ETL LOG ANALYSIS USING SPLUNK

INTRODUCTION

ETL refers to the three processes of extracting, transforming and loading data collected from multiple sources into a unified and consistent database. Typically, this single data source is a data warehouse with formatted data suitable for processing to gain analytics insights. ETL is a foundational data management practice.

Log data is data that machine generate. It comes from a variety of sources including software applications, network nodes, components and data center server, connected devices and sensors, consumer online activities and transactions.

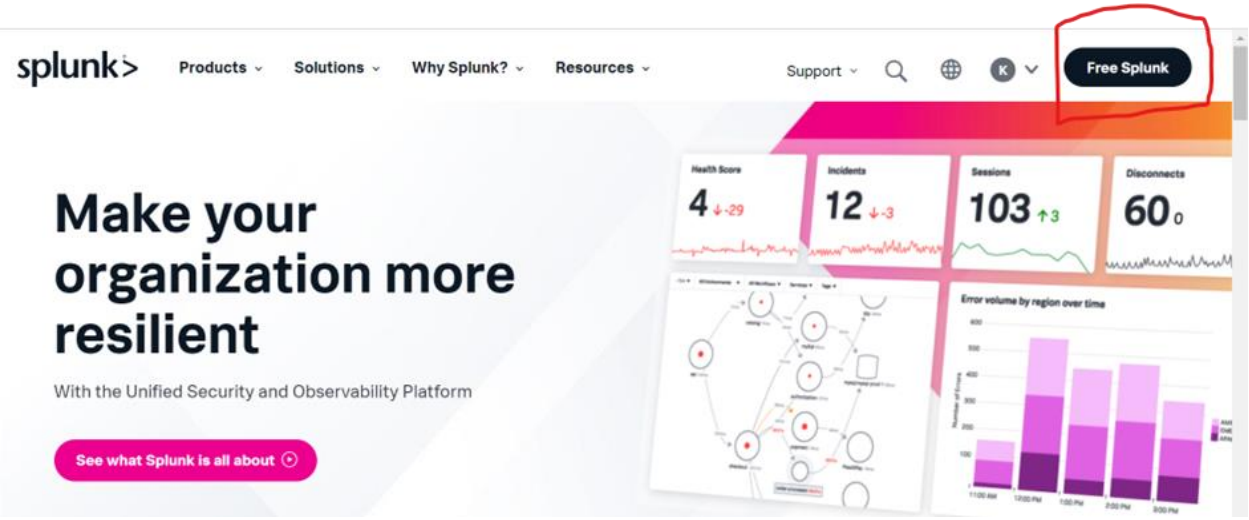
Purpose: To analyze log data for improved security, performance, and operational insights.

Users: IT administrators, security analysts, developers, etc.

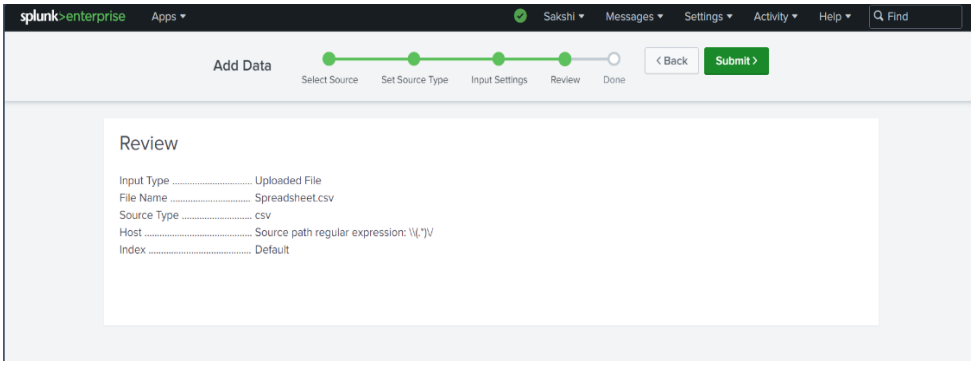
Customers: The various organization using Splunk for log analysis

