HashedIn University

2022

**ETL Track**

**Talend Final Assignment**

23rd May to 26thth May 2022

Prepared For - Linkers, HU Spark IE 22.2

**Track Lead/Helpers (Point of Contact)**

|  |  |
| --- | --- |
| **Name** | **Email** |
| Dhananjay Varma(HU Director) | dhvarma@deloitte.com |
| Ashik Sunny(Track Lead) | assunny@deloitte.com |
| Sachin Prajapati | saprajapati@deloitte.com |
| Kshitij Choudary | kshichoudhary@deloitte.com |
| Mohan Babu | mohanba@deloitte.com |

**Presented By:**



**Prerequisites**

* Download Input file from below given links**:**

Sales Data : Sales Data.csv

Lookup Data : Lookup.csv

* Java 8 should be installed.
* Talend (version – 7.3 or above) setup should be done (follow talend installation doc)

**OVERVIEW**

1. Read the Data from the input file(Given above) in Talend.
2. Do the following **Filtration**.

* Sales\_Channel should be online
* Order\_Date > 01/01/2012

**Note** ⇒ Create a Reject File for the records which don't satisfy above mentioned filters.

1. Read the **lookup file**(Given above) where lookup Order\_id= sales Order\_ID

**Note** -> Get all common rows ( where lookup order id is same as sales order id) in separate file

1. **Transformations** (Do the following transformations on the data which passes through the filter after joining with lookup) – **HINT** (use TMAP)

* Create two files, one with Order\_Priority = M, L and the other with Order\_Priority = H, C (Note -> both the files will undergo transformations)
* If the country is **Ukraine** make order priority H
* **Percentage Cost of  particular Item type for each country** -

 (Unit Sold \* Unit Cost)/ ( Total Unit Sold \* Total Cost) \* 100 and output should be like “ Percentage cost of type fruit in country India  is  2.32%” where type ,fruit and percentage will be changed accordingly. (**Hint** -> use the concept of global variable to calculate total unit sold \* total cost of particular item type)

* **In unit price column** increase unit price by 50%where region is Asia and item type is fruit  and order date is greater than “29/11/14”
* Total Units sold for each country (**Hint** -> use the concept of global variable)
* Add new column in output as Unique Code of type string and get first three characters from Name and last three characters from region and last three numbers from id
* Append **Name**(from lookup)with **Country** where **Region = Asia**
* Find the total Number of unit sold for each item type (use global Variable)

1. Get all the lookup rejects in a separate csv file.
2. Sort the output files and sort it according to the Unit Sold
3. Merge them and insert them into the Database.

**Project Milestones**:

|  |  |  |
| --- | --- | --- |
| **MILESTONES** | | **Weightage(in Points)** |
| Milestone 1 | * Create and Read Config and read the input file Schema from metadata and path from Context. | 15 |
| Milestone 2 | * Create a Read Reject File. * Filter the data according to the given filters and create the Filter reject in the separate file. * Read the lookup File and get the lookup rejects in a separate file. | 20 |
| Milestone 3 | * Do the transformations of data as mentioned above. * Create a mapping reject file. | 50 |
| Milestone 4 | * Create the output files and sort it according to unit sold. * Merge them and Insert it to the database table. | 15 |

**NOTE**:

* Follow the best practices.
* Give the file path through context.
* Make use of for schemas.
* Give sensible names to the files, metadata, jobs used.
* Make database connection through the m metadata etadata.
* Get the **Output** in a separate output files/folder, **Rejects** in rejects folder and **Success** in success folder