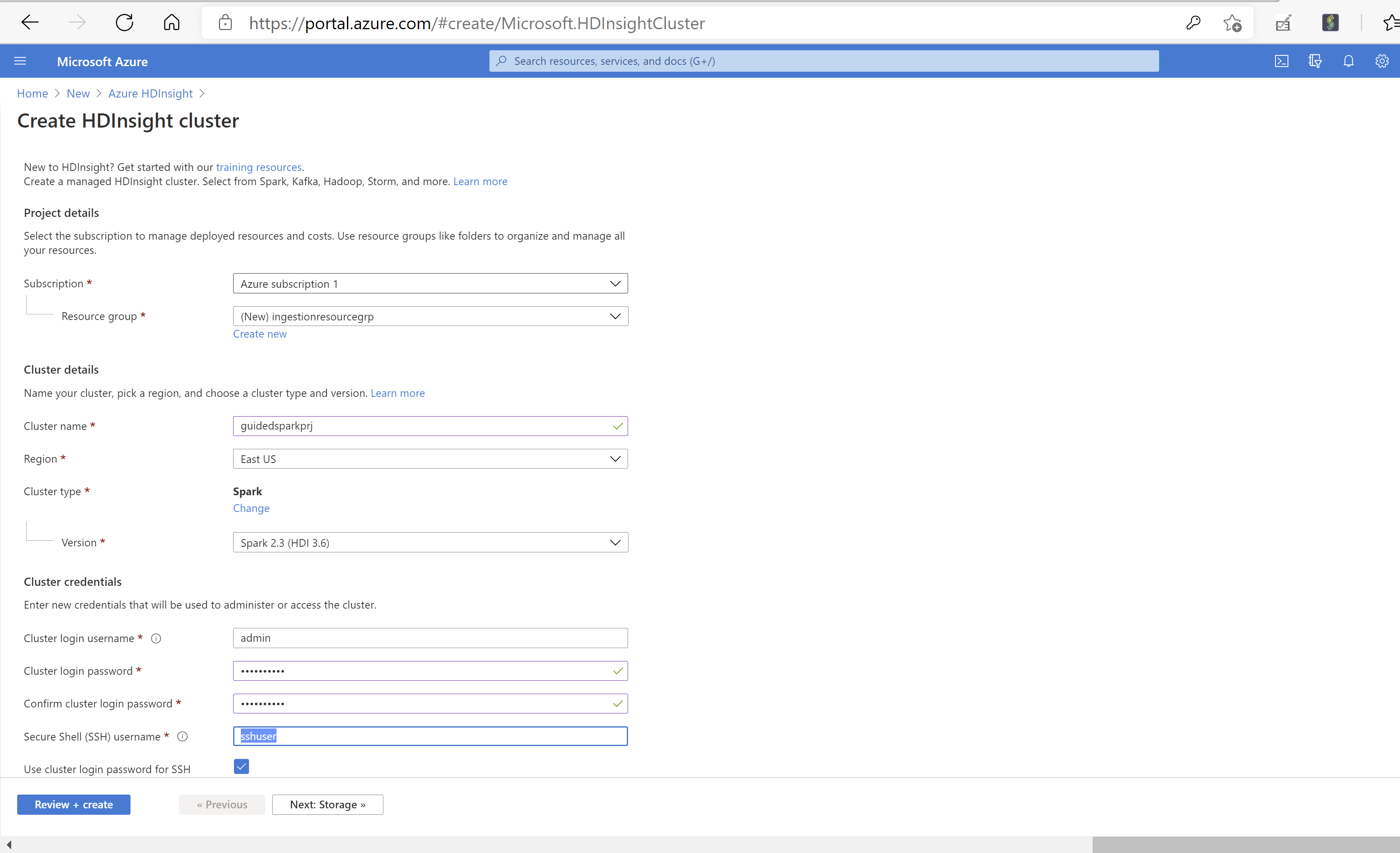
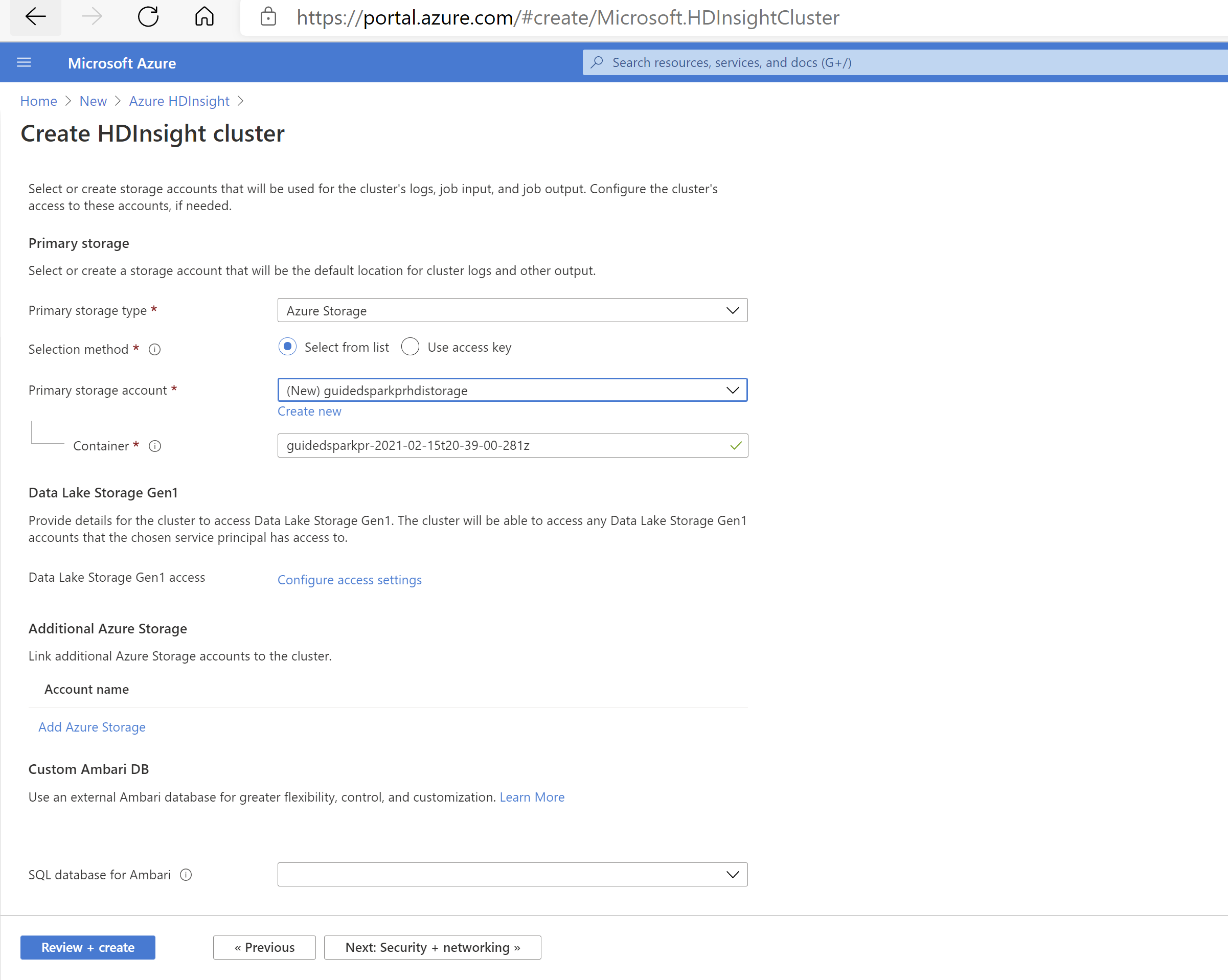
