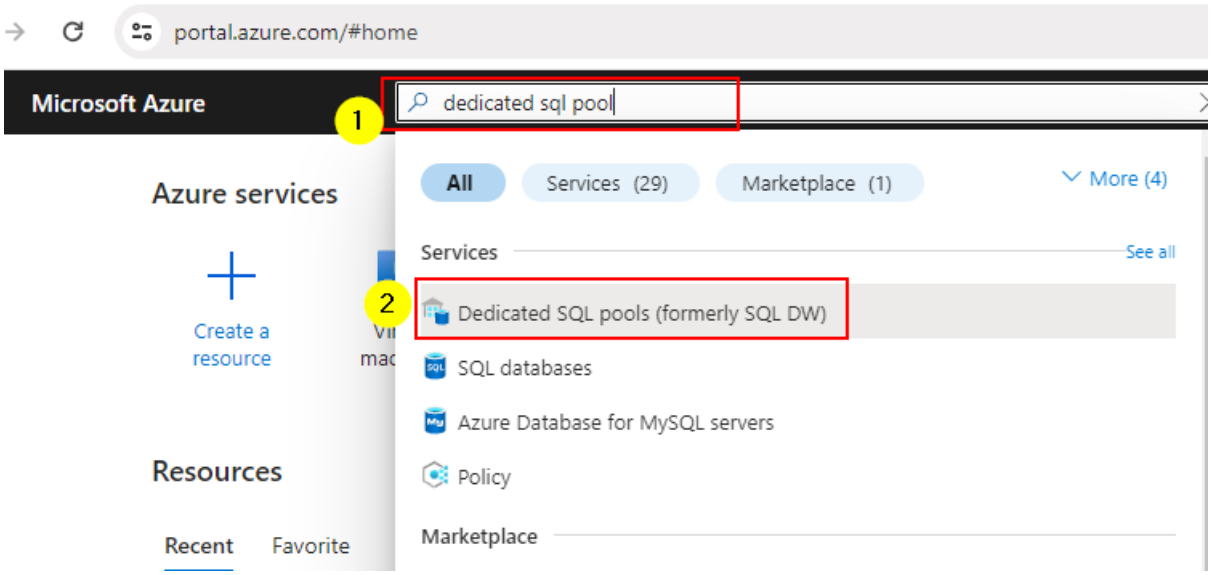
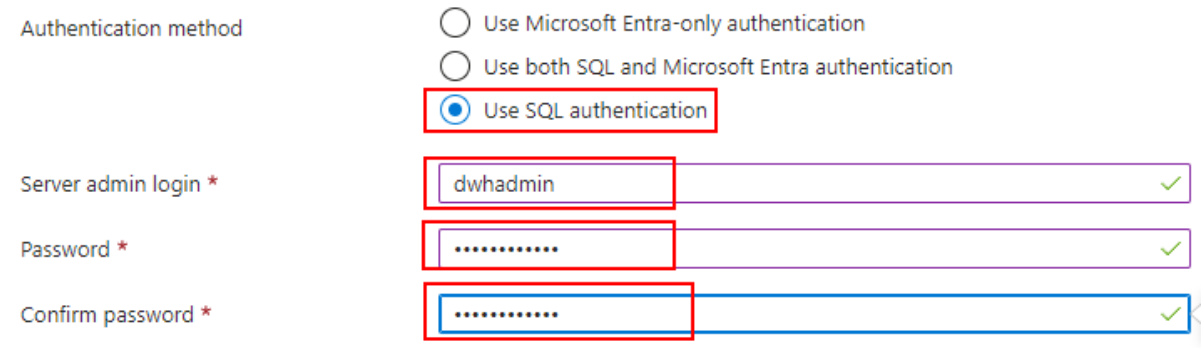
Setup Dedicated SQL Pool in Azure

