Crime Analysis

Project #: 001

**TECHNICAL REQUIREMENT DOCUMENT**

**PREPARED BY:**

**Jaikit Jilka**

**Index**

**Table of Contents**

1. Introduction-------------------------------------------------------------------------------------------3

1.1 Scope-----------------------------------------------------------------------------------------3

1.2 Objective-------------------------------------------------------------------------------------3

1.3 Audience-------------------------------------------------------------------------------------3

2. System Overview-------------------------------------------------------------------------------------4

2.1 Domain Name-------------------------------------------------------------------------------4

2.2 Database Management System Configuration-------------------------------------------4

2.3 Database Support Utilities-----------------------------------------------------------------4

2.4 ETL Software Used-------------------------------------------------------------------------4

2.5 Point of Contact-----------------------------------------------------------------------------5

3. Database Design--------------------------------------------------------------------------------------6

3.1 Logical Design-------------------------------------------------------------------------------7

3.2 Physical Design------------------------------------------------------------------------------7

4. Architecture-------------------------------------------------------------------------------------------9

4.1 Overview of ETL----------------------------------------------------------------------------9

4.2 ETL Logic------------------------------------------------------------------------------------10

5. Reporting-----------------------------------------------------------------------------------------------11

# Introduction

.

The purpose of this document is to provide technical specification for ABC Analyst pvt. ltd. This document will have database logical design as well as physical design.

* 1. **Scope and Purpose**

The project will provide our clients with analysis of crime and fines received for a particular crime. This will help our clients to know which city has more crime rate and can take specific action it.

* 1. **Objectives**

The objective of the project is to provide information regarding crime that has been committed in various cities and will also provide how many fines has been collected for a particular crime. The objectives will be achieved if we complete various milestones on time and successfully.

* 1. **Audiences**

The audience to the system will be the Sponsors for this project, police department.

|  |  |  |
| --- | --- | --- |
| **Sr.No** | **Audience Roles** | **Names** |
| 1. | Owner | Sam Billings |
| 2. | Project Manager | Jaikit Jilka |
| 3. | Crime Analyst | Ashwajeet Gajre |
| 4. | Fines Analyst | Yu Fan, Hamad Amlutairi |

**2. System Overview**

**2.1 Domain Name and Details:**

|  |  |
| --- | --- |
| **System Overview** | **Details** |
| Project Manager | Mr Jaikit Jilka |
| System Name | Crime Analysis |
| System Type | Dataware house and BI Application |
| Operational Status | Operational/ in development/ under modification |

**2.2 Database Management System Configuration:**

The vendor, version and targeted hardware for the database management system are specified below.

|  |  |  |  |
| --- | --- | --- | --- |
| Vendor | Hardware | Version | Comments |
| Oracle | Server | -- | -- |

**2.3 Database Support Utilities:**

|  |  |  |  |
| --- | --- | --- | --- |
| Vendor | Product | Version | Comments |
| Oracle | MySQL | -- | -- |

**2.4 ETL Software Used:**

|  |  |  |
| --- | --- | --- |
| Product | Version | Comments |
| CloverETL | 4.1.3 | Tools Used for ETL Purpose |

**2.5 Point Of Contact:**

|  |  |  |
| --- | --- | --- |
| **Sr.No** | **Contact Person** | **Role** |
| 1 | Mr. Jaikit Jilka | Project Manager, ETL Developer |
| 2 | Mr. Ashwajeet Gajre | BI developer, Database Administrator |
| 3 | Mr. Hamad | CEO, System Analyst |
| 4 | Yu Fan | Business Sponsors |

**3. Database Design**

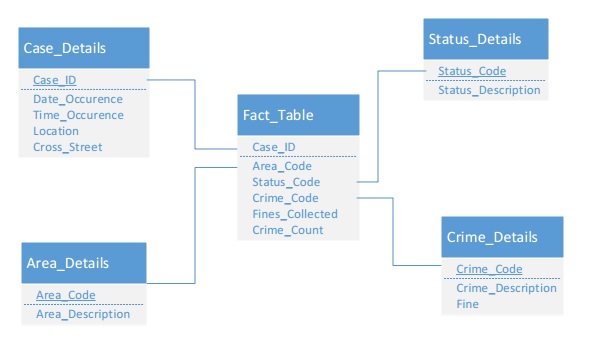
Database design is the process of producing a detailed [data model](https://en.wikipedia.org/wiki/Data_model) of a [database](https://en.wikipedia.org/wiki/Database). This [data model](https://en.wikipedia.org/wiki/Data_model) contains all the needed logical and physical design choices and physical storage parameters needed to generate a design in a [data definition language](https://en.wikipedia.org/wiki/Data_definition_language), which can then be used to create a database. A fully attributed data model contains detailed attributes for each entity.

