**UDF**

UDF’s are user Defined functions which helps with custom processing of data on both PIG and Hive. UDF’s are used for the scenarios where the built in functions cannot be used to solve the use case.

**Hive UDF Example:**

**Example** 1: If a date field is string in a staging table in Hive and we want to create a derived column to mention the "Day (Sunday, Monday etc) against that date a UDF can be written to achieve this.

**Example** 2: IF the data file name as a date and we want to partition the data by the date in the file name, UDF will be required for this.

**Example** 3: Scenario: If a data file has millions of records and the trailer record has a sum check on a particular field. Sum check will be considered pass if there is difference of 0.2% else data load should not be done to the table.

**Example** 4: Scenario: We are sqoop the data from an Oracle source and the source has dates in dd/mm/yy and in hive table we want to save in DDMMYYHHMMSSffffff format. (Add 00 - for hhmmssffffff) UDF needs to be written.

**Example** 5: Scenario: If the data needs to be sorted lexographically (needs to be converted to binary codes and then sorted).

**Example** 6: Scenario: A data file is has Temperature values in Celsius and needs to be converted to Fahrenheit and added as a derived column in the table while uploading data.

**Pig UDF’s**

**Example** 1:

UDF's are used for use cases for image processing.

UDF can be written to convert "Image to a sequence" file and then "Sequence to an image UDF"

**Example** 2:

UDF are used to parse csv files in mainframes based on copybook.

Pig has a built-in loader that can read CSV files, but the parsing has to be done on the CSV in conjunction with the contents of the copybook.

**Example** 3:

Pig is not great at string processing, so small UDF like, stripping quotes “ from the sentence, changing case of strings, are very helpful.