1. Open Azure Portal (https://portal.azure.com/#home)
2. Search dedicated SQL pool in search bar

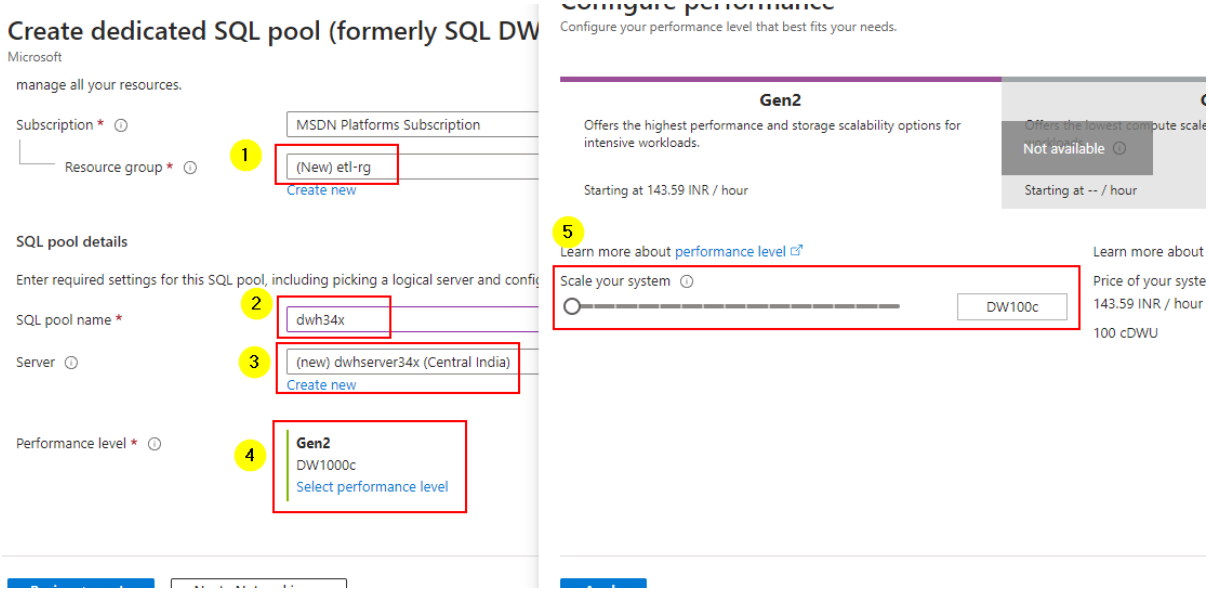


3. On dedicated SQL pool page click create
4. Create a new resource group: etl-rg
5. SQL pool name: dwh34x
6. Create server
    a. Server Name: dwhserverxxxx (replace xxxx with random number)
    b. Location: Central India
    c. Authentication method: Use SQL authentication
        i. Server admin login: dwhadmin
        ii. Password: demo!pass123
        iii. Confirm password: demo!pass123

