

Virtusa Azure Data Engineer Interview Guide – Experienced 3+

Technical Round 1: ADF and Data Transformation

1. Explain the types of triggers in ADF, including schedule, tumbling window, and event-based triggers.

Follow-ups:

- How do tumbling window triggers ensure data consistency in batch processing?
- Can you chain multiple triggers for a single pipeline?

2. Compare Control Flow and Data Flow in ADF.

Follow-ups:

- When would you use Control Flow over Data Flow in a real-world scenario?
- How does Data Flow optimize data transformations for large datasets?

3. How can you increase parallelism in ADF pipelines?

Follow-ups:

- What challenges can arise when using high degrees of parallelism?
- How do partitioning strategies differ between source and sink?

4. How would you migrate 1TB of data using ADF?

Follow-ups:

- How do you ensure fault tolerance during large-scale data migrations?
- What strategies do you use to handle network bottlenecks?

5. What are Assert Transformations, and where are they used?

Follow-ups:

- Can you provide a use case where Assert Transformations helped maintain data quality?
- What are the limitations of Assert Transformations in complex data flows?

6. Differentiate between global and local variables in ADF.

Follow-ups:

- How do you pass global variables between pipelines?
- What scenarios require local variables instead of global ones?

7. Explain the use of Web Activity in ADF.

Follow-ups:

- How do you manage authentication for REST API calls using Web Activity?

- Can you trigger a pipeline from an external application using Web Activity?

Technical Round 2: Data Quality and Complex Integrations

1. How do you merge data from different sources in ADF while maintaining data quality?

Follow-ups:

- How do you handle schema mismatches during merging?
- What techniques ensure deduplication in large datasets?

2. How do you integrate with an on-premises SQL Server without using SHIR?

Follow-ups:

- What are the limitations of using Azure Hybrid Connections?
- How do you secure the connection for sensitive data transfers?

3. How do you delete files older than 30 days using ADF?

Follow-ups:

- How can lifecycle management policies complement ADF for this task?
- What would you do if the files are stored in multiple folders with varying retention policies?

4. Explain how you debug failed pipelines in ADF.

Follow-ups:

- What are the key logs or metrics you analyze first?
- How do you use dependency tracing to identify root causes in pipeline failures?

5. How do you use Azure Databricks notebooks within ADF pipelines?

Follow-ups:

- What configurations are needed to pass parameters to a Databricks notebook?
- How do you monitor the status of a notebook run from ADF?

6. What are common issues faced with REST APIs in ADF, and how do you resolve them?

Follow-ups:

- How do you handle API rate limits in ADF?
- What steps do you take to debug authentication errors in REST API calls?

7. How are Logic Apps used in ADF projects?

Follow-ups:

- How do Logic Apps enhance notification workflows for monitoring pipelines?
- What are the performance considerations when integrating Logic Apps with ADF?

Glassdoor Virtusa Review –

<https://www.glassdoor.co.in/Reviews/Virtusa-Reviews-E29396.htm>

Virtusa Careers –

<https://www.virtusa.com/careers>

Subscribe to my YouTube Channel for Free Data Engineering Content –

<https://www.youtube.com/@shubhamwadekar27>

Connect with me here –

<https://bento.me/shubhamwadekar>

Checkout more Interview Preparation Material on –

https://topmate.io/shubham_wadekar