**Files:**

Connection String:

DefaultEndpointsProtocol=https;AccountName=asadatalakecql;AccountKey=8W36tQAJ1r8WmEJqz1nkmw+R5GtTrX4FmFD7feqM5JncsoOlzz65jS7Mxen8PPg/UlU5B767aH1f+AStPuvpTA==;EndpointSuffix=core.windows.net

**Data Structure:**

1. **Parquet Files – Sales Data**

The files are in subfolders, the general logic to access the data is presented below:

Sample path to the parquet folder for January 1, 2019:

*/wwi-02/sale-small/Year=2019/Quarter=Q1/Month=1/Day=20190101/sale-small-20190101-snappy.parquet*

Files contain the following information:

[TransactionId]

[CustomerId]

[ProductId]

[Quantity]

[Price]

[TotalAmount]

[TransactionDateId]

[ProfitAmount]

[Hour]

[Minute]

[StoreId]

You can reflect the table structure in SQL by running the following query, or you can manually create such.

CREATE TABLE [Your\_DataBase\_Name].[Your\_Table\_name]

(

[TransactionId] [uniqueidentifier] NOT NULL,

[CustomerId] [int] NOT NULL,

[ProductId] [smallint] NOT NULL,

[Quantity] [tinyint] NOT NULL,

[Price] [decimal](9,2) NOT NULL,

[TotalAmount] [decimal](9,2) NOT NULL,

[TransactionDateId] [int] NOT NULL,

[ProfitAmount] [decimal](9,2) NOT NULL,

[Hour] [tinyint] NOT NULL,

[Minute] [tinyint] NOT NULL,

[StoreId] [smallint] NOT NULL

)

WITH

(

DISTRIBUTION = HASH ( [CustomerId] ),

CLUSTERED COLUMNSTORE INDEX

);

1. **CSV - Customer Information**

There is only one CSV file available under:

*/wwi-02/customer-info/customerinfo.csv*

CSV file contain customer information in the following format:

UserName,Gender,Phone,Email,CreditCard

**TIP:** Will you make use of this data? What benefits (if any) can you see in using those data? What is the risk?

1. **JSON**

There is one JSON file that contains all campaigns:

/wwi-02/campaign-analytics/campaignanalytics.json

TIP: Please note json specific structure:

Text, letter

Description automatically generated

In this case Revenue is 15119 USD – field RevenuePart1 PLUS Revenue and target 17269 USD with the same logic behind. Your pipeline needs to take this into consideration.

You can reflect the table structure in SQL by running the following query, or you can manually create such.

CREATE TABLE [wwi\_mcw].[CampaignAnalytics]

(

[Region] [nvarchar](50) NULL,

[Country] [nvarchar](30) NOT NULL,

[ProductCategory] [nvarchar](50) NOT NULL,

[CampaignName] [nvarchar](500) NOT NULL,

[Analyst] [nvarchar](25) NULL,

[Revenue] [decimal](10,2) NULL,

[RevenueTarget] [decimal](10,2) NULL,

RevenuePart1,Revenue,RevenueTargetPart1,Revenue\_Target

[City] [nvarchar](50) NULL,

[State] [nvarchar](25) NULL

)

WITH

(

DISTRIBUTION = HASH ( [Region] ),

CLUSTERED COLUMNSTORE INDEX

);