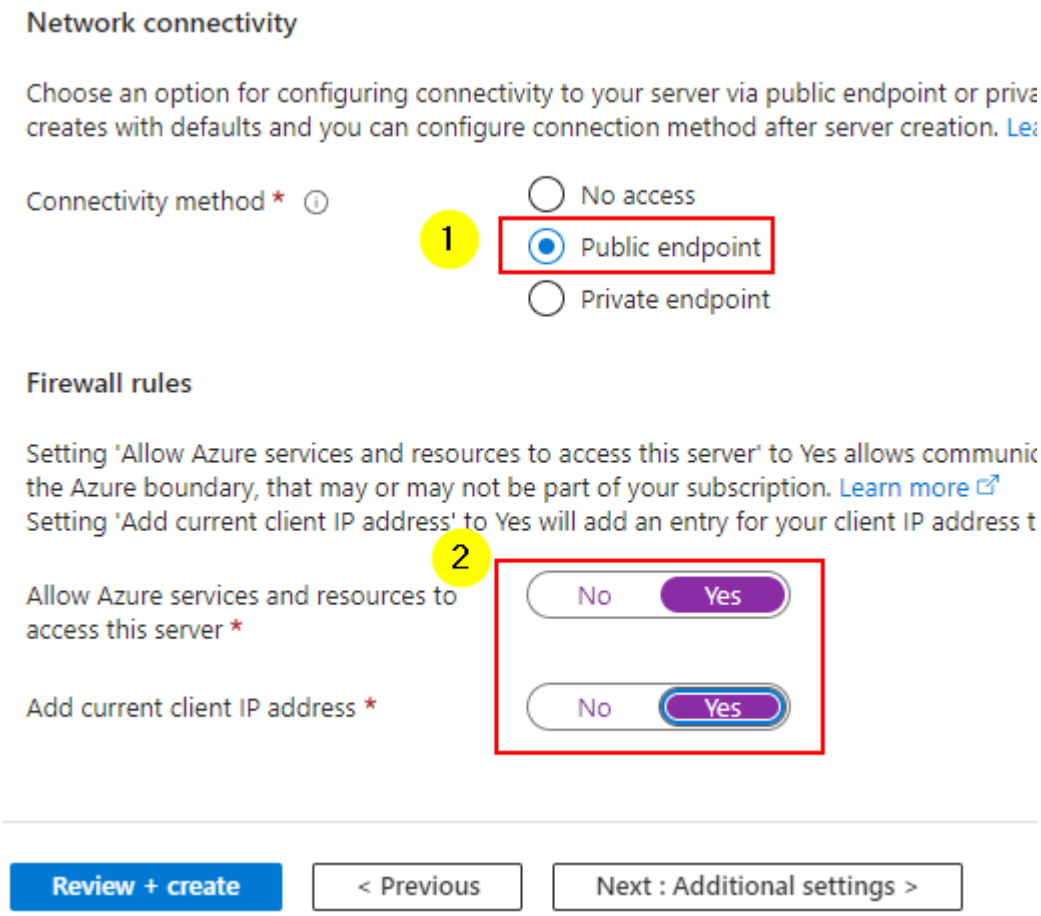


d. Click Ok

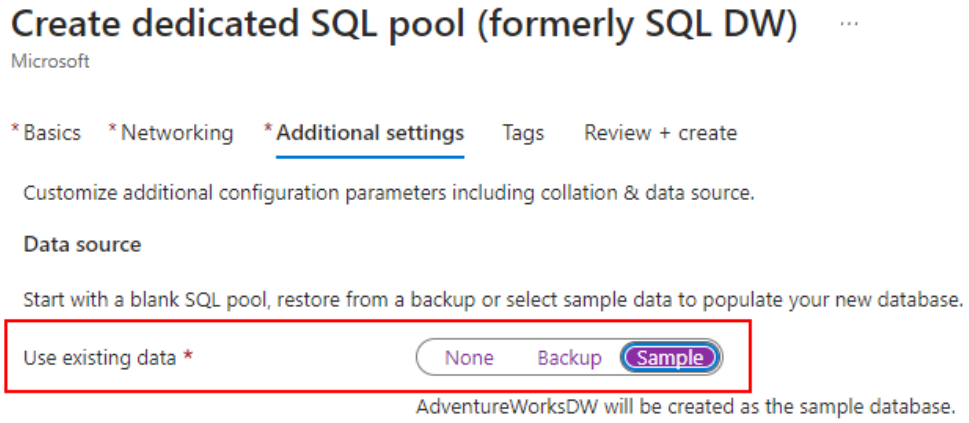
7. Performance level: DW100c



8. Click Next: Networking
9. In Networking Tab:
    a. Connectivity method: Public endpoint
    b. Firewall rule:
        i. Allow Azure services and resources to access this server: Yes
        ii. Add current client IP address: Yes



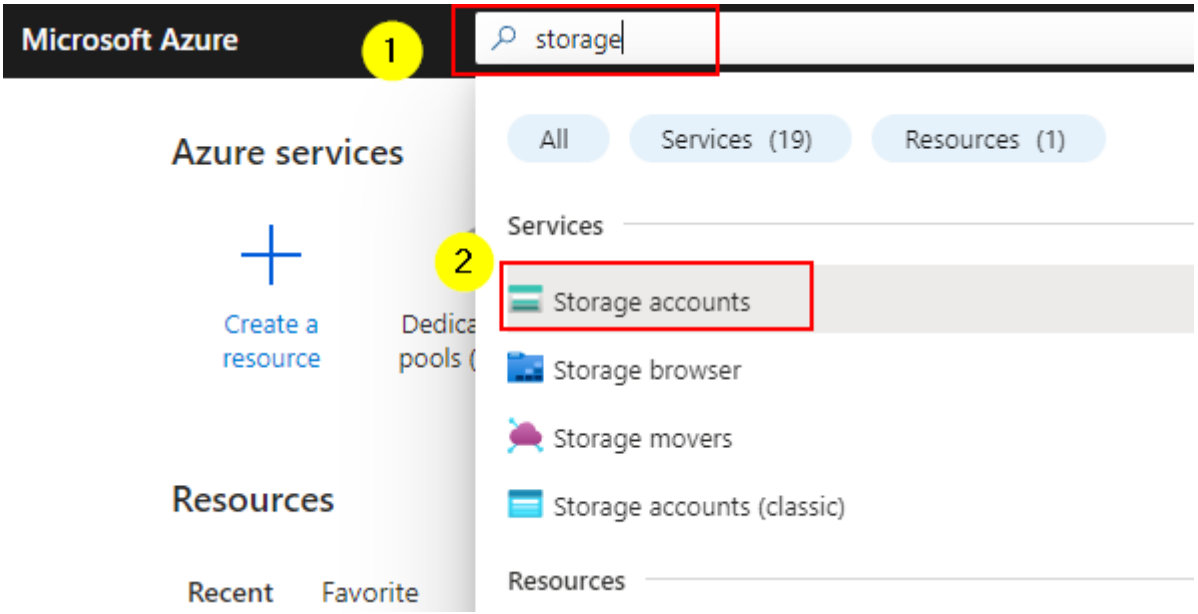
10. Click Next: Additional Setting
11. On Additional setting tab page
    a. Use existing data: Sample



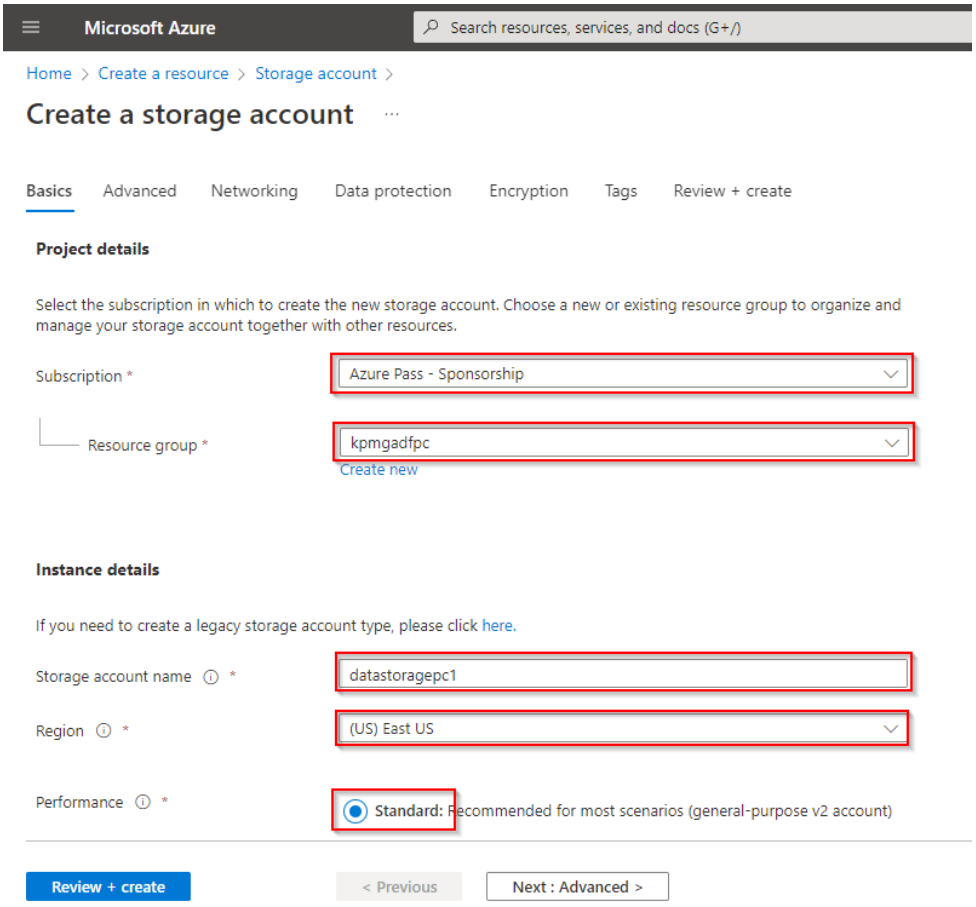
12. Click Review + Create and then Click Create
13. Wait for the deployment too complete.

