91 Azure Data Factory interview questions to hire top engineers

Questions

1. What is Azure Data Factory? Can you explain it like I'm five?

2. Why would someone use Azure Data Factory instead of just writing code to move data?

3. What are the main components of Azure Data Factory? Can you give an analogy to explain it better?

4. What is a Linked Service in Azure Data Factory, and why do we need it?

5. What is the difference between a Dataset and a Linked Service in Azure Data Factory?

6. What is a Pipeline in Azure Data Factory? What does it do?

7. What is an Activity in Azure Data Factory? What are different types of activities you know of?

8. What is the Integration Runtime in Azure Data Factory? What does it do behind the scenes?

9. Can you describe a simple data integration scenario where you would use Azure Data Factory?

10. What are the different types of Integration Runtimes available in ADF? When should each one be preferred?

11. How do you monitor pipelines in Azure Data Factory?

12. What is the difference between a Copy Activity and a Data Flow activity?

13. How can you schedule a pipeline to run automatically in Azure Data Factory?

14. If a pipeline fails, how would you troubleshoot it in Azure Data Factory?

15. What are parameters in Azure Data Factory, and why are they useful?

16. How can you handle errors or exceptions in Azure Data Factory pipelines?

17. What is the purpose of using variables inside pipelines?

18. Explain how you would move data from an on-premises SQL Server to Azure Blob Storage using Azure Data Factory.

19. What are some common data transformation tasks you can perform using Azure Data Factory?

20. Can you explain how to implement incremental data loading using Azure Data Factory?

21. What security considerations should you keep in mind when working with Azure Data Factory?

22. How would you ensure that your Azure Data Factory pipelines are performing optimally?

23. How does Azure Data Factory help with data governance and compliance?

24. What's Azure Data Factory (ADF) in super simple words?

25. Imagine you have LEGO bricks (data) in different boxes. How does ADF help you build something cool with them?

26. Can you name a few things ADF can connect to, like different kinds of data storage?

27. What's a 'pipeline' in ADF? Think of it like a train track.

28. What's an 'activity' in ADF? Like one step the train takes.

29. If a pipeline fails, what do you think you should check first?

30. What's the difference between a source and a sink in ADF?

31. Why is it important to schedule pipelines? (Think of automating tasks).

32. What is integration runtime in Azure Data Factory?

33. How does ADF help in moving data from on-premises to the cloud?

34. What is the copy activity used for?

35. If you need to transform data while moving it, how does ADF help?

36. Can you explain the term 'data integration' in simple terms?

37. What are linked services in ADF, and why are they important?

38. Explain what triggers are in Azure Data Factory.

39. How can you monitor a pipeline's progress in ADF?

40. What's a dataset in ADF, and how is it different from a linked service?

41. How do you handle errors or failures in ADF pipelines?

42. Can ADF connect to different types of databases, like SQL Server and Oracle?

43. What is the use of parameters in Azure Data Factory?

44. How does ADF help in making sure data is consistent and accurate?

45. Explain how version control works with Azure Data Factory.

46. What role does JSON play in Azure Data Factory?

47. What is the use of data flows in Azure Data Factory?

48. How can you handle slowly changing dimensions (SCDs) in Azure Data Factory?

49. Describe a scenario where you would use a tumbling window trigger, and how you would configure it.

50. What are the key differences between a self-hosted integration runtime and an Azure integration runtime, and when would you choose each?

51. Explain how you would implement error handling and logging in an Azure Data Factory pipeline.

52. How do you parameterize datasets and pipelines in Azure Data Factory, and why is this useful?

53. Describe a use case for using the 'Until' activity in Azure Data Factory.

54. How would you secure credentials used in Azure Data Factory pipelines?

55. What are the different types of activities available in Azure Data Factory, and give an example of when you might use each.

56. How can you monitor and troubleshoot Azure Data Factory pipeline executions?

57. Explain how you can implement CI/CD (Continuous Integration/Continuous Deployment) for Azure Data Factory pipelines.

58. How would you handle dependencies between pipelines in Azure Data Factory?

59. Describe how you would use Azure Key Vault with Azure Data Factory.

60. What are some strategies for optimizing the performance of Azure Data Factory pipelines?

61. How can you use Azure Data Factory to load data into a data lake?

62. Explain how you can implement data quality checks within an Azure Data Factory pipeline.

63. How would you design a pipeline to incrementally load data from a source system into a data warehouse?

64. Describe a scenario where you would use a data flow in Azure Data Factory, and explain the benefits of using data flows in that scenario.

65. How can you use Azure Data Factory to call an external API or web service?

66. What are some common challenges you might encounter when working with Azure Data Factory, and how would you address them?

67. Explain how you can version control your Azure Data Factory pipelines.

68. How do you handle slowly changing dimensions (SCDs) in Azure Data Factory pipelines, and what are the trade-offs between different SCD types?

69. Describe a complex data integration scenario you've implemented using Azure Data Factory, highlighting the challenges and your solutions.

70. How would you design a robust error handling and monitoring strategy for a mission-critical Azure Data Factory pipeline?

71. Explain how you can optimize the performance of Azure Data Factory pipelines when dealing with large datasets, including partitioning and data compression techniques.

72. What are some advanced techniques for parameterizing Azure Data Factory pipelines to make them more reusable and adaptable to different environments?

73. How do you implement CI/CD for Azure Data Factory pipelines, and what tools or practices do you recommend?

74. Describe your experience with using custom activities in Azure Data Factory, and provide an example of when you would use them.

75. How do you ensure data quality and consistency when integrating data from multiple sources using Azure Data Factory?

76. What are some strategies for managing dependencies between Azure Data Factory pipelines, and how do you handle pipeline orchestration?

77. Explain how you can leverage Azure Data Factory to implement data lake ingestion and transformation patterns.

78. How would you design an Azure Data Factory pipeline to incrementally load data from a source system into a data warehouse?

79. Describe your experience with using Azure Data Factory to integrate with different types of data sources, such as REST APIs, NoSQL databases, and on-premises systems.

80. How do you approach troubleshooting performance bottlenecks in Azure Data Factory pipelines?

81. What are your preferred methods for securing Azure Data Factory pipelines and data, and how do you implement data masking or encryption?

82. Explain how you can use Azure Data Factory to implement data lineage and auditing capabilities.

83. How do you handle data governance requirements when building data integration solutions with Azure Data Factory?

84. Describe a time when you had to debug a complex issue in an Azure Data Factory pipeline, and what steps you took to resolve it.

85. What are some best practices for designing scalable and maintainable Azure Data Factory pipelines?

86. How do you approach testing Azure Data Factory pipelines, and what types of tests do you perform?

87. Explain your understanding of the integration runtime in Azure Data Factory and different types of integration runtimes available.

88. How do you monitor the cost of running Azure Data Factory pipelines and optimize them for cost efficiency?

89. What are some advanced features of data flows in Azure Data Factory, and how can you use them to solve complex data transformation problems?

90. How do you version control and manage changes to Azure Data Factory pipelines in a collaborative environment?