The term database design can be used to describe many different parts of the design of an overall [database system](https://en.wikipedia.org/wiki/Database_system). Principally, and most correctly, it can be thought of as the logical design of the base data structures used to store the data. In the [relational model](https://en.wikipedia.org/wiki/Relational_model) these are the [tables](https://en.wikipedia.org/wiki/Database_table) and [views](https://en.wikipedia.org/wiki/Database_view). In an [object database](https://en.wikipedia.org/wiki/Object_database) the entities and relationships map directly to object classes and named relationships. However, the term database design could also be used to apply to the overall process of designing, not just the base data structures, but also the forms and queries used as part of the overall database application within the [database management system](https://en.wikipedia.org/wiki/Database_management_system) (DBMS).

The process of doing database design generally consists of a number of steps which will be carried out by the database designer. Usually, the designer must:

* Determine the data to be stored in the database.
* Determine the relationships between the different data elements.
* Superimpose a logical structure upon the data on the basis of these relationships

**3.1 Logical Diagram:**

****

**3.2 Physical Design**

**Source File:**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data type** | **Description** |
| **Case\_ID** | Integer | Unique case id for each case. |
| **Date\_Reptd** | Date | Date when the crime was reported. |
| **DATE.OCC** | Date | Date when the crime actually committed. |
| **Area** | Int | Area code. |
| **Area.Name** | Varchar | It is the name of area. |
| **Crm.cd** | Int | Code of a particular crime |
| **Crm.cd.desc** | Varchar | Description of crime |
| **Status** | Varchar | Status code |
| **Status.desc** | Varchar | Description of the status |
| **Location** | Varchar | Location name |
| **Cross Street** | Varchar | Name of the cross street where crime was committed. |

**Destination Tables:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table Name** | **Fields** | **Data Types** | **Constraint** | **Description** |
| **Status\_Details** | Status\_Code | Varchar(5) | Primary Key | It has various status code and its description |
| Status\_Description | Varchar(50) |  |
|  |  |  |
| **Area\_Details** | Area\_Code | int(10) | Primary Key | It has various area code and its description |
| Area\_Description | Varchar(50) |  |
|  |  |  |
| **Case\_Details** | Case\_Id | Int(10) | Primary Key | It has all the information regarding a case. |
| Date\_Reported | Date |  |
| Date\_Occurence | Date |  |
| Location | Varchar(50) |  |
| Cross\_Street | Varchar(100) |  |
|  |  |  |
| **Crime\_Details** | Crime\_Code | Int(10) | Primary Key | It provides crime code and what is fine for that crime. |
| Crime\_Description | Varchar(100) |  |
| Fines | Int(10) |  |
| **Fact\_Details** | Status\_Code | Varchar(25) |  | This is the fact table and used to calculate fine |
| Area\_Code | Int(20) |  |
| Case\_ID | Int(20) |  |
| Crime\_Code | Int(20) |  |
| Fine | Int(20) |  |
|  |  |  |

**4 Architecture**

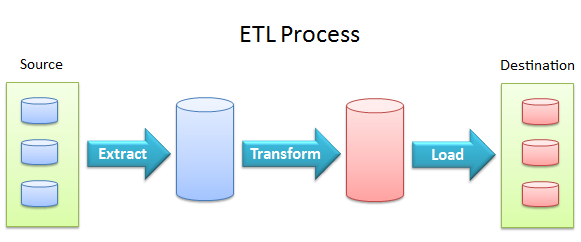
**4.1 Overview of ETL in Data warehouse**

**Introduction:**

ETL is short for extract, transform and load three [database](http://www.webopedia.com/TERM/D/database.html) functions that are combined into one tool to pull data out of one database and place it into another database.

* Extract is the process of reading data from a database.
* Transform is the process of converting the extracted data from its previous form into the form it needs to be in so that it can be placed into another database. Transformation occurs by using rules or lookup tables or by combining the data with other data.
* Load is the process of writing the data into the target database.

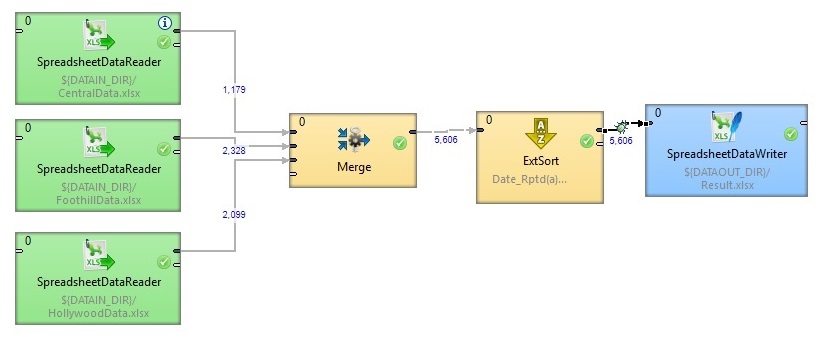
ETL is used to [migrate](http://www.webopedia.com/TERM/D/data_migration.html) data from one database to another, to form [data marts](http://www.webopedia.com/TERM/D/data_mart.html) and [data warehouses](http://www.webopedia.com/TERM/D/data_warehouse.html) and also to convert databases from one format or type to another.