Setup Data Lake in Azure

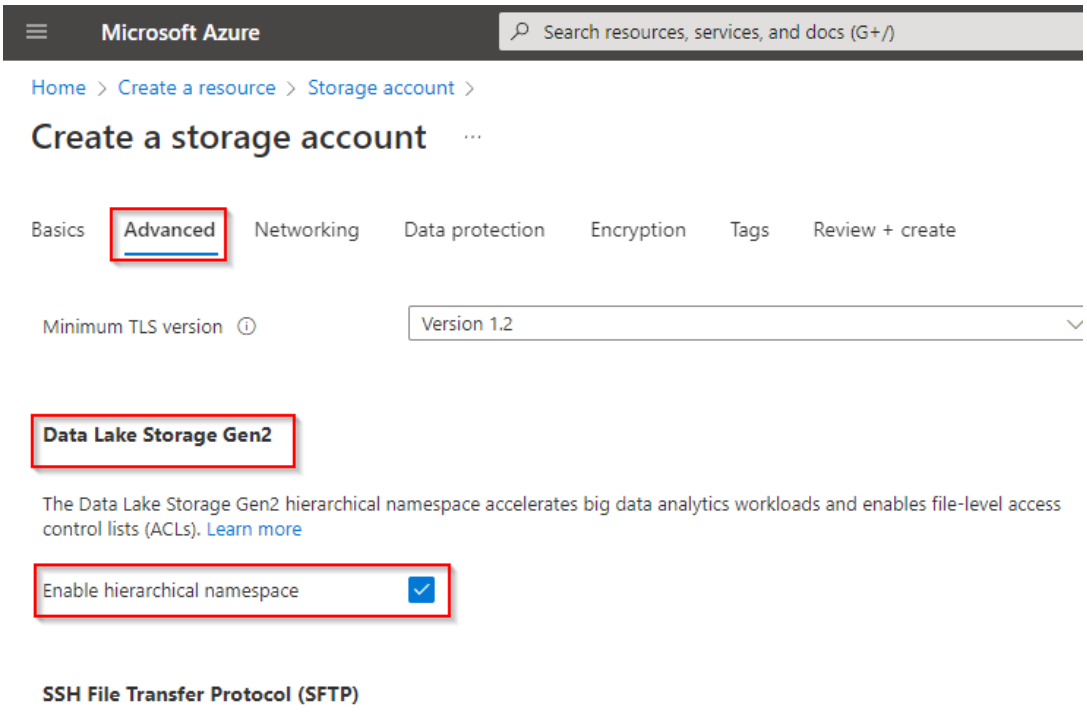
1. Go to Azure home page (https://portal.azure.com/#home)
2. In the search bar search storage and click Storage accounts



3. On storage account page click **+ Create**
4. On **Create a storage account** page in the **Basics** tab, provide the following settings:
    - Under the project details, specify the following settings:
        - **Subscription**: the name of the subscription you are using in this lab
        - **Resource group**: **etl-rg**.
    - Under the Instance details, specify the following settings:
        - **Storage account name**: **datalakexx**, where **xx** are your initials.
        - **Region**: the name of the Azure region which is closest to the lab location. (Central India)
        - **Performance**: **Standard**.
        - **Redundancy**: **Locally-redundant storage (LRS)**

