HashedIn University

2022

**ETL Track**

**Talend Mini 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:**



Sales Data: Use the given sales data

Data cleaning and validation:

1. OrderID should be 9 characters long
2. OrderDate should be in between 2010-2019

Transformations:

Add below columns:

1. Total Revenue (Unit Sold \* Unit Price)
2. Total Cost (Unit Sold \* Cost Price)
3. Total Profit (Total Revenue – Total Cost)

Config file:

It will contain below details.

1. Source File Location
2. Staging File Location
3. Target File Location

Steps:

1. Load config file into the Talend context.
2. Create a Talend Job that will load the source data, add validation checks, and filter it out in a CSV file. Call it staging csv file.
3. After filtering out, use the staging CSV file as source and do the above transformations and generate a final CSV file.
4. Test your ETL if it is working as expected.