Low Level Design (ETL)

INDUSTRY ANALYSIS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Prepared by/ Modified by | | Role | | Start Date – End Date |
| Vanshaj Bhatia | | Team Lead | | 10 Sep 2018 –  20 Sep 2018 |
| Sunny Arora | | SME | |  |
| Santosh Kumar Neela | | Point of Contact | |  |
| Vibhor Gupta | | Designer | |  |
| Nellaiappan,Krubhakar | | ETL Developer | |  |
| Manoj Kumar Rathore | | Data Modeler | |  |
| Reviewed by | | Role | | Date of Review |
| Bhagyashri Gadkari | | Trainer | | 17th Sep,2018 |
| Approved by | | Role | | Date of Approval |
| Shraddha Jadhav | | Batch Mentor | | 20th Sep,2018 |
| Circulation List | To all the ELTP mentors and the applicable trainee teams | | Version number of the template | **1.0** |
| **Version number of the work product** | V1.0 | | | |

Table of Contents

[1. Introduction 5](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304227987)

[1.1 Description of Component 5](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304227989)

[1.2 Document scope 5](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304227990)

[1.3 Dependencies/Assumption 6](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304227991)

[2. Extract Design 8](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304227992)

[3. Tables 15](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304228000)

3.1 [Source Table 15](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304228002)

3.2 [Target Table 16](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304228003)

[4. Sample Data 18](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304228004)

[5. ETL Maps 18](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304228006)

[6. Workflow/ Session Settings 18](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304228009)

[References 19](file:///\\172.21.17.1\Lab5_Informatica\LLD\LLD-ETL.docx#_Toc304228010)

# INTRODUCTION

This document will address the design of the ETL processes to extract data from the Extractstaging environment, transform it and load to the Load staging environment.

#### 1.1 Description Of The Component

ETL stands for Extract Transform and Load. This is a process of updating the data-warehouse. ETL is the automated and auditable data acquisition process from source system that involves one or more sub-process of data extraction, data transportation, data transformation, data consolidation, data integration, data loading and data cleaning. Here, we are importing data which is in the .CSV format from data source to Ab Initio. We will load FACT table using Ab Initio tool.

# 1.2 Document Scope

This document will cover the technical design details related to creating maps and configuring workflows and sessions for given subject area.

This document also includes

* The data source(s) for the target table(s)
* Join logic between source tables
* The structure of the target table(s)
* The ETL maps that will extract the data from the **EXTRACT**  staging environment to the **LOAD** staging environment for final load.

# 1.3Dependencies/Assumptions

No provision shall be given to dynamically change the number of records in source files. Any authorized user will able to modify records for an employee.

**Assumption:**

* It is assumed that the data present in the database has been verified
* Records can be stacked, i.e. you can run multiple times from **EXTRACT** to **LOAD** Schema environments before loading to the final dimension table

# Extract Design

The extract design for Industry Analysis is a straightforward extract, transform, load process.

ETL Steps for **Fact\_Avg\_Sal & Fact\_Industry\_Analysis** is as follows.

1. Import following flat files into Ab Initio:

* Scr01
* Scr02

1. Create Fact\_Avg\_Sal and Fact\_Industry\_Analysis as target table in target designer.
2. Create one to one mapping to load data from source files to the target table.

# SOURCE fILES

## 

## The following source system tables are required.

|  |  |
| --- | --- |
| **Source Tables** | **Schema** |
| Scr01 | Extract |
| Scr02 | Extract |

# Sample

The following are sample records of how the Fact records should look when populated.

# ETL Map

| **Map Name** | **Source Tables** | Target Tables |
| --- | --- | --- |
| **Component Type** |  | **Description** |
| m\_dim\_course\_faculty\_lot\_time | TIME | TIME |
| m\_dim\_course\_faculty\_lot\_time | COURSE | COURSE |
| m\_dim\_course\_faculty\_lot\_time | FACULTY | FACULTY |
| m\_dim\_course\_faculty\_lot\_time | LOT | LOT |
| m\_dim\_batch | BATCH | BATCH |
| m\_dim\_employee | EMPLOYEE | EMPLOYEE |
| m\_fact\_marks | FACT | FACT |

# WorkflowS

|  |
| --- |
| **Workflow Names** |
| wf\_ilearn\_dim\_load  wf\_ilearn\_fact\_load |

# 

# SESSion TaSKS

| **Map Name** | **Task Name** |
| --- | --- |
| **Component Type** |  |
| m\_dim\_batch | tsk\_ilearn\_dim1\_batch |
| m\_dim\_employee | tsk\_ilearn\_dim2\_employee |
| m\_dim\_course\_faculty\_lot\_time | tsk\_ilearn |
| m\_fact\_marks | tsk\_ilearn\_load\_fact |

# Reference Documents

|  |  |
| --- | --- |
| **1** | Reeder, J. Templates for Capturing Business Rules*,* Business Rules Community, April 2001 |
| **2** | Ross, R. Principles of the Business Rule Approach*,* Addison-Wesley, Boston, 2003 http://www.information-management.com/issues/20070401/1079638-1.html . |
| **3** | Debevoise, T. Business Process Management, witch a business rules approach, Implementing the Service Oriented Architecture, ISBN 978-4196-7368-9, 2007 |
| **4** | Von Halle, B., Business Rules Applied, John Wiley & Sons, Inc., 2002 |
| **5** | Holden, G. Reactive and Proactive Business Intelligence, 5 September 2007, http://www.beye-network.co.uk/view-articles/5899 , |

# Revision History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Date** | **Version #** | **Section/Page # changed** | **Details of changes made** |
| 1. | 17th Sep 2018 | V1.0 | URS DOCUMENT |  |
| 2 | 19th May 2014 | V1.1 | HLD DOCUMENT |  |