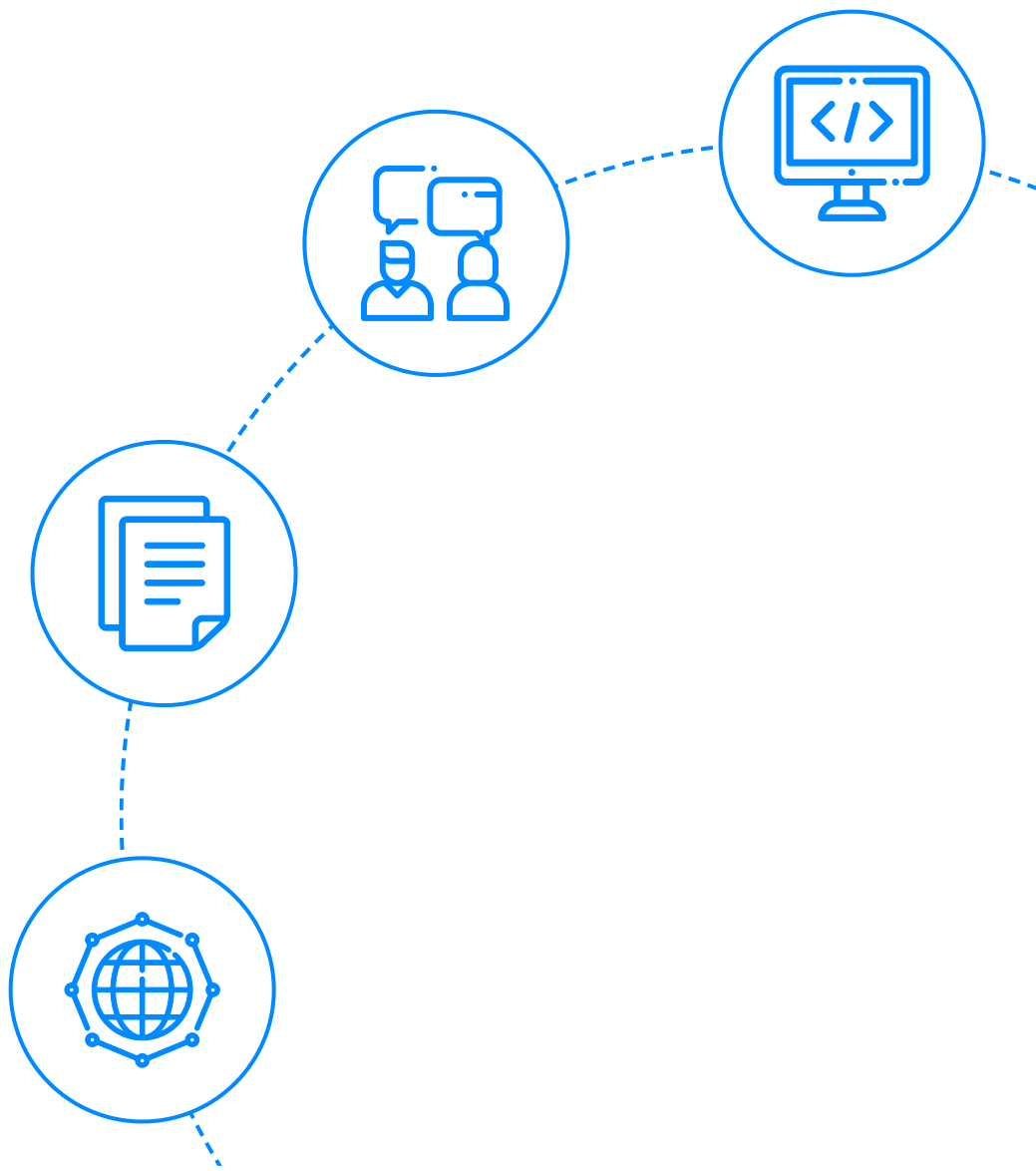




InterviewBit

Azure Data Factory Interview Questions



To view the live version of the page, [click here](#).

© Copyright by Interviewbit

Contents

Azure Data Factory Interview Questions for Freshers

1. What is Azure Data Factory?
2. In the pipeline, can I set default values for the parameters?
3. What is the anticipated length of time needed for the integration?
4. How many times may an integration be run through its iterations?
5. Where can I obtain additional information on the blob storage offered by Azure?
6. Is there a cap on the number of cycles that can be invested in the integration process?
7. How does the Data Factory's integration runtime actually function, though?
8. Provide information regarding the steps required to create an ETL procedure in Azure Data Factory.
9. What exactly are the three different categories of triggers that are available for use with Azure Data Factory?
10. Where can I locate the step-by-step instructions for creating Azure Functions?
11. How can I utilize one of the other 80 dataset types that Data Factory provides to get what I need?
12. What prerequisites does Data Factory's SSIS execution require?
13. What are "Datasets" in the ADF framework?
14. What is the purpose of ADF Service?
15. State the difference between the transformation procedures known as Mapping data flow and Wrangling data flow when it comes to Data Factory?
16. What are Azure Databricks, and Where Can I Find Them?
17. What is meant by "Azure SQL Data Warehouse"?
18. What is Azure Data Lake ?
19. Determining the data sources utilized by the Azure Data Factory

Azure Data Factory Interview Questions for Freshers

(.....Continued)

21. What are some of the advantages of carrying out a lookup in the Azure Data Factory?
22. What sorts of variables are supported by Azure Data Factory and how many different kinds are there?
23. What is the connected service offered by the Azure Data Factory, and how does it operate?
24. What is meant to be referred to when people use the phrase "breakpoint" in conjunction with the ADF pipeline?
25. Please provide a more in-depth explanation of what Data Factory Integration Runtime entails.

