Group E

11/16/19

Incident Management System

Overview

The dataset is an event log for that describes the process of how an incident within the **ServiceNow Solution for Incident Management System**[1] is reported, processed and managed within an IT company. The data is collected from an audit system for one instance of the ServiceNow platform for Year 2016.

When an incident happens, the IT company needs to correct failures as soon as possible to minimize the impact on normal business operation[2].

The business objective is to help the IT company reduce the completion time for incident resolution (Ticket Completion Time). The analytic method investigates into the repetitive incident categories, resolution steps, and task overload during the processes.

Analyze the dataset

The original dataset includes 141.712 events for 24.918 incidents, described by 36 attributes (for each attribute information, see Appendix 1).

I. An incident management process mapping exercise is done to organize attributes into four categories, following the chronological flow of the process. The four process categories were: (1) customer reports incident, (2) information related to the system, (3) processing the incident and (4) close the incident.

Incident Management Process Mapping

1. *Customer reports the incident*
    - *a. OPEN [9], [10]*
    - *b. Each EVENT is recorded with the same INCIDENT NUMBER [1] along with Attributes [2] to [8].*
        - *i. INCIDENT STATUS [2], [3]*
2. *Incident information enters (registered into) the management system*
    - *a. CREATE: Employee files the incident into the system [11], [12]*
    - *b. INCIDENT DESCRIPTION*
        - *i. Contact type, location, (sub)category, customer perception of service availability [15], [16], [17], [18] [19]*
    - *c. Measure URGENCY of the incident [21], [22], [23]*
3. *Processing the incident*

[1] https://www.servicenow.com/products/incident-management.html

[2] Amaral C.A.L., Fantinato M., Reijers H.A., Peres S.M. (2019) Enhancing Completion Time Prediction Through Attribute Selection. In: Ziemba E. (eds) Information Technology for Management: Emerging Research and Applications. AITM 2018, ISM 2018. Lecture Notes in Business Information Processing, vol 346.

   a. *Employee updates the incident in the system*
      i. *<u>UPDATE</u> [13], [14]*
      ii. *<u>ASSIGN</u> the task to group [24], and a responsible [25]*
   b. *Identify the <u>PROBLEM</u> caused the incident [29], [30],[31],[32]*
   c. *<u>RESOLVE</u> the incident [34], [35]*
4. *<u>CLOSE</u> the incident [33], [36]*

II. With limited information on the IT company and its business context, we then extracted **key data facts** from the dataset to build an overview of the incident resolving capacity of this IT company in Year 2016.

a. <u>Incidents in Year 2016</u>

- The total number of incidents is ***24,918***
- The total number of events is ***141,712***
- **There is an average of 6 events per incident**

b. <u>Incident State</u>

- 8 levels controlling:

[2] INCIDENT STATE	#	[3] ACTIVE	#
New		TRUE	
Active			
Awaiting Problem	247		
Awaiting User Info	5471		
Awaiting Vendor	259		
Awaiting Evidence	21		
Resolved			
Closed		FALSE	24,918

- All incidents registered in the system are closed.
- 5471 incidents went through "Awaiting User Info" state.
-

c. <u>Service Level Agreement (SLA) target</u>

- 15,803 out of total 24,918 incidents (63.4%) are **compliant with SLA** after the incident was **CLOSED**.

Detect potential data quality issues

- **Missing Values (Appendix 2)**

  17.53% of the dataset values are missing, in other words, around 900000 cells in the dataset were void of information. The missing cells were mostly grouped in the same attributes; therefore, the missing values do not affect the overall quality of the dataset.

  - **To remove**: *[32]* caused_by, *[31]*vendor, *[30]*rfc, *[29]*problem_id,*[20]*cmbd_ci. There are 5 attributes that have over 98% of unknown information, accounting for 78.8% of all missing values We decided to remove them from our final database.

  - **To keep**: *[11]*sys_created_by, *[12]*sys_created_at, *[19]*u_symptom, *[25]*assigned_to, *[24]*assignment_group
    Though these 5 attributes have a proportion of missing values ranging from 10% to 30%, the valuable information is critical to future data analysis. Missing values are treated as "unknown".

  - The remaining attributes with missing data contain less of 5% of missing values, with most of them missing less than 0.1%. Therefore, we classify the impact from these missing values on our database as negligible and treat missing values as "unknown".

