Stages in load process

|  |  |  |
| --- | --- | --- |
| Step | What it did | What to do if interrupted |
| Load into ETL\_Staging | There was an SSIS package that looped through a control table that held the list of source directories. Each source was checked to ensure all the files were there and they were recent enough. If not that source would be skipped and an error logged. If they were okay then the files would be loaded. If a file failed as it was loaded then the whole source would be rolled back (ie that source would be removed from ETL\_Staging.) Again an error would be logged.  The first step in this process was to runcate ETL\_Staging  Once a source's files had been successfully loaded, then prev was put on the end of the file to indicate that the files had been loaded.  This was pretty quick, about 15 mins. | If this step failed then do the following:  This was a config table that could be used to control whether the step truncated ETL\_Staging tables or not. Turn this on so the tables weren’t truncated and then start the load process from this step. Once this step was complete, you need to remember to turn the switch back off. |
| Load from ETL\_Staging into DW | This was done as a package. A stored proc existed for each table. Because there is no referential integrity between the fact and dimension tables, it didn’t matter what order this was done in. Each table was logged. Also, logging in the stored proc for table that took a long time.  Once all the table were loaded into, some maintenance was done such as identifying the current record and recording where there was fact tables without dimension record.  This was reasonably quick, about 1 hour | This step was designed to be repeatable. So if it failed, it could be run again. Because each table was checked to see if it had changed records, the same data wouldn’t be loaded again. The only thing you had to remember to do was run the IsCurrent stored proc if it hadn’t been run. |
| Load from DW into Presentation | This was done as a package that again ran stored procs. It was time consuming as it implemented a lot of logic. This could take 4-6 hours.  Presentation tables were truncated before being loaded into. This truncation happened just before they were ready to be loaded into. So the table were only depopulated for a very short time (generally seconds) so users would not be impacted | If interrupted then we would probably just not run it again that day. Users would not have data for that day but that was generally not an issue. It could be re-run but would have to do it from the start of the step. |