GURU NANAK INSTITUTIONS

BIG DATA ANALYTICS

UNDER THE GUIDANCE OF
MR.A.KRISHNA
(ASSISTANT PROFESSOR)

TEAM

AMOGH PITTI (LEAD) – 21WJ1A6905
A.SAMEEKSHA REDDY – 21WJ1A6901
KONDAPARTHY THARUN – 21WJ1A6935

PROJECT TITLE

Snowflake for Healthcare Analytics - Leveraging Clinical and Patient Data for Insights and Decision Making

TABLE OF CONTENTS

ABSTRACT

- Healthcare Potential: The effective use of clinical and patient data can improve patient outcomes, optimize healthcare services, and support evidence-based decision-making.
- Snowflake Platform:
  - A cloud-based platform known for scalability, performance, and accessibility.
  - Provides a secure, centralized system for storing and analyzing diverse healthcare data.
- Key Benefits:
  - Helps healthcare providers access actionable insights.
  - Supports strategic, data-driven decisions.
  - Enhances the overall quality of healthcare delivery.
- Project Focus: This project explores how Snowflake can be utilized to analyze clinical and patient data, enabling informed decisions that improve healthcare outcomes.

INTRODUCTION

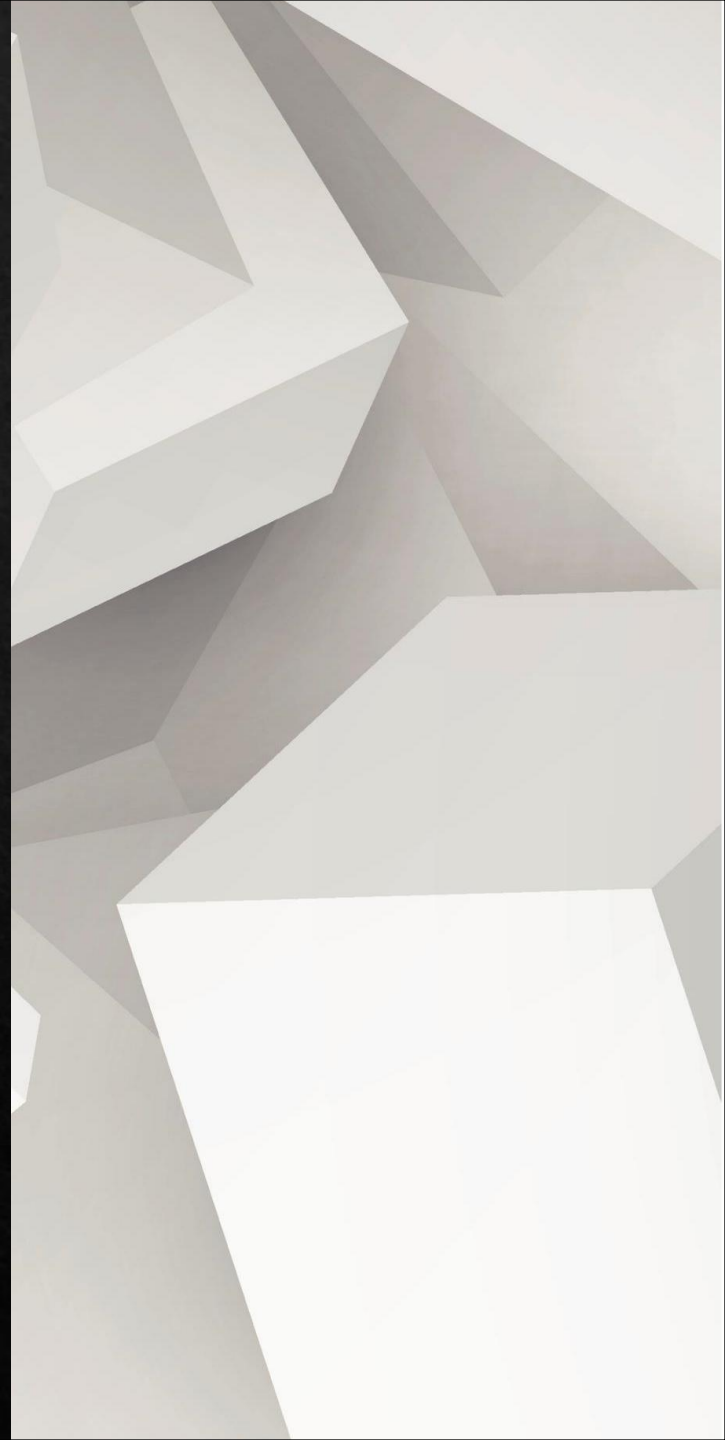
- The Role of Data in Healthcare:
  - Leveraging data is essential for improving patient outcomes and enhancing care delivery.
- Current Challenge:
  - The rapid growth of clinical and patient data makes it difficult to manage, analyze, and extract actionable insights.
- Introducing Snowflake as a Solution:
  - Snowflake provides a scalable, secure, and centralized cloud platform tailored for healthcare analytics.
  - Offers an efficient way to store, manage, and analyze diverse healthcare data.
- Project Goal:
  - This project explores how Snowflake can transform raw healthcare data into valuable insights.
  - Supports evidence-based decisions, ultimately improving patient care and operational efficiency.