Functional Requirements

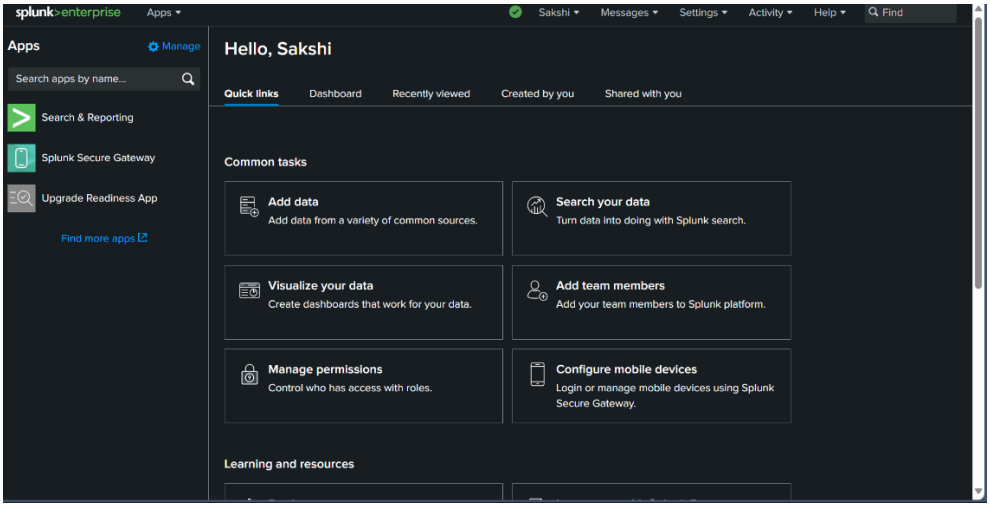
Software Requirements: Splunk which is a paid service, though it has a free-trial option too. I am going to demonstrate how to undertake log analysis using Splunk. First, I will register an account with Splunk at https://www.splunk.com/ . After registration, login and click on Free Splunk.



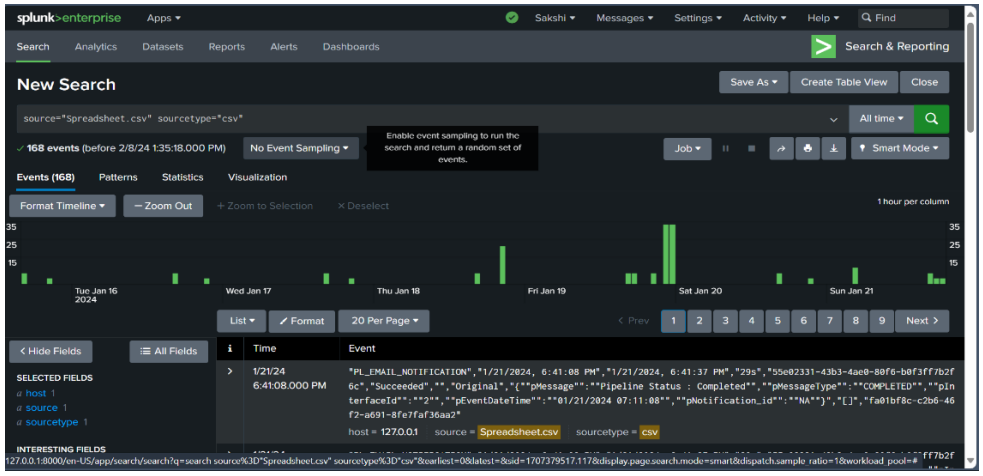
Data ingestion and normalization: In Splunk, it investigate company log data to identify activity origins and transform log data into a standardized format for analysis. Upload spreadsheet.csv via Settings → Add Data → selecting Upload Date → Proceed to Source Type (choosing log type or relying on Splunk detection) → Select host settings (opting for Regular Expression to extract host values) → Review summary and submit .