- **Error: Negative Values**

  There is one attribute which has data with error. The attribute *[2]* incident_state contains 8 levels to describe an incident state, such as New, Active, Awaiting Evidence, Closed…). There are 5 rows with a value of "-100". As the negative values happens for two different incidents, and there is no similar trait observed from these two incidents, we classify "-100" as the error. We replace each "-100" with a reasonable state level in accordance to the before and after events for each incident.

Detect duplicates

There are no duplicate rows found in our dataset.

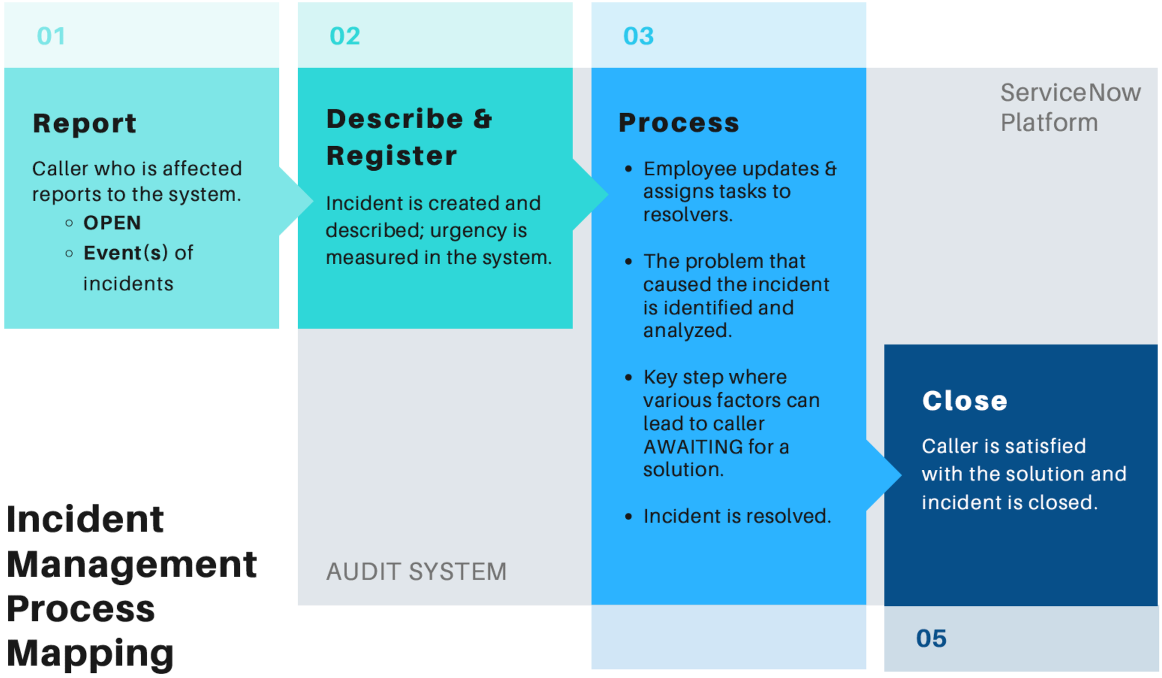
Every row of an incident number describes a particular interaction or update in a corresponding process-aware information system, which is called event. Therefore, an incident number can have multiple events (rows), compiling the same information and updating values to different columns.

Data Warehouse Approach Selection

We select the **multidimensional data warehouse approach** for the incident management system, as it focuses on process and performance. A four-step dimensional process design is taken to build up our rationale.

Four-Step Dimensional Design Process

1) **Select Business Process**
   The event log is a direct output metrics from the incident management process.



01

Report

Caller who is affected reports to the system.
- **OPEN**
- **Event(s)** of incidents

02

Describe & Register

Incident is created and described; urgency is measured in the system.

03

Process

- Employee updates & assigns tasks to resolvers.
- The problem that caused the incident is identified and analyzed.
- Key step where various factors can lead to caller AWAITING for a solution.
- Incident is resolved.

ServiceNow Platform

Close

Caller is satisfied with the solution and incident is closed.

05

AUDIT SYSTEM

Incident Management Process Mapping

2) **Specify the level of data details (Declare the *Grain*[3])**
   The attributes in the event log contain most detailed and atomic information. They are categorized into three main groups according to information. The information categories were 1) People, 2) Time, and 3) Incident Description.

