Case Study 3 — Real-Time Data Lakehouse Architecture

By Kafeel Kamran

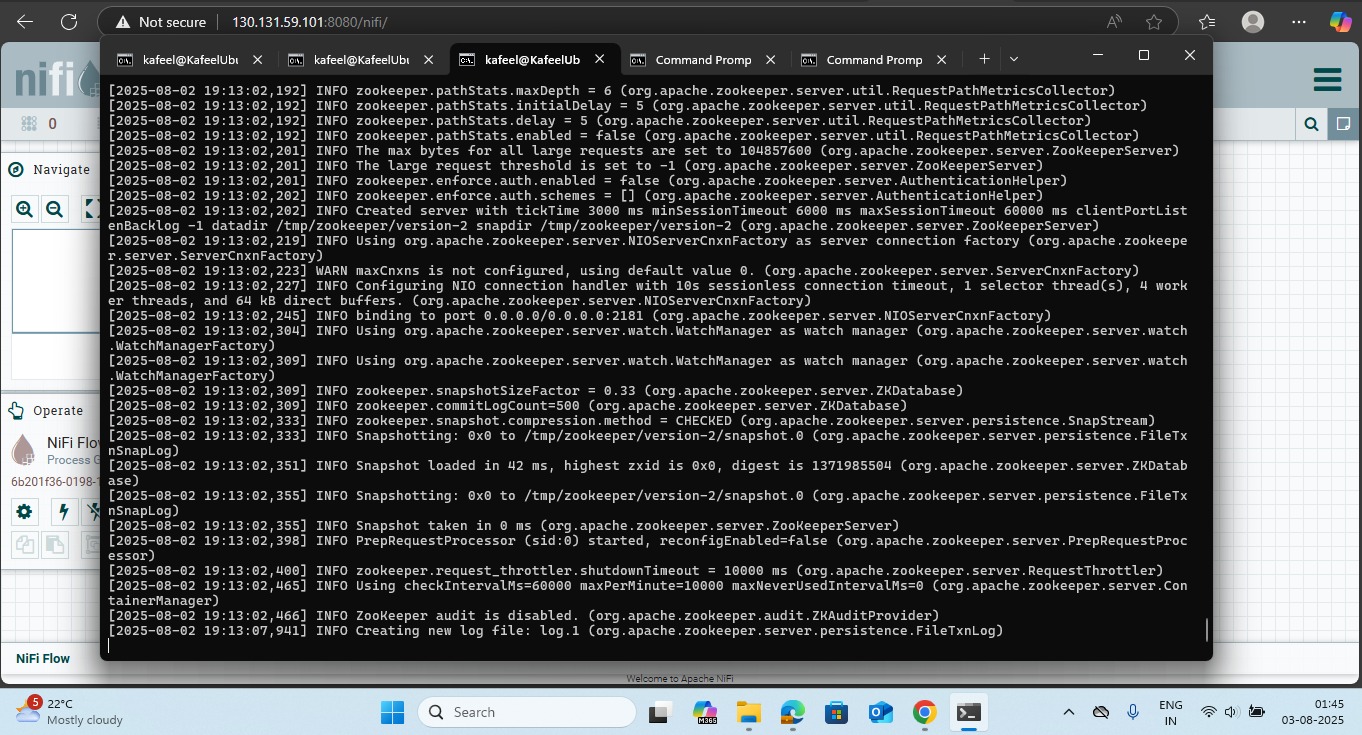
Data: hw\_200.csv,

Sales.csv (https://www.kaggle.com/datasets/karkavelrajaj/amazon-sales-dataset )