Azure Data Factory Interview Questions for Experienced

26. What are the prerequisites that need to be met before an SSIS package can be executed in Data Factory?
27. Is there a limit to the number of Integration Runtimes that may be built or is it unlimited?
28. What is meant by the term "ARM Templates" when referring to Azure Data Factory? Where do we plan to use them?
29. How about discussing the three most important tasks that you can complete with Microsoft Azure Data Factory?
30. How well does Data Factory support the Hadoop and Spark computing environments when it comes to carrying out transform operations?
31. How many individual steps are there in an ETL procedure?
32. Which activity should be performed if the goal is to make use of the results that were acquired by performing a query?
33. Is it possible to communicate with a pipeline run by passing information along in the form of parameters?
34. Do you have any prior experience with the Execute Notebook activity in Data

Azure Data Factory Interview Questions for Experienced

(.....Continued)

39. In the context of the Azure Data Factory, just what does it mean when it's referred to as "copy activity"?
40. Could you explain to me how I should go about planning a pipeline?
41. In which situations does Azure Data Factory seem the best option?
42. Do you have any tips on how to access the information you require by taking advantage of the other ninety dataset types that are accessible in the Data Factory?
43. Can the value of a new column in an ADF table be determined by using an existing mapping column?
44. Where can I find more information on the benefits of using lookup operations in the Azure Data Factory?
45. Please provide any more information that you have on the Azure Data Factory Get Metadata operation.
46. Where did you experience the most difficulty while attempting to migrate data from on-premises to the Azure cloud via Data Factory?
47. Do I have the ability to copy information simultaneously from many Excel sheets?
48. Nesting of loops within loops in Azure Data Factory: yes or no?
49. Are there any particular limitations placed on ADF members?
50. What is Data Flow Debug?

Let's get Started

A cloud-based service that acquires and cleans raw corporate data is provided by Microsoft under the brand name Azure Data Factory. There is a significant need for Azure Data Factory Engineers among a wide variety of businesses. This article contains answers to some of the most frequently asked **Azure Data Factory Interview Questions and Answers**.

The following are examples of the kinds of questions that you should be prepared to answer:



Azure Data Factory Interview Questions for Freshers

1. What is Azure Data Factory?

1. In today's world, there is an abundance of data coming from a wide range of different sources; collectively, this information forms a gigantic mountain of data. Before we can upload this information to the cloud, there are a few things that need to be taken care of first.
2. As a result of the fact that data can come from a broad number of locations, each of which may employ relatively different protocols for transporting or channelling the information, the data itself can take on a large variety of shapes and sizes. Once this information has been uploaded to the cloud or some other specific storage, it is absolutely necessary for us to manage it in the appropriate manner. That is, you will need to make some adjustments to the statistics and get rid of any unnecessary details. Concerning the transfer of data, it is important for us to collect data from a variety of sources, combine it in a single area for storage, and, if required, change it into a more helpful form.
3. A traditional data warehouse is also capable of achieving this goal, albeit with a few significant limitations. When it comes to integrating all of these sources, there are times when we have no choice but to go ahead and construct bespoke programs that handle each of these procedures on an individual basis. This is not only a time-consuming process but also a significant source of frustration. We need to either find means to automate this process or come up with more effective workflows.
4. This entire process may be carried out in a manner that is more streamlined, organized, and controllable with the assistance of Data Factory.

2. In the pipeline, can I set default values for the parameters?

Parameters in pipelines can have default values defined.

3. What is the anticipated length of time needed for the integration?

The integration runtime of Azure Data Factory is the underlying computational architecture that enables the following data integration functionalities across a range of network topologies. These features can be accessed through the Azure portal.

Integration runtimes can be broken down into one of three categories:

1. The Azure Integration Run Time makes it very easy to copy data from one cloud data store to another cloud data storage. The transformations can be handled by any one of a number of various computing services, such as Azure HDInsight or SQL Server.
2. You have the option of employing a piece of software known as Self Hosted Integration Run Time rather than making use of Microsoft's Azure Integration Run Time. However, you must first install it on a host computer, either at your business or on a virtual server located in the cloud. Data can be replicated between an on-premises repository and a cloud-based repository using a self-hosted information repository. It is also able to delegate transformation jobs to several machines that are connected to an intranet. Because all of the on-primitive data sources are protected by a firewall, the Data factory won't be able to access any of them; therefore, we have no choice except to use Self-Hosted IR. If we configure the Azure firewall in a certain way, we can circumvent the need for a self-hosted IR in certain circumstances. This will enable a direct connection to be made between Azure and the data sources that are located on-premises.
3. You will have the ability to natively execute SSIS packages in a managed environment if you use the Azure SSIS Integration Run Time. After that, the Azure SSIS Integration Run Time is utilized to transport the SSIS packages to the data factory.

