**Ansh Ranjan  
Azure Databricks**

**Exercise 1 – Settings up DataBricks and Spark Basics**

**TASK 1: Create a new Azure DataBricks workspace**

1. Go to Azure Portal > Azure DataBricks > Create > Enter details > Review and Create  
   A screenshot of a computer

   AI-generated content may be incorrect.

**TASK 2: Launch a spark cluster and explore databricks interface**

1. Launch your Databricks workspace > Computer side tab > Create Compute  
   A screenshot of a computer

   AI-generated content may be incorrect.
2. Now click on New > Notebook and you will have your notebook ready in your workspace  
   A screenshot of a computer

   AI-generated content may be incorrect.

The **Azure Databricks workspace** provides a unified environment for data engineering, data science, and machine learning. The main parts of the interface include:

**1. Workspace**

* Organize notebooks, libraries, and workflows.
* Create folders and share them with users or groups.

**2. Notebooks**

* Interactive notebooks supporting **Python**, **SQL**, **Scala**, and **R**.
* Run code in cells and visualize data easily.

**3. Clusters**

* Spin up Spark clusters for running jobs or interactive analysis.
* Choose autoscaling and runtime version (with Delta, ML, or GPU support).

**4. Jobs**

* Schedule notebooks or workflows.
* Automate ETL, ML training, or batch jobs.

**5. Data**

* Browse databases, tables, and files.
* Supports Unity Catalogue (if enabled) for secure, centralized governance.

**6. Repos (Git Integration)**

* Connect to GitHub or Azure DevOps to version-control notebooks and code.

**TASK 3: Run Basic Spark commands**

1. Running spark.version command in first cell  
   A screen shot of a computer

   AI-generated content may be incorrect. A screen shot of a computer

   AI-generated content may be incorrect.

Now that we know our databricks workspace is ready we can start performing ETL tasks