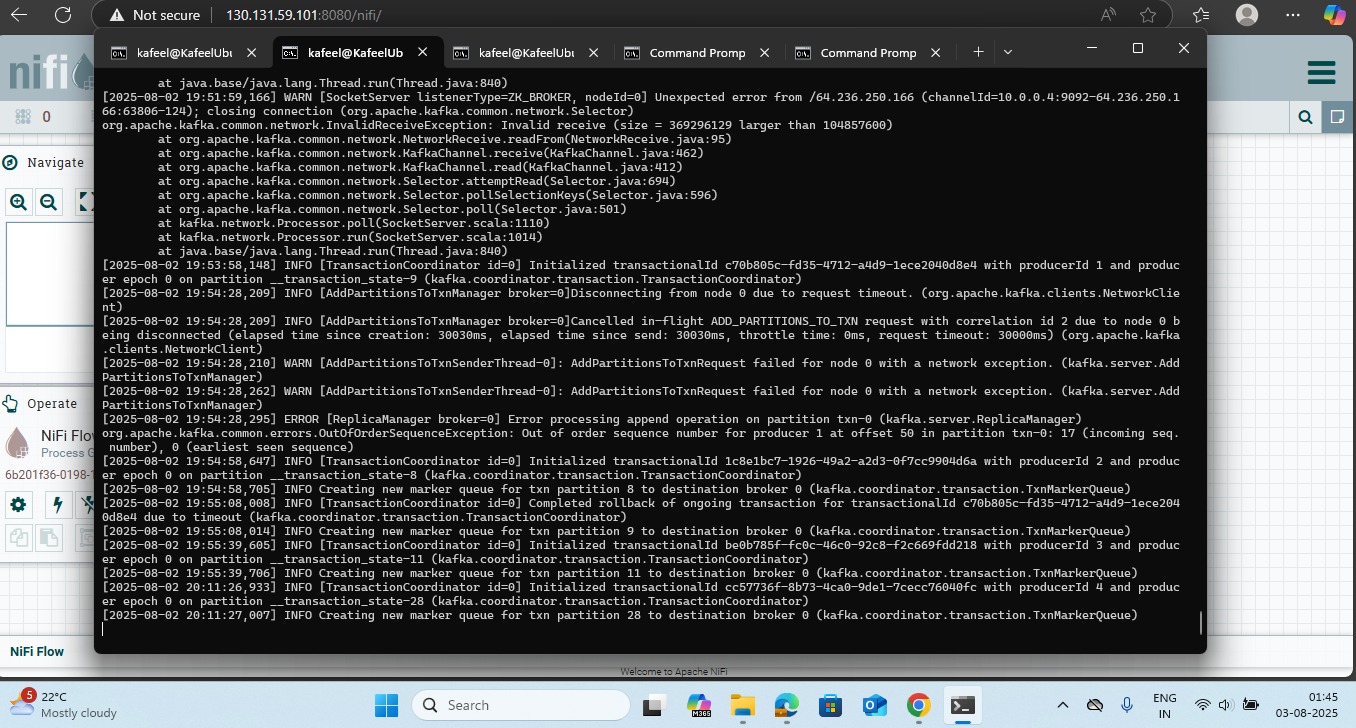
Docs for reference

Basic Environment setup (Nifi + Kafka)