4. How many times may an integration be run through its iterations?

There are no limits placed in any way on the amount of integration runtime instances that can exist within a data factory. However, there is a limit on the number of virtual machine cores that can be utilized by the integration runtime for the execution of SSIS packages for each subscription.

5. Where can I obtain additional information on the blob storage offered by Azure?

With the use of a service known as Blob Storage, vast amounts of data belonging to Azure Objects, such as text or binary data, can be saved. Using Blob Storage, you have the option of retaining the confidentiality of the data associated with your application or making it accessible to the general public. The following are some examples of typical applications of Blob Storage:

1. Providing files to a user's browser in an unmediated fashion.
2. preservation of data with the goal of enhancing accessibility from a remote location.
3. Streaming live audio and video content
4. Examples of data archiving and backup that can be used in the event of a catastrophe.
5. Putting away information so that it can be used at a later time by a service that is hosted either locally or on Microsoft Azure.

6. Is there a cap on the number of cycles that can be invested in the integration process?

In no way is this the case; an Azure data factory can support an unlimited number of integration runtime occurrences simultaneously. However, there is a maximum number of VM cores that can be used by the integration runtime while executing SSIS packages, and this limitation varies depending on the type of subscription. It is essential that you have a solid grasp of these ideas before you start your journey toward earning a certification in Microsoft Azure.

7. How does the Data Factory's integration runtime actually function, though?

Integration Runtime, a safe computing platform, makes it feasible for Data Factory to offer data integration capabilities that are portable across a variety of network configurations. This is made possible by the use of Integration Runtime. Because of its proximity to the data centre, the work will almost certainly be performed there. If you want to Learn Azure Step by Step, you need to be familiar with terminology like this and other key aspects of Azure.

8. Provide information regarding the steps required to create an ETL procedure in Azure Data Factory.

Any data that requires processing in order to be accessed from an Azure SQL server database is first subjected to the processing step, and then the data is saved to the Data Lake Store. The following are the steps that must be taken to construct ETL:

1. To get started, you will need to create a Linked Service for the SQL Server Database that will be used as the source database.
2. Let's imagine for a moment that we are analyzing data from a car database.
3. Construct a Linked Service with the Azure Data Lake Store as its destination at this point.
4. The next thing you need to do is create a dataset in the Data Storage app.
5. Make the necessary preparations for the system and include a phase for copying.
6. Following that, a trigger ought to be included in the pipeline's timetable.

9. What exactly are the three different categories of triggers that are available for use with Azure Data Factory?

1. Utilizing the Schedule trigger helps ensure that the ADF pipeline is executed in accordance with a predetermined timetable.
2. With the assistance of the Tumbling window trigger, the ADF pipeline can be triggered to execute at predetermined time intervals. The current status of the pipeline has been maintained.
3. The Event-based trigger is activated whenever there is a triggering event that is associated with the blob in some way. The addition of a blob to your Azure storage account or its deletion are two instances of actions that fall within this category.

10. Where can I locate the step-by-step instructions for creating Azure Functions?

With Azure Functions, building cloud-based applications require only a few lines of code rather than the traditional tens or hundreds of lines. Because of this functionality, we are able to choose the programming language that best suits our needs. Because the only cost is the time the code is actually executed, pricing is determined on a per-user basis.

It is compatible with a wide variety of programming languages, including F#, C#, Node.js, Java, Python, and PHP, among others. Additionally, the system enables continuous integration and deployment of updates. The development of serverless applications is made possible through the use of Azure Functions apps. By enrolling in Azure Training in Hyderabad, you will have the opportunity to learn everything there is to know about the creation of Azure Functions.

11. How can I utilize one of the other 80 dataset types that Data Factory provides to get what I need?

Existing options for sinks and sources for Mapping Data Flow include the Azure SQL Data Warehouse and the Azure SQL Database, as well as specified text files from Azure Blob storage or Azure Data Lake Storage Gen2 and Parquet files from either Blob storage or Data Lake Storage Gen2.

You will need to make use of the Copy activity to retrieve information from one of the auxiliary connectors so that you may make reference to it. After the data has been staged, you will need to carry out an activity known as a Data Flow if you wish to effectively convert the data.

12. What prerequisites does Data Factory's SSIS execution require?

Either an Azure SQL Managed Instance or an Azure SQL Database must be used as the hosting location for your SSIS IR and SSISDB catalogue.

13. What are "Datasets" in the ADF framework?

The pipeline activities will make use of the inputs and outputs that are contained in the dataset, which contains those activities. A connected data store can be any kind of file, folder, document, or anything else imaginable; datasets frequently represent the organization of information within such a store. An Azure blob dataset, for example, details the blob storage folder and container from which a particular pipeline activity must read data to continue processing. This information is used to determine where the data will be read from.

