Seminar 3

In the lecture, you were given an example from industry relating to the implementation of a DBWH (the background is repeated below to refresh your recollection). This seminar is focusing on the challenges of the data transfer from under pinning systems through to the staging areas.

Scenario

Elements is a company that has a number of subsidiary companies underneath the umbrella organisation. The parent company wishes to collate and report on data that is held in each of the child companies and organisations, the aim of which is to identify any inter-company trends using analysis of customer data. In order to carry out this analysis the company has decided to implement a DBWH that will hold the consolidated data and then report on this via a series of data marts that are to be tailored for departmental needs.

Each of the child organisations runs a different SAP/ERP system, there is no common OS or backend system, data is to be extracted by each organisation on a daily basis into a standardised flat file, delimited format which will be transferred to the warehouse server by batch job overnight. These files are then put through a ‘cleansing process’ and transferred to the staging area of the warehouse for further checks before consolidation into the live warehouse.

The file transfer, file validation, data cleansing, data consolidation and transfer to the live area are all to be done automatically with minimal input from any system administrator.

Chromium

Europe

Latex

HO

Flat file

Flat file

Flat file

Flat file

Staging Area

LiveArea

DM

DM

DM

Clean data

Tasks:

Working in groups of 3 or 4 answer the following:

* Based on the scenario above, and what was covered in the lecture, what do you think the primary challenges facing the project are?
* Where does the DBWH get its data from, will an interface to the live system that permits write access be needed?
* What are concerns that should be considered during the file transfer phase, what would the impact be of those concerns being realised and how could they be addressed or minimised
* Considering the data consolidation phase what are the main challenges faced and how would the impact be minimised or addressed.