   Several attributes that need to be moved to lower level of details are identified (1^{ST} Normal Form), including:
   - all ***EMPLOYEE*** attributes under **Category 01 PEOPLE**
     - EXAMPLE:
       "Resolver 78" → "Resolve" (ACTION) + "78" (EMPLOYEE ID)
   - all attributes under **Category 03 INCIDENT/ BUSINESS AFFECTED**
   - *GROUPASSIGNED* under **Category 01 PEOPLE / EMPLOYEE**

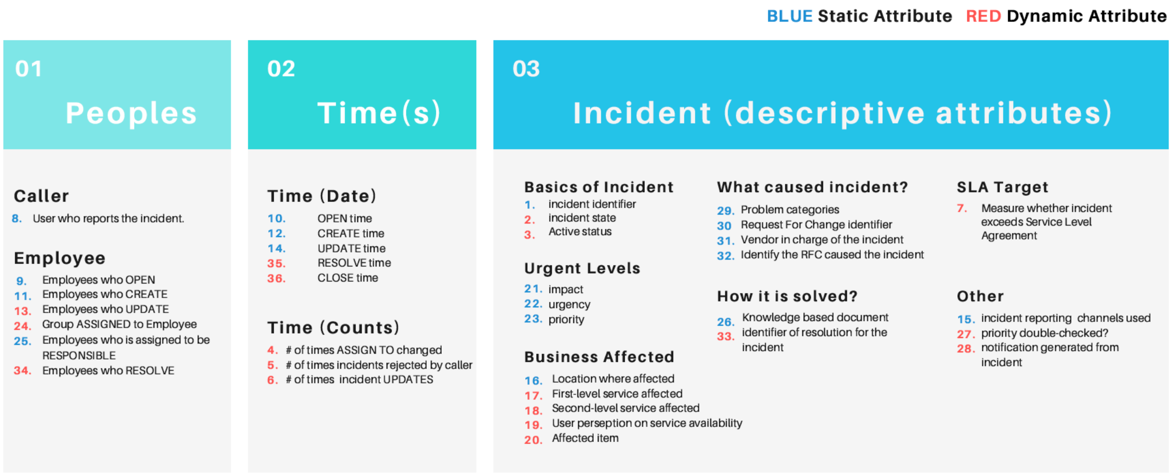
[3] Kimball, R. and Ross, M. (2013) The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (3rd Edition) Indiana: Wiley Publishing.

We further color-coded the attributes by **DYNAMIC** and **STATIC**.

- Dynamic attributes are variables/features that can be altered in a single case and represented in events. In the event log they document the changes of each action taken and are critical for the process analysis.
- Static attributes record one-time information that do not change nor get affected by actions taken in the process, such as identifiers of incident.

Descriptive Categories of Attributes

Incident Management System



BLUE Static Attribute **RED** Dynamic Attribute

01 Peoples

Caller
8. User who reports the incident.

Employee
9. Employees who OPEN
11. Employees who CREATE
13. Employees who UPDATE
24. Group ASSIGNED to Employee
25. Employees who is assigned to be RESPONSIBLE
34. Employees who RESOLVE

02 Time(s)

Time (Date)
10. OPEN time
12. CREATE time
14. UPDATE time
35. RESOLVE time
36. CLOSE time

Time (Counts)
4. # of times ASSIGN TO changed
5. # of times incidents rejected by caller
6. # of times incident UPDATES

03 Incident (descriptive attributes)

Basics of Incident
1. incident identifier
2. incident state
3. Active status

Urgent Levels
21. impact
22. urgency
23. priority

Business Affected
16. Location where affected
17. First-level service affected
18. Second-level service affected
19. User perseption on service availability
20. Affected item

What caused incident?
29. Problem categories
30 Request For Change identifier
31. Vendor in charge of the incident
32. Identify the RFC caused the incident

How it is solved?
26. Knowledge based document
33. identifier of resolution for the incident