14. What is the purpose of ADF Service?

ADF's primary purpose is to handle data replication across local and remote, relational and non-relational data sources. In addition, ADF may replicate data between these different types of sources. Additionally, the ADF Service can be used to modify the incoming data to cater to the requirements of a particular organization. Ingestion of data can be accomplished using ADF Service either as an ETL or an ELT tool. This makes it a vital component of the vast majority of Big Data solutions. Sign up for Azure Training in Hyderabad to gain an in-depth understanding of the several advantages offered by ADF Service.

15. State the difference between the transformation procedures known as Mapping data flow and Wrangling data flow when it comes to Data Factory?

The process of graphically designing and transforming data is referred to as mapping data flow. This application allows you to design data transformation logic in a graphical interface without the need to engage a professional programmer, which is a significant benefit. In addition to this, it is executed as an activity within the ADF pipeline on a scaled-out Spark cluster that is fully managed by the ADF.

On the other hand, a non-programmatic method of data preparation is referred to as "wrangling data flow activity," which is the phrase that is used. Users of spark get access to all of the data manipulation capabilities of Power Query M since spark is compatible with Power Query Online. This gives users more control over the data.

16. What are Azure Databricks, and Where Can I Find Them?

Azure Databricks is an analytics platform that is built on Apache Spark and has been fine-tuned for Azure. It is fast, simple, and can be used in collaboration with others. Apache Spark was conceived and developed in collaboration with its creators. Azure Databricks is a service that combines the most beneficial aspects of Databricks and Azure to enable rapid deployment. This service is designed to assist customers in accelerating innovation. The enjoyable activities and engaging environment both contribute to making collaboration between data engineers, data scientists, and business analysts easier to do.

17. What is meant by "Azure SQL Data Warehouse"?

It is a vast storehouse of knowledge that may be mined for useful insights and utilized to guide management decisions. Using this strategy, the data from numerous different databases that are either located in different physical places or are spread across a network can be aggregated into a single repository.

It is possible to construct an Azure SQL Data Warehouse by merging data from multiple sources. This can be done for a variety of reasons, including the fact that it will make it easier to conduct analyses, generate reports, and make decisions. Because it is a business tool that operates in the cloud and allows parallel processing, it enables you to quickly analyze even the most complex queries on the most extensive data sets. In addition to that, it can be used as a workable alternative for Big Data theories.

18. What is Azure Data Lake ?

The enhanced productivity and reduced complexity of data storage that Azure Data Lake offers are all benefits that can accrue to data analysts, software engineers, and data scientists. It is a cutting-edge method that enables you to carry out tasks like these in a wide variety of programming languages and environments.

The problems that are normally involved with archiving information are eliminated as a result. Additionally, it makes it simple to perform batch, interactive, and streaming analytics. The Azure Data Lake from Microsoft provides capabilities that assist businesses in satisfying their growing requirements and overcoming challenges relating to productivity and scalability.

19. Determining the data sources utilized by the Azure Data Factory

The original or final storage location for information that will be processed or utilized in some manner is referred to as the data source. The format of the data could be anything from binary to text to a file containing comma-separated values to a file containing JSON, and so on and so forth.

It's possible that this is an actual database, but it might also be an image, video, or audio file. One example of a data source is a database. Some examples of databases include MySQL, Azure SQL Database, PostgreSQL, and others. Azure Data Lake Storage and Azure Blob Storage are also examples of data sources.

20. The Auto Resolve Integration Runtime provides users with several benefits; nonetheless, the question remains: why should you use it?

AutoResolveIntegration: The runtime environment will make every effort to carry out the tasks in the same physical place as the source of the sink data, or one that is as close as it can get. Productivity may also increase using the same.

21. What are some of the advantages of carrying out a lookup in the Azure Data Factory?

Within the ADF pipeline, the Lookup activity is utilized rather frequently for configuration lookup. It includes the data set in its initial form. In addition to this, the output of the activity can be used to retrieve the data from the dataset that served as the source. In most cases, the outcomes of a lookup operation are sent back down the pipeline to be used as input for later phases.

To provide a more detailed explanation, the lookup activity in the ADF pipeline is responsible for retrieving data. You may only utilize it in a manner that is appropriate for the process you are going through. You have the option of retrieving just the first row, or you may select to obtain all of the rows in the dataset depending on the query.

22. What sorts of variables are supported by Azure Data Factory and how many different kinds are there?

Variables are included in the ADF pipeline so that values can be temporarily stored in them. Their application is almost entirely equivalent to that of variables in programming languages. There are two types of operations that are used to assign and change the values of variables. These are set variables and add variables.

The Azure data factory makes use of two different categories of variables:

1. In Azure, the pipeline's constants are referred to as system variables. Pipeline ID, Pipeline Name, Trigger Name, etc. are all instances.
2. The user is responsible for declaring user variables, which are then utilized by the logic of the pipeline.

23. What is the connected service offered by the Azure Data Factory, and how does it operate?

In Azure Data Factory, the connection method that is utilized to join an external source is referred to as a "connected service," and the phrase is used interchangeably. It not only serves as the connection string, but it also saves the user validation data.