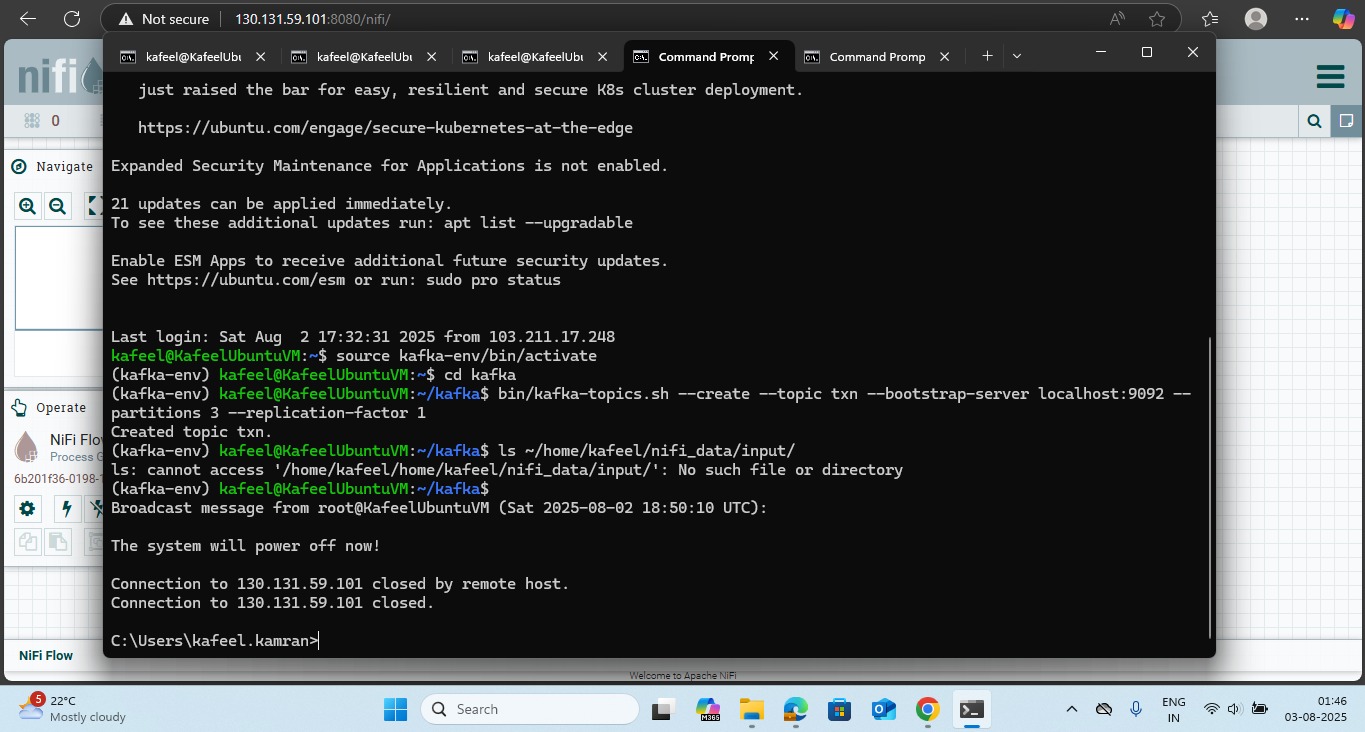
* ***Start and run zookeeper***



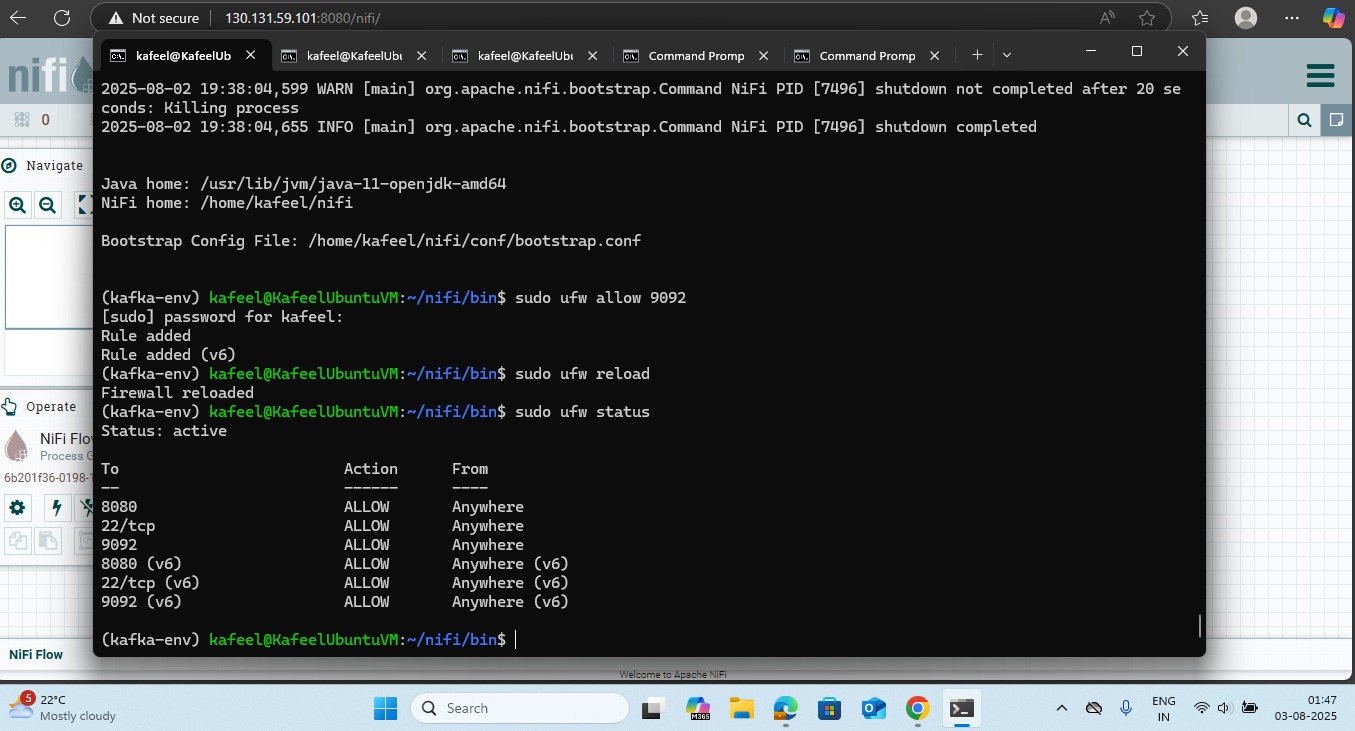
* ***Start and run Kafka server***