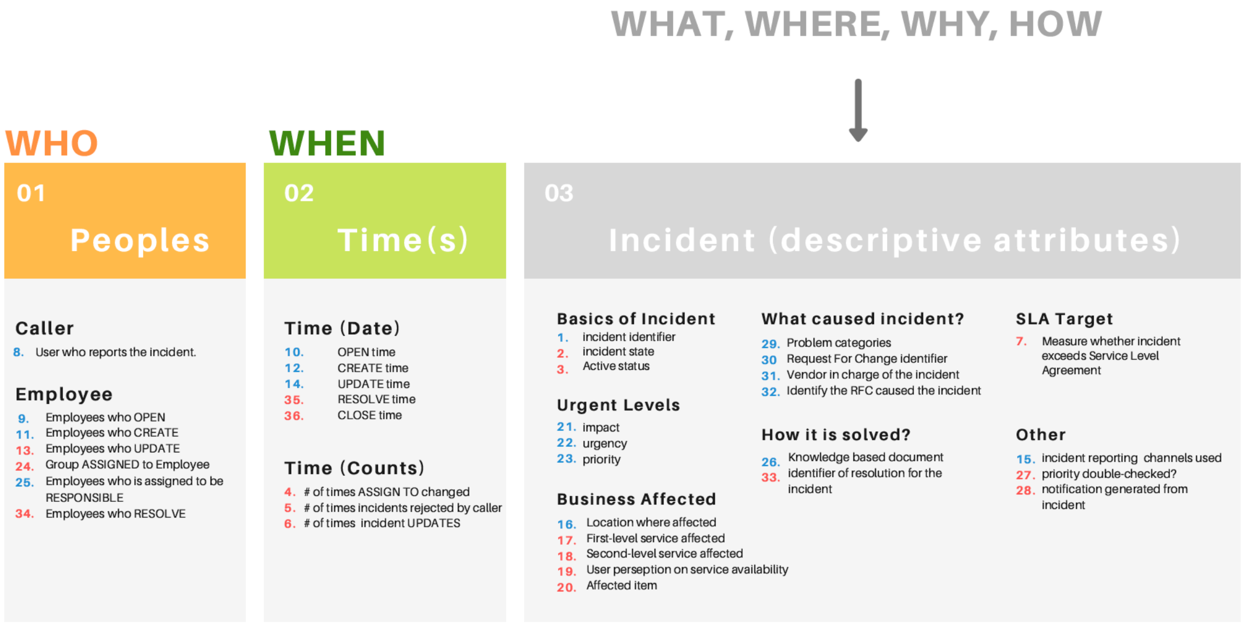
SLA Target
7. Measure whether incident exceeds Service Level Agreement

Other
15. incident reporting channels used
27. priority double-checked?
28. notification generated from incident

3) Identify the Dimensions

To answer "WHAT, WHERE, WHO, WHEN, WHY and HOW", following steps are taken:

- **Category 01 PEOPLE** answers "WHO" and **Category 02 TIME** answers "WHEN"



01 Peoples

Caller
8. User who reports the incident.

Employee
9. Employees who OPEN
11. Employees who CREATE
13. Employees who UPDATE
24. Group ASSIGNED to Employee
25. Employees who is assigned to be RESPONSIBLE
34. Employees who RESOLVE

02 Time(s)

Time (Date)
10. OPEN time
12. CREATE time
14. UPDATE time
35. RESOLVE time
36. CLOSE time

Time (Counts)
4. # of times ASSIGN TO changed
5. # of times incidents rejected by caller
6. # of times incident UPDATES

03 Incident (descriptive attributes)

Basics of Incident
1. incident identifier
2. incident state
3. Active status

Urgent Levels
21. impact
22. urgency
23. priority

Business Affected
16. Location where affected
17. First-level service affected
18. Second-level service affected
19. User perseption on service availability
20. Affected item

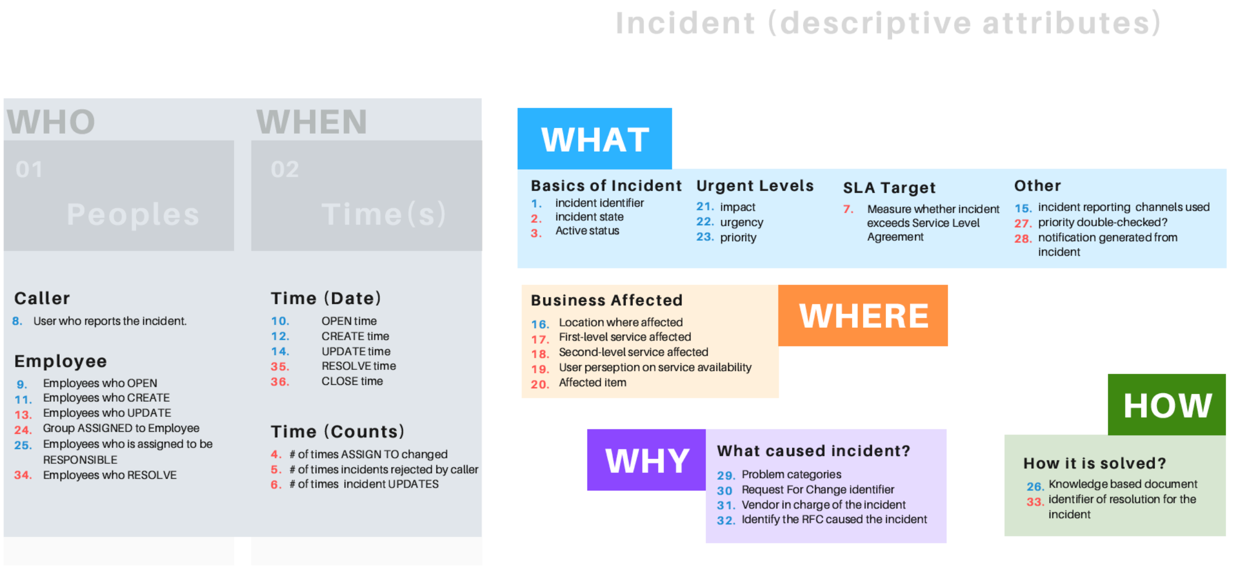
What caused incident?
29. Problem categories
30 Request For Change identifier
31. Vendor in charge of the incident
32. Identify the RFC caused the incident