The connected service can be implemented in two different ways, which are as follows:

1. ARM approach.
2. Azure Portal.

24. What is meant to be referred to when people use the phrase "breakpoint" in conjunction with the ADF pipeline?

The commencement of the testing step of the pipeline is indicated by the placement of a debugging breakpoint. Before committing to a particular action, you can make use of breakpoints to check and make sure that the pipeline is operating as it should.

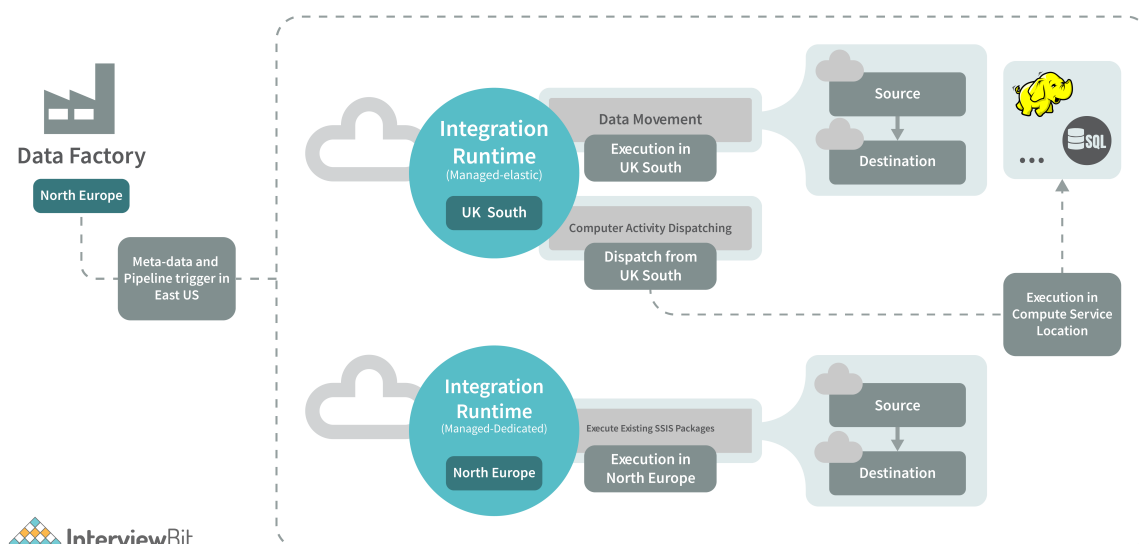
Take the following example into consideration to get a better understanding of the concept: you have three activities in your pipeline, but you only want to debug through the second one. In order to be successful in this endeavour, a breakpoint needs to be established for the second task. By simply clicking the circle located at the very top of the activity, you will be able to add a breakpoint.

25. Please provide a more in-depth explanation of what Data Factory Integration Runtime entails.

The Integration Runtime (IR) is the underlying computing environment that is used when working with Azure Data Factory pipelines. In essence, it links the activities that individuals participate in with the services that they require to participate in those activities.

It offers the computing environment in which the activity is either directly run or dispatched, and as a result, it is referenced by the service or activity that is associated with it. This indicates that the task can be finished regardless of where the closest data storage or computing service is located in the world.

The following configurations are available for Data Factory and its integration runtimes, as shown in the accompanying diagram:



There are three separate integration runtimes available with Azure Data Factory. These runtimes each have their own set of benefits and downsides, which are determined by the user's level of experience with data integration and the desired network setup.

1. You can use Azure Integration Runtime to transport information across different cloud storage services and to trigger activities in other platforms such as SQL Server, Azure HDInsight, and other similar services.
2. The Self-Hosted Integration Runtime is where the action takes place whenever there is a need for data to be replicated between the cloud and private networks. Both the Azure Integration Runtime and the self-hosted integration runtime are the same pieces of software; however, the Azure Integration Runtime is designed to run locally on your computer, whilst the self-hosted integration runtime is designed to run in the cloud.
3. Execution of SSIS packages is made possible by the Azure SSIS Integration Runtime, which offers a managed environment in which to do so. As a result, the lifting and moving of SSIS packages to the data factory is accomplished through the utilization of the Azure SSIS Integration Runtime.

Azure Data Factory Interview Questions for Experienced

26. What are the prerequisites that need to be met before an SSIS package can be executed in Data Factory?

Setting up an SSIS integration runtime and an SSISDB catalogue in an Azure SQL server database or an Azure SQL-managed instance is required before an SSIS package can be executed. This can be done in either of these locations.

27. Is there a limit to the number of Integration Runtimes that may be built or is it unlimited?

The default maximum for anything that may be contained within a Data Factory is 5000, and this includes a pipeline, data set, trigger, connected service, Private Endpoint, and integration runtime. You can file a request to increase this amount through the online help desk if you find that you require more.

28. What is meant by the term "ARM Templates" when referring to Azure Data Factory? Where do we plan to use them?

An ARM template is a file that uses JavaScript Object Notation (JSON), and it is where all of the definitions for the data factory pipeline operations, associated services, and datasets are stored. Code that is analogous to the one used in our pipeline will be incorporated into the template.

