Management of Organizational Data

Professor Yang

Fall Module 1

Kazanci Holding Auditing Department

Date: October 2, 2023

Group 4

Liana Simopoulos

Harshraj Jadeja

Ridvan Kanca

Sahana Reddy

Anton De Franco

Cole Wesley

I. Background

Kazanci Holding is a conglomerate, global energy company headquartered in Turkey with over 10,000 employees and facilities located in four different continents. The company exports its various energy products to 178 countries. The company started manufacturing generators in the 1980s and began constructing power plants and generating electricity in the 1990s. Kazancı Holding also invests in the rental, agriculture, and tourism industries.

Kazanci Holding is operated mainly in 4 different sectors: Energy (power generation), Natural Gas, Electricity and Power Generation. Particularly, our client Kazanci Holding Auditing Department is located in the company's headquarters in Istanbul/Turkey. The auditing department conducts internal audits across all of the company's facilities and for all sectors. We have analyzed Kazanci's internal auditing department to design a database to store the auditing results in a more systematic way.

Our database includes information on the auditors, audit reports, the audit finding, processes, status of audit findings, sectors and time sheet information submitted by the employees. By having this data readily available to the customer, they will be able to identify areas of the auditing process that aren't as efficient or information on the performance of employees themselves.

II. Data

Our data comes from the original source, helping us provide informative and realistic solutions to the client, Kazanci Holding Auditing Department. By directly obtaining data from the company itself, we ensure the highest level of accuracy. This accuracy is pivotal in the energy sector, where even small discrepancies can have significant financial and operational consequences. Our commitment to precise data collection sets a strong foundation for all our analyses and recommendations. Energy investments often involve substantial capital, as our authentic data helps identify and mitigate potential risks. We can provide our client with detailed risk assessments, ensuring they are well-informed and prepared for any challenges they may face.

III. Goals and Objectives

Our objective is to create a normalized database that minimizes data redundancy and to optimize the reporting process for the auditing department and measure its efficiency. As a team, we can assist them in providing a robust, centralized, and secure platform for data management and analysis. This, in turn, enhances decision-making, operational efficiency, regulatory compliance, and the ability to adapt to changing market conditions. The database becomes a valuable asset that supports the company's growth, competitiveness, and sustainability in the dynamic energy sector. The goals of this project are to reduce data redundancy, allowing