How it is solved?
26. Knowledge based document
33. identifier of resolution for the incident

SLA Target
7. Measure whether incident exceeds Service Level Agreement

Other
15. incident reporting channels used
27. priority double-checked?
28. notification generated from incident

- **Category 03 INCIDENT DESCRIPTION** is broken down and reorganized to answer "WHAT, WHERE, WHY and HOW"



Incident (descriptive attributes)

WHO

01 Peoples

Caller
8. User who reports the incident.

Employee
9. Employees who OPEN
11. Employees who CREATE
13. Employees who UPDATE
24. Group ASSIGNED to Employee
25. Employees who is assigned to be RESPONSIBLE
34. Employees who RESOLVE

WHEN

02 Time(s)

Time (Date)
10. OPEN time
12. CREATE time
14. UPDATE time
35. RESOLVE time
36. CLOSE time

Time (Counts)
4. # of times ASSIGN TO changed
5. # of times incidents rejected by caller
6. # of times incident UPDATES

WHAT

Basics of Incident
1. incident identifier
2. incident state
3. Active status

Urgent Levels
21. impact
22. urgency
23. priority

SLA Target
7. Measure whether incident exceeds Service Level Agreement

Other
15. incident reporting channels used
27. priority double-checked?
28. notification generated from incident

WHERE

Business Affected
16. Location where affected
17. First-level service affected
18. Second-level service affected
19. User perseption on service availability
20. Affected item

WHY

What caused incident?
29. Problem categories
30 Request For Change identifier
31. Vendor in charge of the incident
32. Identify the RFC caused the incident

HOW

How it is solved?
26. Knowledge based document
33. identifier of resolution for the incident

  - Due to missing 98% percent of information for all attributes in **Category 03 INCIDENT DESCRIPTION/BUSINESS AFFECTED**, "WHERE" will not be answered in our model.

- **A multidimensional approach is identified:**
  - People
  - Date
  - Incident
    - Description
    - Causes of Incident
    - Resolution

4) Identify the Facts

The incident management process measures how long it takes to resolve an incident. For the IT company, in order to minimize the negative impact to business, the main focus of this process is to reduce the completion time for incident resolution (Ticket Completion Time) as much as possible.

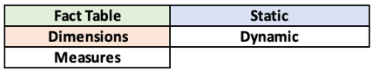
We identified **key business users' requirements** as follows, in accordance to PART II. of dataset analysis on **key data facts**:

- Summary on events and incidents
- Incident State
- Service Level Agreement (SLA) target
- Ticket Completion Time
- Employee Task Assignment & Performance