Once we have determined that the code for our pipeline is operating as it should, we will be able to use ARM templates to migrate it to higher environments, such as Production or Staging, from the Development setting.

29. How about discussing the three most important tasks that you can complete with Microsoft Azure Data Factory?

As was discussed in the previous section's third question, Data Factory makes it easier to carry out three processes: moving data, transforming data, and exercising control.

1. The operations known as data movement do precisely what their name suggests, which is to say that they facilitate the flow of data from one point to another. For example, information can be moved from one data store to another using Data Factory's Copy Activity. Other data stores may also be used.
2. "Data transformation activities" are any operations that modify data as it is being loaded into its final destination system. Stored Procedures, U-SQL, Azure Functions, and so on are just a few examples.
3. Control (flow) activities, as their name suggests, are designed to help regulate the speed of any process that is going through a pipeline. For example, selecting the Wait action will result in the pipeline pausing for the amount of time that was specified.

30. How well does Data Factory support the Hadoop and Spark computing environments when it comes to carrying out transform operations?

The following types of computer environments are able to carry out transformation operations with the assistance of a Data Factory:

1. The On-Demand Computing Environment provided by I ADF is a solution that is ready to use and includes full management. A cluster is created for the computation to carry out the transformation, and this cluster is afterwards removed once the transformation has been carried out.
2. Bring Your Own Equipment: If you already possess the computer gear and software required to deliver services on-premises, you can use ADF to manage your computing environment in this situation.

31. How many individual steps are there in an ETL procedure?

The ETL (Extract, Transform, Load) technique consists of carrying out these four stages in the correct order.

1. Establishing a link to the data source (or sources) is the initial stage. After that, collecting the information and transferring it to either a local database or a crowdsourcing database is the next step in the process.
2. Making use of computational services includes activities such as transforming data by utilizing HDInsight, Hadoop, Spark, and similar tools.
3. Send information to an Azure service, such as a data lake, a data warehouse, a database, Cosmos DB, or a SQL database. This step can also be accomplished by using the Publish API.
4. To facilitate pipeline monitoring, Azure Data Factory makes use of Azure Monitor, API, PowerShell, Azure Monitor logs, and health panels on the Azure site.

32. Which activity should be performed if the goal is to make use of the results that were acquired by performing a query?

A lookup operation can be used to acquire the result of a query or a stored process. The end result might be a single value or an array of attributes that can be utilized in a ForEach activity or some other control flow or transformation function. Either way, it could be a single value.

33. Is it possible to communicate with a pipeline run by passing information along in the form of parameters?

In Data Factory, a parameter is handled just like any other fully-fledged top-level notion would be. The defining of parameters at the pipeline level enables the passage of arguments during on-demand or triggered execution of the pipeline.

34. Do you have any prior experience with the Execute Notebook activity in Data Factory? Does anybody have any idea how to configure the settings for a laptop task?

Through the use of the execute notebook activity, we can communicate with our data bricks cluster from within a notebook. We are able to send parameters to an activity within a notebook by utilizing the baseParameters attribute of that activity. In the event that the parameters are not explicitly defined or specified in the activity, the notebook's default settings are utilized.

35. Which components of Data Factory's building blocks are considered to be the most useful ones?

1. Each activity inside the pipeline has the ability to use the @parameter construct in order to make use of the parameter value that has been provided to the pipeline.
2. By making use of the @coalesce construct within the expressions, we are able to deal with null values in a pleasant manner.
3. The @activity construct makes it possible to make use of the results obtained from one activity in another.

36. Is it possible to use ADF to implement CI/CD, which stands for continuous integration and delivery?

The Data Factory provides full support for CI and CD for your data pipelines by utilizing Azure DevOps and GitHub. As a consequence of this, you are able to construct and roll out new versions of your ETL procedures in stages before delivering the completed product. After the raw data has been converted into a form that a firm can utilize, it should be imported into an Azure Data Warehouse, Azure SQL Azure Data Lake, Azure Cosmos DB, or another analytics engine that your organization's BI tools can reference. This step should take place as soon as possible.

37. In the context of Azure Data Factory, what does the term "variables" mean?

The variables that are used in the Azure Data Factory pipeline serve this storing function. Variables can be accessed within the pipeline in the same way that they can be used in any programming language because they are likewise available there.

Changing or setting the values of variables can be accomplished through the use of the Set Variable and Add Variable actions, respectively. A data factory typically has both continuous and discrete variables in its database.

1. These Azure Pipeline system variables are grouped together under the heading of System variables. The name, ID, and name of any Triggers that are used in Pipelines, etc. These are the things that you need in order to access the system information that might be relevant to the use case that you are working on.
2. The user variable is the second kind of variable, and it is the sort of variable that is defined explicitly in your code and is driven by the pipeline logic.

