DATABASE AND DATA WAREHOUSING

CIS 499 Information System capstone

May 9, 2020

**Database and Data Warehousing**

Decision making is key for every organization in the world now. Companies are run based on decisions that affect their position in business in the market. Decision making is regarded as one of the major functions of management. Some others include planning, staffing, organizing and directing all entities in the business. To make accurate and effective decisions in an organization, all data despite their hierarchy, should be used in the decision and thus the need for data warehousing. This is because all the data that the company stores are kept in a data warehouse and can easily be extracted to do analytics and make decisions (Krishnan, 1).

On Point Analytics will store its data in a data warehouse and with this, the company will be able to monitor the inputs and outputs in the company easily. The goal of On Pint Analytics is to make better decisions than our key competitors and thus establishing their name in the market. With the Information system, the data once collected, will be moved to the data warehouse of the company. The company is expected to filly make use of the 10 TB storage space which will be backed up on the cloud for safety and protection against data loss. Once in the data warehouse, the data will be cleaned and scanned for any viruses which might have attached to it. After that, the data will be formatted and validated according to the system requirements and later on reorganized and summarized for easier decision making practices.

With the uniformly formatted data once it enters the data warehouse, On Point Analytics will ensure there is consistency in the data that they store. These will make it easier for both the employees and corporate decision-makers that might be hired to analyze and share the company's data insight with their stakeholders and colleagues. This will also ensure the data is free of errors and thus accurate to be used. With accurate data, it will also be easier for the company to carry out decision making and make informed decisions important in the running of the company. These will also improve the speed at which On Point Analytics will be able to access the data and thus make decisions on time for analysis. This way, they can easily come up with decisions to improve business and marketing strategies that will see the company defeat its major competitors. With an established warehouse, the company can easily refer to past decisions that are easily available to the team. This way, they can gain access to historical strategies and decisions that could be presently successful.

**Best practices**

The company will be using both on-premise and cloud data warehouses. On Point Analysis Inc. has an on-premise storage space of 10 TB which is set to be increased over the years. This storage will, however, be backed up to the cloud to increase security and reduce the risk of data loss. With the On-premise warehouse, the company will have full control over the database and thus implementing strict and functional data security policies and devices to protect the data. A basic security device here will be a firewall that will be used to filter all the data as it comes in. While it is possible to lose data while storing it at a physical location, the backups available at the cloud will lay a major role for insurance against loss of data. The company won't have to worry about data loss given that they have readily available back up at the cloud. This will also be cost-effective given that the On Point Analytics Inc. will only be required to pay for the storage space that is needed and therefore easier expansion of storage (Lawyer, 2).

The company will also employ the extract-transform-load, ETL workflow, whereby the data will have first to be transformed before being loaded. This way, the company won’t have to do further transformation before analyzing and reporting. This way, the company will only transform data when it is needed and thus won't need to include the transformation logic in the design of the data flow. With the uniformly formatted data, however, the data stored won't face problems such as time wasted in transforming the data at the time when it is needed. This would, therefore, minimize issues with unstructured data (El-Sappagh, 3).

The company will make use of completely custom-built tools, which will ensure the ETL tool is built to suit the needs of the organization and thus more secure. Jobs at On Point Analytics Inc. will be done faster with the ETL tool being able to automatically create tables to store data with just a few entered data. Despite being rather expensive, the company will save greatly on future security updates and additional features given that most will be added at the development stage.

**Database schema**

**Organizational database**

The organization deals with customers who are after getting the best services from On Point Analytics Inc. The company also has a team of well-qualified individuals who work hand in hand across the various departments to ensure the satisfaction of our customers. These departments all fall under the management of the company with some outsourced employees. The information system is set to store details of the customers, employees and the outsourced employees. The system will also have a table for all the services that are offered with the cost and times of doing the services. The services database will store all the details of the services that are offered by the company; these data will include service offered, service id, cost of service, and duration of the service. These databases will be linked to the employees’ database whereby it will be easier to assign the employees various tasks (See Appendix C)

**Employee Database**

At On Point Analytics Inc., The employees are assigned various duties based on their departments. Each employee works under a department manager who oversees the whole department. The database will store the information of every employee which includes but not limited to their ID, name, department. Date of birth, address, email, phone number, and salary. The ID which is unique for every employee will be used as the primary key (See Appendix B)

**Customer database schema**

The list of customers is mainly organizations and a few individuals who are after the services of On Point Inc. The database will be built to keep records of these customers such as their ID, which will be auto generated, name of the organization, service offered, service Id cost, date, their views on the services that have been provided. The customer database will also include a table for the loyal and consistent customers which will be generated based on the number of times they sought the company’s services. Alongside this will be a table for the invoices which are to keep records for the finances of the organization. The table will include the service provided, invoice number, cost and other relevant information on the organization. The customer ID will work as the primary key with the invoice number as a foreign key (See appendix A)

**Entity-Relation (E-R) diagram**

The various databases will have relationships to connect the different entities in the design. These relationships are to be used in making the collection of reports and in carrying out an analysis of data from different departments and sections of the company. For instance, the employees' table will be linked to the department that they belong to. This relationship will be one to many, whereby the department will be the primary key of the relationship, given that it is both tables. The employees' table will then be linked to the customers' table given that each customer will be assigned a team to work on the project. This would hence facilitate the identity of employees who are working on a certain customer’s order.