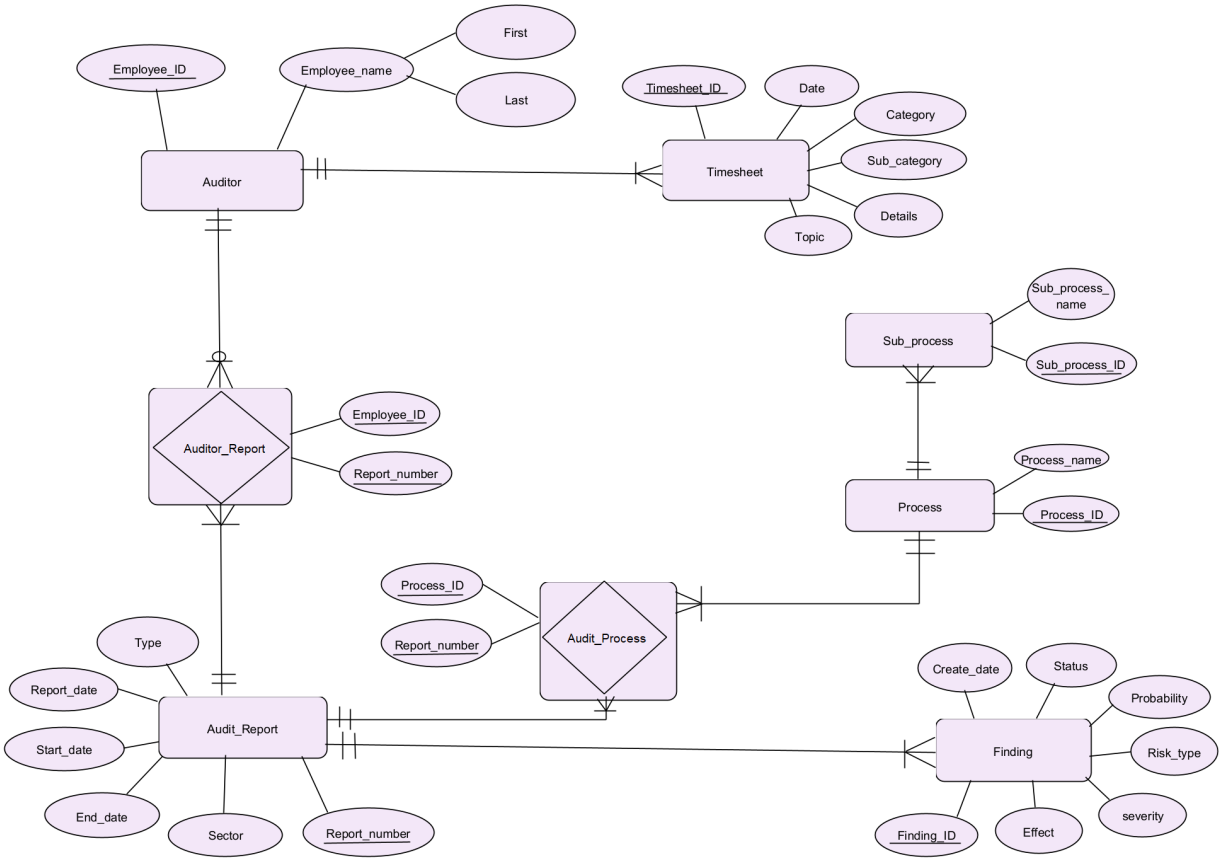
Kazanci to minimize costs and can potentially increase ROI. Additionally, by showing performances of the auditors and business owners (sectors) Kazanci will gain an overall insight on the auditing process, auditing department and findings.

IV. Conceptual Data Model

The data for the auditing department stores information on the audit reports, the audit findings, auditors, the auditors' timesheets, and the audit processes and sub-processes. In the database design there are six source tables and two associative tables used to link the data and minimize redundancy. The entity relationship diagram below the description provides a visual for the database design. The six main tables are Auditor, Audit_Report, Finding, Process, Sub-Process, and Timesheet and the two associative tables are Auditor_Report and Audit_Process. Descriptions of each of the tables and the data contained in the can be found below.

- Auditor
    - Contains the Employee_ID and the Employee name for each of the auditors and each of the auditors is uniquely identified by their Employee_ID. Every auditor submits a timesheet for the activities that they perform and every auditor has worked on zero, one, or many reports.
- Timesheet
    - Each timesheet submitted by an auditor includes a Timesheet_ID, Date that it was submitted, Category and Subcategory for the type of work performed, Details, and the Topic or subject of the work performed. Each of these timesheets is written by only one auditor.
- Audit_Report
    - An audit report contains all the information that was derived from the auditor themselves.Each report is uniquely identified by a Report_number. This table includes the Start and End date of the audit along with the Report date. The Sector for which the report was created and the Type of audit that was performed is also stored in the table. Every audit has at least one or many findings, processes, and timesheets associated with it. Each of these are many-to-many relationships and require the three associative tables to minimize repetitiveness.
- Finding
    - A finding is related to one and only one audit report and an audit report can have one to many findings. Each finding is uniquely identified with a Finding_ID. Each of these findings has a Status to identify any documentation or mitigation steps that need to be taken. There are also Risk_type, Probability of the event occurring, the Severity, and the effect if the event were to occur. Additionally, there is a Create_date associated with each of the findings.

- Sub-Process
    - The sub-process table contains more granular information about the steps taken in an audit. A sub-process is a part of a single process and is uniquely identified by the Sub_process_ID attribute. There is also a name/description that is stored in the table.
- Process
    - There are one to many processes that are performed by auditors that are made up of one to many sub-processes. Each of these are uniquely identified by the Process_ID and have a name associated with the task. The associative tables are used to like the Process to the audit report.
- Auditor_Report
    - Since there can be many auditors that contribute to the audit, there is a table needed to link the auditors to their specific contributions. The Auditor_Report table accomplishes this by using the Employee_ID from the Auditor table and the Report_number from the Audit_Report table.
- Audit_Process
    - Since an audit contains many processes and sub-processes and a process can be used for many different audits, another associative table is needed to identify the specific process that was performed each specific audit. The Audit_Process table provides this linkage by storing the Process_ID and the Report_number.



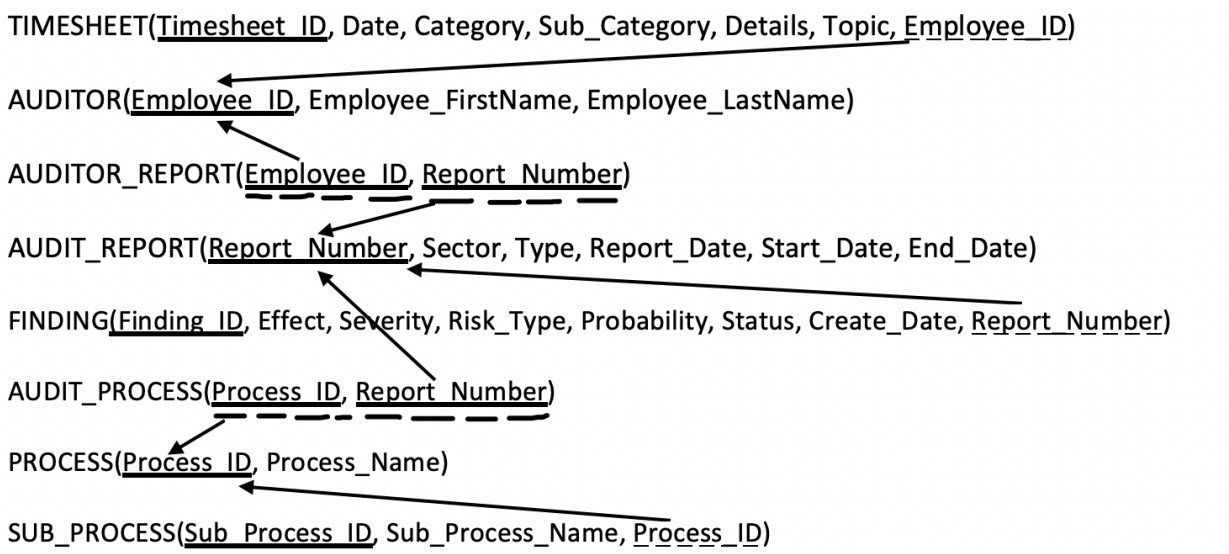
V. Relational Data Model

The relational data model was derived from the entity relationship diagram above and demonstrates the linkage between each of the tables. The primary keys are identified by a solid line under the attribute and the foreign keys are identified by a dashed line. The associative tables will have a solid and a dashed line that will be used to uniquely identify each of the records. The foreign keys are the “many” side of the relationship and used to link data across the tables. Descriptions of the primary and foreign keys can be found below.

- Auditor
    - The auditors are uniquely identified by the Employee_ID attribute which is used a foreign key in the Auditor_Report and the Timesheet tables. We would need to join this table with the Timesheet or Audior_Report tables to find the name of the auditor that submitted a timesheet or a specific report.
- Auditor_Report
    - Each of the auditors' contributions to a specific audit report are recorded using the associative table Auditor_Report. Using a left or right join, this table can be used to filter the Auditor or Audit_Report tables by using either the Employee_ID or

the Report_Number to see the contributions of each employee on a specific report.