EXISTING SYSTEM

- Challenges in Data Management:
  - Managing clinical and patient data is increasingly difficult due to data complexity and volume.
- Reliance on Traditional Systems:
  - Many healthcare providers use outdated systems that cannot handle modern data requirements.
- Data Fragmentation:
  - Data is often stored in disconnected platforms, making consolidation and effective analysis challenging.
- Limited Real-Time Insights:
  - Lack of system integration prevents real-time analysis, resulting in missed opportunities for timely, data-driven decisions.
- Data Security and Compliance Issues:
  - Ensuring data security and regulatory compliance adds further complexity to data management.
- Impact on Innovation and Efficiency:
  - These limitations hinder innovation and prevent healthcare organizations from fully improving patient outcomes and operational efficiency.

PROPOSED SYSTEM

- Centralized Data Repository: Implement a cloud-based platform (e.g., Snowflake) to consolidate and manage clinical and patient data from various sources.
- Real-Time Data Processing: Enable real-time analysis to support timely decision-making and improve patient care.
- Advanced Analytics Tools: Integrate machine learning algorithms for predictive analytics and evidence-based clinical decisions.
- Enhanced Security and Compliance: Ensure robust security measures and compliance protocols to protect sensitive healthcare data.
- User-Friendly Dashboard: Develop an interactive dashboard for healthcare professionals to visualize data insights and monitor key performance indicators.
- Training and Support: Provide comprehensive training and ongoing support for healthcare staff to effectively utilize the new system, ensuring they can leverage data insights to improve patient care.

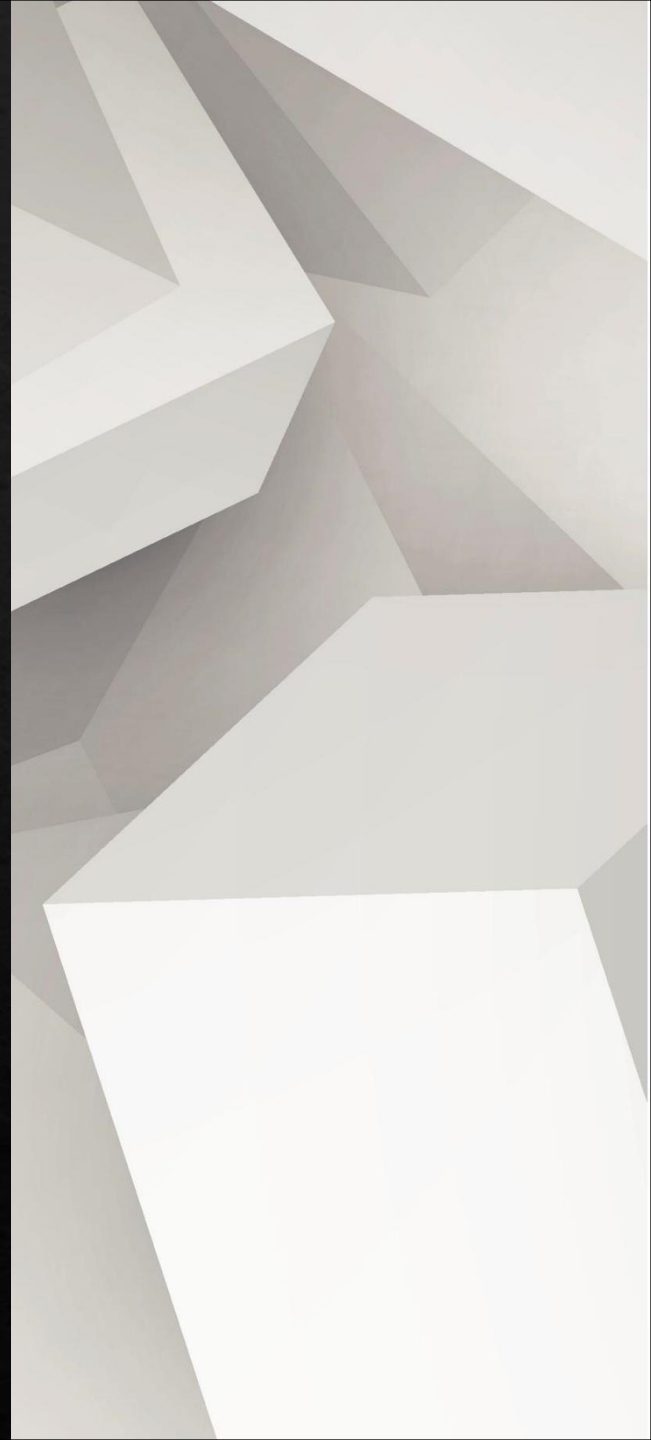


LITERATURE SURVEY

- Advancements in Big Data Analytics (2022): The integration of Big Data with AI and IoT has revolutionized healthcare by enhancing decision-making and addressing operational challenges.
- Challenges in Data Integration (2022): Fragmented healthcare systems highlight the need for unified platforms to enable effective analytics and comprehensive data analysis.
- AI and Machine Learning in Analytics (2023): AI models are increasingly applied to predict patient outcomes, improve diagnostics, and personalize treatments, driving efficiency in healthcare.
- Future Trends in Healthcare Analytics (2023): Real-time data processing and predictive analytics are crucial for streamlining operations and solving complex healthcare challenges.

METHODOLOGY

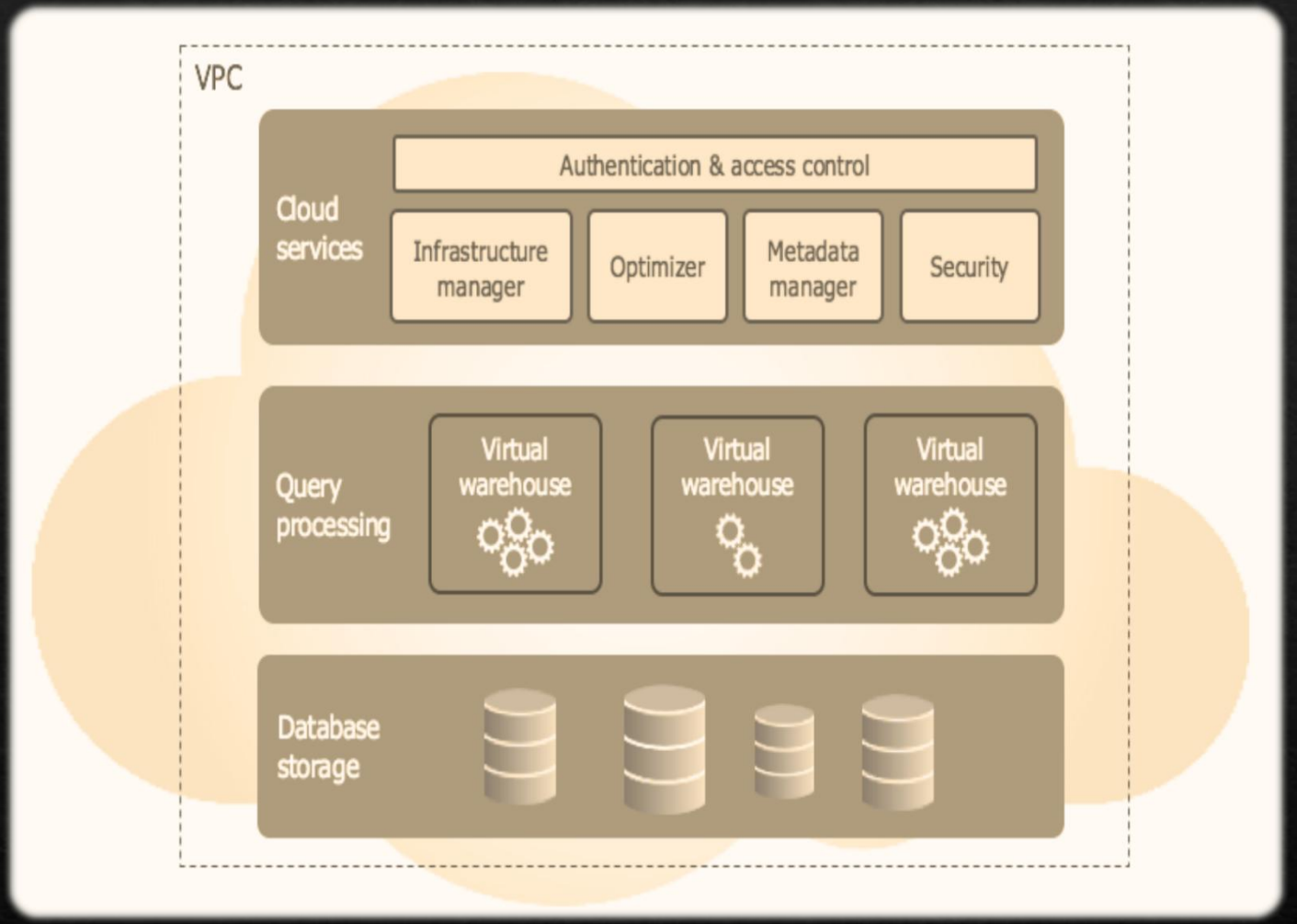
- Data Collection: Review literature on healthcare analytics and identify relevant clinical and patient data sources for integration into Snowflake.
- System Design: Develop a centralized data architecture in Snowflake, defining schemas for managing both structured and unstructured data.
- Analytics Implementation: Integrate machine learning algorithms for data processing and insights, creating workflows for data ingestion and transformation.
- Dashboard Development: Create an interactive user-friendly dashboard for healthcare professionals to access and visualize data insights.
- Testing and Deployment: Conduct system testing for accuracy and compliance, train staff, and deploy the system while collecting user feedback for ongoing improvement.



CHALLENGES AND SOLUTIONS

- Data Fragmentation:
  - Challenge: Healthcare data is scattered across multiple systems, complicating analysis.
  - Solution: Use Snowflake to create a centralized repository for seamless data integration and access.
- Data Security and Compliance:
  - Challenge: Protecting sensitive patient information while ensuring regulatory compliance.
  - Solution: Leverage Snowflake's security features and establish robust compliance protocols to maintain data integrity.
- User Adoption:
  - Challenge: Resistance to new technologies among staff.
  - Solution: Provide comprehensive training and ongoing support to enhance user confidence and proficiency.

SYSTEM ARCHITECTURE



- Cloud Services Layer:-

  Manages key functions like authentication, access control, metadata management, query optimization, and security. Ensures compliance with healthcare standards like HIPAA, while optimizing query performance.

- Database Storage Layer:-

  Centralized, scalable cloud storage stores patient records, clinical data, and other healthcare datasets. The separation of compute and storage allows for cost-effective, flexible scaling.

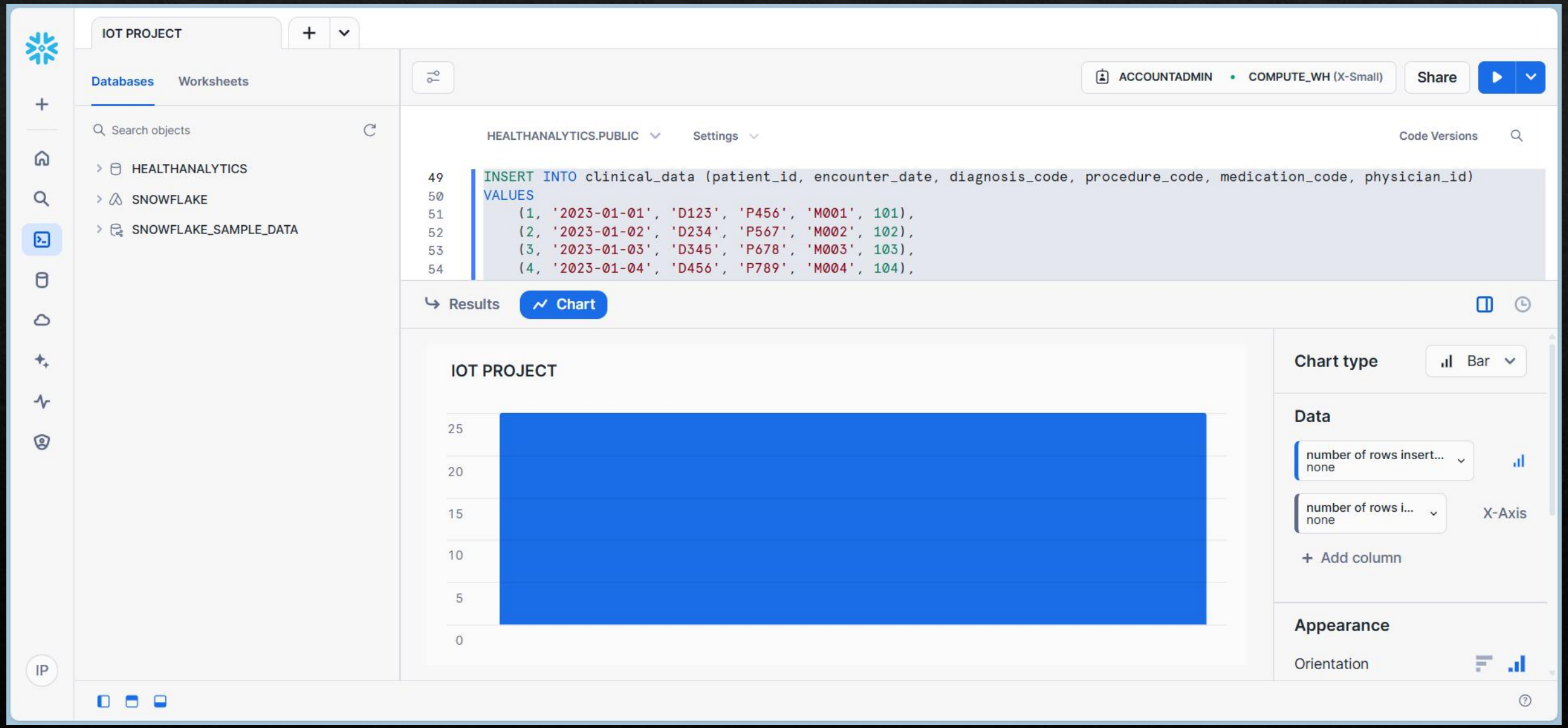
- Query Processing Layer (Virtual Warehouses):-

  Virtual warehouses perform data processing independently, enabling efficient execution of real-time analytics, batch processing, and advanced machine learning tasks. Multiple warehouses support isolated workloads for better performance.

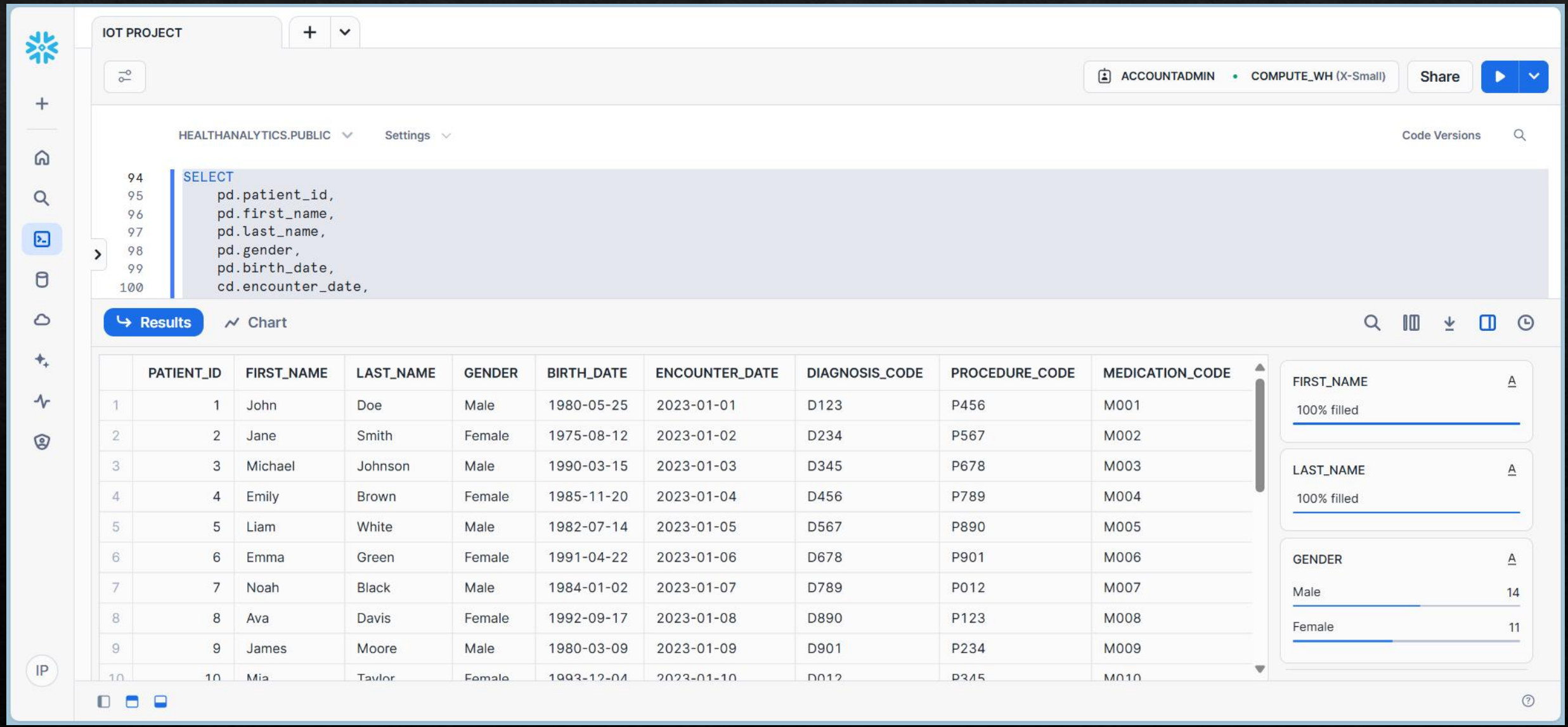
- Insights Delivery:-

  Processed data is delivered through dashboards, reports, and AI models, providing actionable insights to healthcare professionals, researchers, and administrators for decision-making and improving patient care.

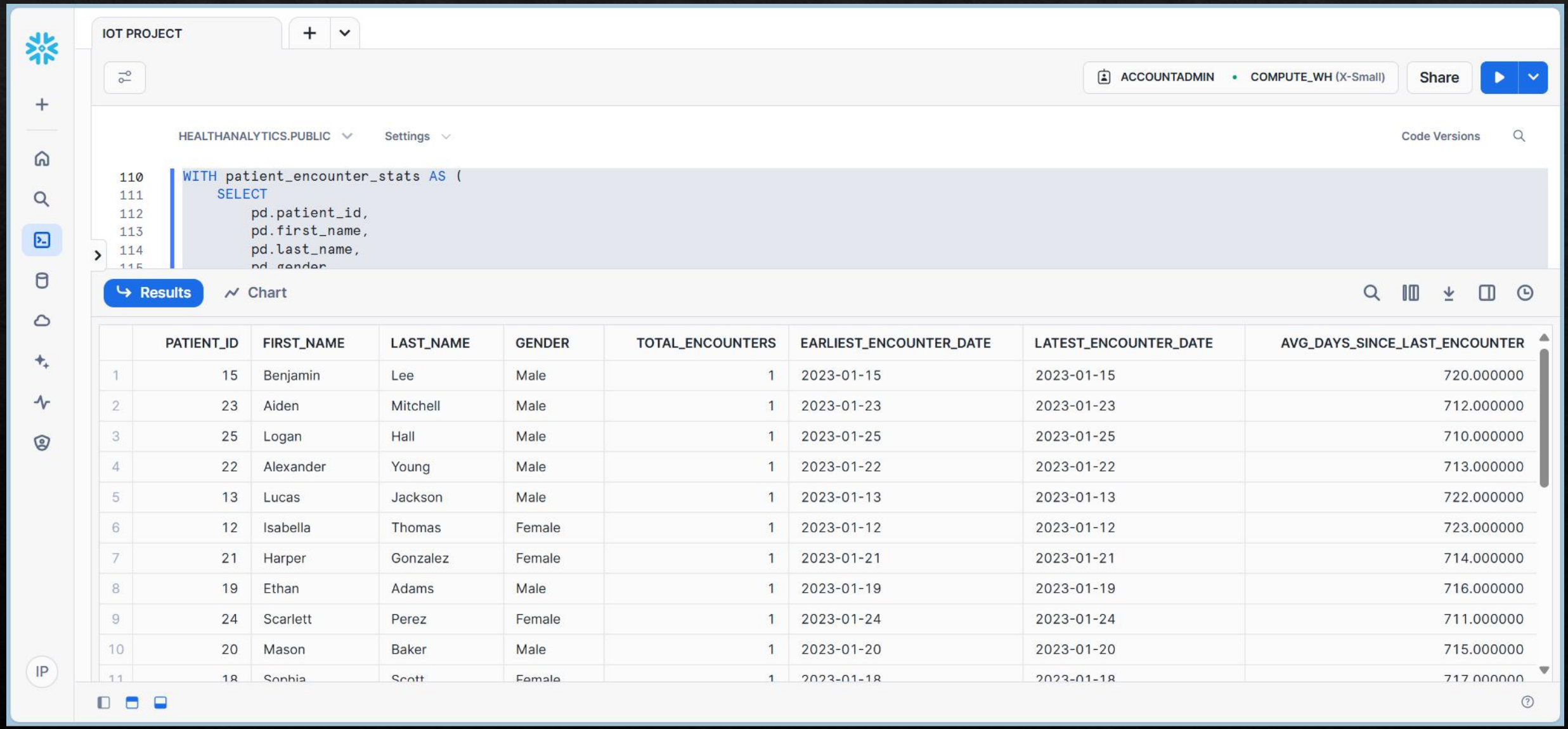
OUTPUT


IOT PROJECT
Databases
Worksheets
Search objects
HEALTHANALYTICS
SNOWFLAKE
SNOWFLAKE_SAMPLE_DATA
ACCOUNTADMIN
COMPUTE_WH (X-Small)
Share
HEALTHANALYTICS.PUBLIC
Settings
Code Versions
49 INSERT INTO clinical_data (patient_id, encounter_date, diagnosis_code, procedure_code, medication_code, physician_id)
50 VALUES
51 (1, '2023-01-01', 'D123', 'P456', 'M001', 101),
52 (2, '2023-01-02', 'D234', 'P567', 'M002', 102),
53 (3, '2023-01-03', 'D345', 'P678', 'M003', 103),
54 (4, '2023-01-04', 'D456', 'P789', 'M004', 104),
Results
Chart
IOT PROJECT
25
20
15
10
5
0
Chart type
Bar
Data
number of rows insert...
none
number of rows i...
none
X-Axis
Add column
Appearance
Orientation
IP

OUTPUT


IOT PROJECT
ACCOUNTADMIN • COMPUTE_WH (X-Small)
Share
HEALTHANALYTICS.PUBLIC
Settings
Code Versions
94 SELECT
95 pd.patient_id,
96 pd.first_name,
97 pd.last_name,
98 pd.gender,
99 pd.birth_date,
100 cd.encounter_date,
Results
Chart
PATIENT_ID FIRST_NAME LAST_NAME GENDER BIRTH_DATE ENCOUNTER_DATE DIAGNOSIS_CODE PROCEDURE_CODE MEDICATION_CODE
1 1 John Doe Male 1980-05-25 2023-01-01 D123 P456 M001
2 2 Jane Smith Female 1975-08-12 2023-01-02 D234 P567 M002
3 3 Michael Johnson Male 1990-03-15 2023-01-03 D345 P678 M003
4 4 Emily Brown Female 1985-11-20 2023-01-04 D456 P789 M004
5 5 Liam White Male 1982-07-14 2023-01-05 D567 P890 M005
6 6 Emma Green Female 1991-04-22 2023-01-06 D678 P901 M006
7 7 Noah Black Male 1984-01-02 2023-01-07 D789 P012 M007
8 8 Ava Davis Female 1992-09-17 2023-01-08 D890 P123 M008
9 9 James Moore Male 1980-03-09 2023-01-09 D901 P234 M009
FIRST_NAME
100% filled
LAST_NAME
100% filled
GENDER
Male 14
Female 11
IP

OUTPUT



	PATIENT_ID	FIRST_NAME	LAST_NAME	GENDER	TOTAL_ENCOUNTERS	EARLIEST_ENCOUNTER_DATE	LATEST_ENCOUNTER_DATE	AVG_DAYS_SINCE_LAST_ENCOUNTER
1	15	Benjamin	Lee	Male	1	2023-01-15	2023-01-15	720.000000
2	23	Aiden	Mitchell	Male	1	2023-01-23	2023-01-23	712.000000
3	25	Logan	Hall	Male	1	2023-01-25	2023-01-25	710.000000
4	22	Alexander	Young	Male	1	2023-01-22	2023-01-22	713.000000
5	13	Lucas	Jackson	Male	1	2023-01-13	2023-01-13	722.000000
6	12	Isabella	Thomas	Female	1	2023-01-12	2023-01-12	723.000000
7	21	Harper	Gonzalez	Female	1	2023-01-21	2023-01-21	714.000000
8	19	Ethan	Adams	Male	1	2023-01-19	2023-01-19	716.000000
9	24	Scarlett	Perez	Female	1	2023-01-24	2023-01-24	711.000000
10	20	Mason	Baker	Male	1	2023-01-20	2023-01-20	715.000000
11	18	Sophia	Scott	Female	1	2023-01-18	2023-01-18	717.000000

FUTURE SCOPE

- **Integration with AI/ML:** *Snowflake can be enhanced with machine learning and AI tools to provide predictive analytics, automated decision-making, and personalized healthcare recommendations.*
- **Improved Real-Time Insights:** *Advanced real-time data ingestion and processing can enable quicker decision-making in critical healthcare scenarios.*
- **Enhanced Data Security:** *Development of advanced encryption techniques and privacy measures to ensure compliance with evolving regulations and protect sensitive patient information.*
- **Scalable Interoperability:** *Facilitate seamless integration with other healthcare systems and platforms for a unified ecosystem.*

CONCLUSION

- **Project Focus**: *Demonstrates the potential of using Snowflake for healthcare analytics by centralizing clinical and patient data.*
- **Key Benefits**:
  - *Improved data integration, security, and real-time processing.*
  - *Enables healthcare providers to make informed, data-driven decisions.*
- **Outcomes**:
  - *Enhances patient outcomes.*
  - *Boosts operational efficiency in healthcare.*
- **Proposed System**:
  - *Ensures scalability and compliance.*
  - *Adapts to the growing complexities of modern healthcare data.*

THANK YOU