Azure Databricks

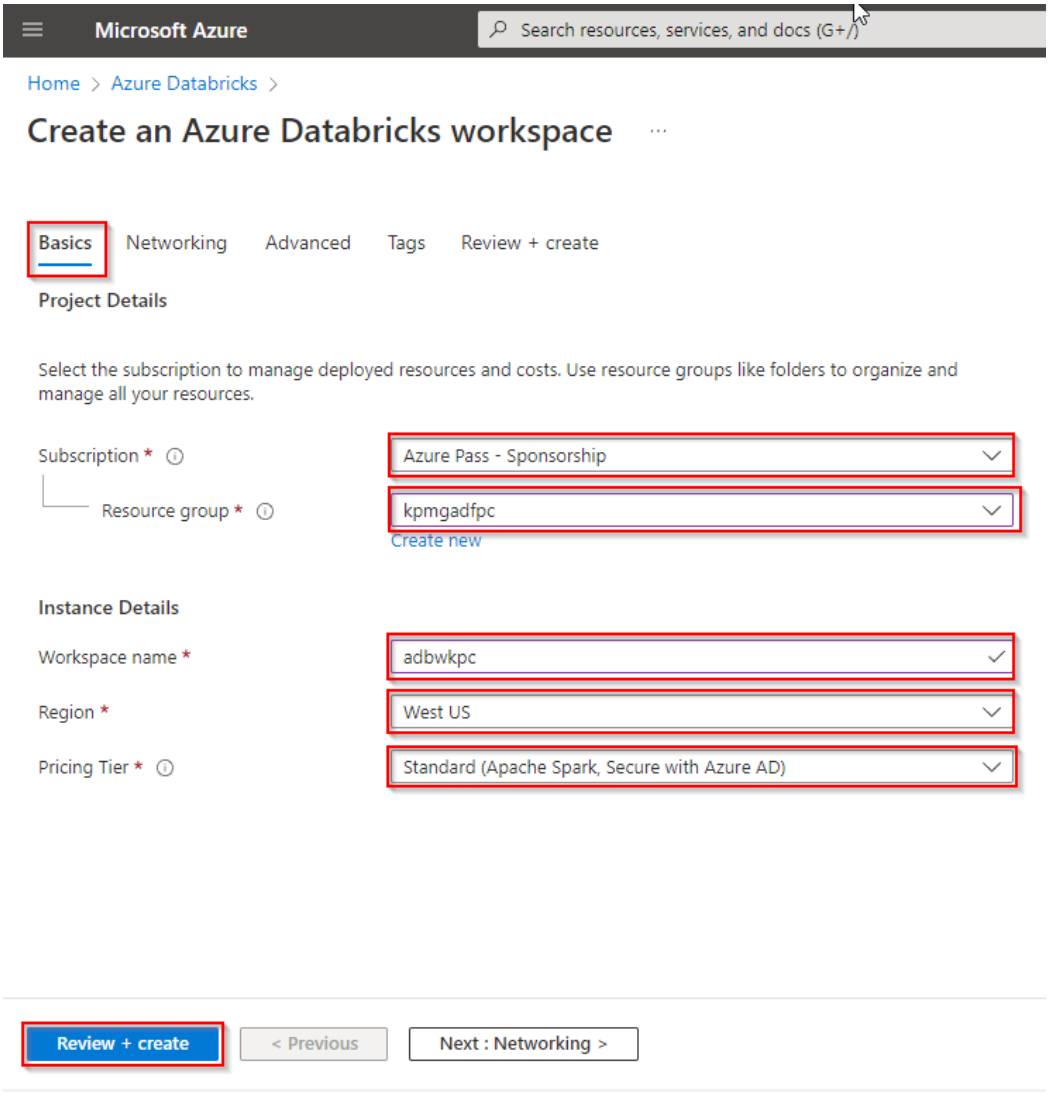
Lab: Setup Azure Databricks

The main tasks for this exercise are as follows:

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

Task 1: Create and configure an Azure Databricks instance.

1. In the Azure portal, at the top left of the screen, click on the **Home** hyperlink.
2. In the Azure portal, click on the **+ Create a resource** icon.
3. In the New screen, click in the **Search services and marketplace** text box, and type the word **Azure databricks**. Click **Azure Databricks** in the list that appears.
4. In the **Azure Databricks** blade, click **Create**.
5. In the **Azure Databricks Service** blade, create an Azure Databricks Workspace with the following settings:
    - **Subscription**: the name of the subscription you are using in this lab
    - **Resource group**: **adbwkxx-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.
    - **Pricing Tier**: **Standard (Apace Spark, Secure with Azure AD)**.



