Sales Representative Promotion: Data Centric Systems

Date: 11/25/2021

Author: Neeraj Reddy Gunda

EXECUTIVE SUMMARY

The pertinent aim of the system is to find out the top 50 company sales representatives with the highest sales in the last four years. This is done so that they can be promoted to higher ranks. The average sales per year, work consistency, the name of the organization, and the overall work experience will be the core areas of focus. The system will help HR avoid assumptions and favouritism when promoting competent sales representatives. Secondly, it will help limit the number of promotion applications by the sales representatives to a small number hence saving time and money. This is because the system gives important factors to consider determining whether a sales representative is eligible for promotion or not.

The core objective- is to develop a system that can solve problems encountered by the HR department and managers when making employee promotion decisions. In the project, we will be developing competence-based analytic which will help the sales representative analyse their competence and evaluate whether they are eligible for promotion or not based on the factors given to consider.

In developing the system, multiple machine learning regression algorithms were evaluated, and the information collected was stored in the MongoDB database. Besides, a simple user interface will be created to help the HR and Managers input data related to sales representatives and obtain the promotion eligibility results.

Contents

1. SYSTEM OVERVIEW

The system is grounded on three crucial components:

- The first component is gathering model data using the strategy of web scraping. In web scraping, several python libraries like requests & Beautiful soup are used. For data pre-processing, pandas, NumPy and matplotlib are some of the libraries used. Sklearn is also used for model training and testing.
- The second component of the system entails training and testing different machine learning models. The common machine learning models used in the system are learning association, classification, prediction, regression is used in examining the accuracy of the dataset used (Bandi & Tulabandhula, 2020).
- The last component is deploying the model in a Web application using UML and MongoDB.

The following are the user class that will use this product:

1. HR departments and managers focus on promoting competent and high-performing sales representatives in their company.
2. Sales representatives were interested in determining their work promotion eligibility in their company.

The system is highly dependent on new data.

There is room for insight and adding more information to support the systems since no existing data can be used.

2. DEFINITIONS AND TERMINOLOGIES

DaaS- Data as Service

RPO- Recovery Point Objective

RTO- Recovery Time Objective

RAID-Redundant Array of Independent Disks

HR- Human Resource

HTML- Hypertext Markup Language

CSS- Cascading Style Sheet

HTTP-Hypertext Transfer Protocol

API- Application Programming Interface

JBO- Just Bunch of Disks

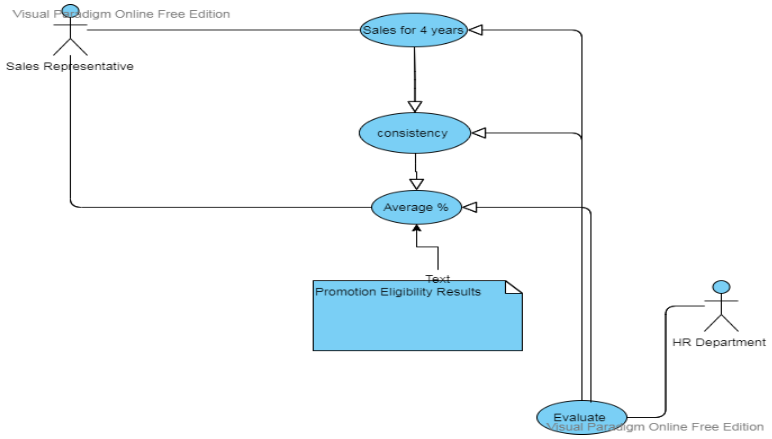
3.OPERATIONAL OVERVIEW

3.1 Usage Scenarios

User: HR Department.
Problem: Needs to find out the top 50 sales representatives eligible for the promotion.
Actions: Evaluate work competence by finding out the highest average sales for 4 years and sales representative consistency.
Results: Sales representatives with high sales and consistence in the last 4 years are selected for promotion.
Failure Case:

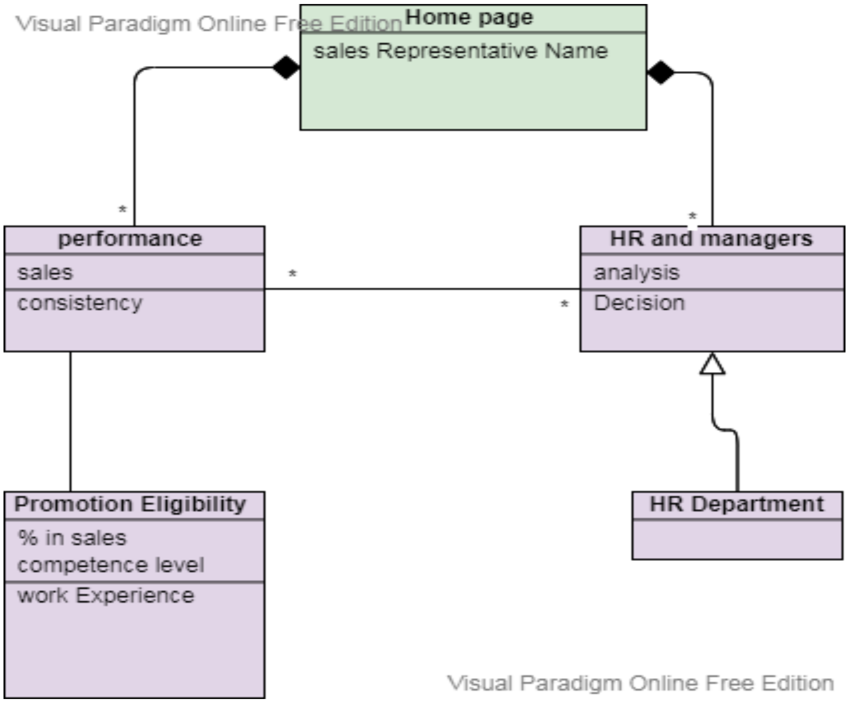
1. If the sales representatives with the highest sales and work consistency cannot be identified: The HR department will not be able to select the right competent sales representatives for promotion.
2. If the analytics are not done correctly: Both competent and incompetent sales representatives will be promoted.

3.2 Usage Cases

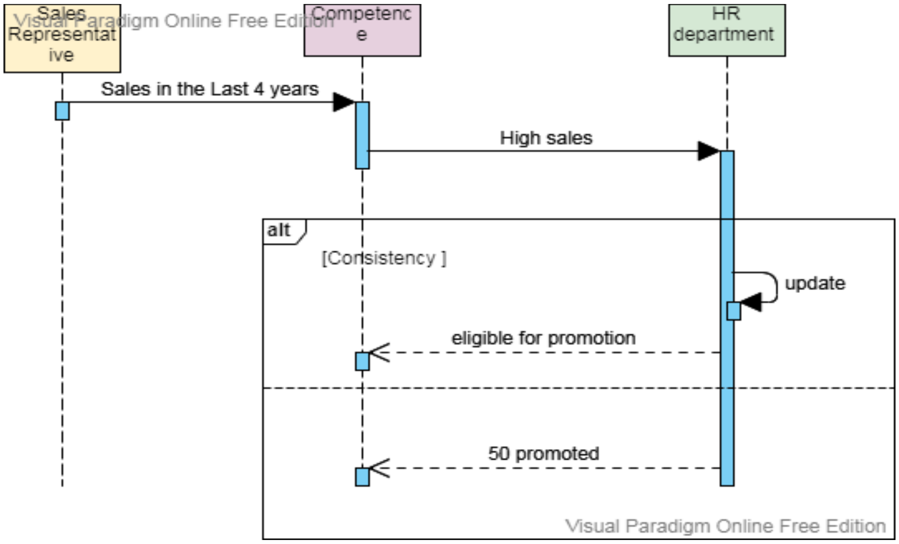

Visual Paradigm Online Free Edition
Sales Representative
Sales for 4 years
consistency
Average %
Text
Promotion Eligibility Results
HR Department
Evaluate
Visual Paradigm Online Free Edition

https://online.visual-paradigm.com/app/diagrams/Class Diagram

Use Case Diagram



https://online.visual-paradigm.com/app/diagrams/Class Diagram



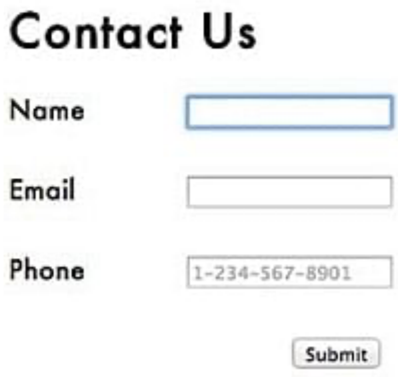
https://online.visual-paradigm.com/app/diagrams/Class Diagram

3.3 Data Policies and Constraints

Personal privacy and consent is required when evaluating the data used in the system. The policy is set to ensure that personal information is made confidential to avoid negative repercussions associated with exposing individual information. Database constraints liked to data integrity should be made a key area of focus (Bandi & Tulabandhula, 2020). Integrity constraints of rules is set to ensure that entity integrity, referential integrity and domain integrity of the data structure is preserved. Precisely, the data policies that will be set in the system will revolve around personal data privacy and data integrity.

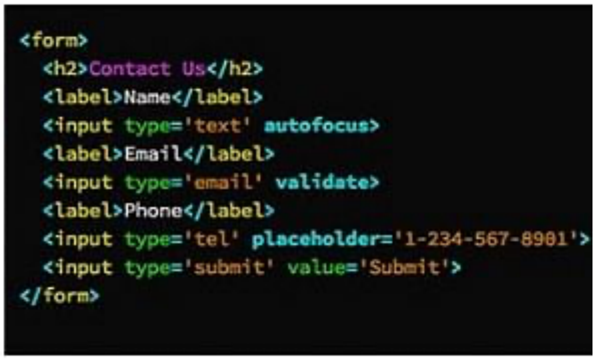
4 DATA INTERFACE DESIGN

The data interface design being provided by the system is simple. The sales promotion prediction system is build using different webpage components that are developed by using HTML and CSS. The HTML and CSS is used in building, running and operating applications on webpage. The contents of the HTML and CSS used in the website are shown in the screenshots below.

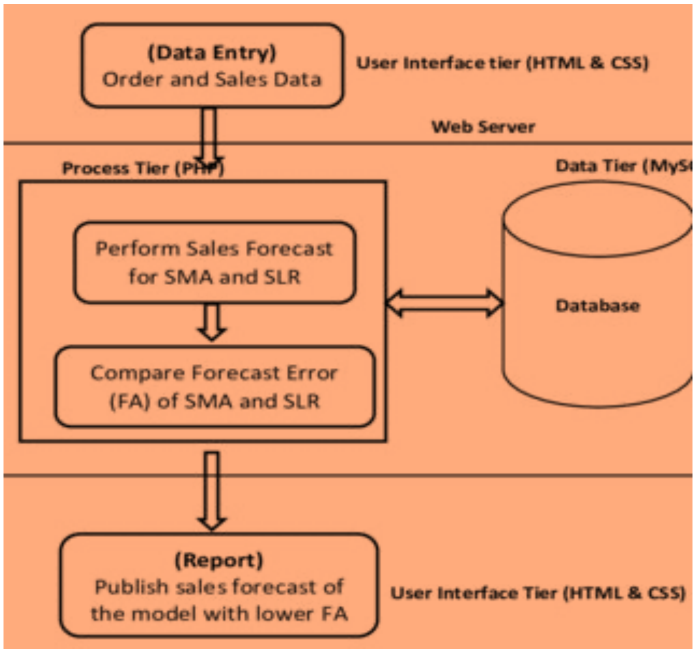


Contact Us

Name	
Email	
Phone	1-234-567-8901



```
<form>
  <h2>Contact Us</h2>
  <label>Name</label>
  <input type='text' autofocus>
  <label>Email</label>
  <input type='email' validate>
  <label>Phone</label>
  <input type='tel' placeholder='1-234-567-8901'>
  <input type='submit' value='Submit'>
</form>
```



(Data Entry)
Order and Sales Data
User Interface tier (HTML & CSS)
Web Server
Process Tier (PHP)
Data Tier (MyS
Perform Sales Forecast for SMA and SLR
Database
Compare Forecast Error (FA) of SMA and SLR
(Report)
Publish sales forecast of the model with lower FA
User Interface Tier (HTML & CSS)

4.1 Interface Data Structure

The main data structure used in designing the system is the employee performance prediction tool (Qin et al., 2020). The tool focuses on analysing data from the website dataset to bring the relevant information needed by HR in making promotion decisions.

4.2 Interface Operations

The interface operations used in designing the system are diverse. The main operations are add*, enter, press, ok, and output.* The datasets attributes such as the average sales per year, work consistency, the name of the organization, and the overall work experience are added to the promotion prediction tool (Ye et al., 2020). After adding the attributes, enter and ok icon is pressed to predict the output.

4.3 Interface Protocol

Interface protocol entails the processes of data delivery and interpretation. The HR department is the user in the system who needs to be provided with adequate data for interpretation. Therefore, HTTPS needs to be implemented by the system to ensure a strong secure connection to the user (HR department) (Bandi et al., 2020). In addition, since the HR department will be required to log in to their website and review the sales of individual sales representatives for the last 4 years, HTTPS will be useful in ensuring sensitive data from the sales representatives is secured.

4.4 Interface Reliability

The data interface is asynchronous as it entails streaming data from the sales representatives' company websites. In guaranteeing the reliability of data transfer, several factors are considered important. First, accuracy in collecting data is made a priority. HR department experts with excellent prowess in data collection are given the responsibility of collecting sales representative data from the websites. Secondly, a strong internet connection is guaranteed to ensure easy data accessibility and faster transfer. Therefore, for the interface to be reliable, the concerned parties should focus on network latency (Ye et al., 2020). Furthermore, high bandwidth is needed to overcome the network traffic for reliable data transfer. Ultimately, interface reliability in data transfer can be guaranteed by eliminating spyware and viruses because they result in slowed internet connection speed.

4.5 Interface Security

Since the data being transferred is sensitive, interface security should be made a top priority. First, a firewall should be set up on the website. The firewall will help in protecting the company network by controlling internet traffic coming and leaving the company. Besides APIs using HTTP and transport layer security TLS, are used in providing interface security in the system (Yu et al., 2021).

5 DATA MODEL DESIGN

5.1 Type of Data Model

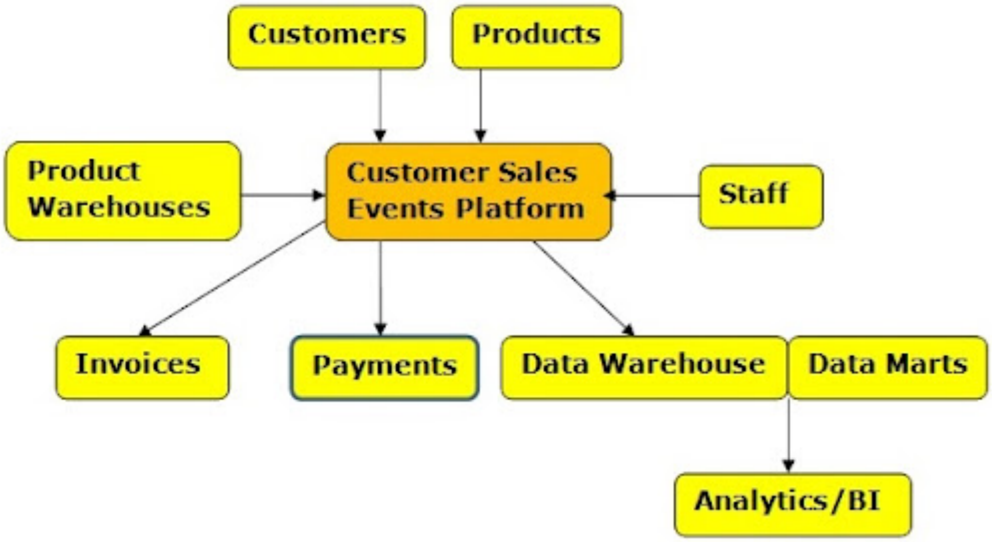
The type of data model used in the sales representative promotion system is relational. This is because the database is built in a structured manner that clearly defines the database relationships (Bandi et al., 2020). For instance, the sales for each sales representative for four years is determined by work consistency. Afterward, the average sales are calculated, and the promotion eligibility results are given as output to the HR department.

5.2 Data Structures

The data structures used in the systems revolve around the conceptual data model and logical data model explained below.

- The conceptual data model shows the relationship of concepts between the sales representatives and the company to aid in predicting promotion eligibility. The concept of average and work consistency is profoundly determined by the conceptual data model.
- A logical data model is also used in showing the relationship existing between the sales representative and the HR department. From the sales analysis, work consistency, average sales for four years, and ultimately the eligibility promotion results, logic is significant in determining the results. Therefore, the data structure used in the system is built under both conceptual data model and logical data model.

The following conceptual data model explains how performance of the sales staffs is analysed:



The payment and the sled made by the sales staff is analysed to get their performance score as shown in the data structure model diagram above.

5.3 Data Integrity Constraints

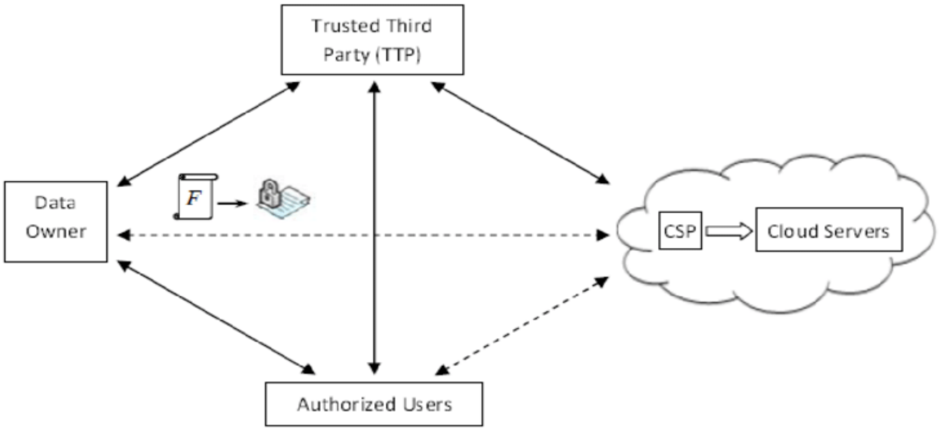
The data model uses error-checking and validation methods in ensuring data integrity is maintained.

- Erro-checking test is conducted in the data transfer and interpretation purposely to measure the ability to identify errors in a data. For instance, a wrong time and content are presented in the database, and error-checked and detected.
- Moreover, the validation method works by integrating valid input in the data model. The *validate input* works by verifying the unknown and known data and ultimately validating it to ensure that the data is accurate. Besides, the data model also removes duplicate data to ensure data integrity is preserved.

6 CAPACITY PLANNING

6.1 Storage

The system requires different storage needs to ensure that it performs effectively. Based on the data structure and data type, the disk space or storage size needed for the system is 50GB (He et al., 2018). This is achieved from the five fields with a data type size of 10GB. The fields include; the HTML server, CSSS, average sales for the sales individuals, and 2 records of the performance of the sales individual for the last 4 years. The indexing structures for the system accumulate to 95% since it is a data warehouse. Concisely, the initial storage for the system is calculated using the method: Data type size * a number of fields (10GB*5) =50. Ultimately, storage fundamentals such as the Block I/O, Redundant Array of Inexpensive Disks (RAID), and Just Bunch of Disks (JBOD). These fundamentals are important in helping create data redundancy of the system. RAID is the main storage fundamental that the system will be relying on most of the time. RAID 5 – striping with parity (popular) is the level of the RAID that is considered. This is because the data redundancy will be created effectively in the system. The data storage process is conducted as shown in the diagram.



6.2 Data Retention

- The data in the system will be retained for 1 year. This is because the system will only be applicable in the company for 1 year and end when the promotion ends. Since data security is a core factor of cooperation, retaining the data for 1 year will not be difficult. The data will be safe and no risks issues will be detected.

6.3 Processing

- The processing capacity of the system is sizable to ensure that data access and update requests are complete promptly and effectively. In addition, peak utilization is used as an essential tool in ensuring the responsiveness of the system is high.
- The HR department and IT department will hold hands to ensure that there are enough resources to process the average number of concurrent requests being made to the system. Concisely, the entire processing capacity of the system is built on enough resources guaranteeing effectiveness and high sustainability to the users.

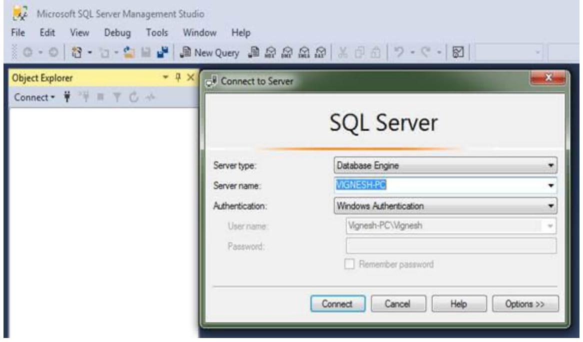
6.4 Network

- Based on data access and update, the system needs a network bandwidth of 10 million bits/second (Mbps). This bandwidth will help easy and faster and access of information. 10 million bits/second (Mbps) provides the strongest network for the system.

7 MONITORING

The monitoring capabilities for the sale promotion system will revolve around the monitoring metrics, the useful information about the system derived from the monitored information, and ultimately the monitoring tools used to ensure system performance is being observed (HUANG et al., 2018).

The monitoring metrics applied by the system are observable Architecture. Therefore, to ensure that the sale promotion system is observed, several monitoring tools were initiated. The common monitoring tools used in the project were SQL Server Management Studio, Oracle Administrator, and MonYog (MariaDB). The main reason behind choosing these monitoring tools is that these tools are effective and will aid in reducing database downtime and improve problem resolution times. Again, the tools are effective in noticing database problems before they occur. A screenshot of how the SQL Server Management Studio will work in monitoring the system performance. Besides, in ensuring the monitoring metrics are working properly, the recovery point objective (RPO) for the system should be 60 minutes and the recovery time objective (RTO) should be 12 hours. The monitoring process should be built on backups to avoid cluster failures that might lead to loss of the critical information in the system.



Server type: Database Engine
Server name: VIGNESH-PC
Authentication: Windows Authentication
User name: Vignesh-PC\Vignesh
Password:
□ Remember password

8 PERFORMANCE CHARACTERISTICS

The Daas needs to meet some specific performance goals to ensure it is effective and promising to the company. One of the key performance goals is ensuring that the system is providing the best performance for the users by tuning. Another performance goal that the system should meet is ensuring efficient use of resources. Storage, network, memory, and processing should be utilized to ensure the performance of the system is at a high level. Tuning the system will help inefficient usage of the resources. When the usage of the resources is not balanced, the performance of the Daas system will be altered. Besides, the system should meet the goal of allowing unexpected traffic spikes. This will help avoid unexpected system interference.

8.1 Query Performance Tuning

Tuning the query performance will involve different steps.

- The first step is defining the targets. Every target will have a specified response time evaluated through the sales performance.
- The second step is translating the targets to metrics being calculated in the sales representative's performance. This is key in ensuring that monitoring metrics are established to determine when to tune.
- The third step is establishing a performance baseline to determine the number of clients needed per unit.
- The fourth step of tuning the performance queries is constructing test scenarios to determine the resource limit and design.

These steps are simplified and explained in the diagram below;

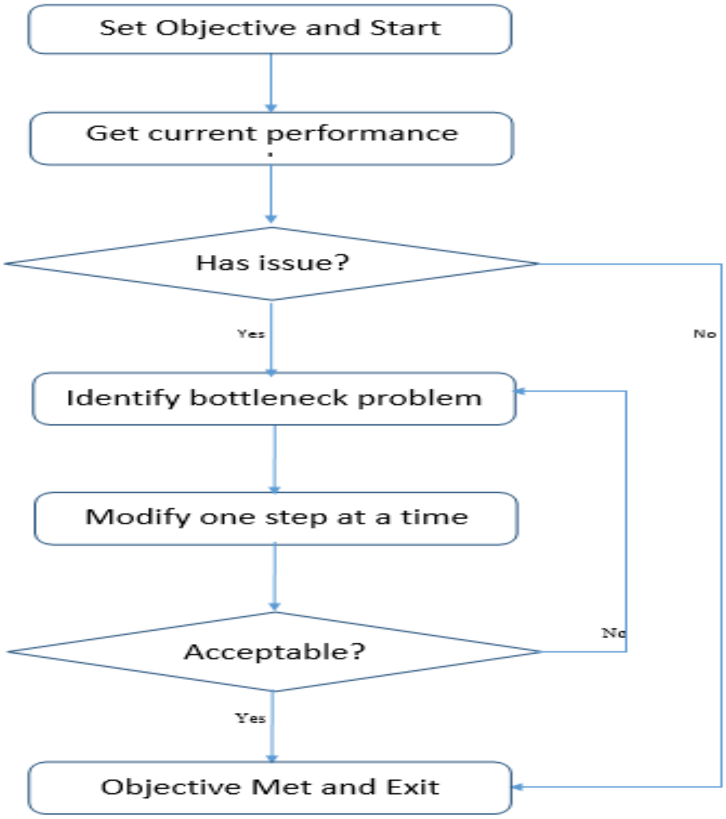


Fig: Performance tuning steps with a flow chart

8.2 Data Structure Performance Tuning

Data Structure Performance Tuning will involve the following

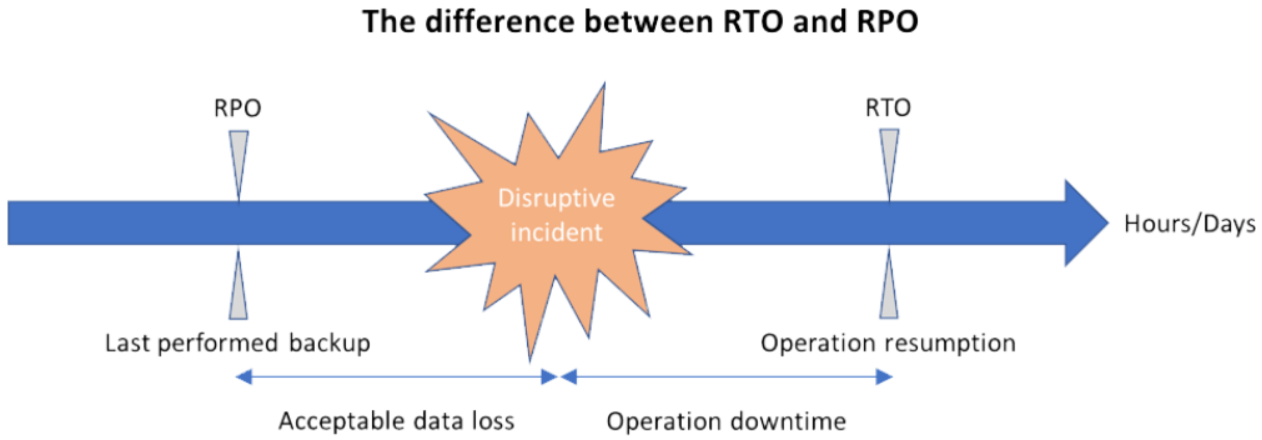
- Check if the service is dedicated to the application. This step will help ensure that the server memory matches the data structure.
- Focusing on normalization and denormalization to factor out the data structure that suits the tuning process. Every tool used in a data structure needs to have a primary key, therefore, this step helps in determining the right primary key for tuning.
- Achieving table data and removal of duplicate or unused indexes.
- Ultimately, use explain statement to determine how the data structures are executed.

9 HIGH AVAILABILITY DESIGN

9.1 High Availability Requirements

- The desired recovery point objective (RPO) for the system is 6 hours. This means that the system is prepared for backup after every 6 hours. In case of unforeseen failures such as hardware failures and power failures, the system will be able to recover the system information for the shortest time period (Mendonça et al., 2019).
- On the other hand, the desired recovery time objective (RTO) for the system is 24 hours. RTO will be enough for the system to recover the lost data and resume its functions.

The system works with 6 hours backup which ensures that the disruptive incident is identified, and solution made. The data lost in the starts to be recovered at the first minute all the way to the 6hours point. 6 hours gives enough time for the system to recover lost data. However, the system is given another 24 hours to ensure that everything is at its normal state. In addition, K8 clusters can be implemented during the backup process. The backup container rely on K8 clusters to run through virtual, physical, cloud-based, and on-premises phase. Precisely, the K8 design use containerization to ensure the backup process undergoes virtual, physical, cloud-based, and on-premises stages to achieve the RTO and RPO of the system.
The diagram below shows the differences between RTO and RPO estimation in the system.



9.2 High Availability Implementation

The system is designed in a manner that it is able to meet the Recovery Point Objective (RPO) and the Recovery time objective (RTO).

Active standby clustering is implemented to ensure high availability configuration. To manage high availability configuration, the system will be using containerization and Kubernetes. First, the container is designed by packaging a software code with its dependencies to create a single executable running consistently in the system. The software code design will be abstracted from the operating system to ensure that the RPO and RTO are properly monitored. In addition, containerization will be essential in the system because it can run write once and run anywhere for practical applications. Containerized applications have diverse advantages in a system.

The key aspects of Kubernetes that is focused on when designing the system include:

- K8s workloads, Kubernetes Secrets
- Kubernetes Service Accounts
- Kubernetes Metrics Service (Metrics API)
- Pod Autoscaling.

Such aspects are crucial in ensuring better performance in the system is achieved. In addition, these aspects of Kubernetes will help in recovering from classic failures.

10 BACKUP AND RESTORE DESIGN

The desired recovery point objective (RPO) for the system is 6 hours. This means that the system is prepared for backup after every 6 hours. In case of unforeseen failures such as hardware failures and power failures, the system will be able to recover the system information for the shortest time period (Hamadah & Aqel, 2019). On the other hand, the desired recovery time objective (RTO) for the system is 24 hours. 24 hour RTO will be enough for the system to recover the lost data and resume its functions. The system backups are done once a day to guarantee high data security and avoid losing valuable data. Afterward, the backups are stored in an external hard disk. In the event of a disaster, the system is able to meet RPO and RTO by frequent backups (Kassim et al., 2018). The log backups and database backups are able to meet the RPO and RTO of the system.

11.DATA SECURITY

11.1 Authentication

The sale representative promotion system will rely on the key security terms to ensure that only the authorized users are capable of accessing the organization's information (Ghaffar et al., 2020). The IT department will operates on the basis of controlling the system access purposely to create a safe working environment (Hamadah & Aqel, 2019). First, the system approves authorization for logical access of the information in a unique manner. To authenticate the users the system will uses passwords. However, only encrypted representations of the password are used. Precisely, to ensure that security is met by the system, the HR department will focus on training the old and new employees about strategies to use to ensure data security in the system. In addition, the system will meet the security requirement by employing experts in data security. The experts will foresee the activities conducted in and out of the organization purposely to block any new interference. Ultimately, the security requirement will be met by using strong passwords that are only accessible by authorized hands.

11.2 Access Control

The access control mechanism of the system works hand in hand with authentication. First, to guarantee data security, the system has disabled and restricted external executables. In addition, the system has enforced authorized access to all PKI private keys. This is to ensure that only the authorized hands can access data.

Besides, the system has implemented a policy articulating that database files are limited to relevant processes and authorized users. Ultimately, there are restrictions set in case of any configuration changes.

11.3 Encryption

Encryption is fundamental both for data at rest and data in transit. The key objective of the system is to maintain information integrity and confidentiality. This can only be achieved by encrypting data at rest and the data in transit (Pawar et al., 2021). NIST FIPS 140-2 is used by the system to validate cryptographic operation during encryption. Therefore, the system needs encryption for both data at rest and in transit to guarantee high information integrity and confidentiality.

11.4 Auditing

The system auditing mechanism is grounded on detailed organization information. For any audit failure, the system has the capability of shutting down. This is to ensure that no data is lost during the system failure. Additionally, all the audit information is protected from unauthorized read access to ensure that the auditing process is properly done. A centralized configuration of the content to be captured in the audit is provided as a core audit mechanism of the system. Ultimately, when security and permissions are modified, the system is set to generate audit records. The audited data is stored in the records of the audit table. Only the authorized teams are permitted to view the audit information. Managers, the audit team, and the IT department team is authorized to view the audit information.

12 References

Bandi, N., & Tulabandhula, T. (2020). Off-Policy Optimization of Portfolio Allocation Policies under Constraints. *arXiv preprint arXiv:2012.11715*.

Ghaffar, Z., Ahmed, S., Mahmood, K., Islam, S. H., Hassan, M. M., & Fortino, G. (2020). An improved authentication scheme for remote data access and sharing over cloud storage in cyber-physical-social-systems. *IEEE Access*, *8*, 47144-47160.

Hamadah, S., & Aqel, D. (2019, April). A proposed virtual private cloud-based disaster recovery strategy. In *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* (pp. 469-473). IEEE.

He, S., Cheng, B., Wang, H., Xiao, X., Cao, Y., & Chen, J. (2018, April). Data security storage model for fog computing in large-scale IoT application. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 39-44). IEEE.

https://online.visual-paradigm.com/app/diagrams/#LBank%20ATM

HUANG, R., KANG, H., SHI, J., & ZHU, Y. (2018). Probability forwarding strategy based on interface reliability in named data networking. *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, 01.

Kassim, M., Sahalan, M. M., & Uzir, N. I. (2018). Framework Architecture on High Data Availability Server Virtualization for Disaster Recovery. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *10*(1-5), 163-169.

Mendonça, J., Lima, R., Queiroz, E., Andrade, E., & Kim, D. S. (2019, June). Evaluation of a backup-as-a-service environment for disaster recovery. In *2019 IEEE Symposium on Computers and Communications (ISCC)* (pp. 1-6). IEEE.

Pawar, A. B., Ghumbre, S. U., & Jogdand, R. M. (2021). Privacy preserving model-based authentication and data security in cloud computing. *International Journal of Pervasive Computing and Communications*.

Qin, X., Luo, Y., Tang, N., & Li, G. (2020). Making data visualization more efficient and effective: a survey. *The VLDB Journal*, *29*(1), 93-117.

Sample-Data-Management-Policy-Structure-1.pdf (culturehive.co.uk)

Ye, C., Wang, H., Zheng, K., Gao, J., & Li, J. (2020). Multi-source data repairing powered by integrity constraints and source reliability. *Information Sciences*, *507*, 386-403.

Yu, C., Chen, S., Wang, F., & Wei, Z. (2021). Improving 4G/5G air interface security: A survey of existing attacks on different LTE layers. *Computer Networks*, 10853z.