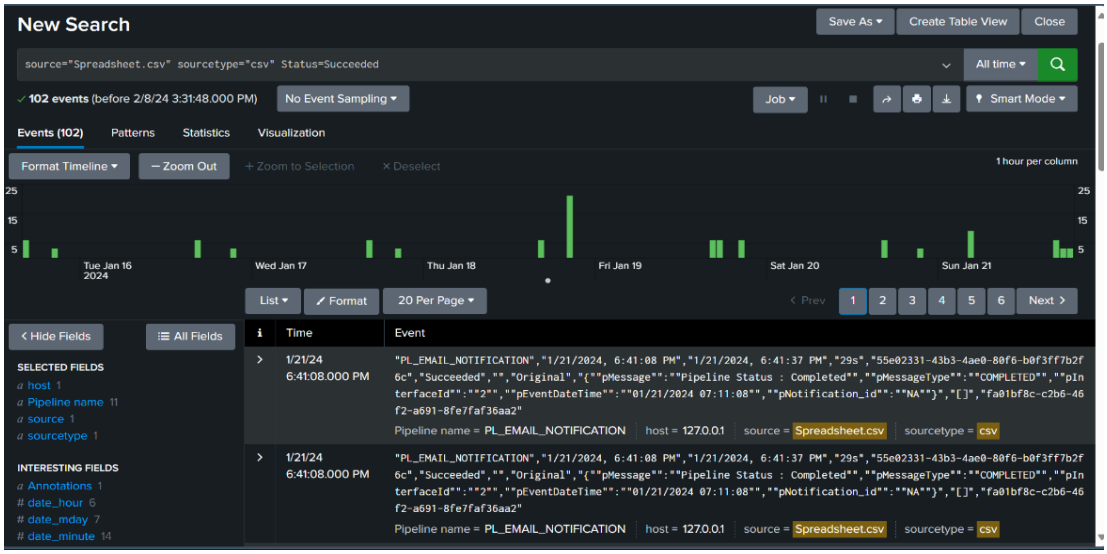
Data indexing: Efficiently store and index data for fast retrieval. Initiate search for Splunk generated queries.



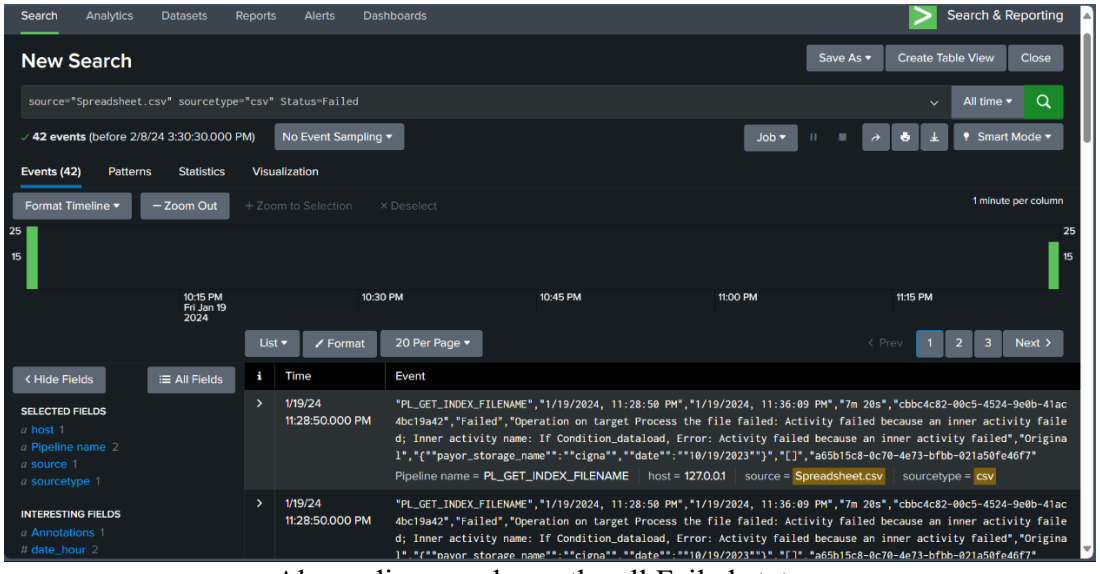
Data search: Perform advanced search queries to identify patterns and anomalies.



