Azure data factory is Microsoft’s cloud version of extract-transform-load (ETL), extract-load-transform (ELT) tool which can be used for integration and transformation of any data. Whether they’re on on-premises or Azure or on another public cloud such as AWS & GCP this tool allows integration with almost zero code effort.

Using Azure Data Factory, you can create and schedule data-driven workflows (called pipelines) without any code. Pipeline can ingest data from any data source where you can build complex ETL processes that transform data visually with data flows or by using compute services such as Azure HDInsight Hadoop, Azure Databricks, and Azure SQL Database.

Integration Runtime:

The integration runtime (IR) is the compute infrastructure to provide the following data integration capabilities across different network environment.

There are three types of Integrated runtime. They are:

1. Azure IR
2. Self-Hosted IR
3. Azure-SSIS

**Azure, Self-Hosted**

Perform data flows, data movement and dispatch activities to external compute.

**Azure**

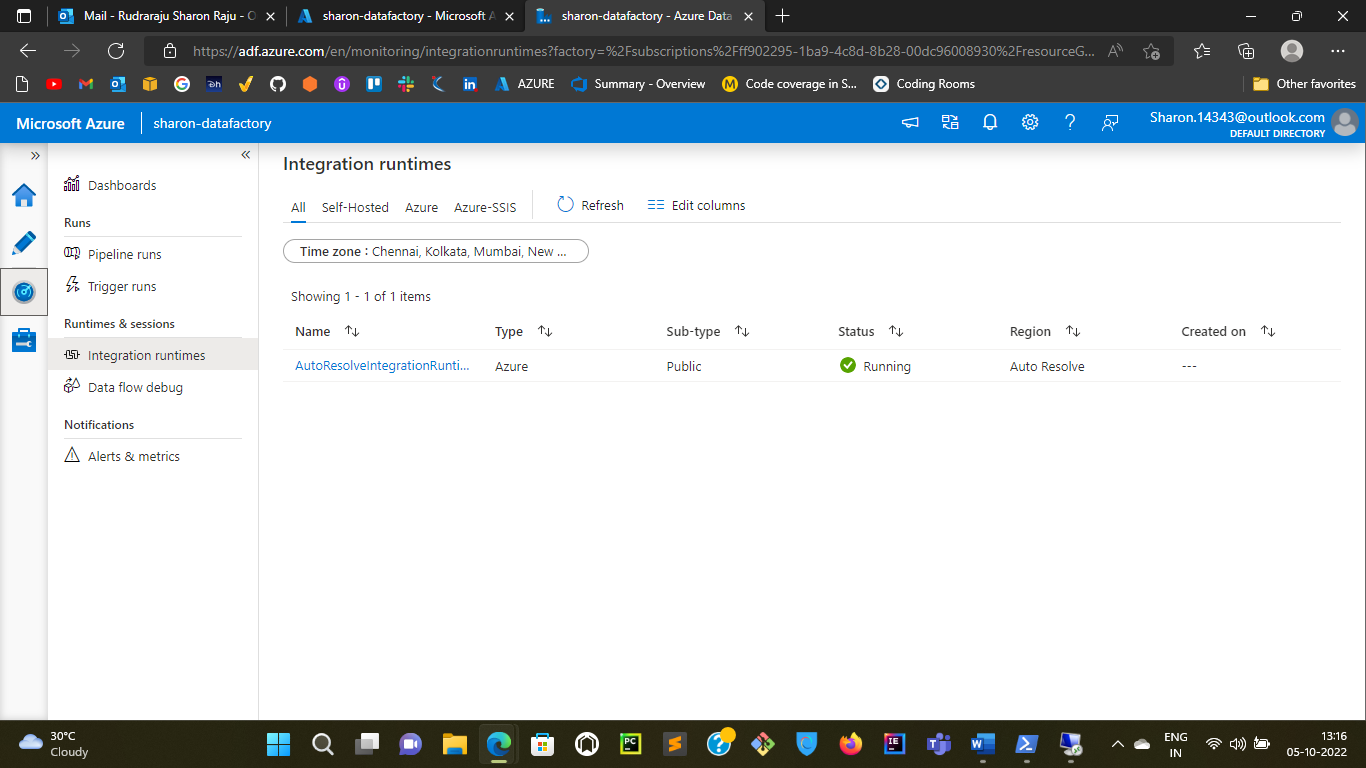