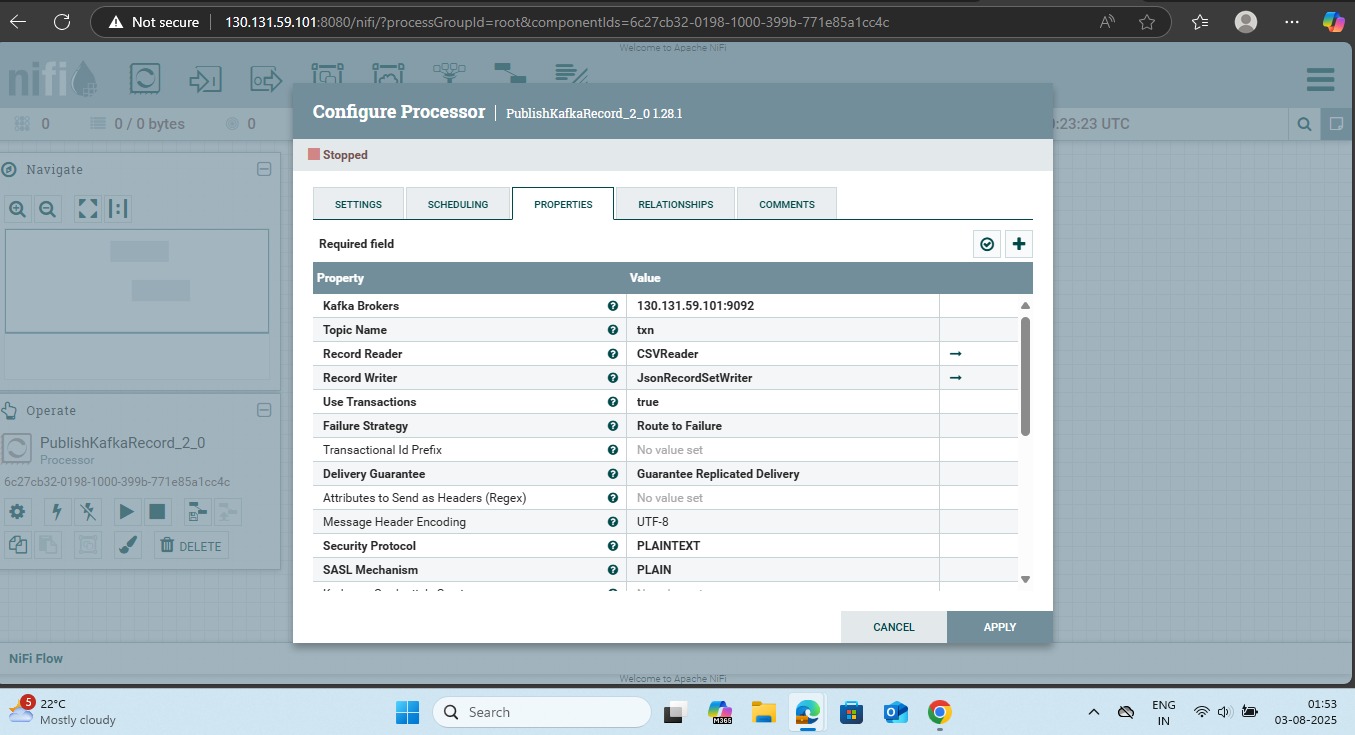
* ***Create a topic ‘txn’ from VM***



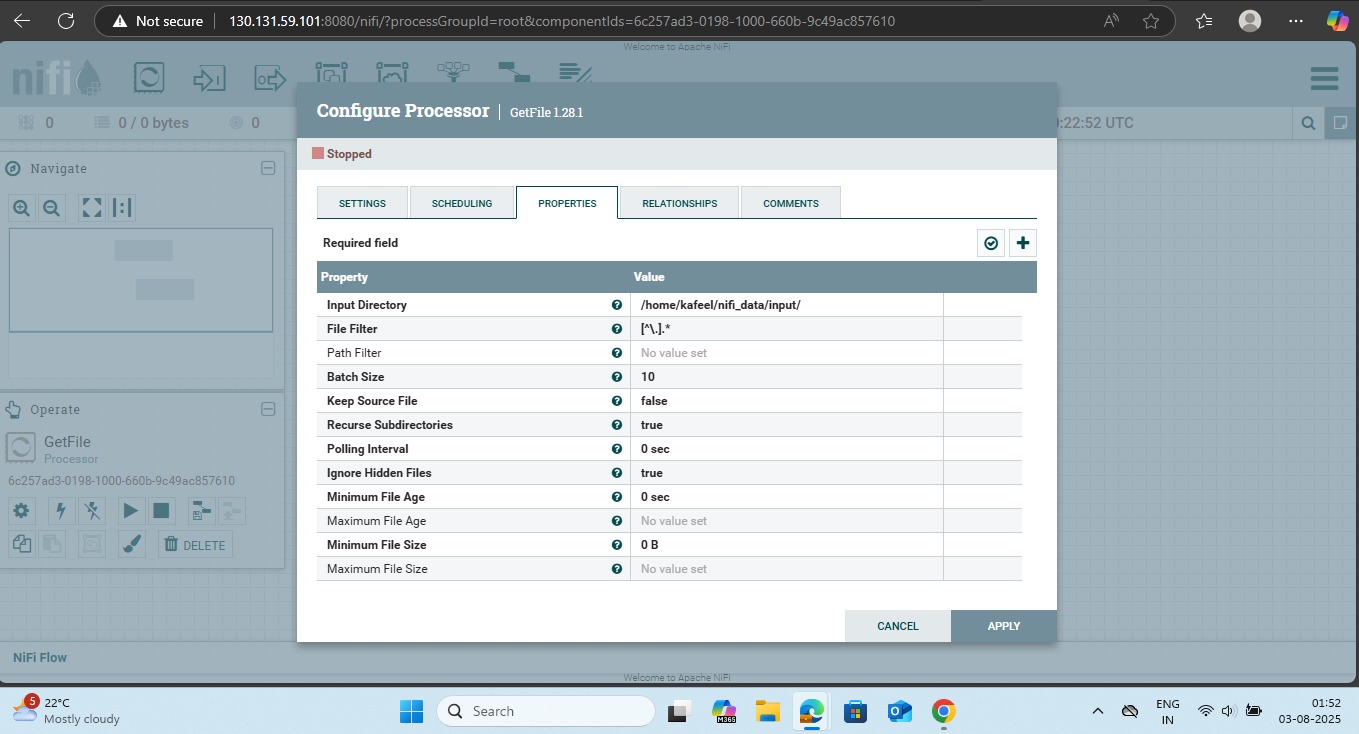
* ***Configure and run Nifi***



