ETL from Colleague to Data Stores and Warehouse

Colleague Portion:

Within Colleague we have a Data Orchestrator Target file defined for each extract. Each target file has the tables and fields defined that will be extracted from Colleague. In some cases, for the data store extract, there are some filters set for the fields. The tables used by the data stores are divided into separate jobs, two that run on a nightly basis and one that runs weekly. The weekly refresh contains information that does not change as frequently. The extract for the data warehouse has only one job and it is run nightly. The extract for the data warehouse also requires that two processes, UGSF and USCC are run prior to the extract. These processes are for refreshing data in stored computed columns that are used in I-Strategy.

Both processes are automated within Colleague and do not normally require intervention. They are set to run at intervals that are intended to not have overlap with the next process to prevent contention and causing some tables not to be updated. Each morning I check the results to verify that the processes have run okay. Occasionally we get bad birthdays kicking out and those are sent to Enrollment Services to correct. The records will then get processed after they have been corrected.

The purpose for the Data Orchestrator process is to extract the data from the Unidata database and convert it to SQL format and send it to the odware server to be stored in its corresponding table. The field sizes and data types defined in Colleague are used when converting to SQL and the data must meet the SQL format or an error will be kicked out. For example, if the data type is numeric (integer, decimal, etc) and the data coming from Unidata is not numeric, we will get an error or in the case of the bad birthdates, if we get a 3 digit year, it will not convert to SQL and kicks out. We will also get an error if the data coming from Unidata is longer than the defined field size. Unfortunately, Unidata does not restrict the data to the defined size, but it will not work when converting to SQL. For the data stores we have the option to change the data size, but that is not the case for the data warehouse.

SQL Portion:

On the SQL side, the data going into the data stores is updated when the process runs and then is available for Views, Queries, etc to be run against it. Our main use is via Views created in SQL and then used by reports created in Visual Studio and run by users via the SSRS web reporting site.

For the data warehouse, the initial process is the same with the data going into source tables. From there additional processing occurs on the odware server. This processing determines the dimensions and measures and aggregates the information and updates and processes the data in the cubes that are used by I-Strategy. This provides refreshed numbers for the I-Strategy reports that will be available from the Active Campus portal.