38. What is a "data flow map"?

Visual data transformations are referred to as mapping data flows when working in Azure Data Factory. Because of data flows, data engineers can construct logic for altering data without having to write any code at all. After the data flows have been generated, they are then implemented as activities inside of the scaled-out Apache Spark clusters that are contained within Azure Data Factory pipelines. The scheduling, control flow, and monitoring elements that are currently available in Azure Data Factory can be utilized to operationalize data flow operations.

The method of data flow mapping is highly immersive visually and does away with the necessity for any form of scripting. The execution clusters in which the data flows are carried out are managed by the ADF, which enables the data to be processed in a manner that is massively parallel. Azure Data Factory is responsible for all of the coding tasks, including the interpretation of code, the optimization of pathways, and the execution of data flow operations.

39. In the context of the Azure Data Factory, just what does it mean when it's referred to as "copy activity"?

The copy operation is one of the most extensively used and generally popular operations in the Azure data factory. The procedure that is known as "lift and shift" is useful in situations in which it is necessary to copy information from one database to another. You can make modifications to the data as you copy it. For instance, before you transmit it to your target data source, you can decide to lower the number of columns in the source txt/csv file from 12 to 7. You can change it in such a way that the target database receives only the required number of columns after the transfer.

40. Could you explain to me how I should go about planning a pipeline?

You can set up a pipeline's schedule by utilizing either the time window trigger or the scheduler trigger. Pipelines can be programmed to run on a timed basis on a periodic basis or in cyclical patterns according to the wall-clock calendar schedule of the trigger (for example, on Mondays at 6:00 PM and Thursdays at 9:00 PM).

There are now three different kinds of triggers that can be utilized with the service, and they are as follows:

1. A state-preserving periodic trigger, the tumbling window trigger is used in this game.
2. A time-based trigger that triggers a specified pipeline at a time that has been predetermined is referred to as a Schedule Trigger.
3. One category of triggers is known as "event-based," which indicates that they respond in some way to any occurrence that takes place, such as when a file is copied into a blob. Pipelines and triggers are mapped to one another in a way that is many-to-many (except for the tumbling window trigger). It is conceivable for a single trigger to launch many pipelines, or it is possible for numerous triggers to initiate a single pipeline. Both scenarios are viable.

41. In which situations does Azure Data Factory seem the best option?

The utilization of the Data Factory is essential at this time.

1. When dealing with massive amounts of data, it is likely required to implement a cloud-based integration solution such as ADF. This is because a data warehouse needs to be created.
2. Not everyone in the team is a coder, and some members may discover that graphical interfaces make it simpler to analyze and manipulate data.
3. When raw business data is located in many places, both on-premises and in the cloud, we need a unified analytics solution such as ADF to analyze it all in one place.
4. We would like to minimize the management of our infrastructure to a minimum by making use of methods that are widely utilized for the transportation of data and the processing of it. Because of this, going with a managed solution such as ADF is the choice that makes the most sense.

42. Do you have any tips on how to access the information you require by taking advantage of the other ninety dataset types that are accessible in the Data Factory?

Data can originate from a wide variety of Azure services, including Azure SQL Database, Azure Synapse Analytics, delimited text files stored in an Azure storage account or Azure Data Lake Storage Gen2, and Parquet files stored in blob storage or Data Lake Storage Gen2. Source and sink data can be combined. In order to transform data coming from external connectors, first stage the data using the Copy action, and then perform an activity in the Data Flow category.

43. Can the value of a new column in an ADF table be determined by using an existing mapping column?

The logic that we specify can be used to generate a new column, and this is done by deriving transformations within the mapping data flow. When developing a derived column, we have the option of creating a brand-new one from scratch or making changes to an existing one. You can recognise the new column by giving it a name in the textbox labelled Column.

If you choose a different column from the menu, the one you currently have selected will be removed from your schema. Select the textbox that corresponds to the derived column, and then press the Enter key on your keyboard to get started crafting an expression for it. You can either manually enter your reasoning or make use of the expression builder to construct it.

44. Where can I find more information on the benefits of using lookup operations in the Azure Data Factory?

In the ADF pipeline, the Lookup activity is typically utilized for configuration lookup most of the time due to the ready availability of the source dataset. In addition to this, the output of the activity can be used to retrieve the data from the dataset that served as the source. In most cases, the outcomes of a lookup operation are sent back down the pipeline to be used as input for later phases.

In order to retrieve data, the ADF pipeline makes heavy use of lookup operations. You may only utilize it in a manner that is appropriate for the process you are going through. You have the option of retrieving either the first row or all of the rows, depending on the dataset or query you choose.

45. Please provide any more information that you have on the Azure Data Factory Get Metadata operation.

The Get Metadata operation can be used to access the metadata associated with any piece of data that is contained within an Azure Data Factory or a Synapse pipeline. We can use the results from the Get Metadata activity in conditional expressions to validate or utilize the metadata in subsequent actions. This can be done by using the Get Metadata activity.

It takes a dataset as an input and then generates descriptive data based on that dataset as an output. The supported connectors and the metadata that may be retrieved for each one are outlined in the table that can be found below. It is possible to accept metadata returns that are up to 4 MB in size.

46. Where did you experience the most difficulty while attempting to migrate data from on-premises to the Azure cloud via Data Factory?

