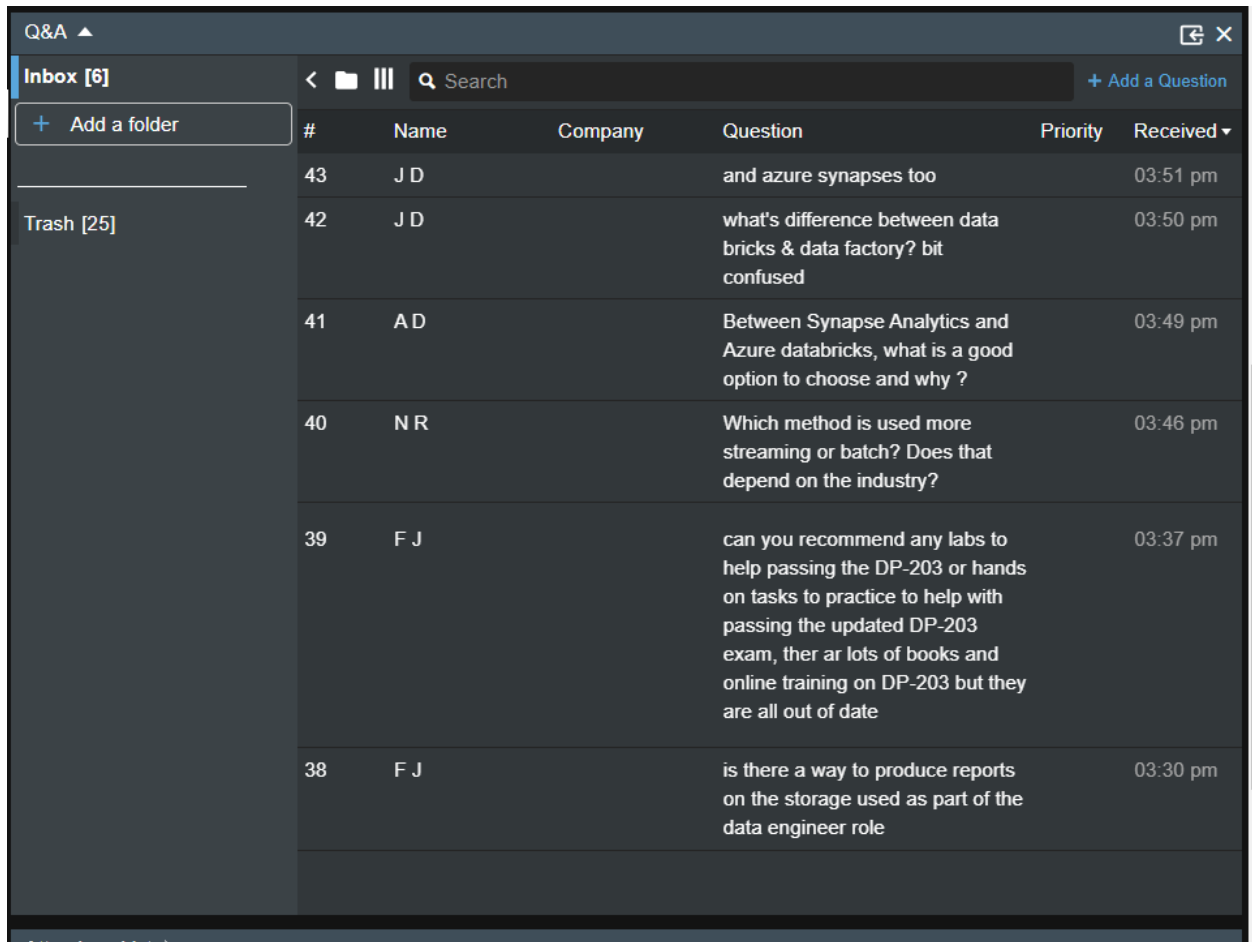


DP-203 OLT, June 16, 2023  
Reza Salehi

Questions:



The screenshot shows a Q&A application interface. On the left is a sidebar with 'Inbox [6]' and 'Trash [25]'. The main area displays a table of questions. The table has columns for '#', 'Name', 'Company', 'Question', 'Priority', and 'Received'. The questions are numbered 38 to 43. Questions 38 and 39 are visible in the screenshot.