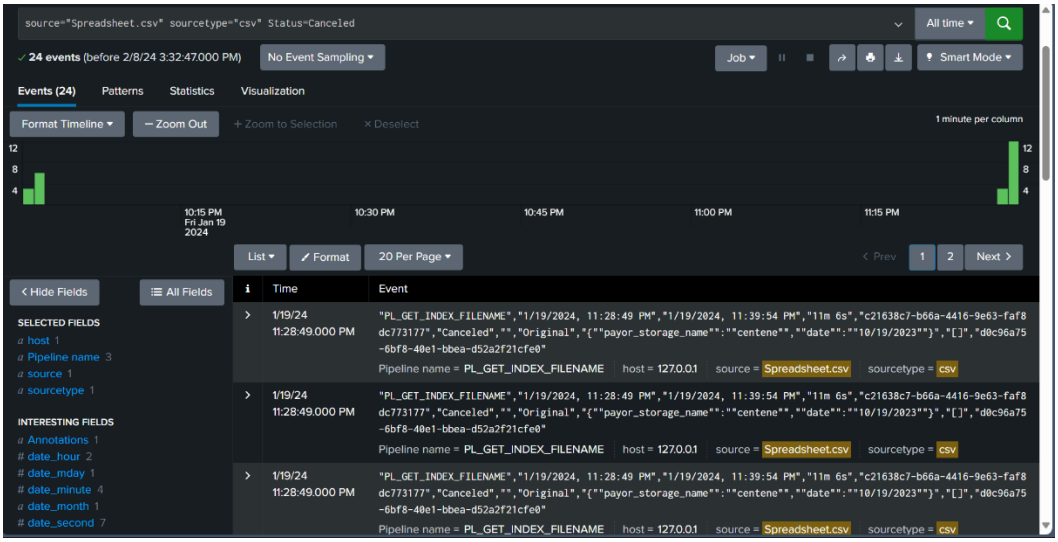
Now "Start Searching" to access query-generated page revealing 168 events. Investigate under "Interesting Fields" tab, example, select "Duration" to spot suspicious activities, analyze Run ID access, explore other events for diverse investigation avenues including user analysis and file request tracking. Identify command types and statuses (SUCCEEDED, FAILED, CANCELED).



Above diagram shows the all success status.