Within the context of our ongoing transition from on-premises to cloud storage, the problems of throughput and speed have emerged as important obstacles. When we attempt to replicate the data from on-premises using the Copy activity, we do not achieve the throughput that we require.

The configuration variables that are available for a copy activity make it possible to fine-tune the process and achieve the desired results.

1. If we load data from servers located on-premises, we should first compress it using the available compression option before writing it to cloud storage, where the compression will afterwards be erased.
2. After the compression has been activated, ii) it is imperative that all of our data be quickly sent to the staging area. Before being stored in the target cloud storage buckets, the data that was transferred might be uncompressed for convenience.
3. Copying Proportion, The use of parallelism is yet another alternative that offers the potential to make the process of transfer more seamless. This accomplishes the same thing as employing a number of different threads to process the data and can speed up the rate at which data is copied.
4. Because there is no one size that fits all, we will need to try out a variety of different values, such as 8, 16, and 32, to see which one functions the most effective.
5. It may be possible to hasten the duplication process by increasing the Data Integration Unit, which is roughly comparable to the number of central processing units.

47. Do I have the ability to copy information simultaneously from many Excel sheets?

When using an Excel connector within a data factory, it is necessary to specify the sheet name from which the data is going to be loaded. This method is unobtrusive when dealing with data from just one or a few sheets, but when dealing with data from tens of sheets or more, it can become tiresome because the sheet name needs to be updated in the code each time.

By utilizing a data factory binary data format connector and directing it to the Excel file, we can avoid having to manually insert the sheet names into the spreadsheet. Using the copy action, you will be able to simultaneously copy the information that is located on each of the sheets that are contained within the file.

48. Nesting of loops within loops in Azure Data Factory: yes or no?

Nesting loops are not directly supported by any of the activities that use the for each or till looping structures in the data factory. On the other hand, we have the option of utilizing an execute pipeline activity that incorporates a for each/until loop activity. By carrying out the aforementioned steps, we will be able to successfully implement nested looping by having one loop activity call another loop activity.

49. Are there any particular limitations placed on ADF members?

Azure Data Factory provides superior tools for transmitting and manipulating data, and these tools can be found in its feature set. However, you should be aware that there are certain limitations as well.

1. Because the data factory does not allow the use of nested looping activities, any pipeline that has such a structure will require a workaround in order to function properly. Here is where we classify everything that has a looping structure: actions involving the conditions if, for, and till respectively.
2. The lookup activity is capable of retrieving a maximum of 5000 rows in a single operation at its maximum capacity. To reiterate, in order to implement this kind of pipeline design, we are going to need to use some additional loop activity in conjunction with SQL with the limit.
3. It is not possible for a pipeline to have more than forty activities in total, and this number includes any inner activities as well as any containers. To find a solution to this problem, pipelines ought to be modularized with regard to the number of datasets, activities, and so on.

50. What is Data Flow Debug?

It is possible to do data flow troubleshooting in Azure Data Factory and Synapse Analytics while simultaneously monitoring the real-time transformation of the data shape. The versatility of the debug session is beneficial to both the Data Flow design sessions as well as the pipeline debug execution.

Conclusion

This article was created to aid you in getting ready for an interview with Azure Data Factory. It is a service that helps in the creation, storage, and management of data that is stored on the cloud. It functions similarly to a database, but much more quickly, and it can store significantly more information. Use it to organize and keep track of enormous data sets, such as photo albums or video collections, with its built-in organization and tracking features.

You will need to have at least a fundamental familiarity with the Azure Data Factory before you can begin to make preparations for the interview. To get started, you have to become familiar with its inner workings. Second, you'll require experience in environmental construction. The final step entails the development and dissemination of your apps.

Interview Preparation Resources

- [Azure Architecture](#)
- [Azure Interview Questions](#)
- [Azure Databricks Interview Questions](#)

Links to More Interview Questions

[C Interview Questions](#)

[Php Interview Questions](#)

[C Sharp Interview Questions](#)

[Web Api Interview Questions](#)

[Hibernate Interview Questions](#)

[Node Js Interview Questions](#)

[Cpp Interview Questions](#)

[Oops Interview Questions](#)

[Devops Interview Questions](#)

[Machine Learning Interview Questions](#)

[Docker Interview Questions](#)

[Mysql Interview Questions](#)

[Css Interview Questions](#)

[Laravel Interview Questions](#)

[Asp Net Interview Questions](#)

[Django Interview Questions](#)

[Dot Net Interview Questions](#)

[Kubernetes Interview Questions](#)

[Operating System Interview Questions](#)

[React Native Interview Questions](#)

[Aws Interview Questions](#)

[Git Interview Questions](#)

[Java 8 Interview Questions](#)

[Mongodb Interview Questions](#)

[Dbms Interview Questions](#)

[Spring Boot Interview Questions](#)

[Power Bi Interview Questions](#)

[Pl Sql Interview Questions](#)

[Tableau Interview Questions](#)

[Linux Interview Questions](#)

[Ansible Interview Questions](#)

[Java Interview Questions](#)

[Jenkins Interview Questions](#)