- Audit_Report
    - Each audit report is uniquely identified by a Report_Number and has various attributes including the Start_Date, End_Date, and the Report_Date of the audit itself. The Sector attribute specifies a branch that the audit was performed on and the Type tells the user the subject of which was being audited.
- Audit_Process
    - An audit process is uniquely identified by a Process_ID and a Report_Number. This is an associative table like the Auditor_Report table previously discussed. This will allow one to query the data and see which processes were performed on a specific audit or steps that have yet to be taken in the entirety of the audit process.
- Timesheet
    - Each timesheet is uniquely identified by a Timesheet_ID. Additionally there is a Date, Category and Subcategory explaining the activities performed, the Topic, and Details of the activities performed and the time spent. There is also an Employee_ID associated with each timesheet allowing the user to find information about the employee who submitted the timesheet.
- Process
    - A process is an activity that is performed during a specific audit and is uniquely identified by a Process_ID. There can be many processes performed during a single audit or a single process could potentially not be part of a particular audit. Each process has a name describing the process. This table can be connected to the Audit_Report table indirectly using the Audit_Process table.
- Sub_Process
    - A Sub-Process is part of a single process and many sub-processes can be a part of a single process. Therefore, the Process_ID from the process table is a foreign key in this table, used to link each of the sub-processes to the parent process. This table also has a Sub_Process_ID to uniquely identify the entity and a Sub_Process_Name to describe the activity performed.
- Finding
    - A finding is the outcome of a specific audit report. There can be one to many findings in an audit but one finding belongs to one and only one audit report. Each finding is uniquely identified by the Finding_ID attribute and can be linked to the Audit_Report table using the Report_Number attribute. This would allow one to look at all the findings for a particular report.



Our database has been designed and organized to meet the criteria of the Third Normal Form (3NF). Each table within the database is equipped with a well-defined primary key, allowing for the unique identification of each row. Furthermore, we have ensured that all columns contain atomic, indivisible values, eliminating the presence of multivalued or composite data. To comply with the principles of 3NF, we have addressed partial dependencies, where non-key attributes rely solely on specific portions of the primary key, and transitive dependencies, where non-key attributes depend on other non-key attributes, through appropriate table structuring.

The database implementation includes robust data integrity constraints, such as unique and foreign key constraints, assuring data accuracy and consistency. The comprehensive adherence to 3NF principles demonstrates our commitment to optimizing data organization and integrity within our database.

VI. Database Implementation

The sample data provided by Kazanci Holding was from two queries from their existing database: one detailing specific audit reports and the other detailing timesheet data submitted by the auditors themselves. Before the data could be loaded into the database, the attributes needed to be separated into their respective tables and formatted appropriately. The data was manually separated first into separate sheets for each entity and then Tableau Prep was used to clean and reformat the data. Composite attributes such as the auditors' names on one report were first split into separate columns and then pivoted into a singular row. The names were then split into first and last names and assigned a primary key using the ROW_NUMBER() function in Tableau prep. Aggregate functions removed any duplicates to ensure that these numbers could be used as

primary keys in the database. These keys were then added as foreign keys to associative tables and as foreign keys to remove any redundancies. Next, the data was assigned as an output to a CSV file to be loaded into the tables.

The SQL file Group#4CreateTables.sql details the DDL statements that were executed to create the six main tables and the two associative tables needed to store the data. The primary keys of each of the tables were defined to make sure when the data was loaded that the identifiers were unique. Lastly, the data was loaded into each of the tables using the output CSV files from Tableau Prep and the data input wizard in MySQL. Few manual adjustments were needed to ensure that the field names were mapped properly and then the data was ready to be queried and analyzed using the sql file Group#4queries.sql.

VII. Conclusion

In conclusion, our objective to create a normalized database with minimal data redundancy has been achieved. Kazanci holdings will now be able to more efficiently query appropriate information to make better informed business decisions. The SQL queries we have created based on the business requirements answer critical questions that we anticipate Kazanci having and questions the company has requested. As a result, our optimized database and efficient set of queries will enhance the internal auditing department's capabilities going into the future.