

# Data Mapping Document

## Data Integration Strategy Overview

### KPIs and Business Requirements

Before the ETL process, we listed the following KPIs and other business requirements for the Incident Management System (IMS), taking into consideration of the business needs:

- Main **Key Performance Indicators (KPIs)** are service-level agreement (SLA) and the complete time for incident resolution (“ticket completion time”).
- **Overall metric** for an effective incident resolution performance include:
  - Short ticket complete time
  - Less repetitive resolution steps
  - Fast identify the problem that caused the incident
  - Efficient and accurate assignment
  - Low attrition rate
- To investigate on factors that impact KDPs and follow the metric, the **targets** are identified:
  1. How long it takes the incident to be resolved, measured by
    - a. [10] the time incident case is opened;
    - b. [36] the time incident case is closed.
  2. How incident is solved, measured by
    - a. [26] whether a *knowledge* base document was used to resolve the incident;
    - b. [33] Close\_code: identifier of the resolution of the incident.
  3. Delay in the process, measured by
    - a. [4] Reassignment counts: number of times the incident has the ASSIGN TO changed;
    - b. [5] Reopen counts: numbers of the incident resolution rejected by the caller.
    - c. [6] Modify counts: numbers of incidents updates until the moment.
- Data Latency:

It is critical to consider a streaming oriented ETL system according to the main business objectives of the IMS. The incident source data should be delivered to our DW/BI system instantaneously.

## Main Business Constraints

The event log generated by the ServiceNow Platform only records closed incident cases. For an incident with multiple reopens and resolutions, only the last resolved time is recorded. Therefore, the business requirement to investigate into each resolution time in such a multi-reopened incident case cannot be fulfilled. Alternative solution to track is through monitoring the change of status from “active” to “resolved” and the corresponding time recorded by “updated at”.

In addition, once a case is closed, no other event will be generated under the same incident number if the same user reports again the same problem. Reopen in this situation cannot be tracked based on the event log system for the ServiceNow Platform.

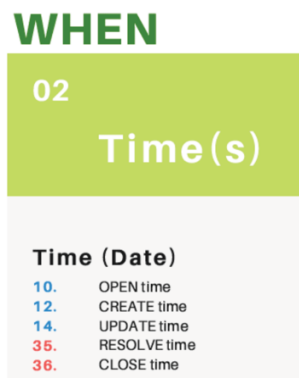
## Extracting

The source data for this assignment is one-time historic data captured from the audit system of ServiceNow Solution for Incident Management System of an IT company. The data is generated in a single system and ready to be extracted by ETL tools.

## Value-added Cleaning and Conforming

- **Special Dimensions Manager (Date/Time)**

The Data Dimension is designed to extract and restore all the date data from the source data, which has been identified in the previous Dimensional Design Process, under the description for “When”.



- **Employee**

In the second step of the Dimensional Design Process, we made the assumption that for employee data, a lower level of details needs to be done, “Created by 1” → “Create” (function) + “1” (Employee #). However, after the first ETL process, we became aware of the uncertainty of such number association with Employee’s role. We decided to keep the “Created by 1” and only used digits that are extracted to reorganize 4 columns of employees into one column.

