1. Why should one use Azure Key Vault when working in the Azure environment? What are the pros and cons? What are the alternatives?

The Azure Key Vault should be used to securely store sensitive information in credentials store which only the ADF service or Administrators can read from. If Credentials need to be rotated ADF Linked Service will not need to be modified. When migrating the ADF pipeline from Dev to Test to production no change is necessary. The alternatives would be to hard-code the credentials and risk sensitive connection details and passwords getting leaked or lost.

2. How do you achieve loop functionality within an Azure Data Factory pipeline? Why would you need to use this functionality in a data pipeline?

You can use a ForEach Containter to execute a pipeline for each item from a Lookup Container

3. What are expressions in Azure Data Factory? How are they helpful when designing a data pipeline? Please explain with an example.

Expressions allow you to setup up dynamic properties by referencing parameters that are set at run-time. This allows the Pipeline to remain unchanged whenever parameters like connection strings or passwords change, and it also allows you to scale and re-use the logic especially during loops.

4. What are the pros and cons of parametrizing a dataset’s activity in Azure Data Factory?

Pros: Allows you to scale, and loop through files dynamically.

Cons: Requires more planning and time on the initial setup, and it can be more difficult to troubleshoot issues with parameters during the run.

5. What are the different supported file formats and compression codecs in Azure Data Factory? When will you use a Parquet file over an ORC file? Why would you choose an AVRO file format over a Parquet file format?

The following file formats and compression codecs are supported in Azure Data Factory:

* Avro format.
* Binary format.
* Delimited text format.
* Excel format.
* JSON format.
* ORC format.
* Parquet format.
* XML format.
* Gzip format

Parquet files are best for nested data.

ORC format is best suited for Hadoop/Hive.

Avro file format is good for storing data in a data lake staging area, since it is very readable and easy to debug.

**Notes / Screenshots from Azure Data Factory Hands-On Mini-Project:**

ADF SQL Instance Name: **adf-dan-dev-sqlserver**

Login: **dwilde**

PW: **\*\*\*\*\*\*2017!**

ADF Data Factory Name: **adf-dan-dev-datafactory**

Managed Identity Object ID: **a6008469-2f4c-44ea-80c0-85745bb7fdff**

Access Key: DefaultEndpointsProtocol=https;AccountName=adfdandevstorage;AccountKey=liFEGFzJ4s9YaKd3XqA/IRw3PlFSiD42Otn+1b+lXbmVuee7iRYkh32SNs/9xdCVv/TJk4nTjf9NlZPy6CfMnQ==;EndpointSuffix=core.windows.net

SQL Connection String:

Server=tcp:adf-dan-dev-sqlserver.database.windows.net,1433;Initial Catalog=adf-dan-dev-sqldb;Persist Security Info=False;User ID=dwilde;Password= \*\*\*\*\*\*2017!;MultipleActiveResultSets=False;Encrypt=True;TrustServerCertificate=False;Connection Timeout=30;

Azure Key Vault Linked Service:

Graphical user interface, text, application, email

Description automatically generated

ERROR with SFTP Connection (see screenshot below):

I just could not get the SFTP Connection to work no matter what I did. I went through the instructions in detail 3 times, and each time I ran into this error with the WWISftp Linked Service:

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text

Description automatically generated

There appears to be some typos in the lesson here – the name is listed as “WWICustomerSftpToBlob”, but we are supposed to be configuring the Orders SFTP

Also, there are 2 steps in the screen shot below that appear to be the same thing (Setup Copy Activity Source) – this all made the exercise very confusing and time consuming to complete

A picture containing diagram

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

The columns from my dataset set did not match the example screenshots, and did not have headers – the columns were also out of order:

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

A screenshot of a computer

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application, Word

Description automatically generated