

The background is a dark blue gradient with a subtle pattern of white dots. Overlaid on the left side are several concentric circles and arcs, some with degree markings ranging from 140 to 260. Some of these circles have arrows indicating a clockwise direction.

EXTRACT, TRANSFORM & LOAD

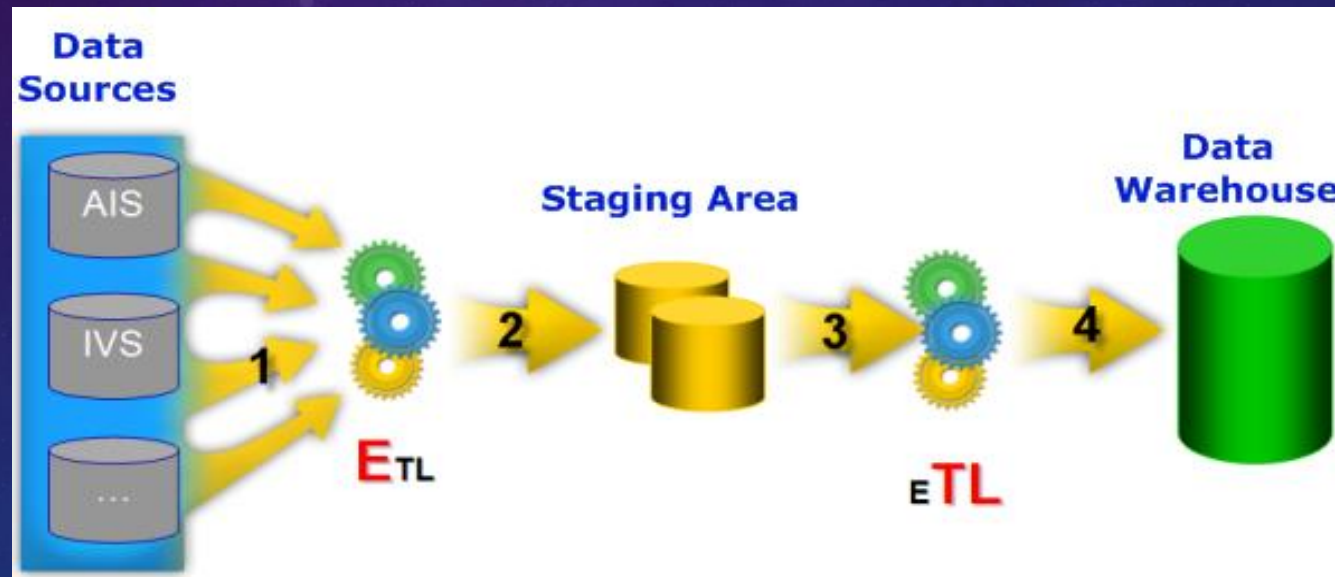
GAP ANALYSIS

ASSIGNMENT 5

MOMODOU LAMIN KEITA

WHAT IS EXTRACT TRANSFORM & LOAD(E.T.L)

- E.T.L stands for extraction, transformation and loading. It is a process in data warehousing responsible for pulling data out of the source systems and placing it into a data warehouse. All three(3) steps in the process work in unison as one tool.



WHAT IS E.T.L?

- {Extract} :extracting data from a source SAP, E.R.P, operational systems or archive systems which are the primary source of data for the data warehouse.
- {Transform} : which may involve joining together data from multiple sources, cleaning, filtering, validating, and applying business rules
- {Load}: Loading the extracted and transformed data into the data-warehouse