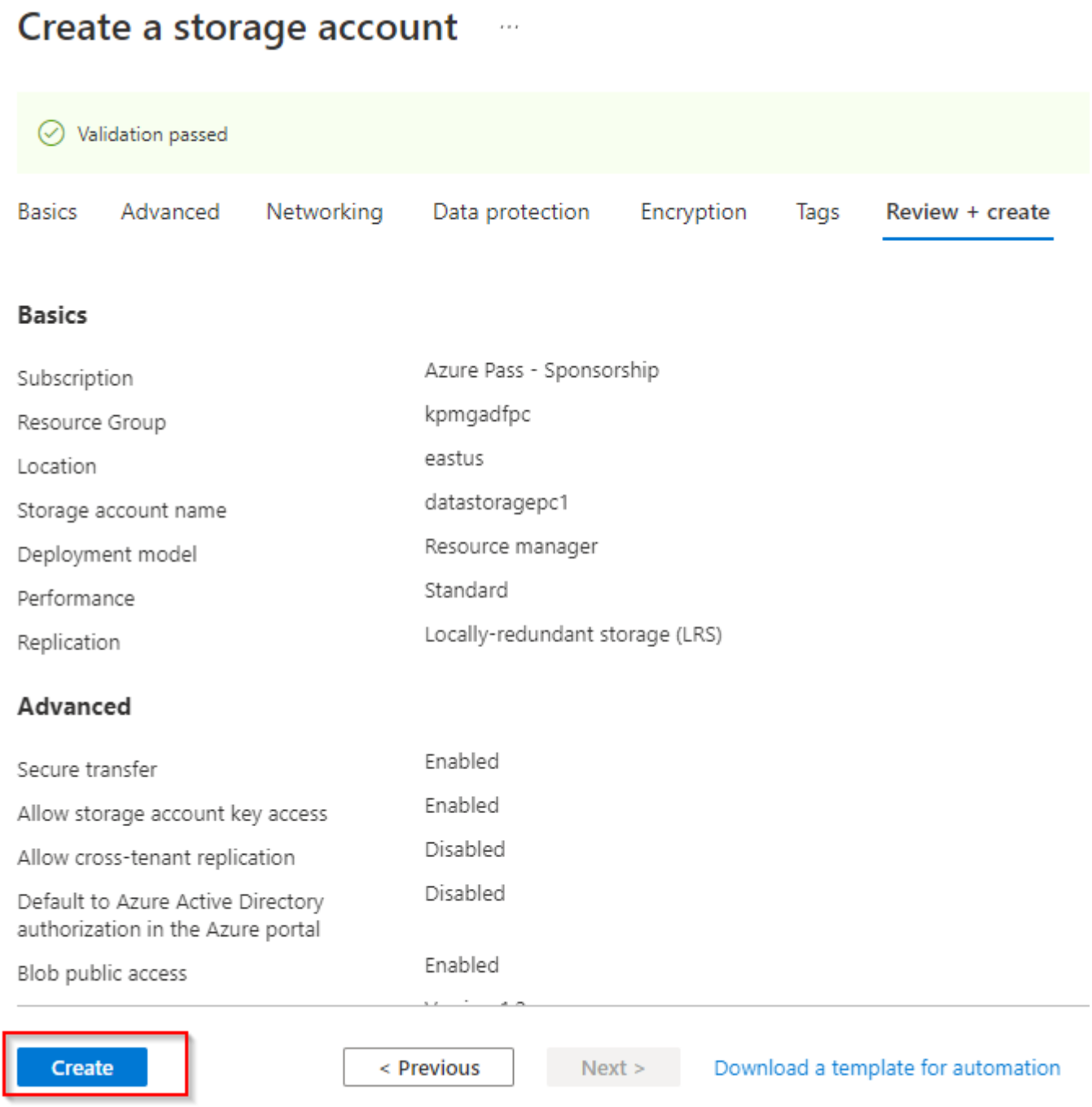


- Select **Advanced** tab in create storage account
  - Under Data Lake Storage Gen2 details
  - Enable **hierarchical namespace**: mark is Select



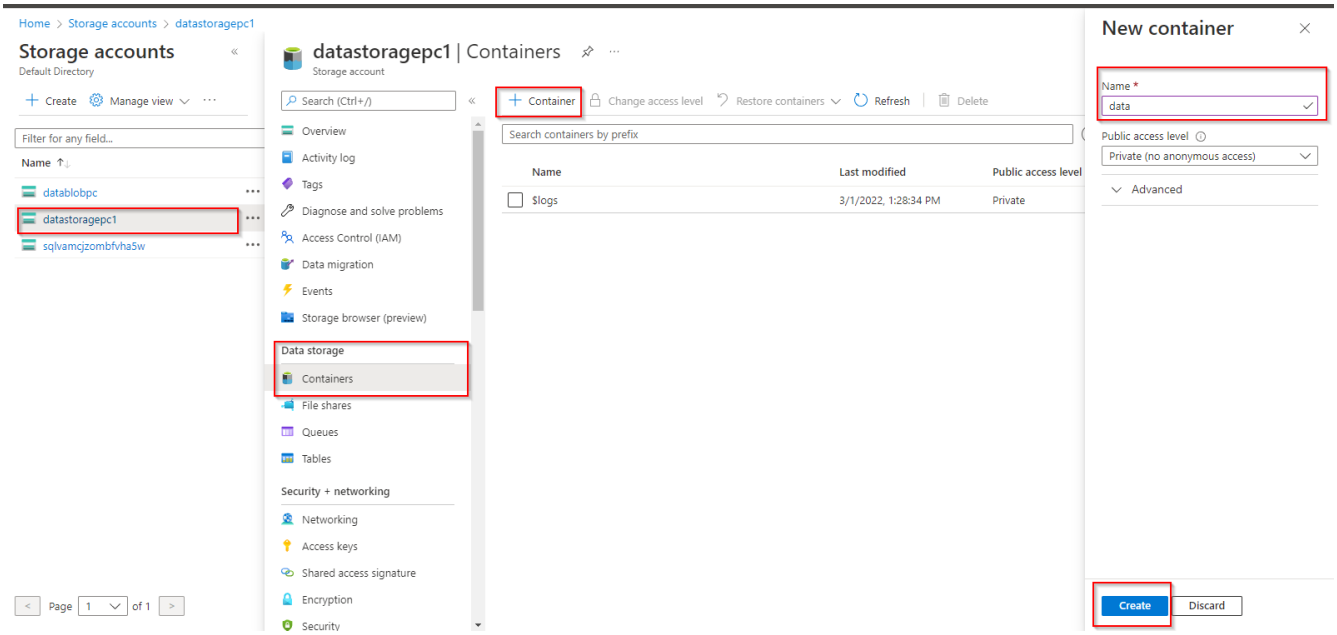
5. In the **Create storage account** screen, click **Review + create**.

