Question 1 : What is Azure Data Factory?

ADF is a cloud based integration service , which allows you to create datadriven workflows in cloud for orchestration and automating data movement , data transformation..

we can create and schedule data driven workflows using adf also called pipeline.

that can ingest the data from various sources.

it can process and transform the data using compute services

such as HDInsight Hadoop, Spark, Azure Data Lake Analytics, and Azure Machine Learning

Question 2 : Windows Azure Storage?

We have manly 4 type of storage in azure

i) Queue
ii) Table
iii) BLOBs (Binary Large Objects)
iv) VHD (Windows Azure Drives)

Question 3 : What is the integration runtime?

IR is a compute infrastruture that ADF use to provide various data integration capabilities accross various network

we have 3 types integration run time

i) Azure integration runtime - we can copy the data between cloud data stores and dispacth it to compute services for transformation.

ii) Self hosted integration runtime - as the name signifies that we can install it on premise machine and virtual machine in a virtual network.

SHIR can run copy activity between public cloud data stores and data store in a private network

It can also dispatch transformation activities against compute resources in a private network. XXX

We use Self Hosted IR because Data factory will not be able to directly access on-primitive data sources as they sit behind a firewall.

iii) Azure SSIS integration runtime

Question 4 : What is the limit on the number of integration runtimes?

There is hard code limit on the number of integration runtime instance in ADF.

but there is a limit on VM cores that intergration runtime can use per subscription

Question 5 : What is blob storage in Azure?

Azure blob storage is a service which is used for storing large amount of data of diffrent types

We can use blob storage for sharing the data publicly
or store application data privately

Common uses of blob storage.
i) storing files ,like audio vedio
ii) storing data as a backup, data disaster recovery,
iii) data analytics

Question 6 : What are the steps for creating ETL process in Azure Data Factory?
While to trying to extract data from any sources such as sql server, if something has to be processed.
then it will be processed and is stored in the Data Lake Store.
Steps for Creating ETL
Create a Linked Service for source data store which is SQL Server Database
Create a Linked Service for destination data store which is Azure Data Lake Store
Create a dataset for Data Saving
Create the pipeline and add copy activity
Schedule the pipeline by adding a trigger

Question 7 : What are the top-level concepts of Azure Data Factory?
Pipeline - It acts as a carrier in which we have various processes taking place.
activity - Activities represent the processing steps in a pipeline.
datasets - it is data structure which holds our data
linked services - These store information that is very important when it comes to connecting an external source.

Question 8 : How can I schedule a pipeline?
You can use the
i) scheduler trigger
ii) time window trigger
to schedule a pipeline.
The trigger uses a wall-clock calendar schedule,
which can schedule pipelines periodically or in calendar-based recurrent patterns (for example, on Mondays at 6:00 PM and Thursdays at 9:00 PM)

Question 9 : Can I pass parameters to a pipeline run?
Yes we can pass paramters to a pipeline run
i) Define the parameters at the pipeline level

ii) pass argument/value as you execute the pipeline run on demand or by using a trigger

Question 10 : Can I define default values for the pipeline parameters?
You can define default values for the parameters in the pipelines.

Question 11 : Can an activity in a pipeline consume arguments that are passed to a pipeline run?
Each activity within the pipeline can consume the parameter value that’s passed to the pipeline and run
with the @parameter construct.

Question 12 : Can an activity output property be consumed in another activity?
An activity output can be consumed in a subsequent activity with the @activity construct

Question 13 : How do I gracefully handle null values in an activity output?
You can use the @coalesce construct in the expressions to handle the null values gracefully.

Question 14 : Which Data Factory version do I use to create data flows?
Use the Data Factory V2 version to create data flows

Question 15 : Explain the two levels of security in ADLS Gen2?

Question 16 : What is table storage in Windows Azure?

Question 17 : What is Azure Functions?
Azure funtions is a soluntion for executing a peice of code or funtions in the cloud.
we can also select the programming language we want to use.
we pay only the time we excute our code.
It supports continuous deployment and integration.

Question 18 : What is Azure HdInsight Cluster?
Azure HdInsight is a cloud service that makes easy, fast and cost effective to process masive amount of data using open source framework.
like Hadoop, Spark, Hive, LLAP, Kafka, Storm and R.

Question 19 : What is Azure Data Lake
Azure data lake is a service which is provided by microsoft azure to store massive amount of data for unlimited time.
Azure data lake is divided into 2 parts

i) Data Analytics - is a service that you use to run big data jobs on the data that is stored in data storage.

ii) Data storage - we can store any amount of data and unlimited time

note: - Data lake is a huge repository where you can store your data in object or named format

Question 20 : What is SQL Azure database?

SQL Azure database is just an approach to get associated with cloud services where you can store your database into the cloud.

Question 21 : How to stop a running slice?

If you need to stop the pipeline from executing, you can use Suspend-AzDataFactoryPipeline cmdlet.

Currently, suspending the pipeline does not stop the slice executions that are in progress.

Once the in-progress executions finish, no extra slice is picked up