Million's O' Boxes Data

PROJECT DESCRIPTION

You have been tasked with the structuring and loading of box shipment data found in the attached CSV files. Please note the measures the client has requested, as some of the measures do not appear natively in the files and must be computed. Furthermore, the data will be processed by another user after your ETL, so you will need to provide several CSV outputs (provided) as well as the OLAP design. Additionally, the company currently pays insurance on each shipment, which is computed as the max of the insurrance rates related to the beginstorecd and endstorecd. The company has suggested a plan to implement a region based insurrance payment on all shipments. Your team needs to determine whether the company should switch to the new method of payment or not and provide reasoning for this decision.

OLAP DESIGN

- All fields should be included in your OLAP design.
- The following are to be considered measures from the .csv files:
  - shipments.csv
    - weightofshipment
    - shipmentinsurrancecost (note, this is the max of the rate associated with the begin and end locations as stated in the stores.csv file under the insurrancecost column)
    - basevalue
  - boxes.csv – The following could be modeled in the fact table or a dimension
    - costtomake
- *All dimensions should be treated as a Type 1 for the purposes of this project.*

ADDITIONAL MEASURES NEEDED

- The following are additional measures that should be included and must be computed from other values from the .csv files:
  - currentcostofshipment = ((weightofshipment * .27) + costtomake)
  - valueofshipment = basevalue * 15.08
  - TotalValue = valueofshipment – (currentcostofshipment * shipmentinsurrancecost)
- The following value does NOT needed to be added to the fact table, but will be necessary for answering the business question and creating the ETL outputs
  - TotalValueNewInsurranceMetric = valueofshipment – (currentcostofshipment * estimateinsurrancecost)

ETL OUTPUTS (Note, these are based on sales)

- Unique counts:
  - Unique count of boxes in shipments (boxcd, boxtxt, count_of_instances)
  - Unique count of shipments by region (regioncd, region_name, count_of_instances)
  - Unique count of shipments by store for stores with nightloading (Yes) (storenumber, storetxt, count_of_instances)
  - Unique count of shipments by store for stores with nightloading (No) (storenumber, storetxt, count_of_instances)
- All Data (sorted by date) (DateOfshipment, boxtxt, region_name, storetxt, [All Measurement fields])

DELIVERABLES

To complete this project, you will need to provide the following items:

- (4 points) Word doc containing
  - OLAP Draw.io design
  - ETL Draw.io design
  - Brief 1 paragraph answer to the following question:
    - Which insurrancecost method do you suggest for the organization? (to use the max calculation method as currently done, or the region based estimation). In your answer, you must back-up your suggestion with data. This data can be computed using a group by control to another output in your ETL, or a chart provided in your answer where computations were done in Excel. For the purposes of your computation, you may assume that any issurance cost (no matter the source of the rate) under the max rate method is attributed to the region associated with the beginstorecd field in the shipment.csv file.
- (4 points) An OLAP database created in MySQL.
  - Create an OLAP design with needed foreign keys
  - Submit a database dump file created through workbench.
- (4 points) Pentaho ETL files
  - Include a “job” file and all associated “transformation” files
  - Connection strings & file paths should be set by variable (failure to provide will result in a loss of a letter grade on the assigment.
  - Submit these files via a .zip file.
  - **NOTE, your pentaho project should load the OLAP database AND produce the required flat file outputs (CSVs)**.

NOTES/HINTS

- Output file formats for CSVs are provided. Please use them as a guide for creating your .csv outputs.
- Be careful of your calculations. Remember an Integer * Decimal = Integer.