E.T.L TOOLS

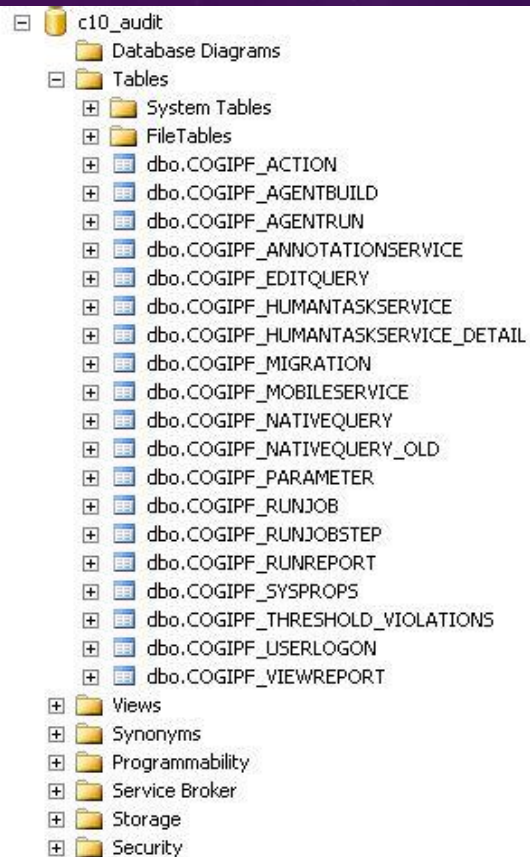
- IBM Websphere DataStage (Formerly known as Ascential DataStage and Ardent DataStage)
- Informatica Power Center
- Oracle ETL
- Ab Initio
- Pentaho Data Integration - Kettle Project (open source ETL)
- SAS ETL studio
- Cognos Decision stream
- Business Objects Data Integrator (BODI)
- Microsoft SQL Server Integration Services (SSIS)

BUSINESS REQUIREMENTS

Design and build a warehouse to house measures and facts of business intelligence log database. The data source is a single database and the resulting warehouse should house measures and facts that allows the client to build reports That:

- Determine which reports are used the most and which are not being used over time
- Identify year over year usage trends
- Determine the time of day and/or day of week that the system is used the most
- Determine how much each user runs reports and when they run them

DATA SOURCE – C10_AUDIT



The tables required to meet the business requirements have been identified. `dbo.COGIF_RUNREPORT` is the driving table. The tables `dbo.COGIF_PARAMETER` and `dbo.COGIF_USERLOGON` provide additional data for building measures for our Dimensions.

DESTINATION - REPORT_AUDIT



The destination ware-house contains four(4) dimension tables , a fact single fact table, a table to store rows of data that error out when building the fact table and another table that supports the fact table.

- DimDate
- DimPerson
- DimReport
- DimTime
- FactRunReport
- FactReportRun_LookUpErrors
- FactReportRun_MaxDateTime

EXTRACTION AND TRANSFORMATION

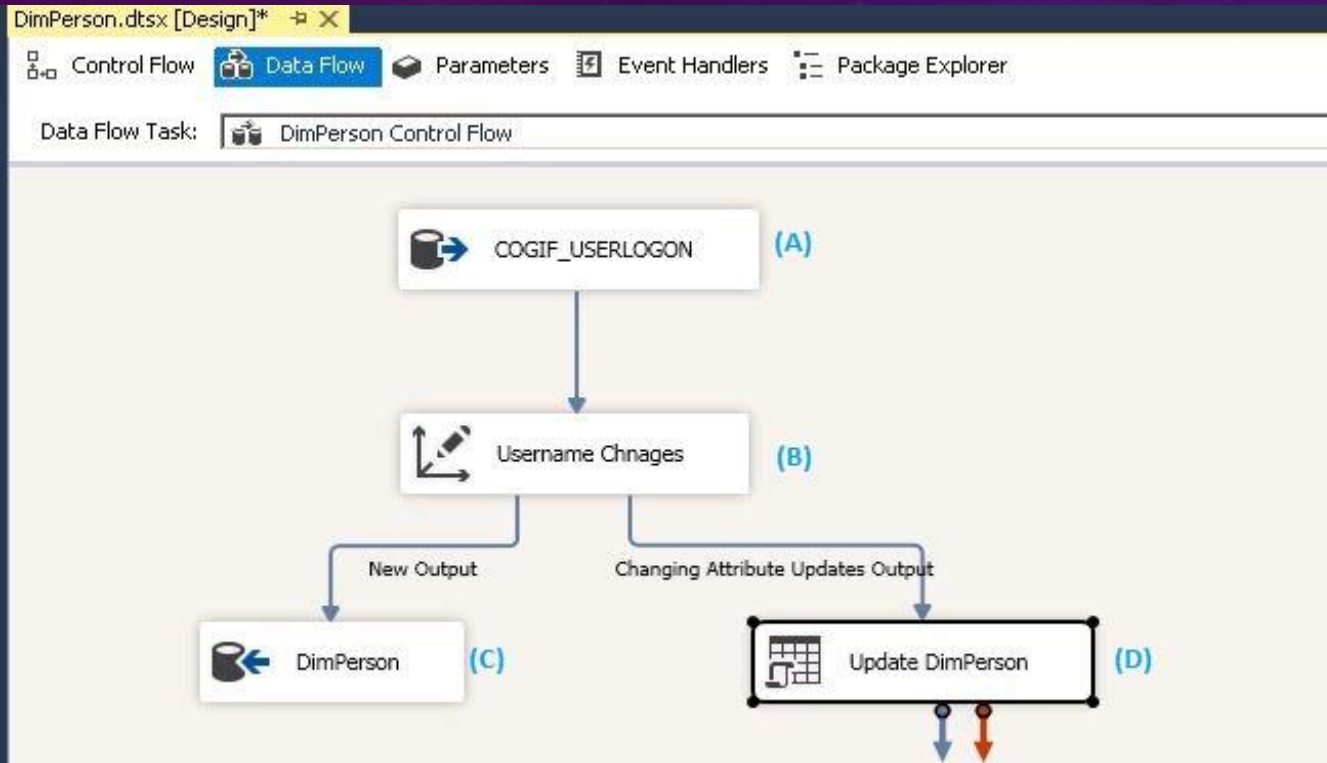
DimPerson

DimReport

DimTime

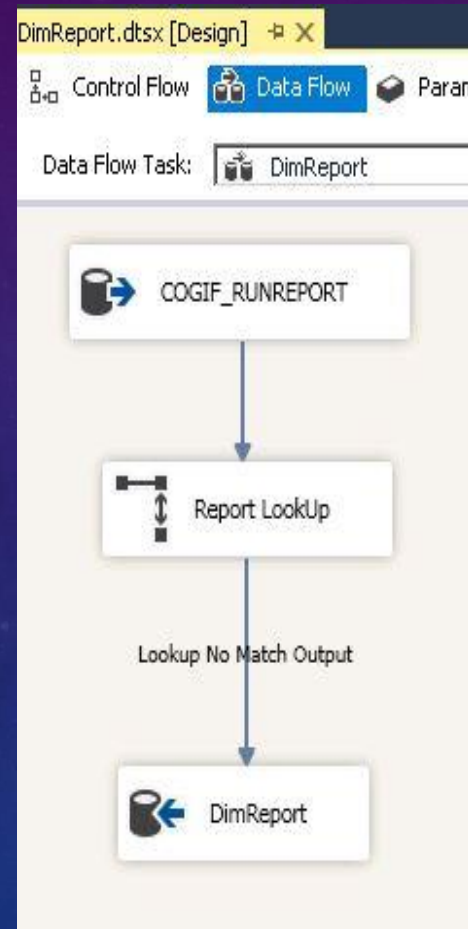
FactRunReport