ETL PROCESS

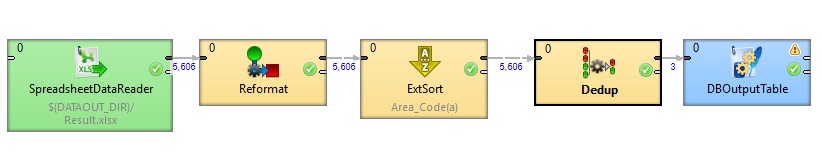
**4.2 ETL Logic**

**1. Merge**

****

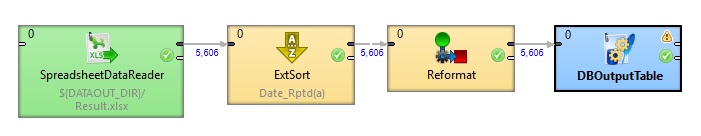
In this ETL graph we merge data from various data source and then sort it based on the date and then add it to excel file.

1. **Area**



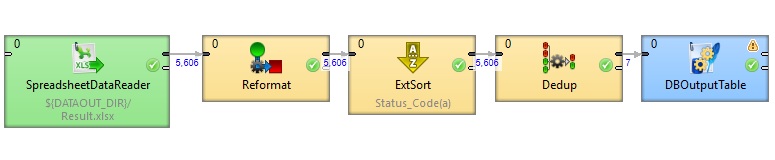
In this graph, the data is first reformatted and then sorted on area code. The duplicate data is finally deleted using dedup component and added to the database table.

1. **Case**

****

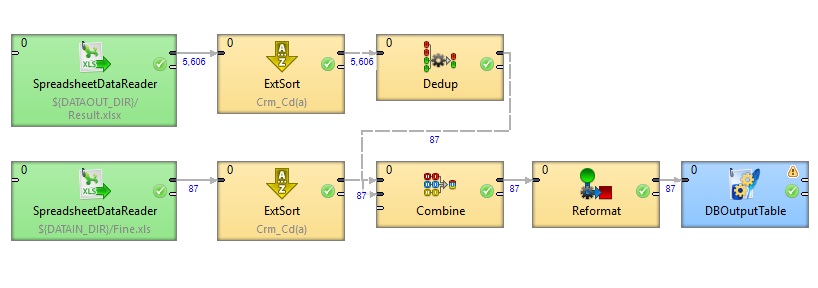
The data is first sorted and then reformatted. In reformation, fields which are required are taken and added to the database table.

1. **Status**

****

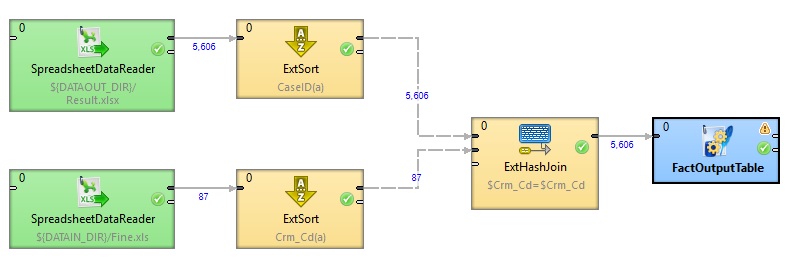
In this Particular ETL Graph, only status code and status description are added to the database table. The data is reformatted and then duplicate status code is deleted.

1. **Crime**



In this graph data has to be added to database table from two different datasource files. Here we used combine component to combine the data and added to database table.

1. **Fact table**

****

In this graph, Data has to be joined from two different data source. We used ExtHashJoin component to execute innerjoin on two data source and join the data as required and add it to the database table.

**5. Reporting**

**Tableau:**

* Tableau is groundbreaking data visualization software created by Tableau Software. Tableau connects easily to nearly any data source, be it corporate Data Warehouse, Microsoft Excel or web-based data. Tableau allows for instantaneous insight by transforming data into visually appealing, interactive visualizations called dashboards. This process takes only seconds or minutes rather than months or years, and is achieved through the use of an easy to use drag-and-drop interface.
* Tableau can be connected to live data sources, either local files (e.g., Excel) or databases (e.g., Oracle, SQL Server, Teradata), old versions of Tableau could be constrained by the underlying data sources, both in the speed of analytics and the volume of data that could be processed.
* Tableau was configured either for local usage (e.g., files, spreadsheets) or to query data directly in data marts or other data stores. Both modes of usage could give rise to latency problems. Running natively on a PC using local data sources.
* Because of data engine Tableau is capable of processing very large data sets locally – up to about a billion rows – and because the majority of that data is held in memory, it produces results rapidly. Additionally, it is still capable of accessing data in data marts. Organizations with fast, highly performant databases can still access those directly with Tableau.