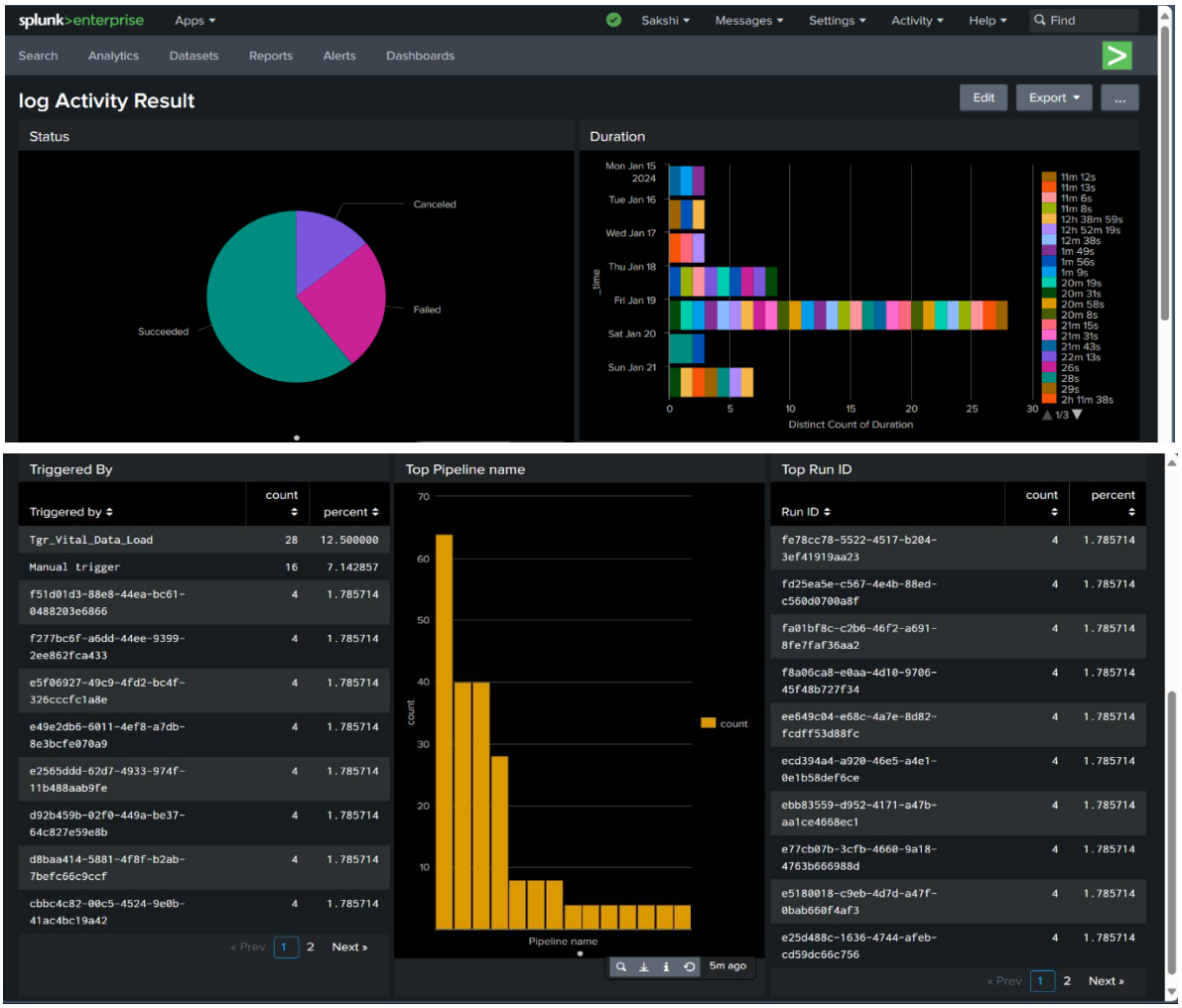
Above diagram shows the all Failed status.



Above diagram shows the all Cancelled status.

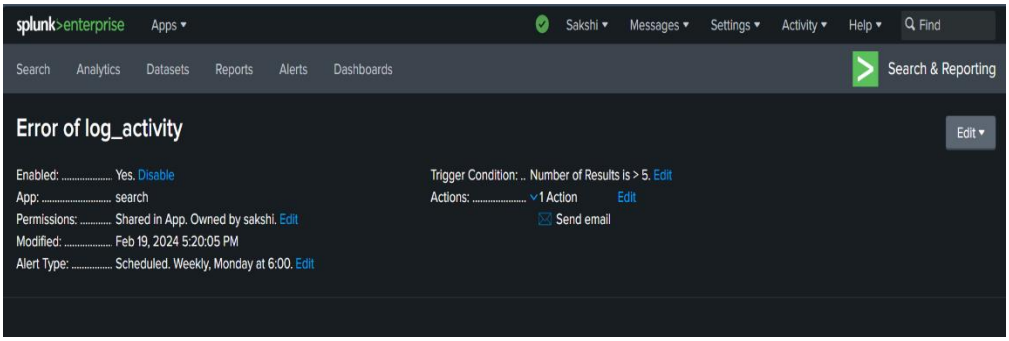
Splunk can be used to gather all the aspects from various logs and connect them together. Splunk is a comprehensive data analytics platform that excels in gathering, analyzing, and correlating data from diverse sources to provide actionable insights for organizations across various domains including IT operations, security, compliance, and business intelligence.

