**Data Warehouse Structure**

To ensure the scalability, flexibility, and cost-effectiveness of our proposed data warehouse structure, I would recommend leveraging Google Cloud Platform (GCP) as the cloud-based data warehousing platform. GCP offers a comprehensive suite of services that align well with our data strategy goals. Here's why GCP is an excellent choice:

*1. Scalability:* GCP's infrastructure is designed for seamless scalability. With services like BigQuery, we can dynamically adjust resources to handle varying data workloads, ensuring optimal performance without manual intervention.

*2. Managed Services:* GCP provides fully managed data warehousing services like BigQuery, which abstract the complexities of infrastructure management. This allows our team to focus on extracting insights from data rather than managing hardware.

*3. Performance:* BigQuery, GCP's data warehousing solution, offers high-performance querying and analytics. It leverages Google's infrastructure to execute complex queries rapidly, enabling timely data analysis.

*4. Integration:* GCP offers seamless integration with various data sources. BigQuery supports data ingestion from diverse sources, making it easy to consolidate our existing dataset and integrate new data sources.

*5. Cost-Efficiency:* With a pay-as-you-go pricing model, GCP ensures cost-efficiency. We only pay for the resources we consume, making it a cost-effective solution for managing and analysing data.

*6. Security and Compliance:* GCP provides robust security measures and compliance certifications, ensuring data protection and regulatory compliance. This aligns well with our commitment to maintaining the security of sensitive data.

**Data Modelling**

**1. Structured Data**: The data modelling layer transforms raw data from various sources into structured entities, such as dimensions and facts, fostering better data organization and minimizing redundancy.

**2. Optimized Querying**: By implementing star or snowflake schema data models, we can enhance query performance. These models enable the creation of efficient joins and aggregations, leading to faster and more accurate query results.

**3. Semantic Understanding**: Data models establish semantic meaning between different data elements, aiding analysts and users in understanding the data's context and relationships.

**Data Warehouse Architecture**

**1. Data Ingestion and Storage: Cloud Storage**

Raw data from various sources will be ingested into Google Cloud Storage. This serves as a centralized repository for storing the unprocessed data securely and cost-effectively.

**2. Data Transformation: Dataproc and Compute Engine**

*- Dataproc's HDFS Cluster:* Data will be copied to Dataproc's Hadoop Distributed File System (HDFS) cluster for scalable data processing and transformation.

*- Google Compute Engine:* Complex ETL tasks requiring custom scripts or tools can be executed on Google Compute Engine instances.

**3. Data Warehousing and Analysis: BigQuery**

Processed data will be loaded into Google BigQuery, our fully managed data warehousing solution. BigQuery offers powerful SQL-based querying capabilities, enabling analysts to retrieve insights quickly and efficiently.

4. Data Visualization: Looker

**Appendix**

The Proposed Architecture Diagram:

