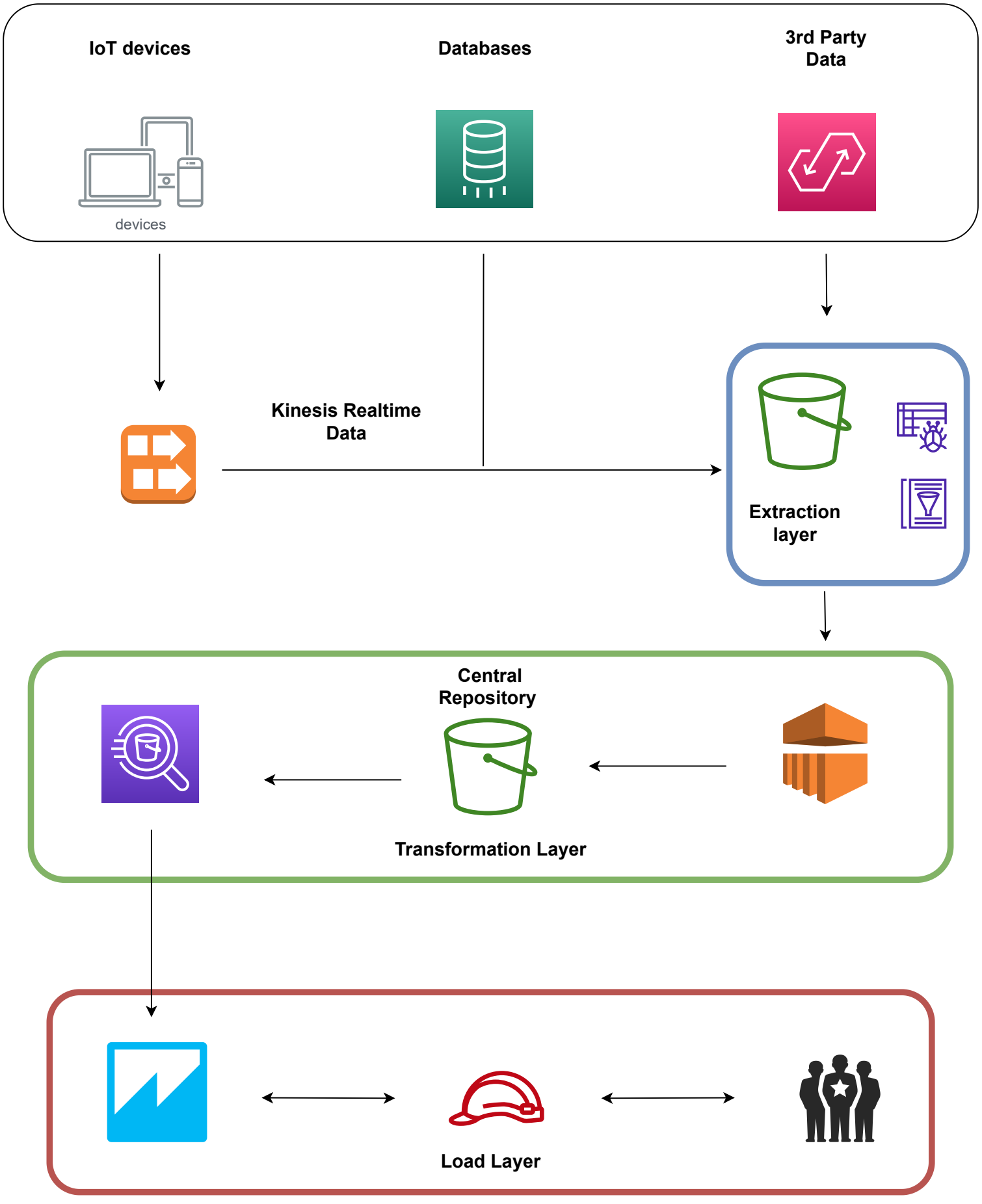
IoT devices
Databases
3rd Party Data
devices
Kinesis Realtime Data
Extraction layer
Central Repository
Transformation Layer
Load Layer

Data Ingestion:

IoT Devices via Kinesis: Using Amazon Kinesis Streams for real-time IoT data provides a scalable and efficient way to ingest and process streaming data. This approach ensures you can handle and process large volumes of data in real time.
Databases via DMS: AWS Database Migration Service (DMS) is a reliable choice for migrating data from databases into your data lake. It helps you maintain data consistency and accuracy during the transfer, preserving historical records.
Third-Party Data via AppFlow: Amazon AppFlow simplifies the data transfer from external sources into your data lake. Its ease of integrating various data sources makes it an efficient choice for supplementing internally generated data with external insights.

Extraction Layer:

S3 Bucket, Glue Crawler, and Glue Catalog: Placing the extracted data in an S3 bucket within the "extraction layer" is a solid choice. AWS Glue Crawler automates discovering and cataloging the data's schema, making it query-ready. AWS Glue Catalog stores the metadata, enhancing data discovery and understanding.

Transformation Layer:

EMR for Data Transformation: Using Amazon EMR (Hadoop) for data transformation is suitable, especially since your organization is already familiar with Apache Hadoop-based software. EMR offers the necessary processing power and versatility to apply complex business logic and formatting.
Central Repository: Storing transformed data in a separate S3 bucket, the "central repository," is a good practice for compliance, auditing, and maintaining a clear distinction between raw and processed data.
Athena for Querying: Utilizing Amazon Athena to query the transformed data is a cost-effective and efficient choice. Athena's serverless nature simplifies querying without requiring infrastructure setup.

Load Layer:

Quicksight for Visualization: Integrating Amazon QuickSight for data visualization and analytics allows the BI team to create insightful dashboards. QuickSight's integration with Athena enables direct querying of data.
IAM Roles for Access Control: Utilizing IAM roles ensures proper access control, limiting permissions to necessary actions. This enhances security while allowing the BI team to perform their tasks.