Data visualization: By using splunk, I Create an interactive dashboards and visualizations to present insight.

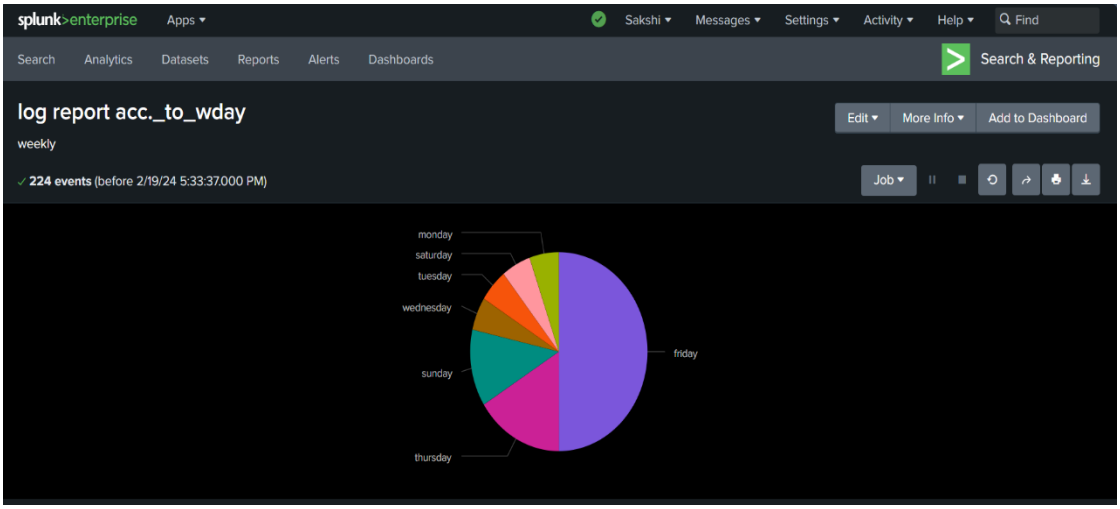


Leveraging Splunk enables the creation of alerts to promptly notify users of critical events, enhancing real-time response capabilities. Additionally, Splunk facilitates the generation of comprehensive reports to disseminate valuable insights to stakeholders, fostering informed decision-making and proactive measures.

Data alerting: In this setting-up alerts to notify users of critical events, For Example-



Data reporting: Generate reports to share insights with stakeholders. For Example-



Non-Functional Requirements

Performance: Fast data processing and query response times.

Security: Ensure data confidentiality, integrity, and availability.

Reliability: Minimize downtime and ensure data availability.

Usability: Intuitive user interface and easy-to-use features.

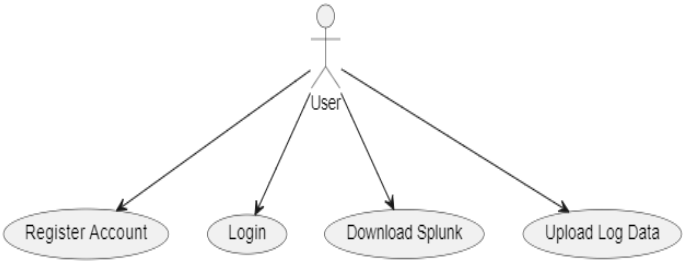
Maintainability: Ease of updating and maintaining the system.