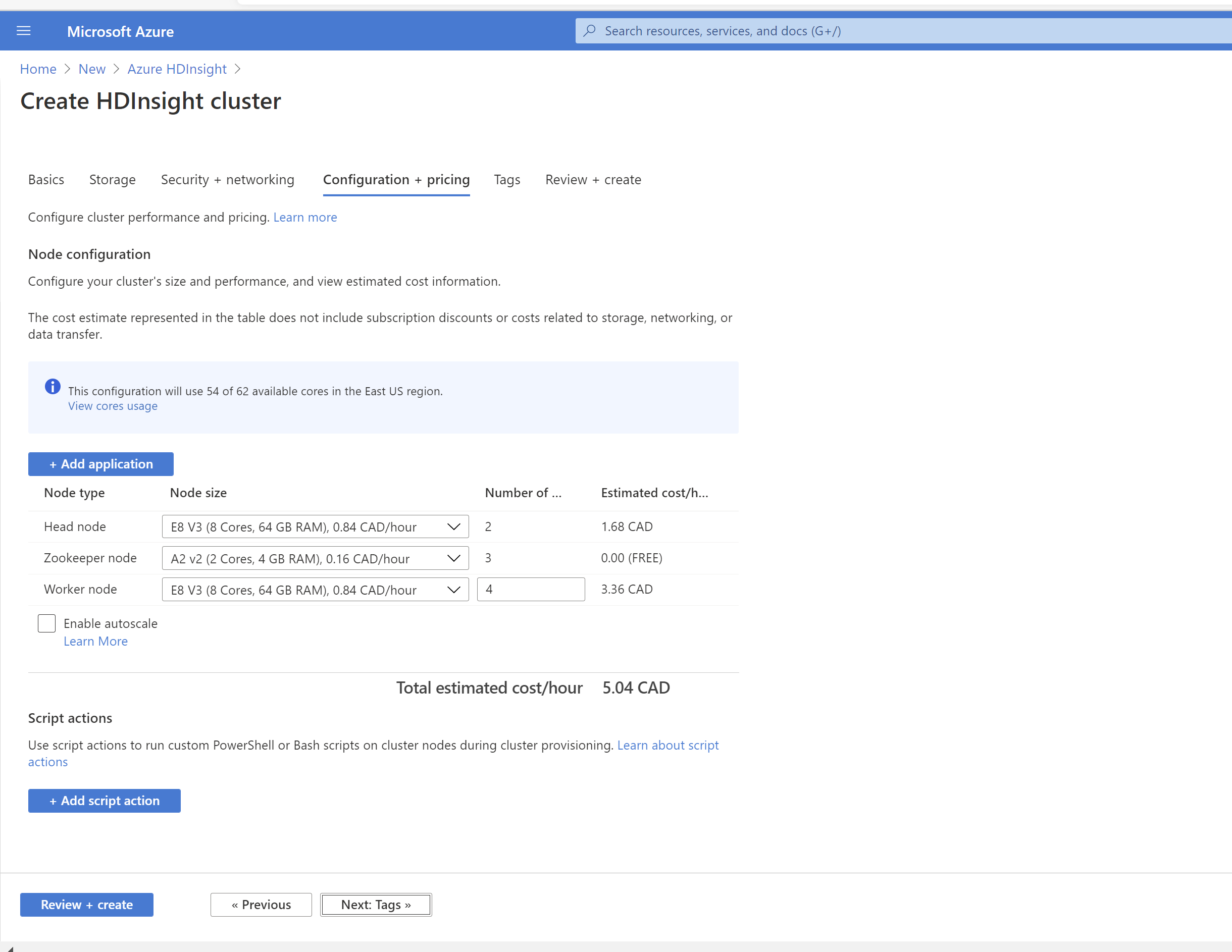
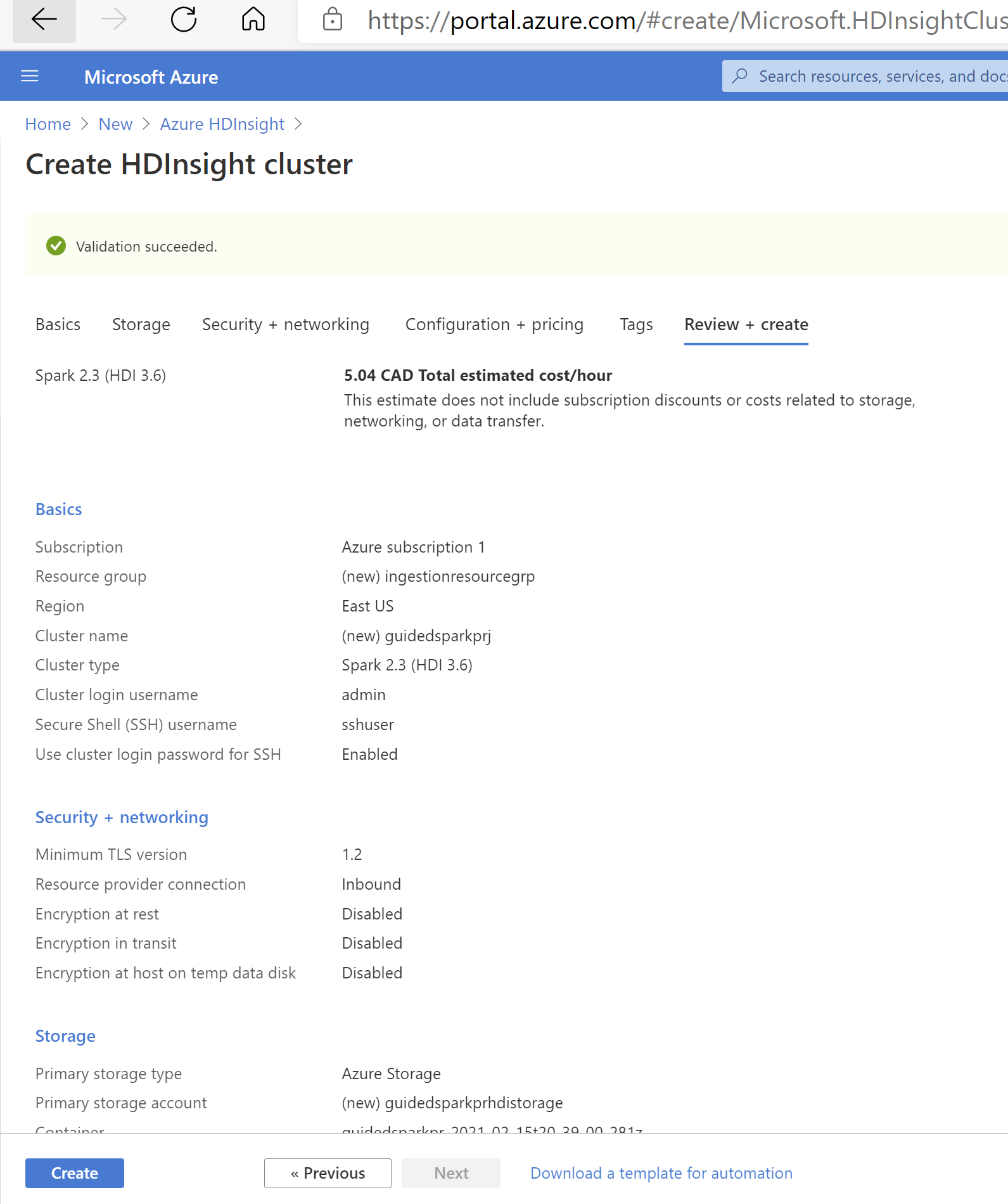
Creating HD Insight Cluster