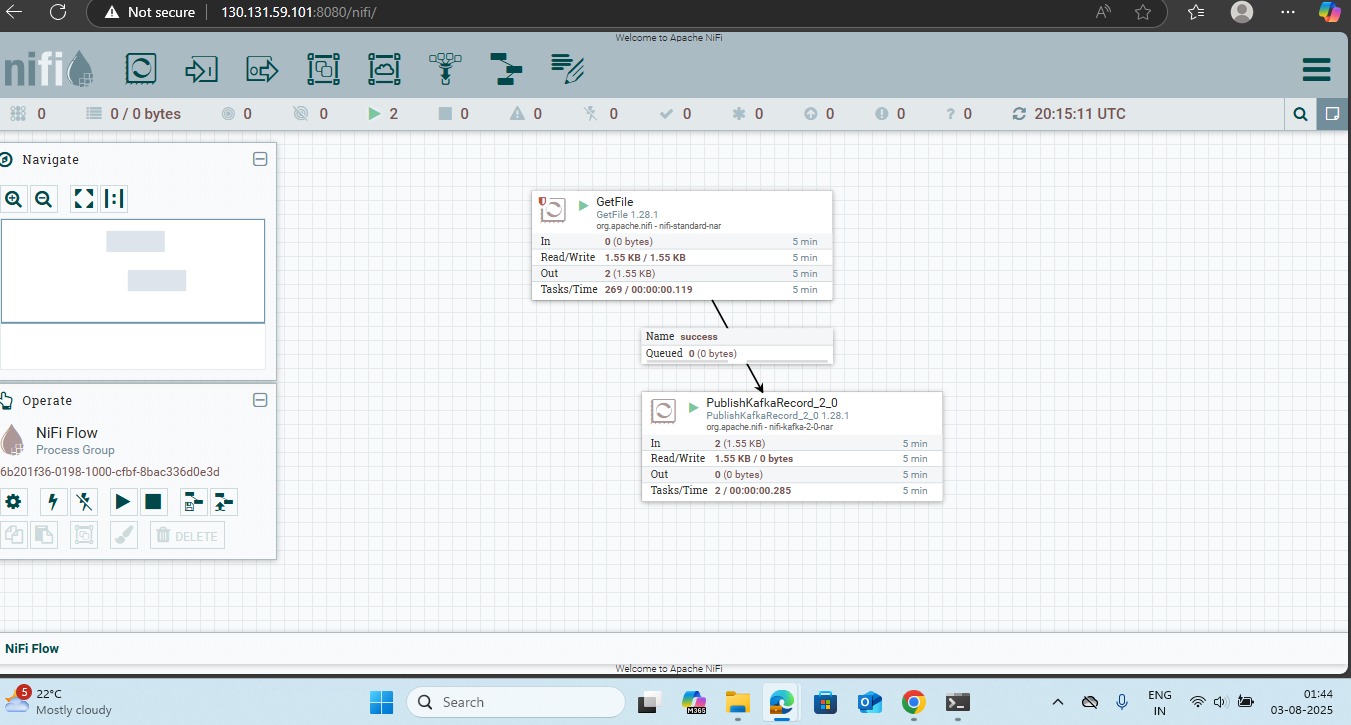
* ***KafkaPublish\_2.0 Config***



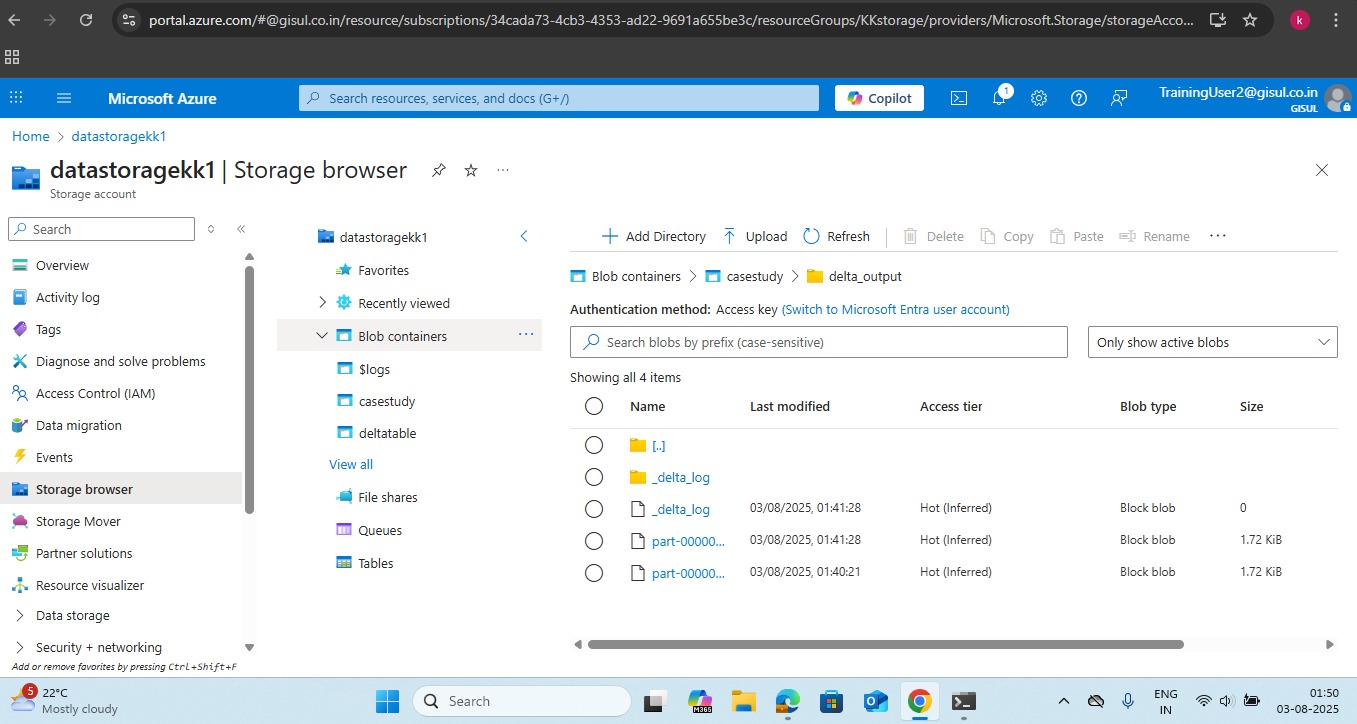
* ***Getfile (input dir config)***



* ***Nifi Ingestor Processor (Getfile --> PublishKafka\_2.0)***



* ***Storage Account***



* ***Storage browser – Bronze layer***