## 2<sup>nd</sup> Version Schema

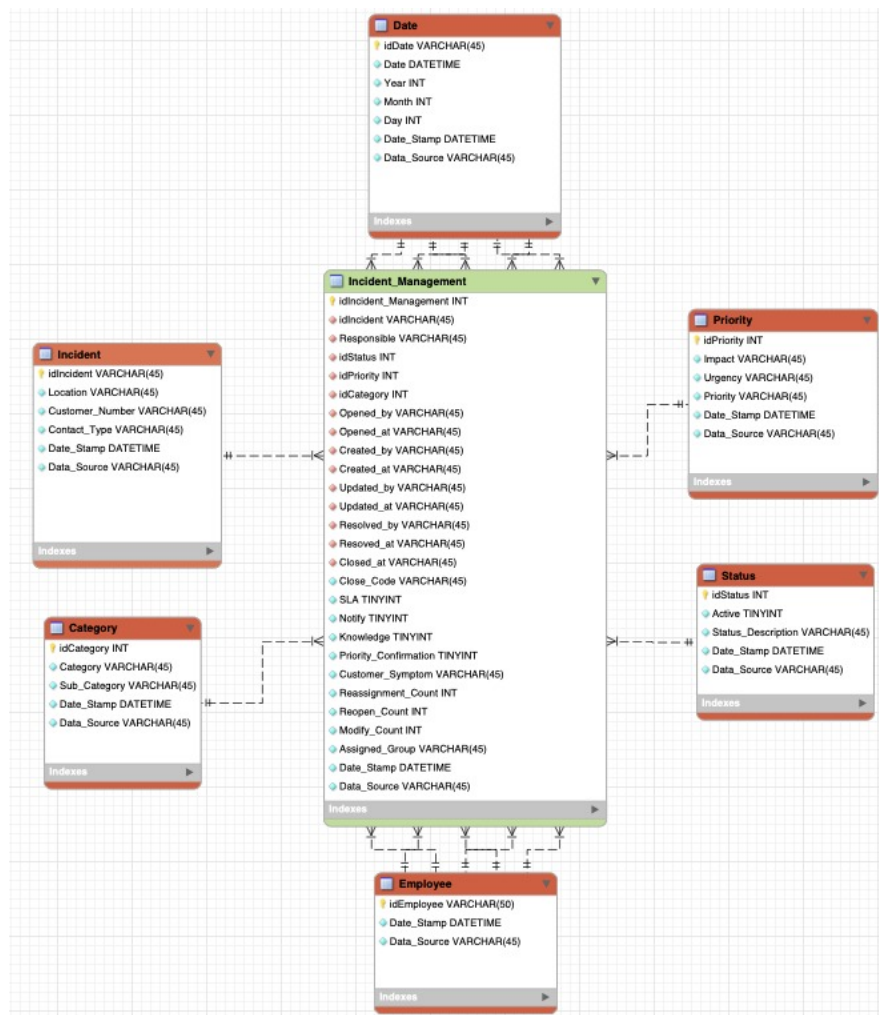
The 1<sup>st</sup> Version Schema developed during the dimension design process has been updated during the entire ETL process. Two major changes are the structure between **DATE DIMENSION** and the **FACT TABLE**, and **EMPLOYEE DIMENSION** and the **FACT TABLE**.

- DATE DIMENSION

Due to the inconsistent date format in our dataset, such as “d/m/16 hh:mm”, “dd/m/16 hh:mm”, “dd/mm/16 hh:mm”, “d/mm/2016 hh:mm” which need long ETL transformation twice before formatting into DATETIME “yyyy/dd/mm hh:mm:ss”. Therefore, the Primary Key of the **DATE DIMENSION** idDATE has data type of varchar. DAY OF THE WEEK and WEEK OF THE YEAR are removed as they are not relevant for our key business requirements identified above.

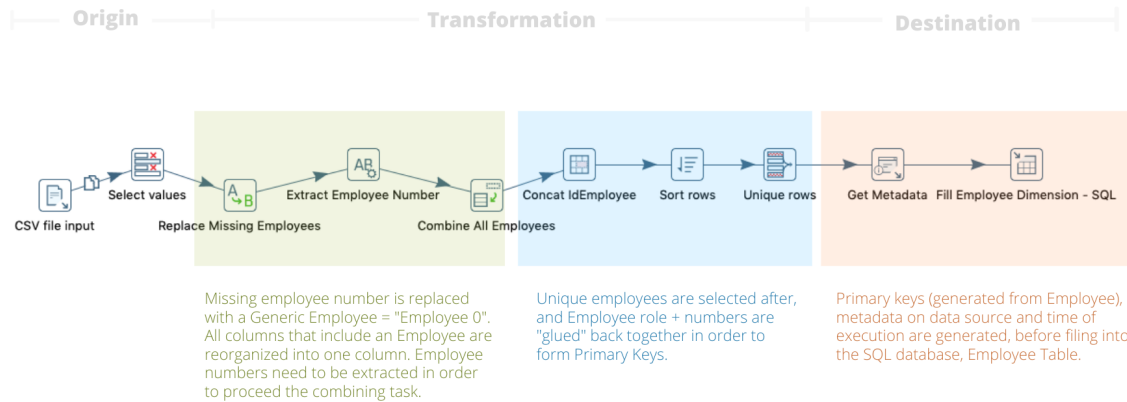
- EMPLOYEE DIMENSION

Corresponding to the “employee section in Value-added Cleaning and Conforming” discussed above, the Primary Key of the **EMPLOYEE DIMENSION** is changed from a numeric ID “1”, to the employee information itself “created by 1”.