EXTRACT TRANSFORM AND LOADING WITH SSIS

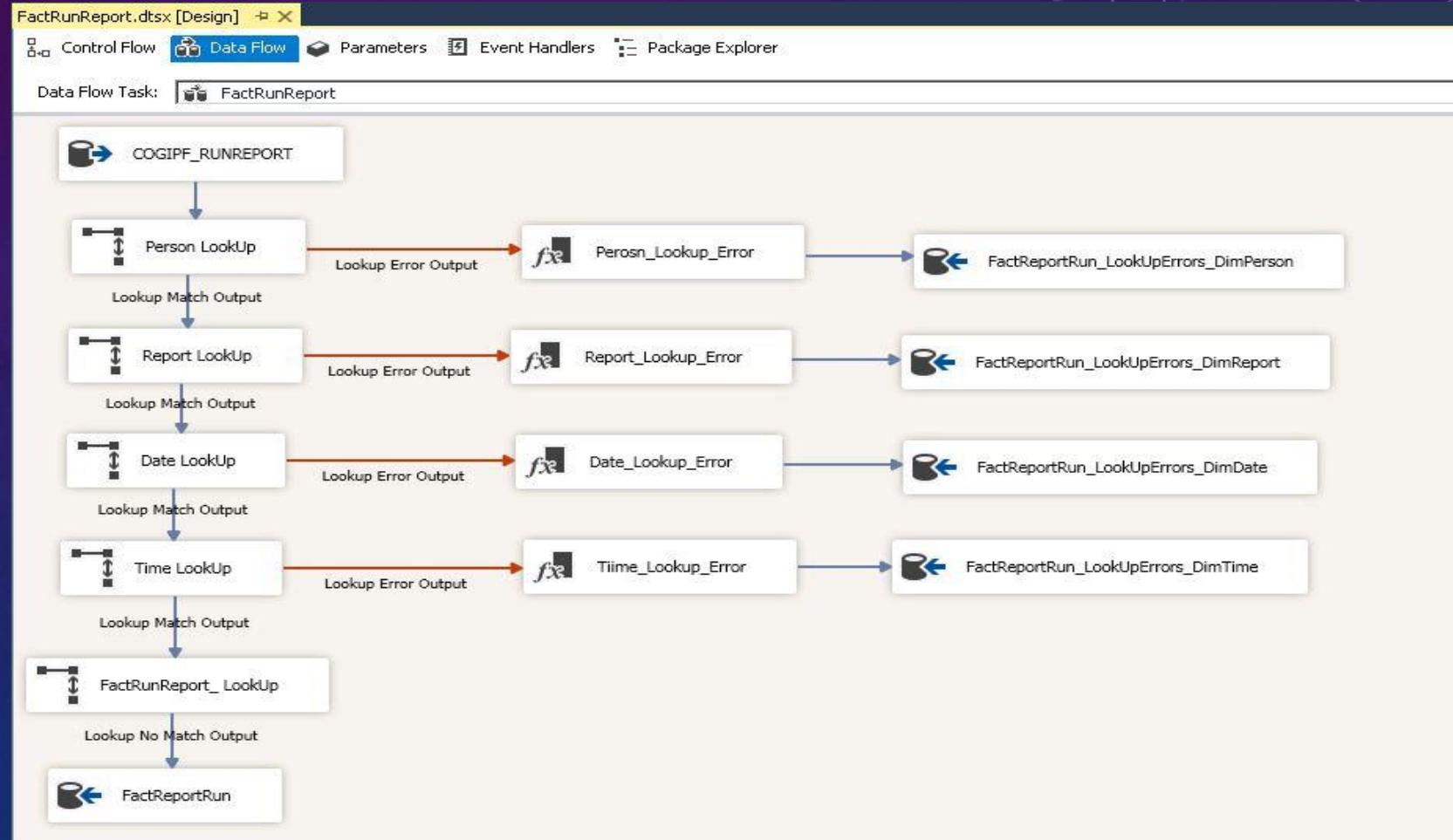


DimPeron DataFlow

DIMREPORT SSIS PACKAGE



FACTRUNREPORT



MASTER RUNREPORT