6. After the validation of the **Create storage account*** screen, click **Create**.



7. **Note**: The creation of the storage account will take approximately 90 seconds while it provisions the disks and the configuration of the disks as per the settings you have defined.

8. Once the storage account created create click on **Go to resource**, under **Data storage** blade click on **Containers** then click on **+ Container** and give name **data** then click on Create button.



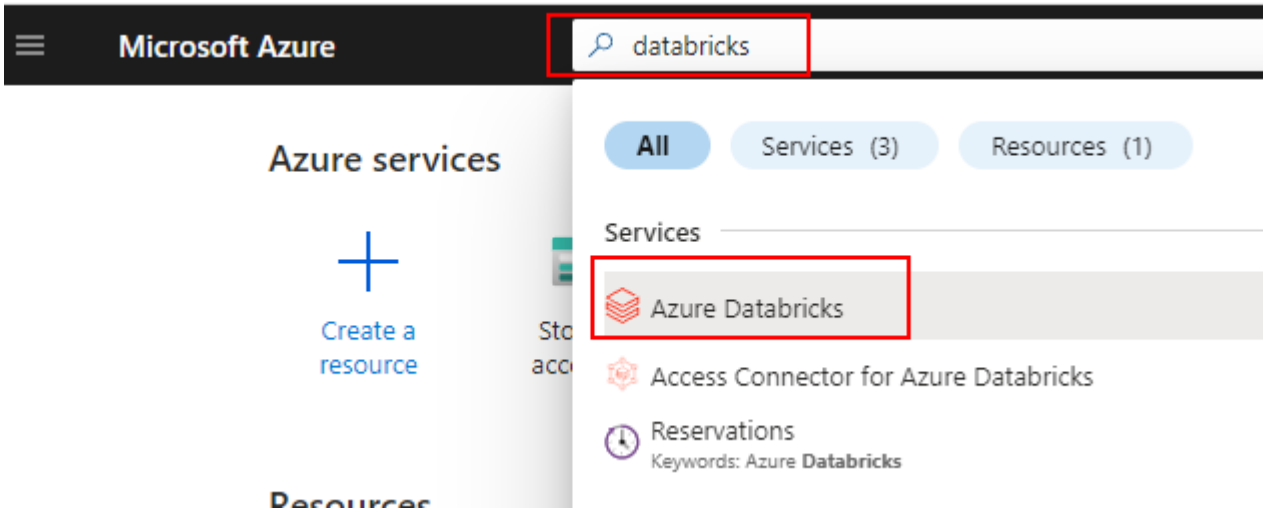
Setup Azure Databricks

Tasks:

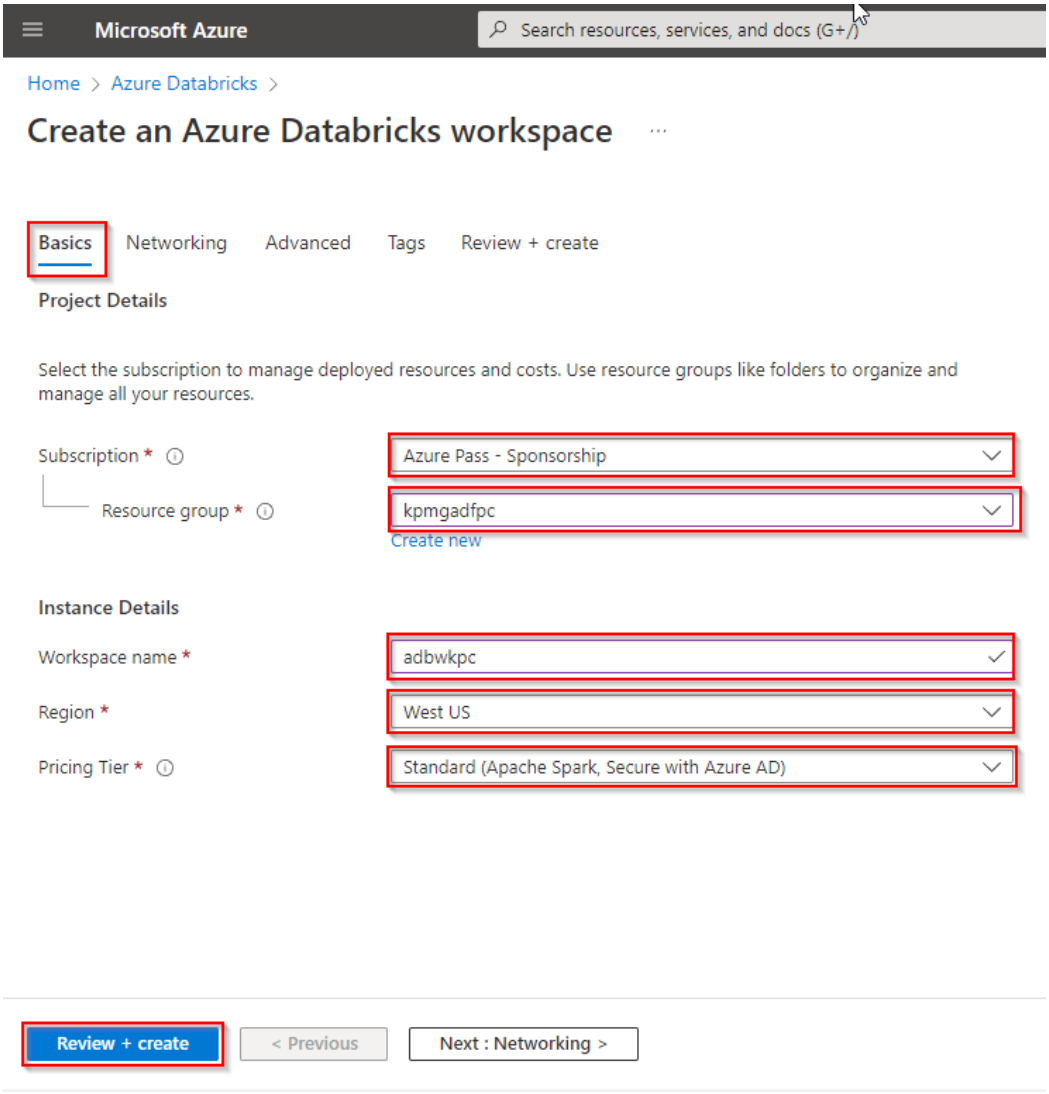
1. Create and Configure Azure Databricks Instance.
2. Create a Spark Cluster in Azure Databricks.

Task 1: Create and configure an Azure Databricks instance.

1. In the Azure portal, at the top left search for databricks, click on Azure Databricks.



2. In the **Azure Databricks** page, click **+ Create**.
3. In the **Create Azure Databricks workspace page**, Under Basic tab provide the following settings:
    - **Subscription**: the name of the subscription you are using in this lab
    - **Resource group**: **adbwk-rg**, where **xx** are your initials.
    - **Workspace name**: **adbwkxx**, where **xx** are your initials.
    - **Region**: the name of the Azure region which is closest to you. (Central India)
    - **Pricing Tier**: **Standard (Apace Spark, Secure with Microsoft Entra ID)**.



4. In the **Create Azure Databricks workspace page**, click **Review +create**. Then click **Create**

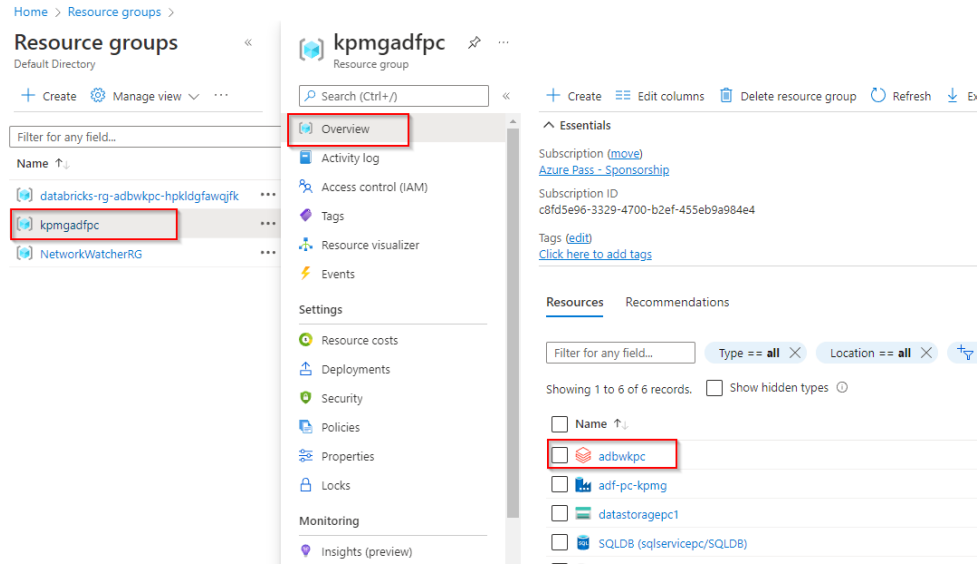
   Note: The provision will take approximately 3 minutes. The Databricks Runtime is built on top of Apache Spark and is natively built for the Azure cloud. Azure Databricks completely abstracts out the infrastructure complexity and the need for specialized expertise to set up and configure your data infrastructure. For data engineers, who care about the performance of production jobs, Azure Databricks provides a Spark engine that is faster and performant through various optimizations at the I/O layer and processing layer (Databricks I/O).

5. Confirm that the Azure Databricks service has been created.

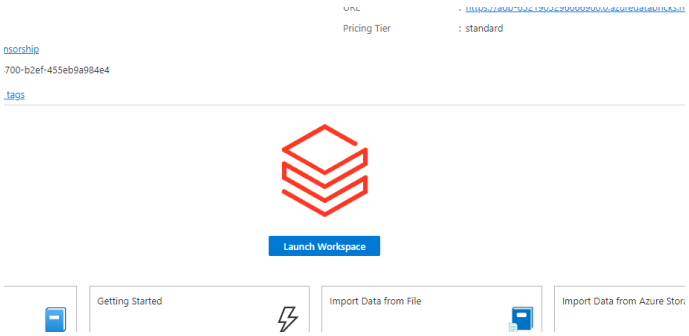
6. In the Azure portal, navigate to the **Resource group** screen.

7. In the Resource groups screen, click on the **adbwk-rg** resource group, where **xx** are your initials.

8. In the **adbwk-rg** screen, click **adbwkxx**, where **xx** are your initials to open Azure Databricks. This will open your Azure Databricks service.

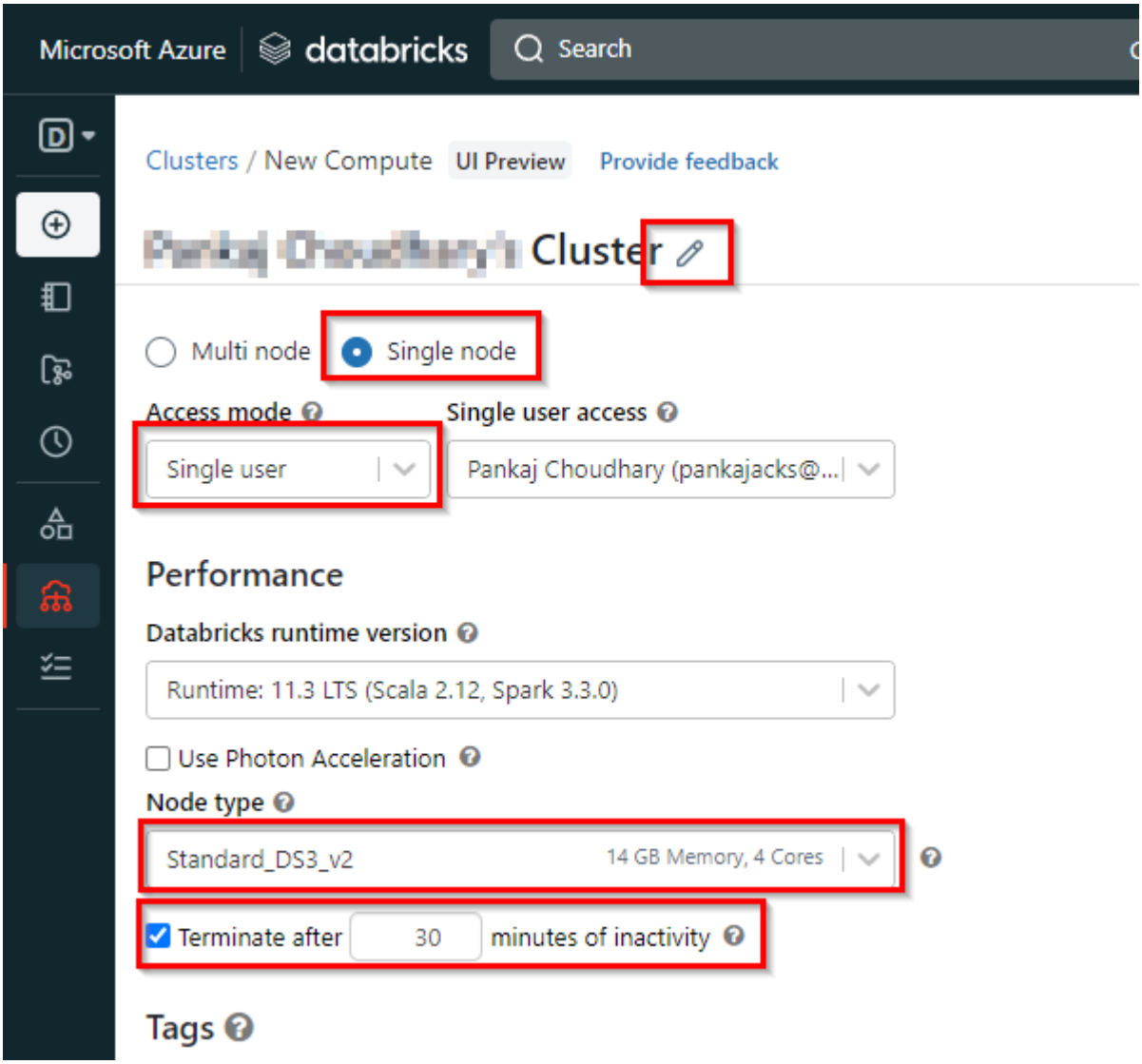


9. Click on **adbwkxx** and click on **Launch Workspace**



Task 2: Create a Spark Cluster in Azure Databricks.

1. Once you Launch the Workspace.
2. Under **Compute Tab**, click **Create Cluster**.
3. In the **Create Cluster** screen, under New Cluster, create a Databricks Cluster with the following settings, and then click on **Create Cluster**:
    - **Cluster name**: **My Cluster**
    - **Cluster Mode**: **Single Node**
    - **Databricks Runtime Version**: **Runtime: 11.3 LTS (Scala 2.12, Spark 3.1.2)**
    - Make sure you select and set the **Terminate after 30** minutes of inactivity check box. If the cluster isn't being used, provide a duration (in minutes) to terminate the cluster.
    - Leave all the remaining options to their current settings.



4. In the **Create Cluster** screen, click on **Create Cluster** and leave the browser screen open.

   Note: The creation of the Azure Databricks Spark cluster instance will take approximately 5-8 minutes.