# Data Mapping Process (ETL Process)

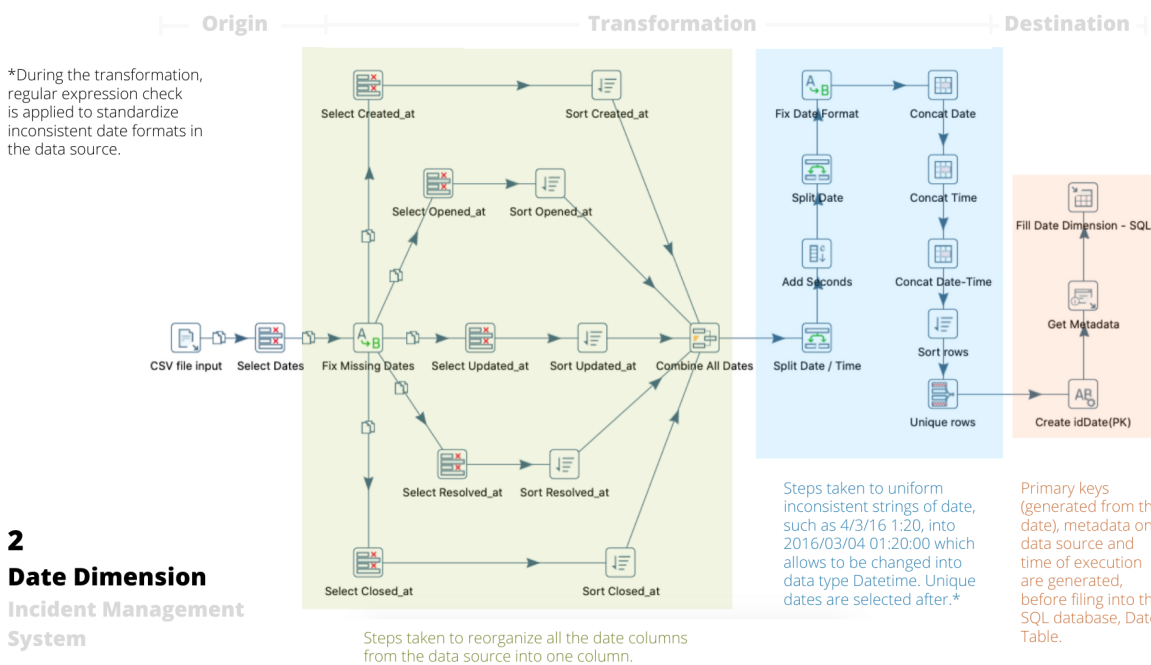
6 Dimensions + 1 Fact Table + 1 Job



1

## Employee Dimension

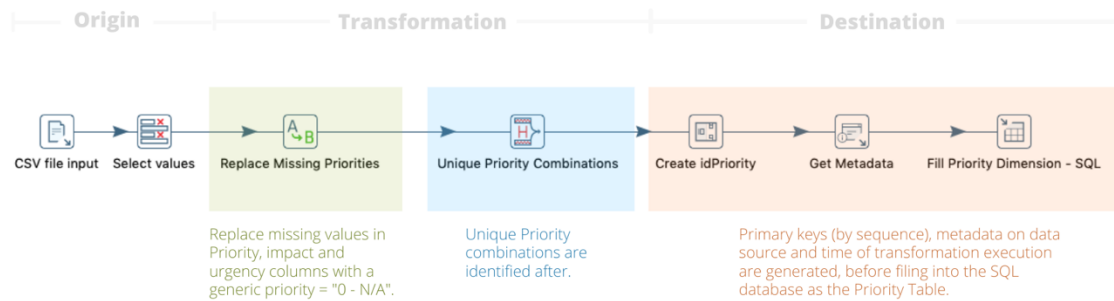
Incident Management System



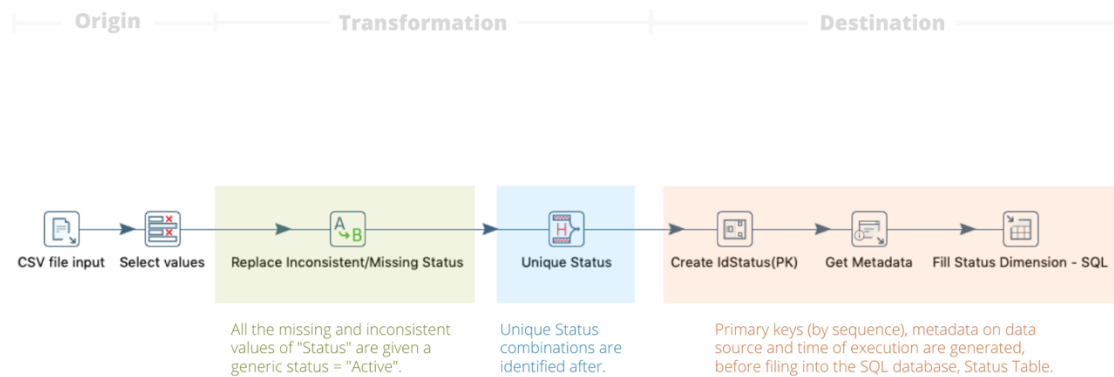
2

## Date Dimension

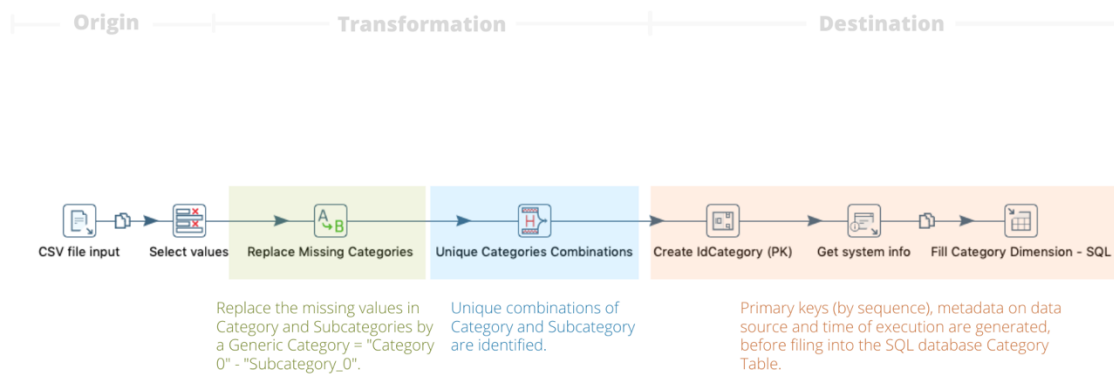
Incident Management System



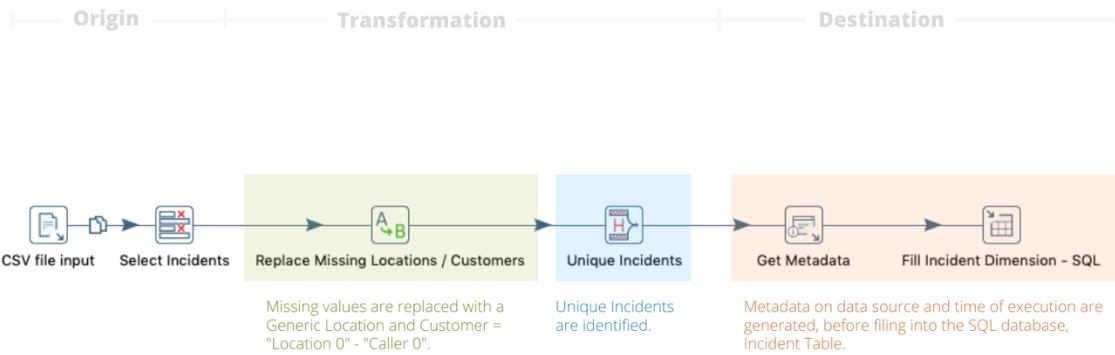
### 3 Priority Dimension Incident Management System