Some entities, like the customers' table and the employee table, will be considered as strong in the E-R diagrams. In the employees' table, the employee’s address will be considered as a weak attribute, given that they are not of much importance like the employee id. (See appendix D)

**Data Flow Diagram**

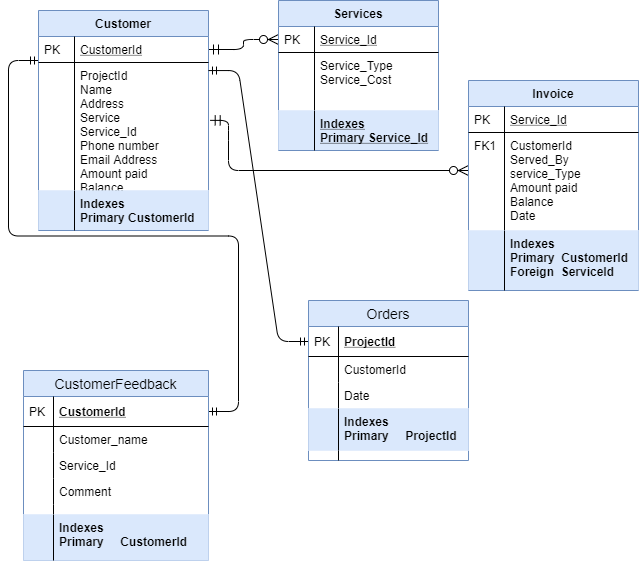
The Flow of data in the company will start by the customer placing an order. By placing an order, it means requesting the services of the company. The client can either pay the full amount and given a receipt or pay the first installment. This project once accepted, will be assigned to a team of employees with a specialty in the type of project. The process goes through different phases to the final step, where the finalization of the project. The client feedback is then collected and if satisfactory, the client would be guided to clear the balance before an invoice being generated. All this information would then be backed up to the cloud for increased security (See appendix E).

**References**

1. Krishnan, K. (2013). *Data warehousing in the age of big data*. Newnes.
2. Lawyer, J., & Chowdhury, S. (2004, January). Best practices in data warehousing to support business initiatives and needs. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the* (pp. 9-pp). IEEE.
3. El-Sappagh, S. H. A., Hendawi, A. M. A., & El Bastawissy, A. H. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University-Computer and Information Sciences*, *23*(2), 91-104.

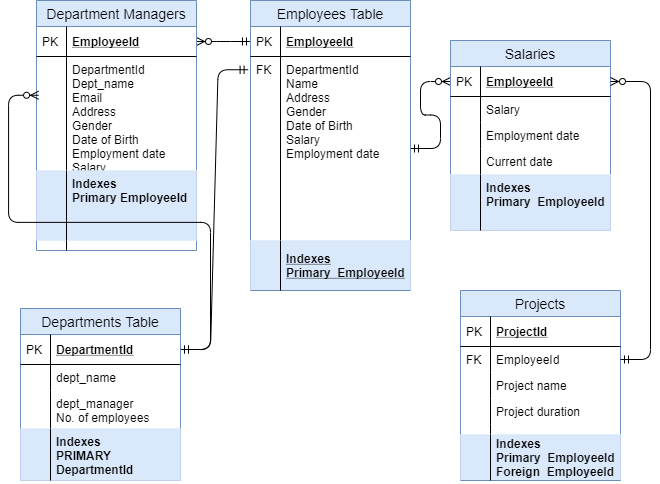
**Appendix A**

**The database schema for customers**



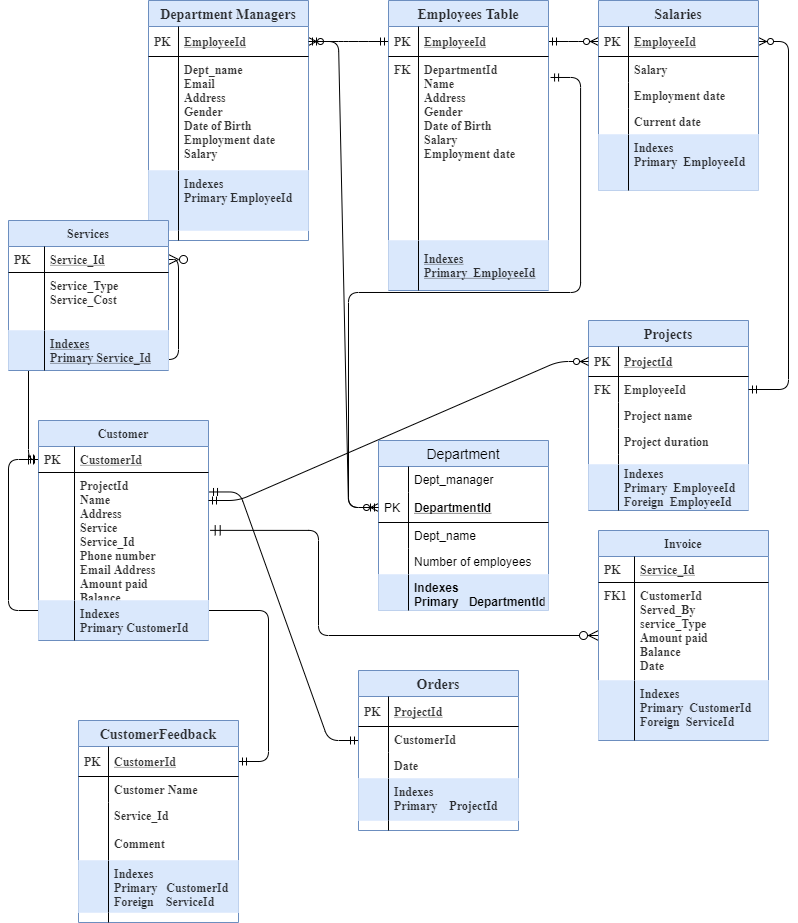
**Appendix B**

**The database schema for Employees**



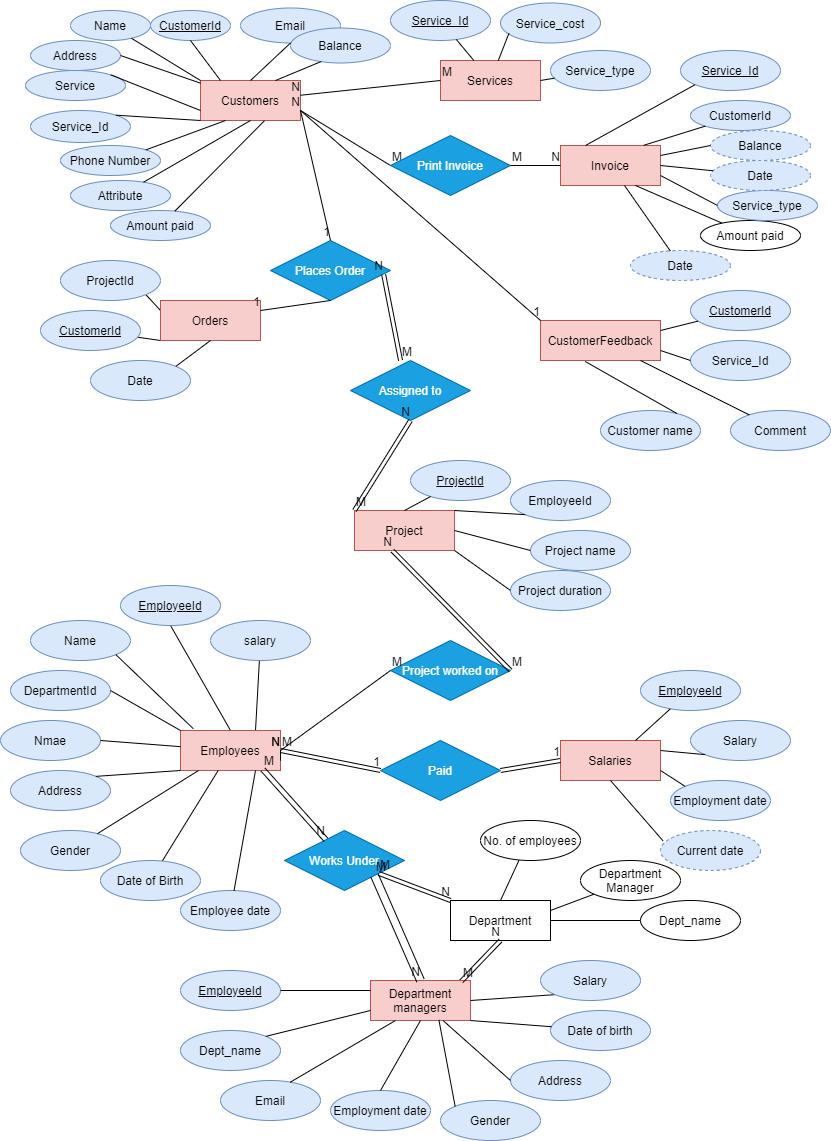
**Appendix C**

**Combined Database Schema for On Point Analytics Inc.**



**Appendix D**

**Entity-Relation (E-R) Diagram for the Company**



**Appendix E**

**Data Flow Diagram**