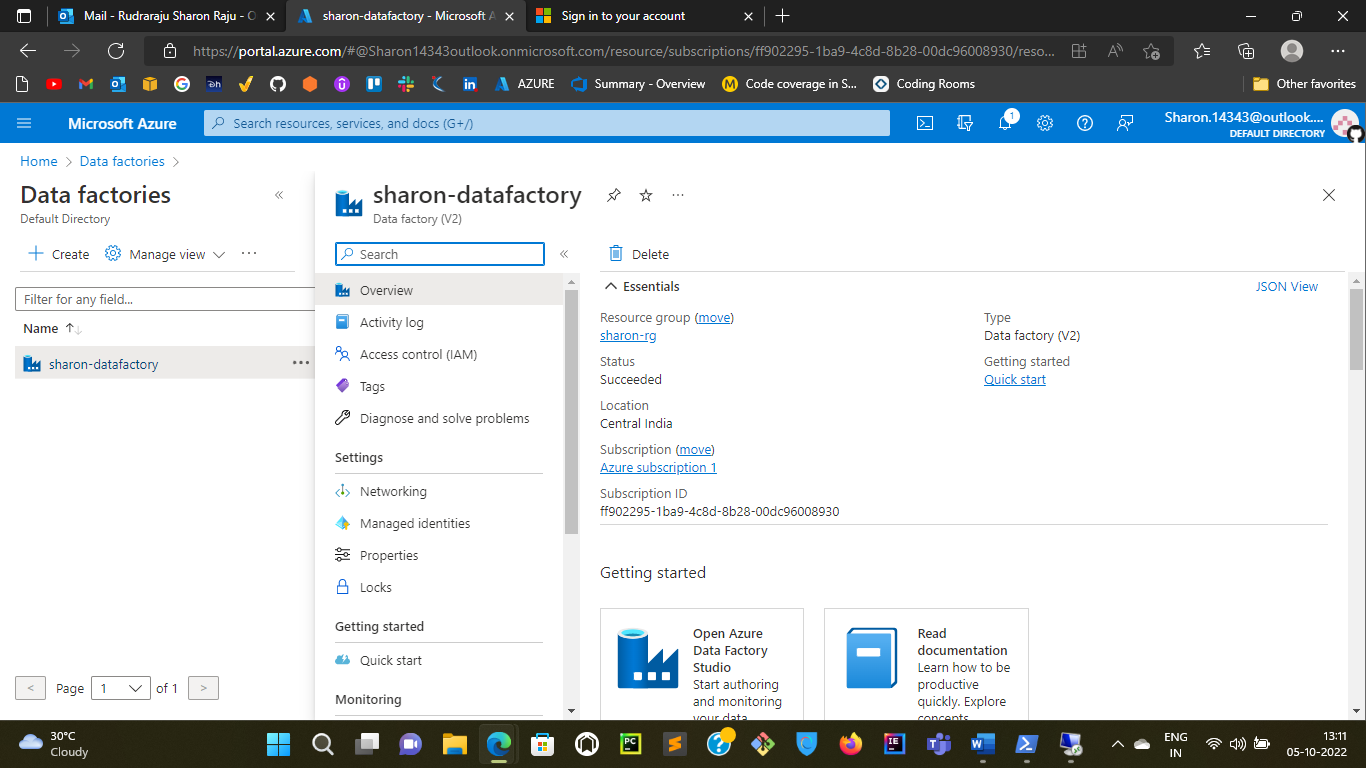
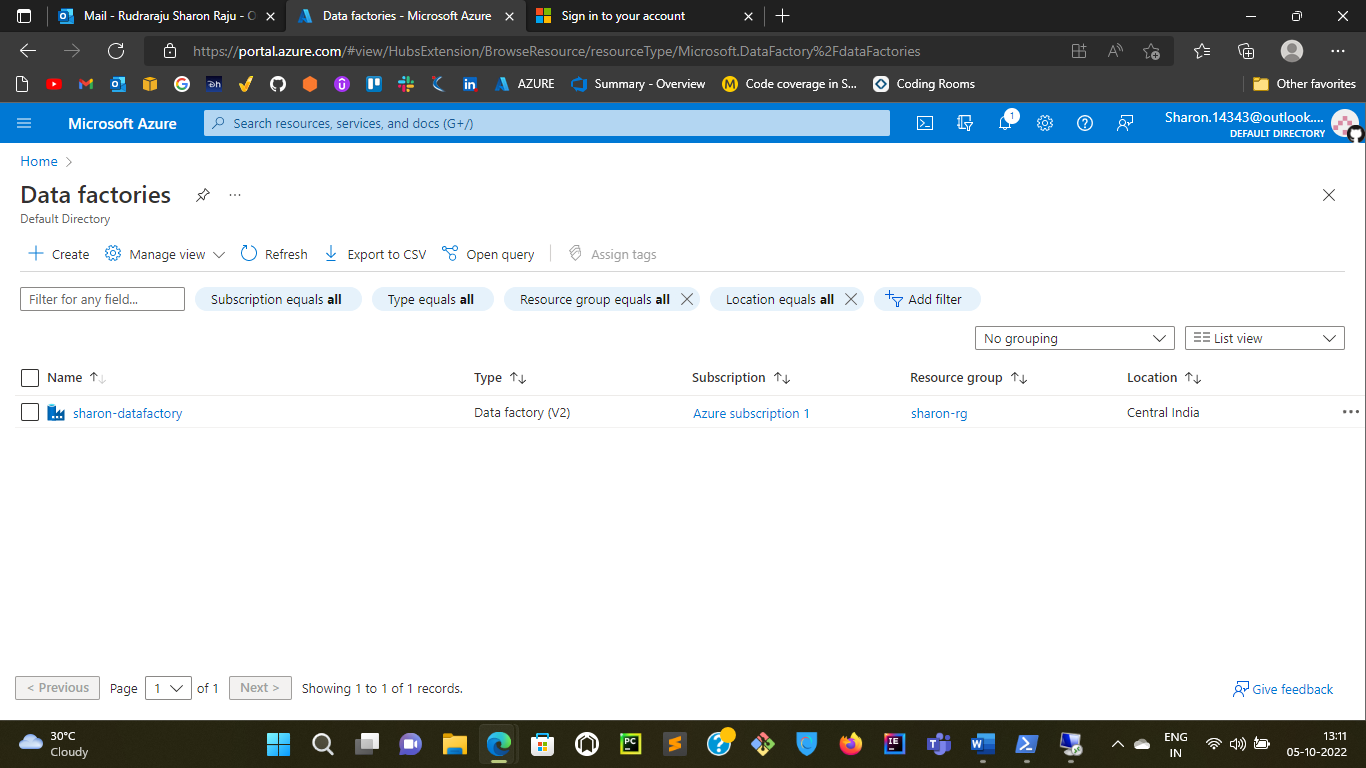
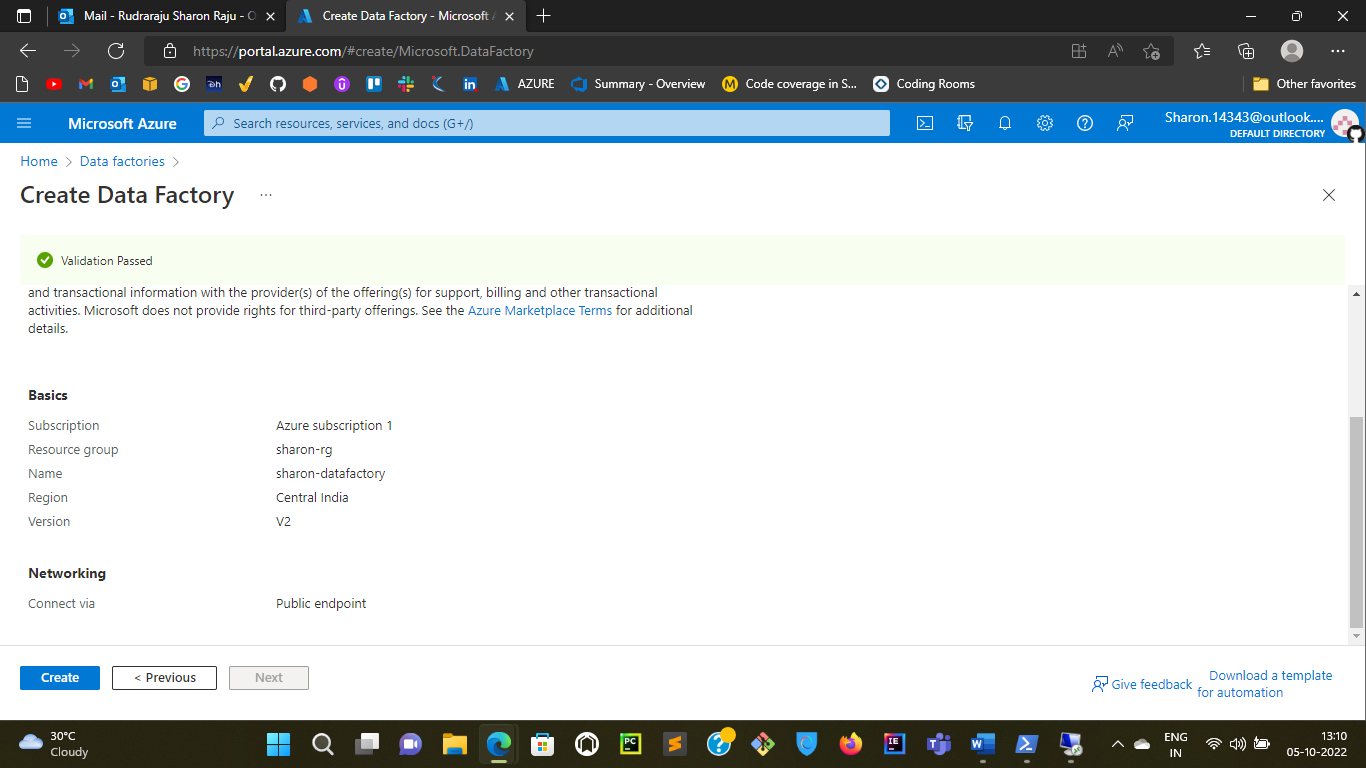
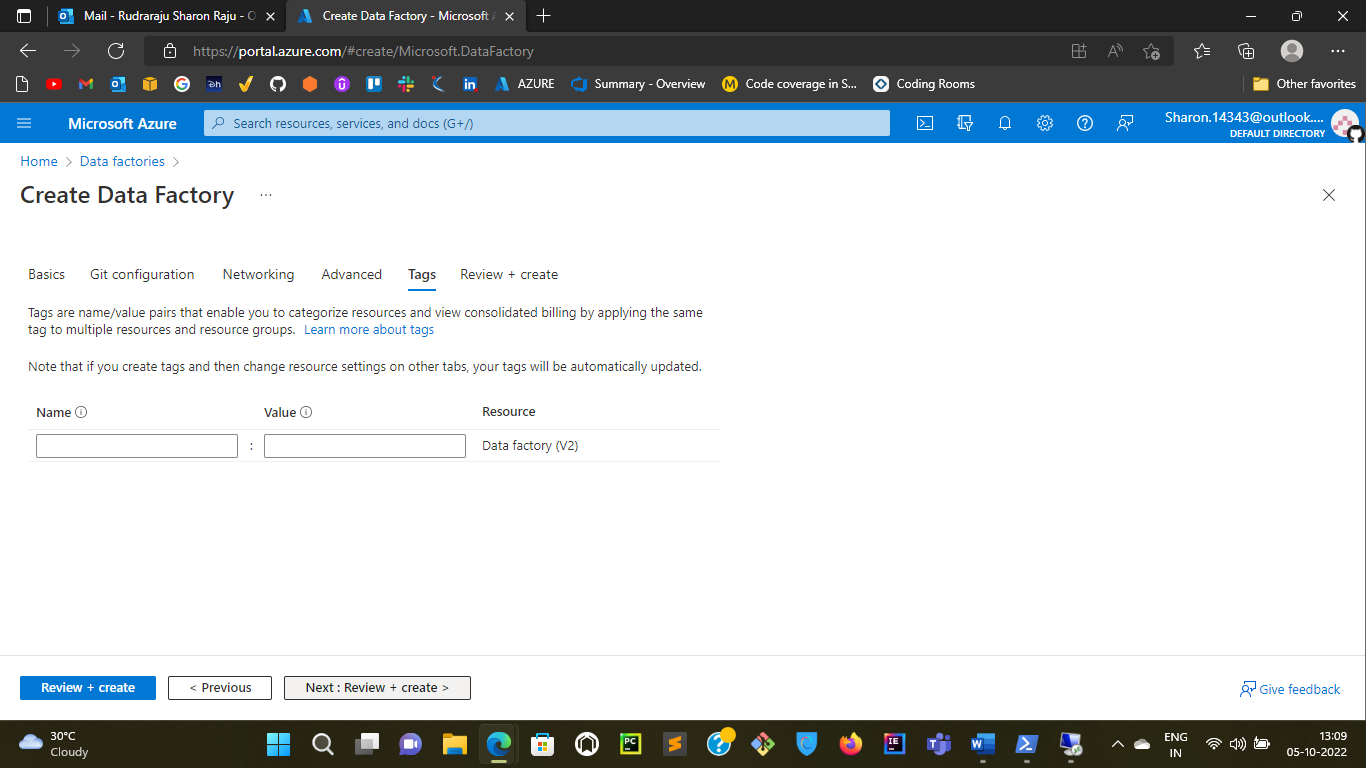
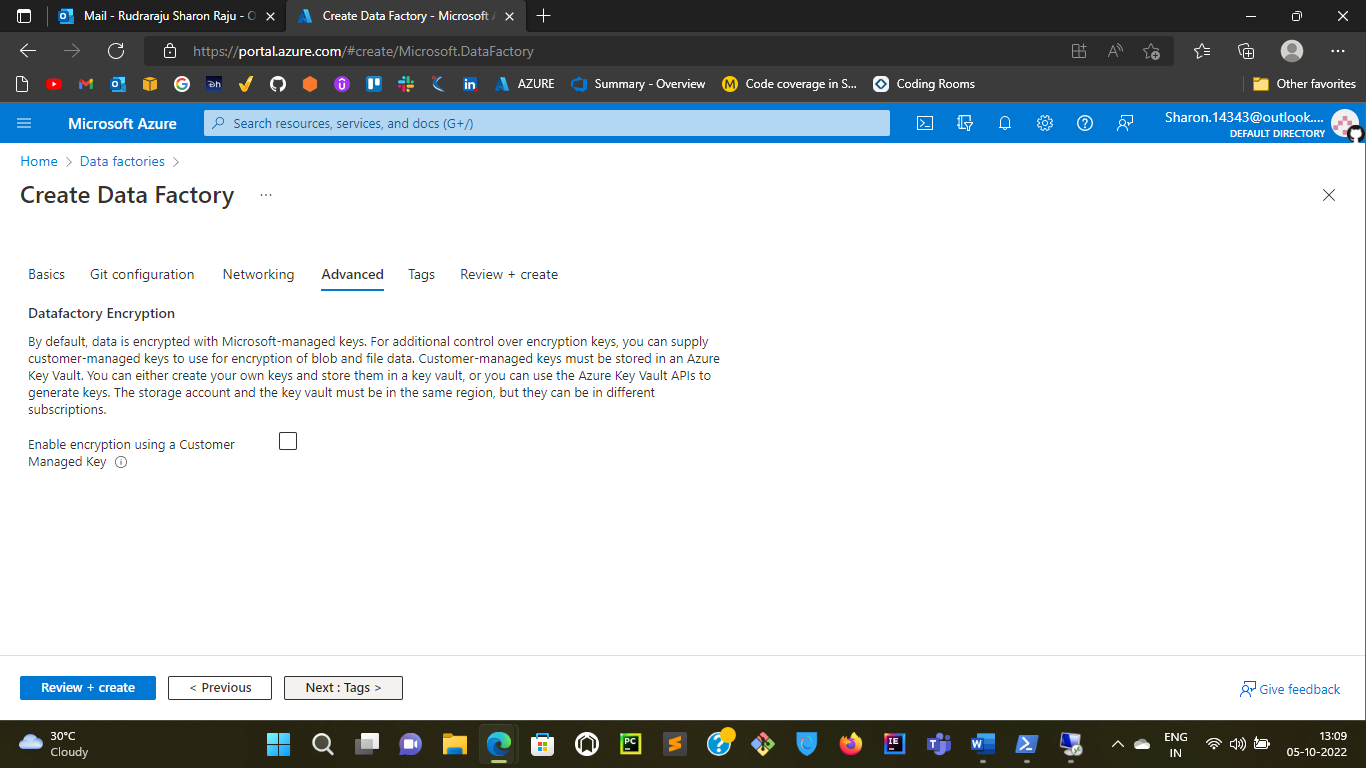
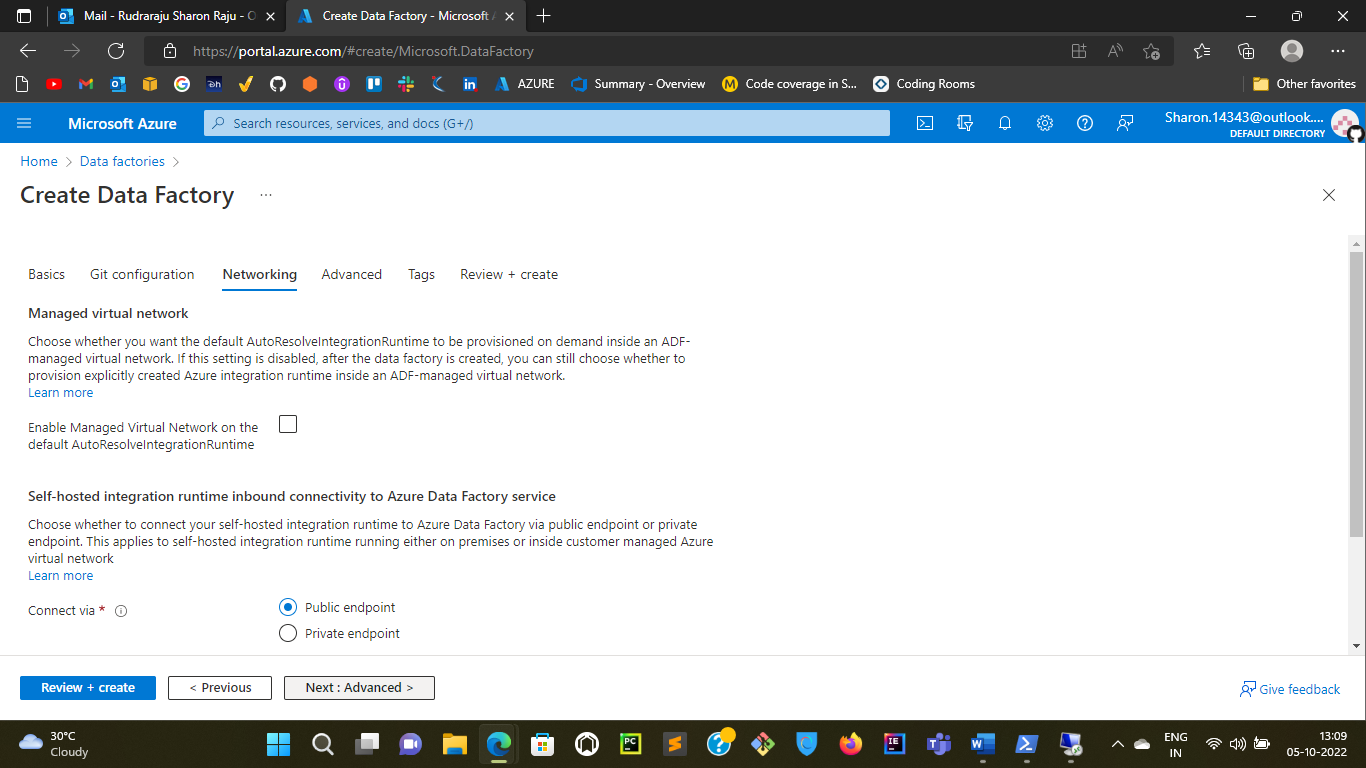
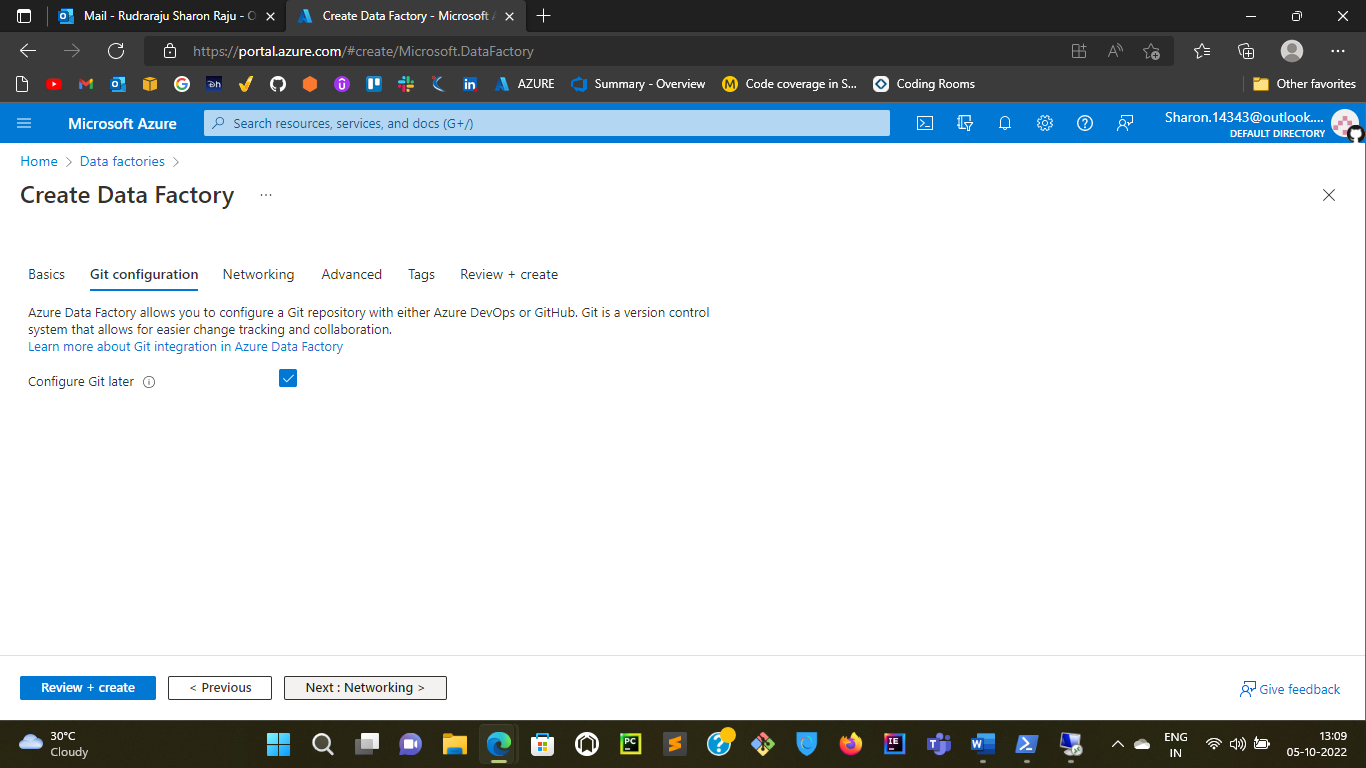
Use this for running data flows, data movement, external and pipeline activities in a fully managed, serverless compute in Azure.

**Self-Hosted**

Use this for running data movement, external and pipeline activities in an on-premises / private network by installing the integration runtime.

**Azure-SSIS**

Lift-and-shift existing SSIS packages to execute in Azure.



Linked service defines the connection information to a data store or compute.

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store.

Pipelines contains activities that put together to complete a task.