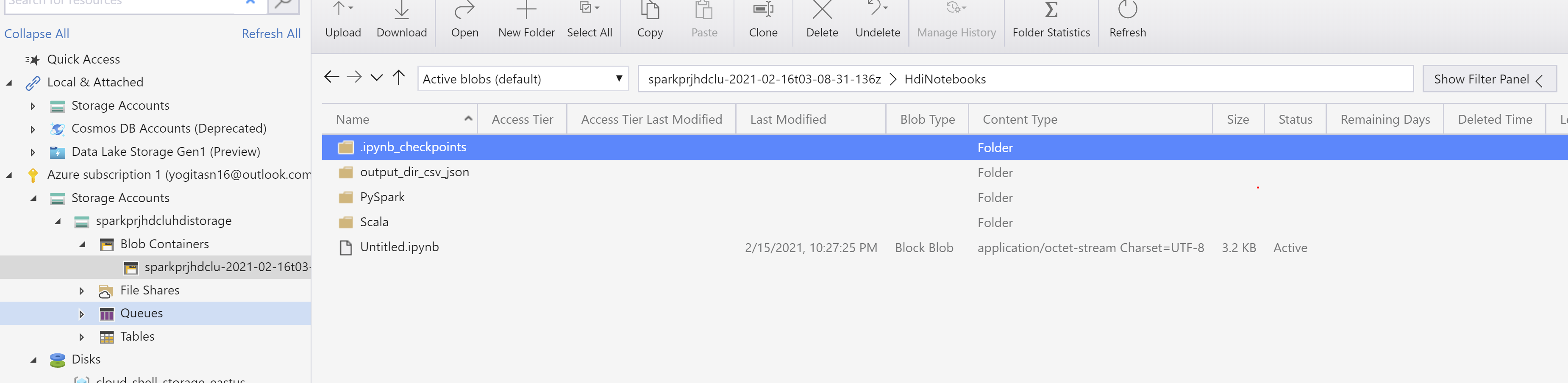




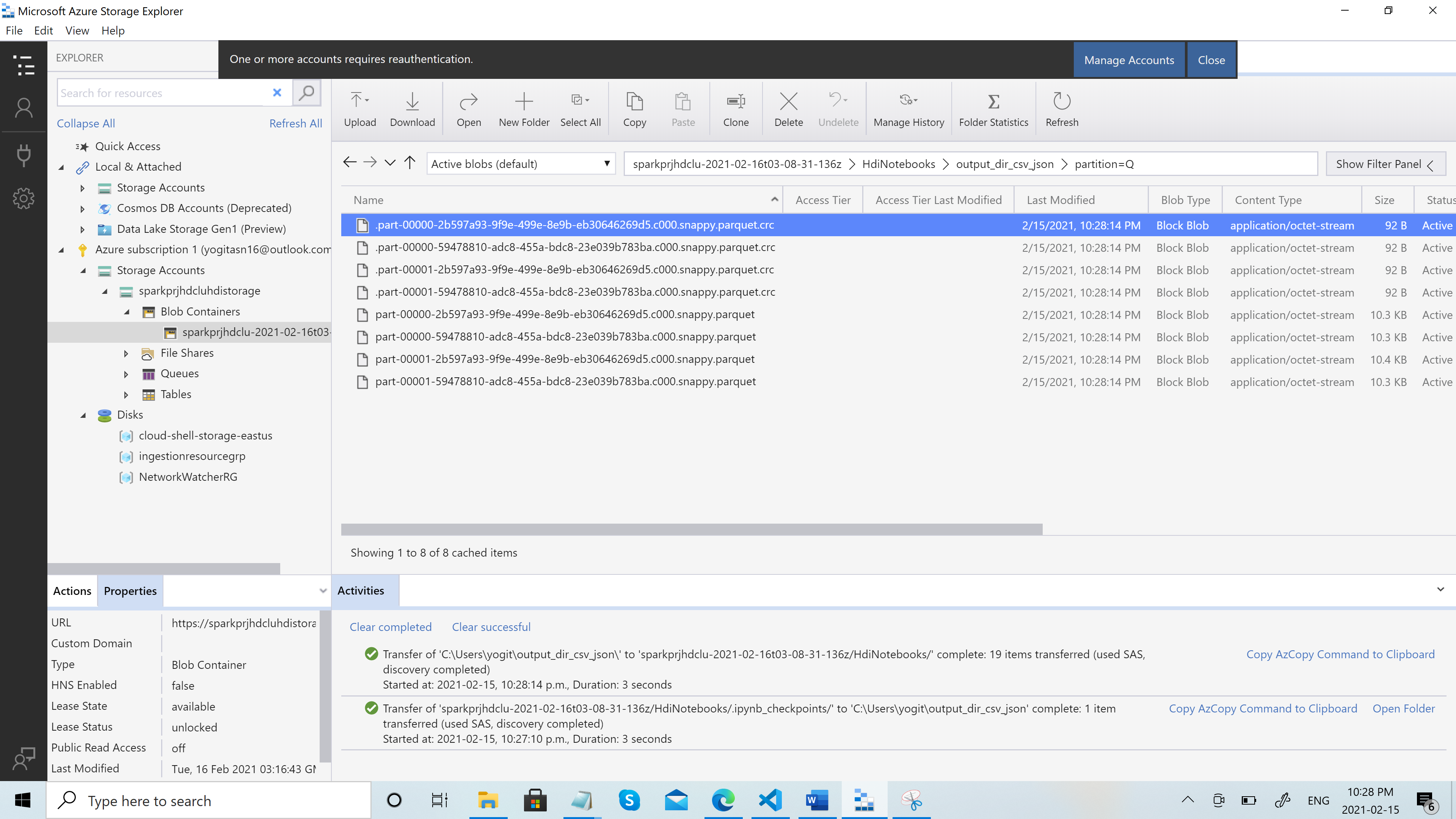