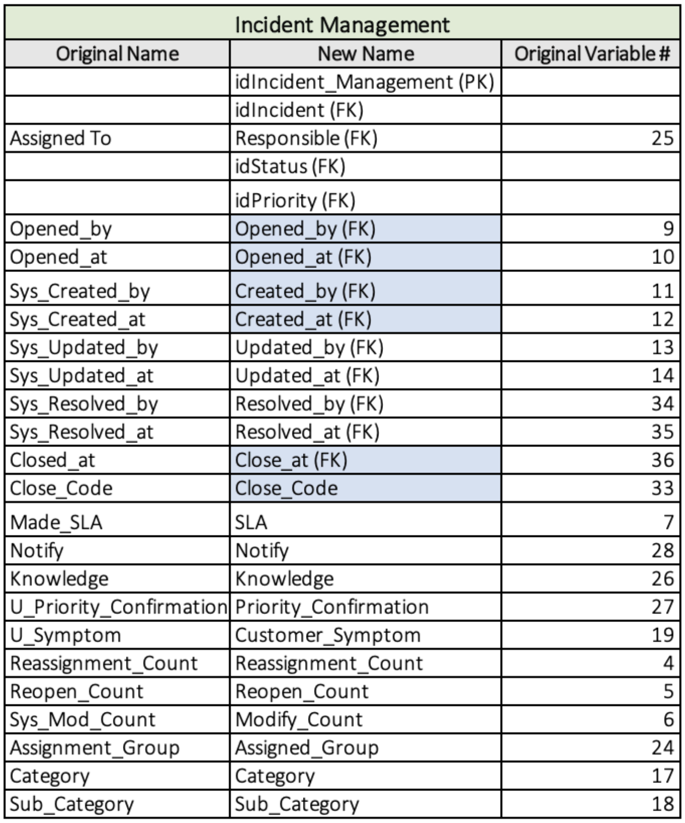
Data Warehouse Design

1) Identify the entities based on multidimensional approach

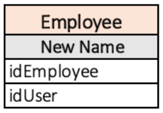
a. Fact Tables & Dimension Tables



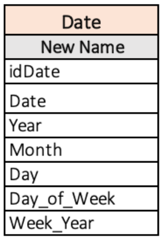
Fact Table	Static
Dimensions	Dynamic
Measures	



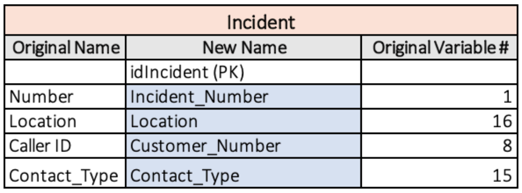
Incident Management		
Original Name	New Name	Original Variable #
	idIncident_Management (PK)	
	idIncident (FK)	
Assigned To	Responsible (FK)	25
	idStatus (FK)	
	idPriority (FK)	
Opened_by	Opened_by (FK)	9
Opened_at	Opened_at (FK)	10
Sys_Created_by	Created_by (FK)	11
Sys_Created_at	Created_at (FK)	12
Sys_Updated_by	Updated_by (FK)	13
Sys_Updated_at	Updated_at (FK)	14
Sys_Resolved_by	Resolved_by (FK)	34
Sys_Resolved_at	Resolved_at (FK)	35
Closed_at	Close_at (FK)	36
Close_Code	Close_Code	33
Made_SLA	SLA	7
Notify	Notify	28
Knowledge	Knowledge	26
U_Priority_Confirmation	Priority_Confirmation	27
U_Symptom	Customer_Symptom	19
Reassignment_Count	Reassignment_Count	4
Reopen_Count	Reopen_Count	5
Sys_Mod_Count	Modify_Count	6
Assignment_Group	Assigned_Group	24
Category	Category	17
Sub_Category	Sub_Category	18



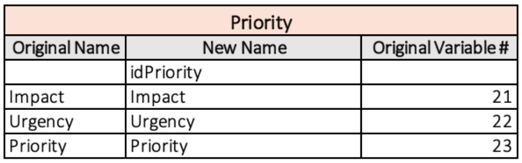
Employee
New Name
idEmployee
idUser



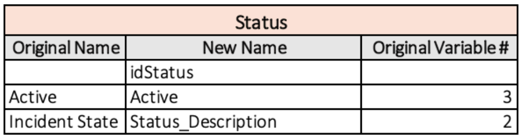
Date
New Name
idDate
Date
Year
Month
Day
Day_of_Week
Week_Year



Incident		
Original Name	New Name	Original Variable #
	idIncident (PK)	
Number	Incident_Number	1
Location	Location	16
Caller ID	Customer_Number	8
Contact_Type	Contact_Type	15



Priority		
Original Name	New Name	Original Variable #
	idPriority	
Impact	Impact	21
Urgency	Urgency	22
Priority	Priority	23



