**Delta File Format**

**Introduction to Delta Format**

Delta Lake is an open-source storage layer that brings **ACID transactions**, **scalability**, and **schema evolution** to big data workloads. It extends **Apache Parquet** with **transactional capabilities**, allowing for reliable, scalable, and performant data lake solutions.

**Why Use Delta Format?**

* **ACID Transactions**: Ensures data consistency even in concurrent read/write operations.
* **Schema Evolution**: Allows adding/removing columns without breaking pipelines.
* **Time Travel**: Enables rollback to previous versions of data.
* **Optimized Performance**: Built-in indexing and compaction improve query speeds.
* **Data Reliability**: Supports upserts and deletes, preventing data duplication.
* **Integration with Big Data Tools**: Compatible with **Apache Spark, Databricks, Azure Synapse, Azure Data Factory, and Power BI**.

A screenshot of a computer

Description automatically generated Create a Pipeline

Got to synapse –> Integrate-> Click in + to create a pipeline, then drag and drop dataflow activity. Then click on dataflow debug

Then click on add source and get source data from ADLS.

A screenshot of a computer

Description automatically generated Go to source options and give the file path.

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated Then add derived column to add Hash column then give this hash function expression and click on save and finish.

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated Add alter row activity to add condition for Upsert which can do action for updating and insertion.

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated Publish pipeline and run

A screenshot of a computer

Description automatically generated Go to Data tab and click on ADLS storage and container and go to the folder and right click and select new sql query

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated Now changing file A screenshot of a computer

Description automatically generated Publish and run the pipeline.

Pipeline ran successfully A screenshot of a computer

Description automatically generated Now to go ADLS and right click on folder to get a query to run  
A new record got inserted and old record got updated

A screenshot of a computer

Description automatically generated