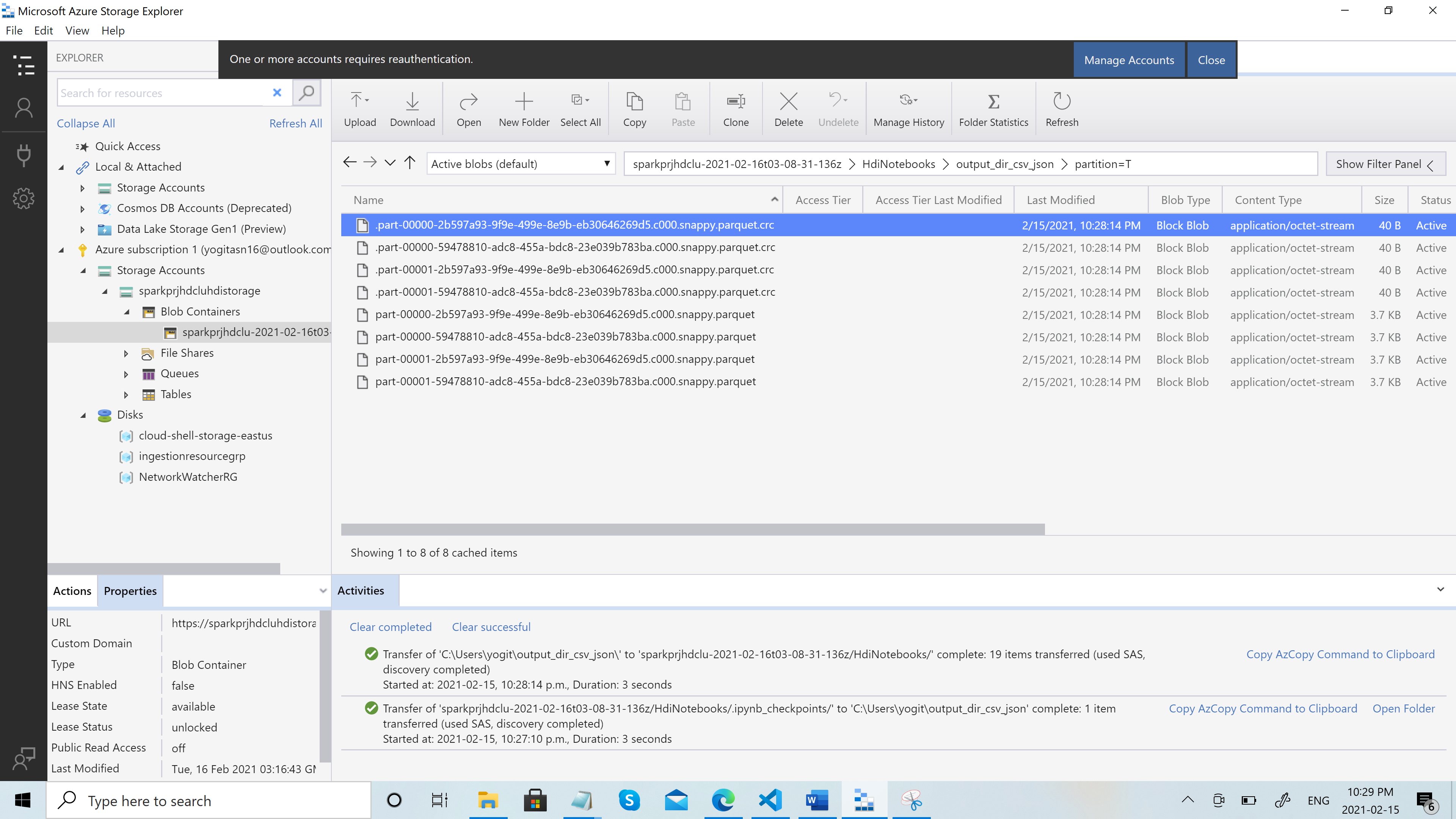
Upload the parquet files created from the last step of Guided