Status		
Original Name	New Name	Original Variable #
	idStatus	
Active	Active	3
Incident State	Status_Description	2

c. Metrics

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

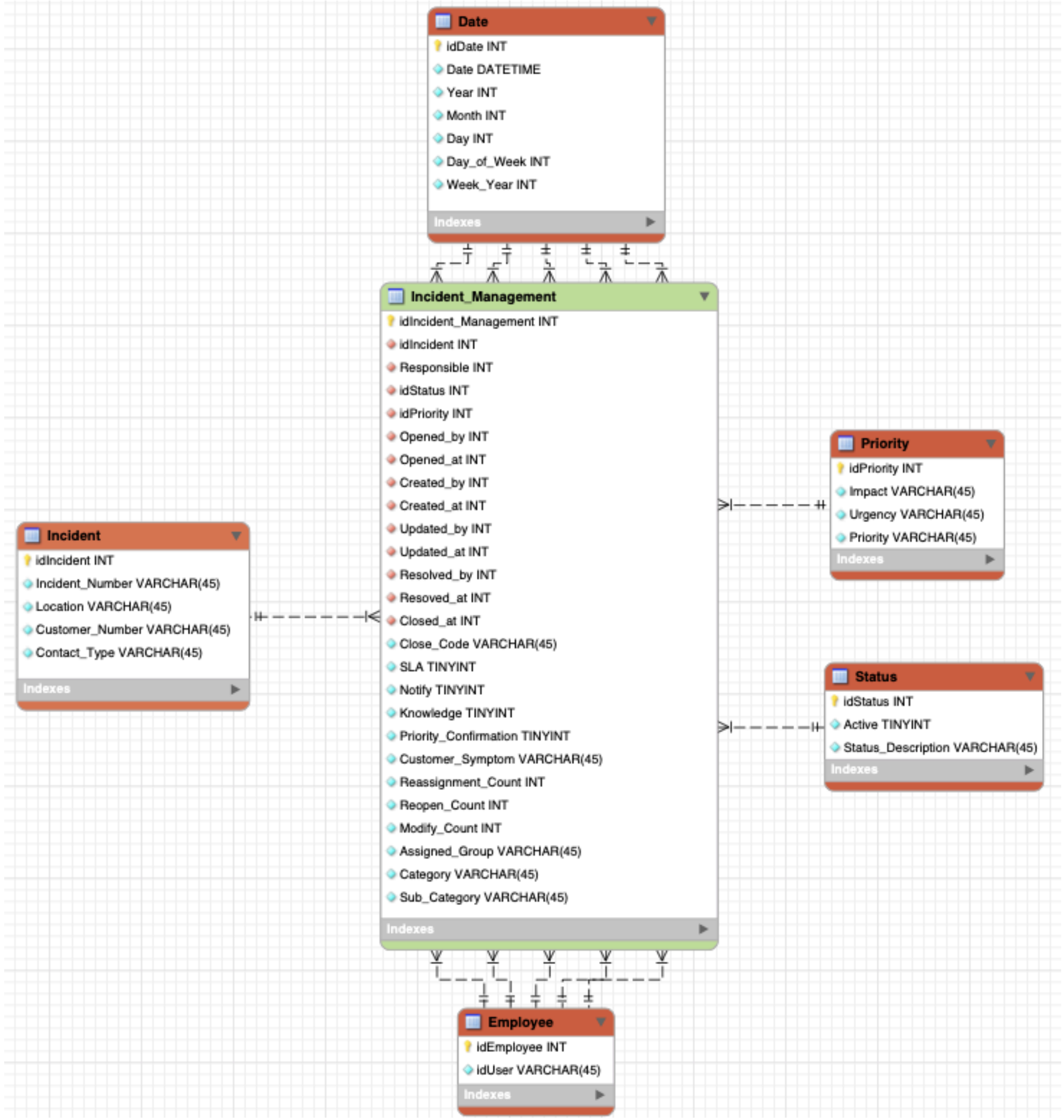
2) How we arrive at a Star Schema

Following the multidimensional approach, **dimension tables** such as **Employee Dimension**, **Date Dimension** and **Incident Dimension** are first defined. By understanding the unique values associated to the incidents, we further identify **Priority Dimension** and **Status Dimension**.

The **fact table** is designed to store all interactions that affect the incident in the time frame (from New to Close status). Various attributes at the needed granularity level are included in the fact table, ready to be aggregated to present key user's requirements.

The time ticket reduction is the **main business objective**, and the first dimension fits such n objective. Therefore, a **star schema** is achieved.

