# **DP 200 - Implementing a Data Platform Solution**

## Lab 3- Enabling Team Based Data Science with Azure Databricks

### Exercise 1: Explain Azure Databricks

**Question 1.**

Outline how Databricks can be used to meet an AdventureWorks digital transformation requirements as outlined by the case study.

|  |  |
| --- | --- |
| **Requirement** | Reason |
| AdventureWorks would like to take their data analytics further and start to utilize predictive analytics capabilities | Databricks is meant for ETL and their machine learning library (MLIB) is dedicated to handling the text recommendation that they want |
| You are also assessing the tooling that can help with the extraction, load and transforming of data | They mention data factory for orchestrating the movement, but advanced etl could be done via databricks if needed |

**Question 2.**

From the lessons learned from the course and labs so far, what would be a candidate data source for the Azure Databricks instance you are configuring?

|  |  |
| --- | --- |
| **Candidate data source** | Reason |
| Our data lake storage gen2 | It can be used to store datawarehouse data, product catalog data, etc… The major sources that we would want to pull from that they’re looking to expand on |