### 4 Status Dimension Incident Management System

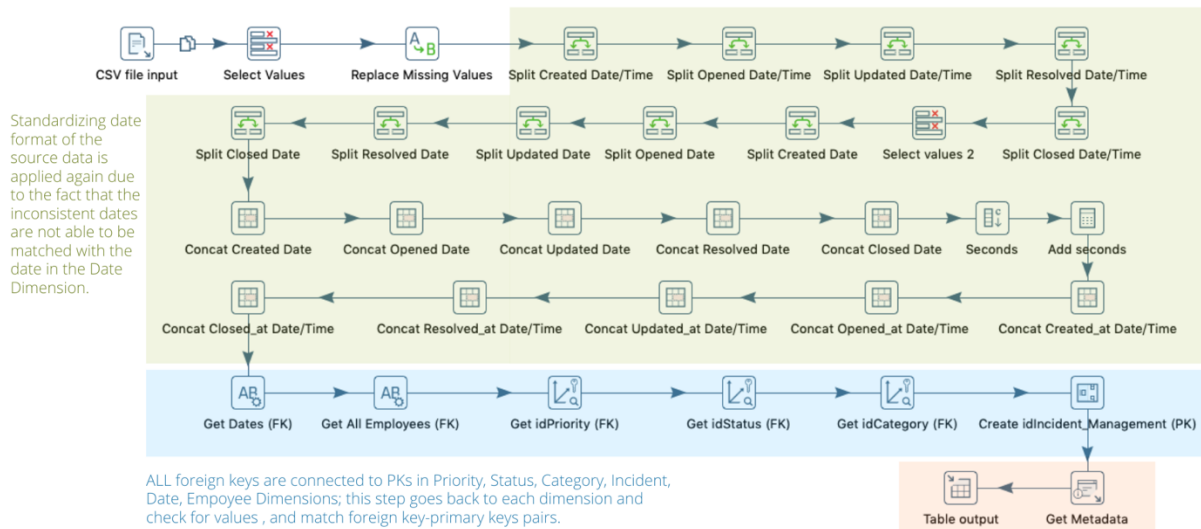


### 5 Category Dimension Incident Management System



## 6 Incident Dimension

### Incident Management System



## 7 Incident Management Fact Table

The green check means that the fact table load will only be executed when the previous dimensions have been loaded successfully.



## Job Incident Management System

# Data Quality Tracking & Metadata Approach

## Data Quality

Data Quality check for source data is executed within the Transformation process, and a quality screens system is designed following three categorical approach, which includes columns, structure and business rule.

- **Quality Screens System:**
  - We have configured the schema to only ask for valid values, with each field NOT NULL.
  - Before loading the fact table, a data validation (already set in the look-up) on structures and relationship between columns in two tables is taken place, to test the foreign keys and primary keys.
  - For our one-time historic dataset, all the opened cases are closed, and all the closed cases was opened. However, a future quality screen check on the business rule can be implemented, such as:
    - An incident can only be closed when it has been opened in the IMS.

## Metadata

For this Data Integration assignment, since the ETL process is fed with a single-sourced one-time historic dataset with constant metadata, the audit system is built to include the simple timestamp and source in each dimension.

However, to consider an overall metadata strategy for the future system loading more data sources, a new dimension of Audit System can be created directly attached to the fact table. Under the Audit Dimension data source, data integration, and ETL process are captured and stored. The future metadata audit system can be designed as below:

