Ideation Phase

Defining the Problem Statements

Date	26-09-2023
Team ID	4162
Project Name	Data Warehousing with IBM Cloud Db2 Warehouse

Data Warehousing with IBM Cloud Db2 Warehouse

Problem Definition and Design Thinking Introduction:

The task at hand is to data mine a diabetes data warehouse with Db2 warehouse in IBM cloud. To collect some of data set based on diabetes from different sources and combining into one data warehouse using ETL tool.

In this document, we will outline the problem statement, the steps involved in solving it, and the design thinking approach that will guide our project.

Problem Statement:

Objective: To bring together data from various sources, perform advanced data integration and transformation, and provide data architects with the tools to explore, analyse, and deliver actionable data for informed decision-making.

Data: We have multiple datasets containing various features for diabetes prediction (such as glucose, insulin, BMI, pregnancies etc). These data will be used to integrate and transform into a robust data warehouse.

Key Challenges:

1. Data Warehouse Structure: Define the schema and structure of the data warehouse to accommodate various data sources.

2. Data Integration: Identify data sources and design a strategy to integrate data seamlessly into the data warehouse.

3. ETL Processes: Plan and implement ETL processes to extract, transform, and load data into the warehouse.

4. Data Exploration: Design queries and analysis techniques to empower data architects to explore and analyse data.

5. Actionable Insights: Focus on delivering actionable insights by enabling informed decision-making based on data.

Design Thinking Approach

1) Empathize:
- Begin by identifying the key stakeholders and users of the diabetes data warehouse. This might include healthcare professionals, researchers, and data analysts.
- Conduct interviews, surveys, and observations to gain a deep understanding of their needs, challenges, and pain points related to diabetes data analysis.

2) Define:
- Based on the insights gained from the empathize stage, define a clear problem statement and specific goals for the diabetes data warehouse project.
- For example, the problem statement could be: "Improving access and analysis of diabetes patient data for healthcare professionals and researchers." **3) Ideate:**
- Organize brainstorming sessions with a cross-functional team that includes data scientists, IT specialists, and domain experts.
- Generate a wide range of ideas for data mining techniques, features, and functionalities that could address the identified problem and goals.

4) Prototype:
- Develop a high-level conceptual design of the diabetes data warehouse using Db2 Warehouse on IBM Cloud. This should include data models, architecture diagrams, and a rough outline of the user interface.
- Consider how to integrate various data sources, such as electronic health records, wearable devices, and research databases.

5) Test:
- Create a prototype or mock-up of the diabetes data warehouse and share it with representative users from the empathize stage.
- Collect feedback on the design, usability, and functionality. Use this feedback to refine the prototype and make necessary adjustments.

- Validate the Data Mining Results
- Verify the accuracy and relevance of the data mining results by comparing them to known diabetes research findings and clinical expertise.
- Ensure that the data warehouse meets performance and scalability requirements.

6) Develop:
- Once you have a validated prototype, start developing the diabetes data warehouse using Db2 Warehouse on IBM Cloud.
- Implement data mining algorithms and data preprocessing techniques to extract valuable insights from the diabetes data.

7) Iterate:
- Data mining and analytics are iterative processes. Continuously gather user feedback and monitor the performance of the data warehouse.
- Make iterative improvements based on user needs, changing data sources, and emerging best practices in diabetes research.

8) Deploy and Monitor:
- Deploy the diabetes data warehouse in a production environment on IBM Cloud.
- Set up monitoring and alerting systems to ensure the system is running smoothly and securely.

9) Train
- Provide training and documentation to users to help them effectively utilize the diabetes data warehouse.
- Establish a support system to address user questions and issues.

10)Evaluate Impact

- After the data warehouse has been in use for some time, evaluate its impact on diabetes research, patient care, and healthcare outcomes.
- Use metrics and key performance indicators to measure the success of the project and identify areas for further improvement.

Conclusion

In this document, we've outlined our approach to solving the problem of data mining a diabetes prediction model with db2 warehouse. We've defined the problem, identified key challenges, and laid out a design thinking approach that involves empathizing with users, defining objectives, ideating potential solutions, prototyping, testing, implementing, and iterating.

By following this approach, you can ensure that the diabetes data warehouse meets the needs of its users and delivers valuable insights for diabetes research and patient care while using Db2 Warehouse on IBM Cloud as the underlying technology infrastructure.