Capstone: Data Ingestion process



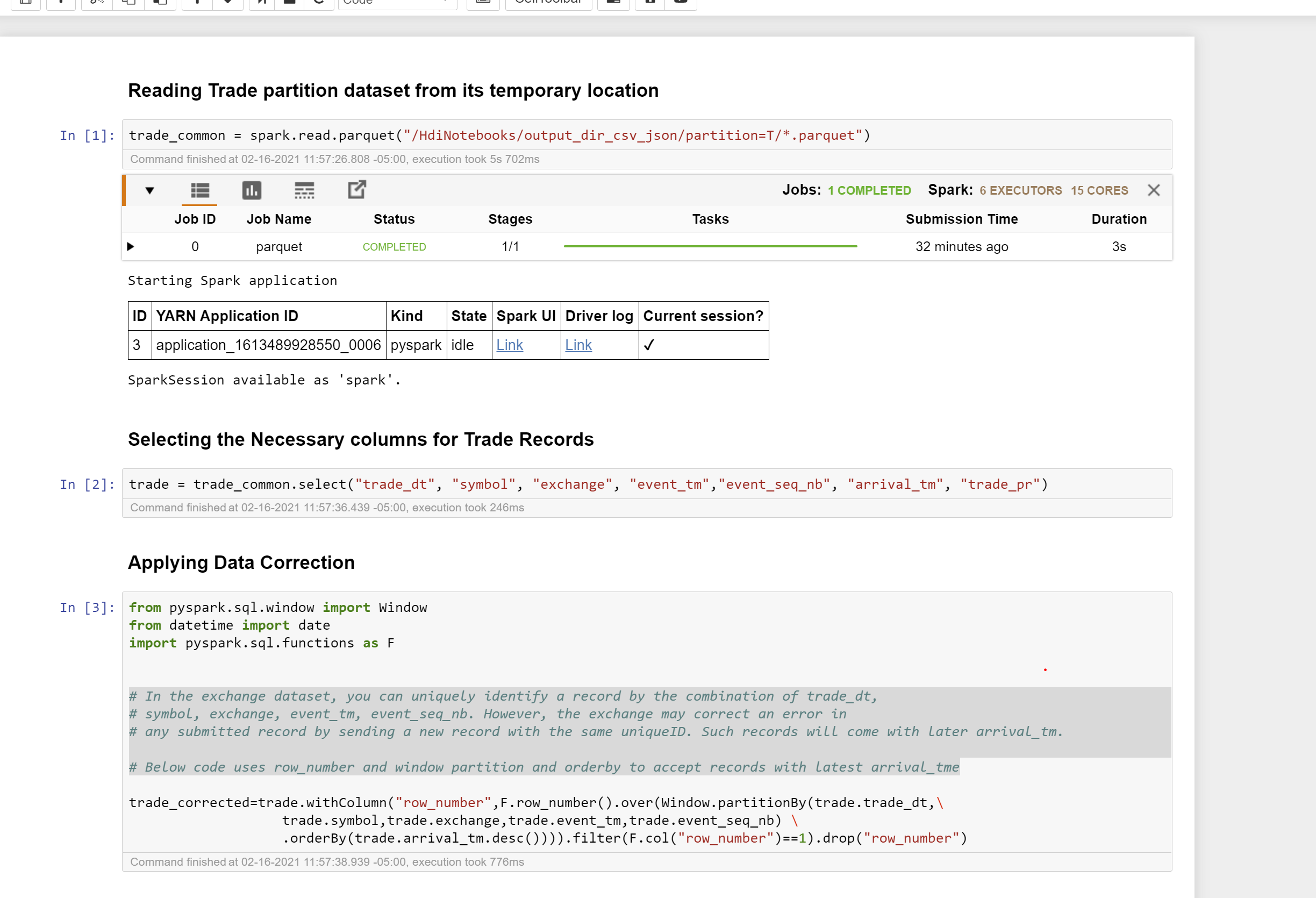
Files in the ‘Q’ partition

