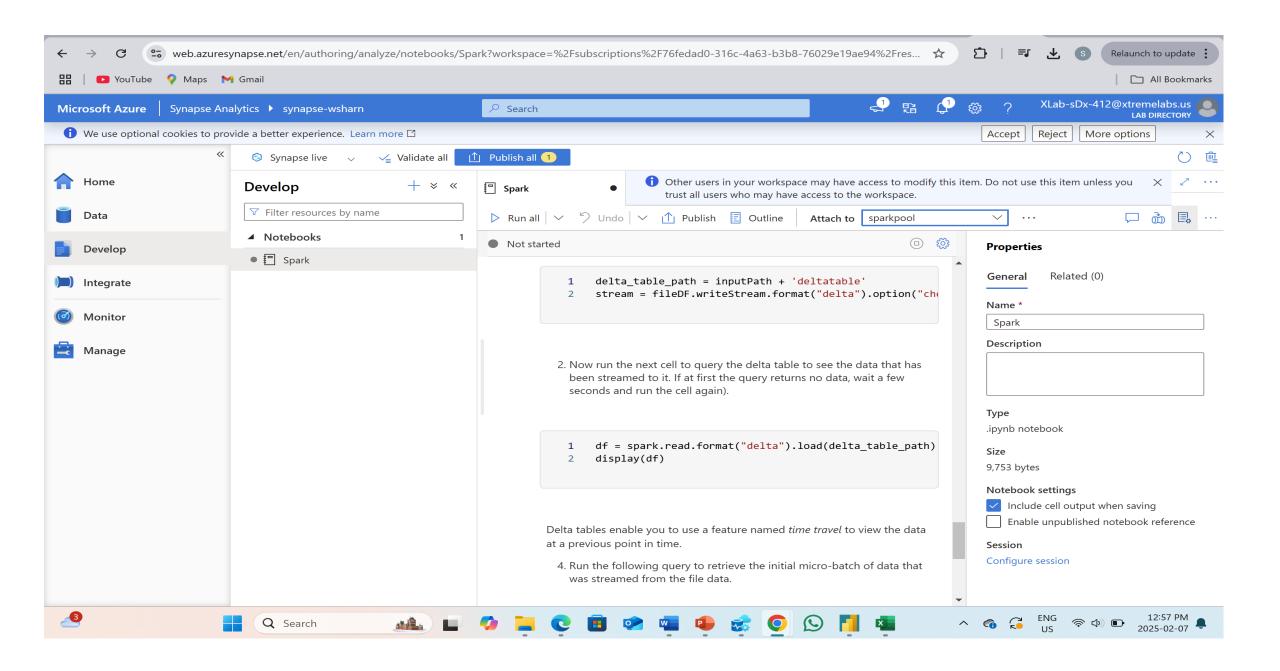
CREATION OF SYNAPSE WORKSPACE:

YouTube   Maps   Gmail

All Bookmarks

Microsoft Azure

Search resources, services, and docs (G+/)

Copilot

XLab-sDx-412@xtremel...
LAB DIRECTORY (XTREMELABS.O...

Home >

Microsoft.Azure.SynapseAnalytics-20250207123851 | Overview
Deployment

sss

Delete   Cancel   Redeploy   Download   Refresh

••• Deployment is in progress

Deployment name	: Microsoft.Azure.SynapseAnalytics-202502071...	Start time	: 7/2/2025, 12:43:04 pm
Subscription	: XLabs-412	Correlation ID	: e6b9af83-92cf-4495-998c-6ef0c8400d2a
Resource group	: synapse-rg		

∨ Deployment details

Resource	Type	Status	Operation details
synapse-wsharn	Synapse workspace	Created	Operation details
sharndatalake/default/sharnf	Microsoft.Storage/storageAccou	Created	Operation details
sharndatalake	Storage account	OK	Operation details

Give feedback

Tell us about your experience with deployment

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 >

Search

ENG
US
12:44 PM
2025-02-07

CREATING DATAPOOL:

Develop

Notebooks   1

- Spark

Spark

Run all | Undo | Publish | Outline   Attach to sparkpool

● Not started

```
1  delta_table_path = inputPath + 'deltatable'
2  stream = fileDF.writeStream.format("delta").option("ch
```

2. Now run the next cell to query the delta table to see the data that has been streamed to it. If at first the query returns no data, wait a few seconds and run the cell again).

```
1  df = spark.read.format("delta").load(delta_table_path)
2  display(df)
```

Delta tables enable you to use a feature named *time travel* to view the data at a previous point in time.

4. Run the following query to retrieve the initial micro-batch of data that was streamed from the file data.

Properties

General   Related (0)

Name *
Spark

Description

Type
.ipynb notebook

Size
9,753 bytes

Notebook settings
☑ Include cell output when saving
☐ Enable unpublished notebook reference

Session
Configure session

Microsoft Azure | Synapse Analytics ▸ synapse-wsharn

XLab-sDx-412@xtremelabs.us
LAB DIRECTORY

Synapse live ⌄   Validate all   Publish all 1

Develop

Filter resources by name

◢ **Notebooks**　　　　　1

　● 📖 Spark

📖 Spark　　　●

ℹ Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

⊘ Cancel all ⌄ | ↶ Undo ⌄ | ⬆ Publish | ▦ Outline　Attach to [sparkpool ⌄]

✅ Ready

```
3    # Create a folder
4    inputPath = '/data/'
5    mssparkutils.fs.mkdirs(inputPath)
```

[*]   ✳ 3 min 38 sec - Waiting

Use Spark Structured Streaming to query a stream of data

1. Run the cell below to create a streaming dataframe that reads data from the folder based on a JSON schema that includes the name of the device and its status.

```
1    from pyspark.sql.types import *
2    from pyspark.sql.functions import *
3
4    # Create a stream that reads data from the folder, usi
5    jsonSchema = StructType([
6      StructField("device", StringType(), False),
7      StructField("status", StringType(), False)
```

Properties

General　　Related (0)

Name *

[Spark]

Description

[　　　　　　　　　　　　　]

Type

.ipynb notebook

Size

16,058 bytes

Notebook settings

☑ Include cell output when saving

☐ Enable unpublished notebook reference

Session

Configure session

Microsoft Azure   🔍 Search resources, services, and docs (G+/)   Copilot   XLab-sDx-412@xtremel...
LAB DIRECTORY (XTREMELABS.O...

Home >

synapse-rg
Resource group

🔍 Search

- Overview
- Activity log
- Access control (IAM)
- Tags
- Resource visualizer
- Events
- Settings
- Cost Management
- Monitoring
- Automation
- Help

+ Create   ⚙ Manage view ⌄   🗑 Delete resource group   🔄 Refresh   ⬇ Export to CSV

⌄ Essentials

Resources   Recommendations

Filter for any field...   Type equals **all** ✕   Location equals **all** ✕   + Add filte

Showing 1 to 3 of 3 records.   ☐ Show hidden types ⓘ

☑ Name ↑↓

- ☐ sharndatalake
- ☑ sparkpool (synapse-wsharn/sparkpool)
- ☐ synapse-wsharn

< Previous   Page 1 ⌄ of 1   Next >

Delete Resources ✕

The selected resources along with their related resources and contents will be permanently deleted. If you are unsure of the selected resource dependencies, navigate to the individual resource page to perform the delete operation. More details of the resource dependencies are available in the manage experience.

Resources to be deleted (1)

Name	Resource type	
sparkpool (synapse-wsharn/sparkpool)	Apache Spark pool	Remove

Enter "delete" to confirm deletion *

deletes

Delete   Cancel

ENG US   1:04 PM 2025-02-07

THANK YOU!!!!!!!!!