# Data Set and How To Retrieve

1. Go to website: <https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings>
2. Download data: Video\_Game\_Sales\_as\_at\_Dec\_2016.csv

# Loading Data Set into HDInsight

1. Load your csv file in HDInsight using the Azure Storage Tool. Put the data in a container called Sales
2. Create a schema for your data from the raw data.

Using your schema, load the data from Azure Storage into Hive. Call the table, “sales”.

1. Validate that the schema is correct using the “table view” tab.
2. Inspect the data were loaded properly using the SQL query interface