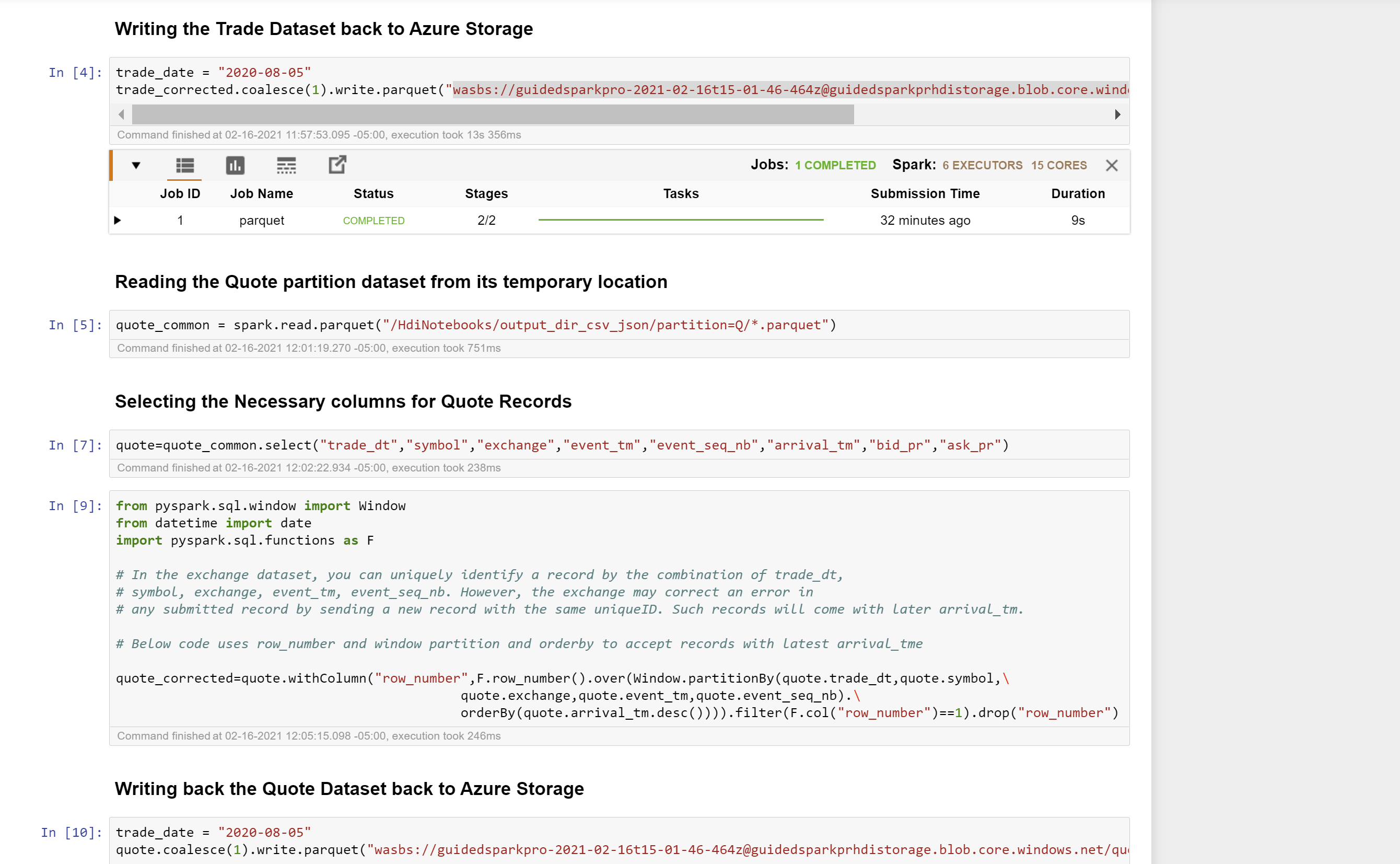


Files in ‘T’ partition

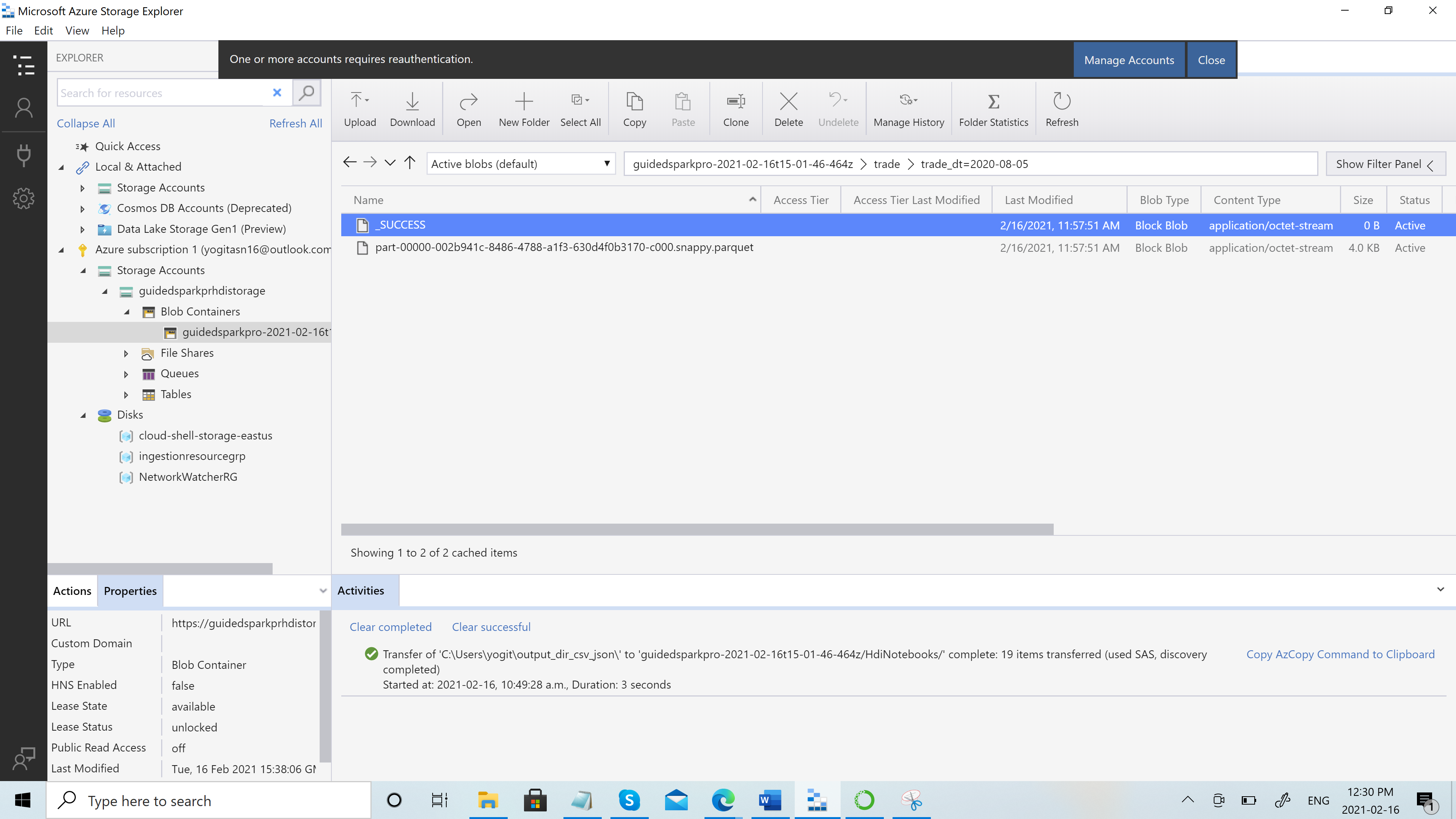


PySpark code in HDInsight that will work on the raw trade and quote records, apply correction, and save it back to Azure storage





Final Trade dataset



Final Quote dataset

