**Day 11**

1. Data engineering- is the process
   * + - 1. Building data pipeline for different sources of data,
         2. Collect and store,
         3. process data and clean it on the basis on requirement,
         4. ensuring data quality for good insights
2. A pipeline must be built to collect data from various places
3. Data integration-

get data from various sources,

mix the data well(convert it into the same format)(ensure consistent),

make data available for insights

1. 2 data source SQL and Excel how do we combine if nature of data is different, and if we somehow put it in one place then how do we extract the columns relevant to us – azure data factory is an application that can be used for Extract, Transform and Load the data.

DAY 12

1. Data factory[activity: how to copy data file to another container in a storage account]:
   1. needs a linked service which serves as the connection string between the storage account and the data factory,
   2. now we build our datasets to perform the functions we need for example in this we need to copy from one storage container to another
   3. finally a pipeline where we use the “copy activity” and select the source and sink for it (which are basically the datasets we made)
   4. we need to then validate>publish>trigger pipeline to see our changes
2. Data lakes used for analytics and storing data

DAY 14

1. If condition helps us to put respective conditions for the pipeline for example if want to copy data if the condition is true, we can put that activity in that action.
2. Metadata is an activity that tells us about data itself we can edit the field list and arguments to figure what do we need exactly.
3. In metadata we can either choose items or child items where child items refer to all the items while items refers to a specific item
4. IAM(Identity&Access Management) helps us to give access to others and helps us set access permissions.