#	Name	Company	Question	Priority	Received
43	J D		and azure synapses too		03:51 pm
42	J D		what's difference between data bricks & data factory? bit confused		03:50 pm
41	A D		Between Synapse Analytics and Azure databricks, what is a good option to choose and why ?		03:49 pm
40	N R		Which method is used more streaming or batch? Does that depend on the industry?		03:46 pm
39	F J		can you recommend any labs to help passing the DP-203 or hands on tasks to practice to help with passing the updated DP-203 exam, ther ar lots of books and online training on DP-203 but they are all out of date		03:37 pm
38	F J		is there a way to produce reports on the storage used as part of the data engineer role		03:30 pm

Answers:

- **#38:** Yes. You can either use Azure Storage Insights, or take a look at Azure Storage Metrics:
  - <https://learn.microsoft.com/en-us/azure/storage/common/storage-insights-overview>
  - <https://learn.microsoft.com/en-us/azure/storage/blobs/monitor-blob-storage-reference#account-level>
- **#39:** You can check out the following:

- The “self-paced” learning paths here  
<https://learn.microsoft.com/en-us/certifications/exams/dp-203/>
- Also, Microsoft has very good training and quick starts for each service. I added a few examples below. You can easily find them in the service documentation:
  - <https://learn.microsoft.com/en-us/azure/databricks/getting-started/etl-quick-start>
  - <https://learn.microsoft.com/en-us/azure/databricks/getting-started/data-pipeline-get-started>
  - <https://learn.microsoft.com/en-us/azure/databricks/getting-started/free-training>
  - <https://learn.microsoft.com/en-us/azure/synapse-analytics/get-started/create-workspace>
  - <https://learn.microsoft.com/en-us/azure/synapse-analytics/get-started-analyze-sql-on-demand>
  - <https://learn.microsoft.com/en-us/azure/synapse-analytics/get-started-analyze-data-explorer>
  - <https://learn.microsoft.com/en-us/azure/synapse-analytics/get-started-analyze-spark>
  - <https://learn.microsoft.com/en-us/azure/synapse-analytics/get-started-analyze-sql-pool>
  - <https://learn.microsoft.com/en-us/azure/synapse-analytics/get-started-analyze-storage>
  - <https://learn.microsoft.com/en-us/azure/synapse-analytics/get-started-pipelines>
  - <https://learn.microsoft.com/en-us/azure/synapse-analytics/get-started-visualize-power-bi>
  - <https://learn.microsoft.com/en-us/azure/stream-analytics/stream-analytics-quick-create-portal>
  - <https://learn.microsoft.com/en-us/azure/stream-analytics/create-cluster>
  - <https://learn.microsoft.com/en-us/azure/stream-analytics/quick-start-build-application>
  - <https://learn.microsoft.com/en-us/azure/stream-analytics/stream-analytics-test-query>
- **#40:** Both are used very commonly, and it depends on the project/use-case. We need to analyze both streaming and batch data.
- **#41:**
  - Both are good options for processing data. It depends on the talent which you have in your company.
  - Use Azure Synapse if data engineers are comfortable with SQL.
  - Choose either one if you need to use Spark.
  - Also keep in mind that Azure Synapse can persist data (it is a data warehouse) but Azure Databricks is used to process big data in memory.

- Compare the pricing and choose the one which is more cost effective.
- Also see the following  
<https://learn.microsoft.com/en-us/answers/questions/587071/difference-between-synapse-and-databricks>
- **42 & 43:**
  - It is hard to answer this question in a few sentences. ADF is a data integration service. You can use a visual designer to create data pipelines to extract, transform and load data. Azure Databricks is a managed Spark cluster which can be used to process big data in memory. You need to write code (Python, Scala, etc.) to analyze your data.
    - <https://learn.microsoft.com/en-us/azure/data-factory/introduction>
    - <https://learn.microsoft.com/en-us/azure/databricks/introduction/>
- Thank you!