6. In the **Azure Databricks Service** blade, click **Review +create**. Then click **Create**

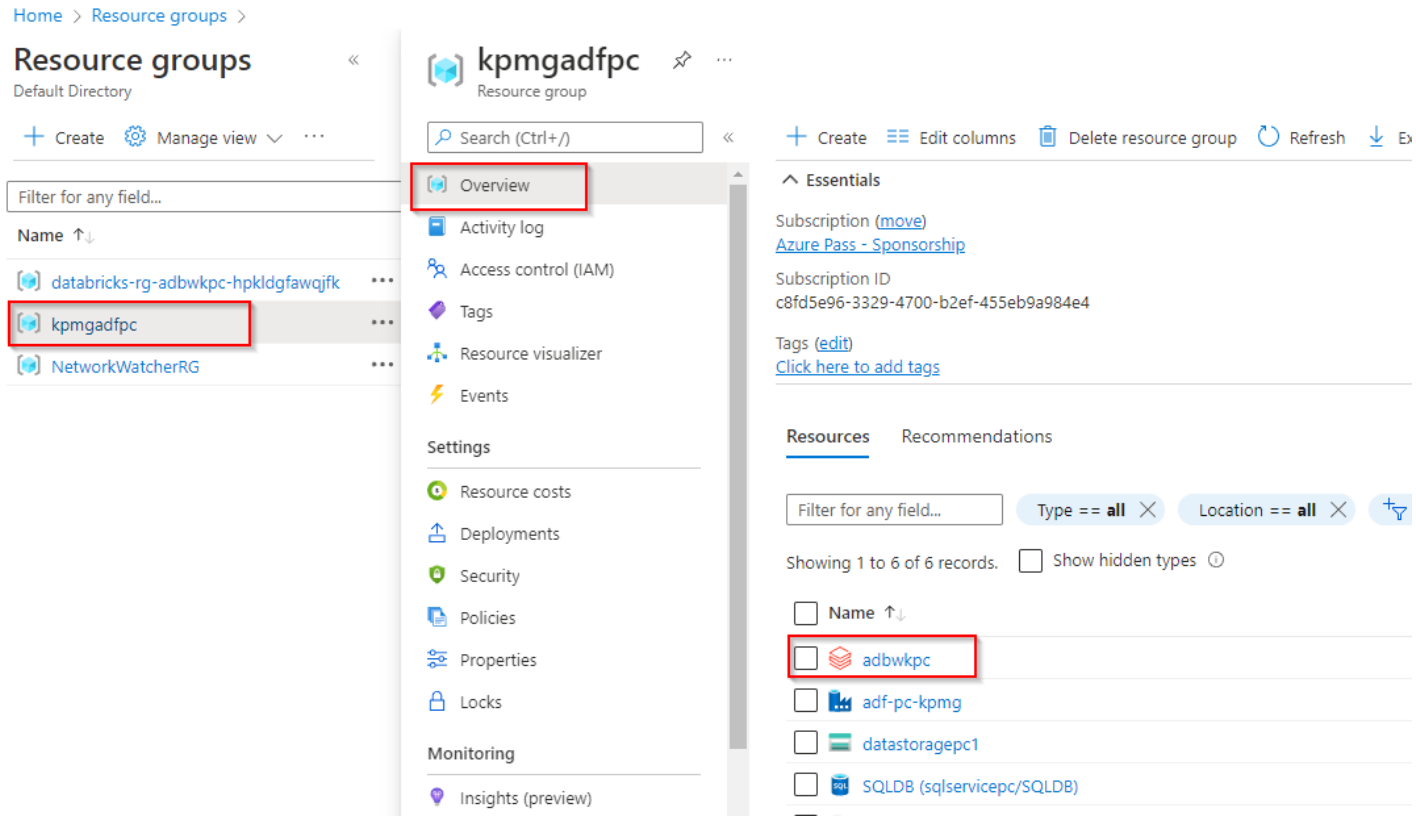
   Note: The provision will take approximately 3 minutes. The Databricks Runtime is built on top of Apache Spark and is natively built for the Azure cloud. Azure Databricks completely abstracts out the infrastructure complexity and the need for specialized expertise to set up and configure your data infrastructure. 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).

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

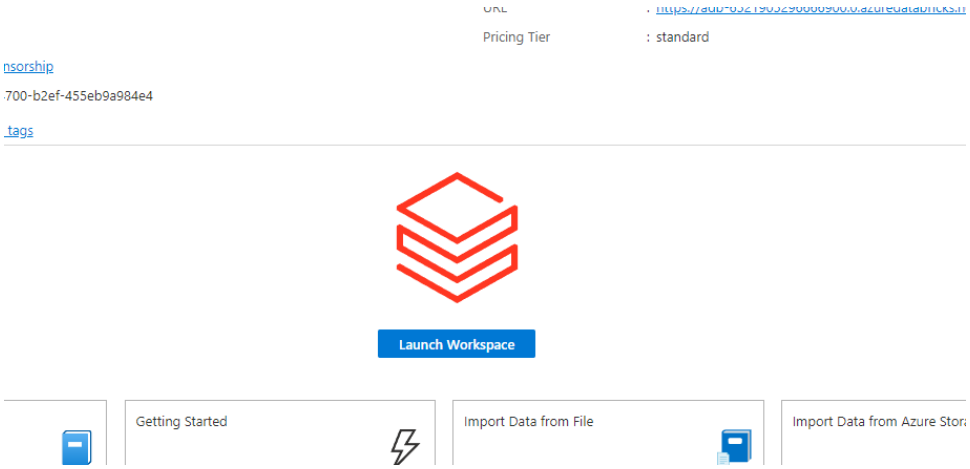
8. In the Azure portal, navigate to the **Resource group** screen.