3) MYSQL Workbench Design (*IMS Star Schema Group E.mwb* attached)



Date
idDate INT
Date DATETIME
Year INT
Month INT
Day INT
Day_of_Week INT
Week_Year INT
Indexes

Incident_Management
idIncident_Management INT
idIncident INT
Responsible INT
idStatus INT
idPriority INT
Opened_by INT
Opened_at INT
Created_by INT
Created_at INT
Updated_by INT
Updated_at INT
Resolved_by INT
Resoved_at INT
Closed_at INT
Close_Code VARCHAR(45)
SLA TINYINT
Notify TINYINT
Knowledge TINYINT
Priority_Confirmation TINYINT
Customer_Symptom VARCHAR(45)
Reassignment_Count INT
Reopen_Count INT
Modify_Count INT
Assigned_Group VARCHAR(45)
Category VARCHAR(45)
Sub_Category VARCHAR(45)
Indexes

Priority
idPriority INT
Impact VARCHAR(45)
Urgency VARCHAR(45)
Priority VARCHAR(45)
Indexes

Incident
idIncident INT
Incident_Number VARCHAR(45)
Location VARCHAR(45)
Customer_Number VARCHAR(45)
Contact_Type VARCHAR(45)
Indexes

Status
idStatus INT
Active TINYINT
Status_Description VARCHAR(45)
Indexes

Employee
idEmployee INT
idUser VARCHAR(45)
Indexes

Appendix 1. Attribute Information[4]

1. number: incident identifier (24,918 different values);
2. incident state: eight levels controlling the incident management process transitions from opening until closing the case;
3. active: boolean attribute that shows whether the record is active or closed/canceled;
4. reassignment_count: number of times the incident has the group or the support analysts changed;
5. reopen_count: number of times the incident resolution was rejected by the caller;
6. sys_mod_count: number of incident updates until that moment;
7. made_sla: boolean attribute that shows whether the incident exceeded the target SLA;
8. caller_id: identifier of the user affected;
9. opened_by: identifier of the user who reported the incident;
10. opened_at: incident user opening date and time;
11. sys_created_by: identifier of the user who registered the incident;
12. sys_created_at: incident system creation date and time;
13. sys_updated_by: identifier of the user who updated the incident and generated the current log record;
14. sys_updated_at: incident system update date and time;
15. contact_type: categorical attribute that shows by what means the incident was reported;
16. location: identifier of the location of the place affected;
17. category: first-level description of the affected service;
18. subcategory: second-level description of the affected service (related to the first level description, i.e., to category);
19. u_symptom: description of the user perception about service availability;
20. cmdb_ci: (confirmation item) identifier used to report the affected item (not mandatory);
21. impact: description of the impact caused by the incident (values: 1-“High; 2-Medium; 3-“Low);
22. urgency: description of the urgency informed by the user for the incident resolution (values: 1-“High; 2-“Medium; 3-“Low);
23. priority: calculated by the system based on 'impact' and 'urgency';
24. assignment_group: identifier of the support group in charge of the incident;
25. assigned_to: identifier of the user in charge of the incident;
26. knowledge: boolean attribute that shows whether a knowledge base document was used to resolve the incident;
27. u_priority_confirmation: boolean attribute that shows whether the priority field has been double-checked;
28. notify: categorical attribute that shows whether notifications were generated for the incident;
29. problem_id: identifier of the problem associated with the incident;
30. rfc: (request for change) identifier of the change request associated with the incident;
31. vendor: identifier of the vendor in charge of the incident;
32. caused_by: identifier of the RFC responsible by the incident;
33. close_code: identifier of the resolution of the incident;
34. resolved_by: identifier of the user who resolved the incident;
35. resolved_at: incident user resolution date and time (dependent variable);
36. closed_at: incident user close date and time (dependent variable).

[4] Amaral C.A.L., Fantinato M., Reijers H.A., Peres S.M. (2019)

Appendix 2. Missing Values

Attribute Name	Nº Missing values	% versus total nº of values
caused_by	141469	99.83%
vendor	141468	99.83%
cmdb_ci	141267	99.69%
rfc	140721	99.30%
problem_id	139417	98.38%
sys_created_by	53076	37.45%
sys_created_at	53076	37.45%
u_symptom	32964	23.26%
assigned_to	27496	19.40%
assignment_group	14213	10.03%
opened_by	4835	3.41%
resolved_at	3141	2.21%
closed_code	714	0.50%
resolved_by	226	0.16%
subcategory	111	0.078%
category	78	0.055%
location	76	0.054%
caller_id	30	0.021%