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

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



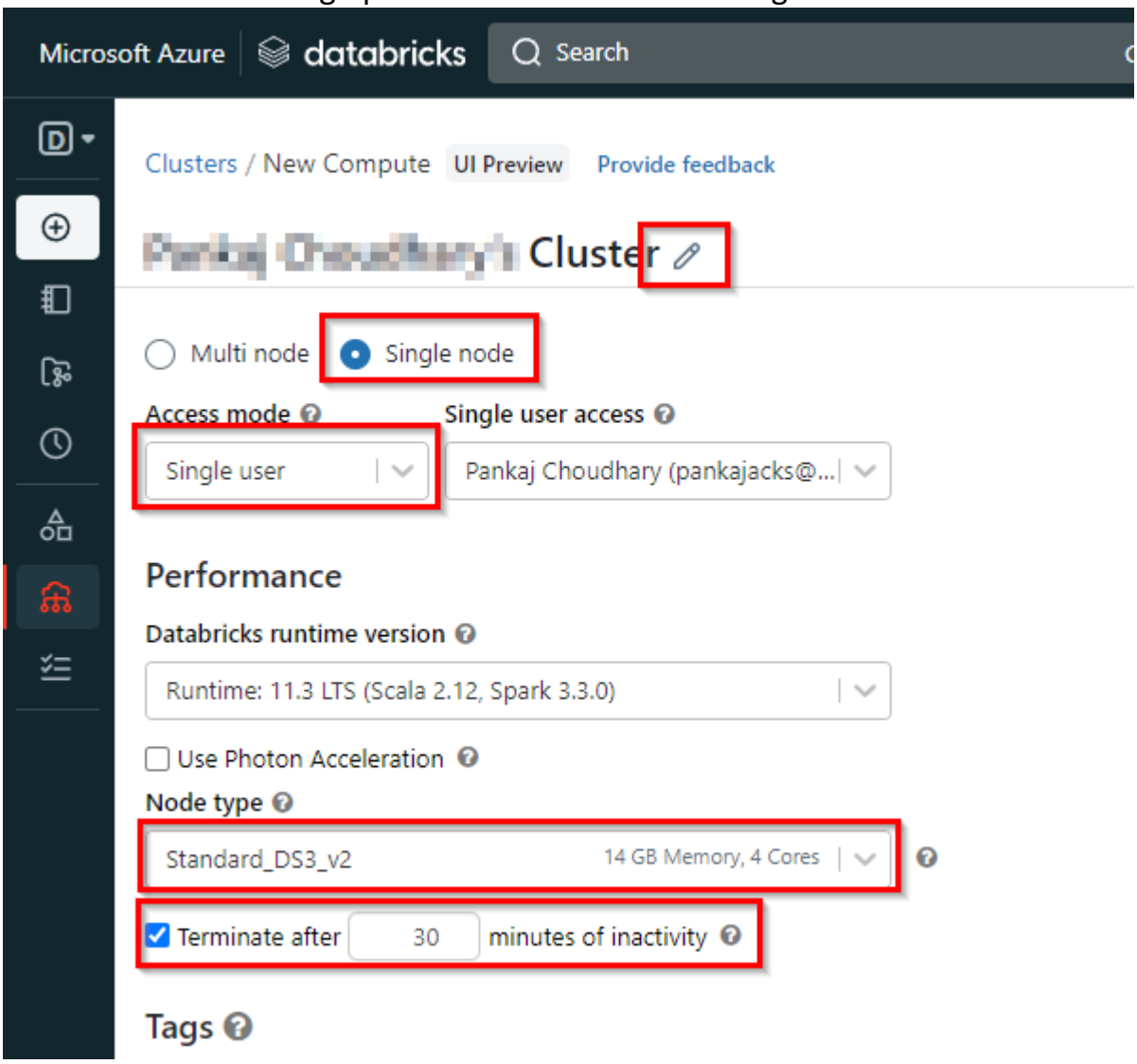
11. Click on **adbwkxx** and click on **Launch Workspace**



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**: **Test Cluster**
    - **Cluster Mode**: **Single Node**
    - **Databricks Runtime Version**: **Runtime: 11.x 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 instance will take approximately 5-8 minutes as the creation of a Spark cluster is simplified through the graphical user interface.