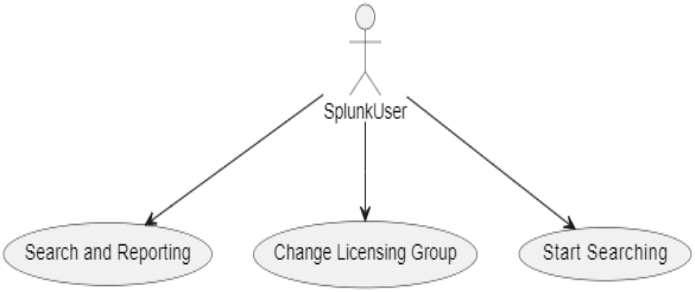
Monitoring and Reporting: Provide comprehensive monitoring and reporting capabilities to track system performance, identify issues, and generate actionable insights for optimization.

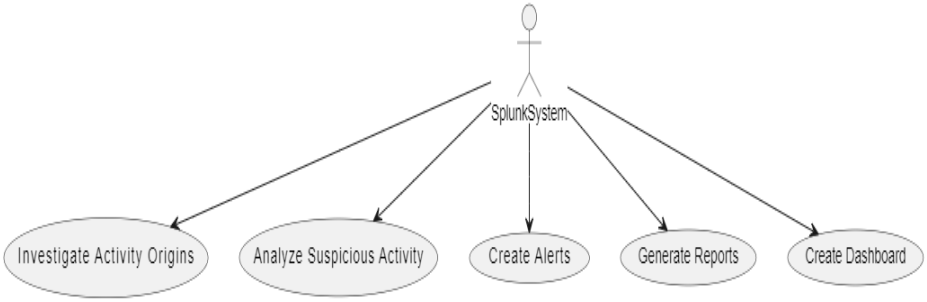
And etc…

UML Diagrams

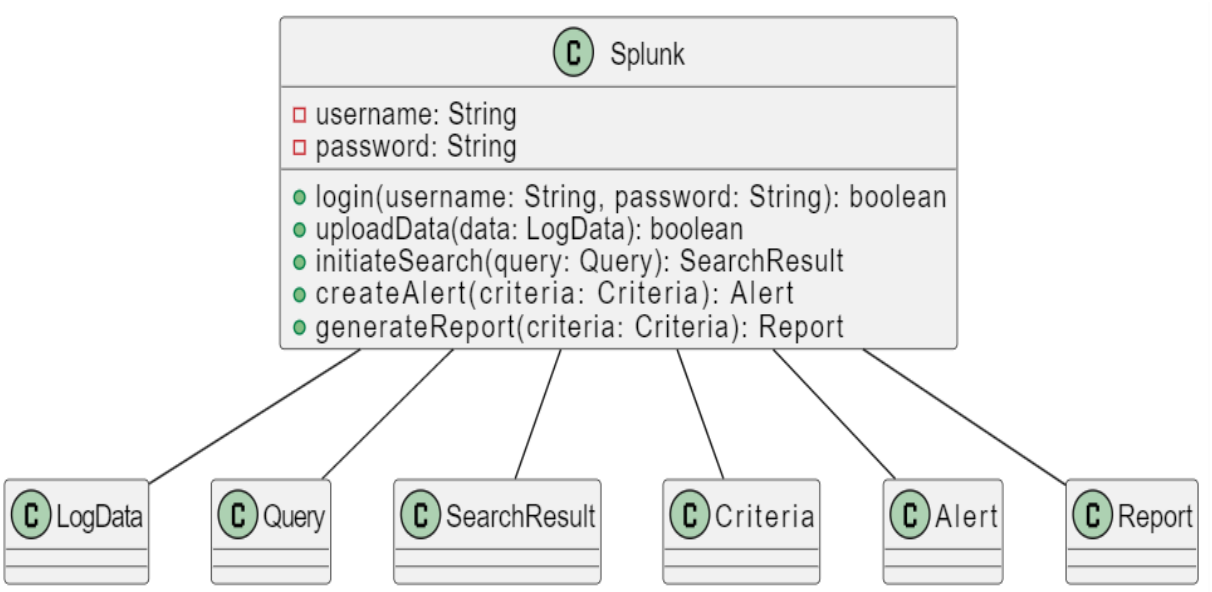
Use case diagram: It show the relationships between actors and use cases.



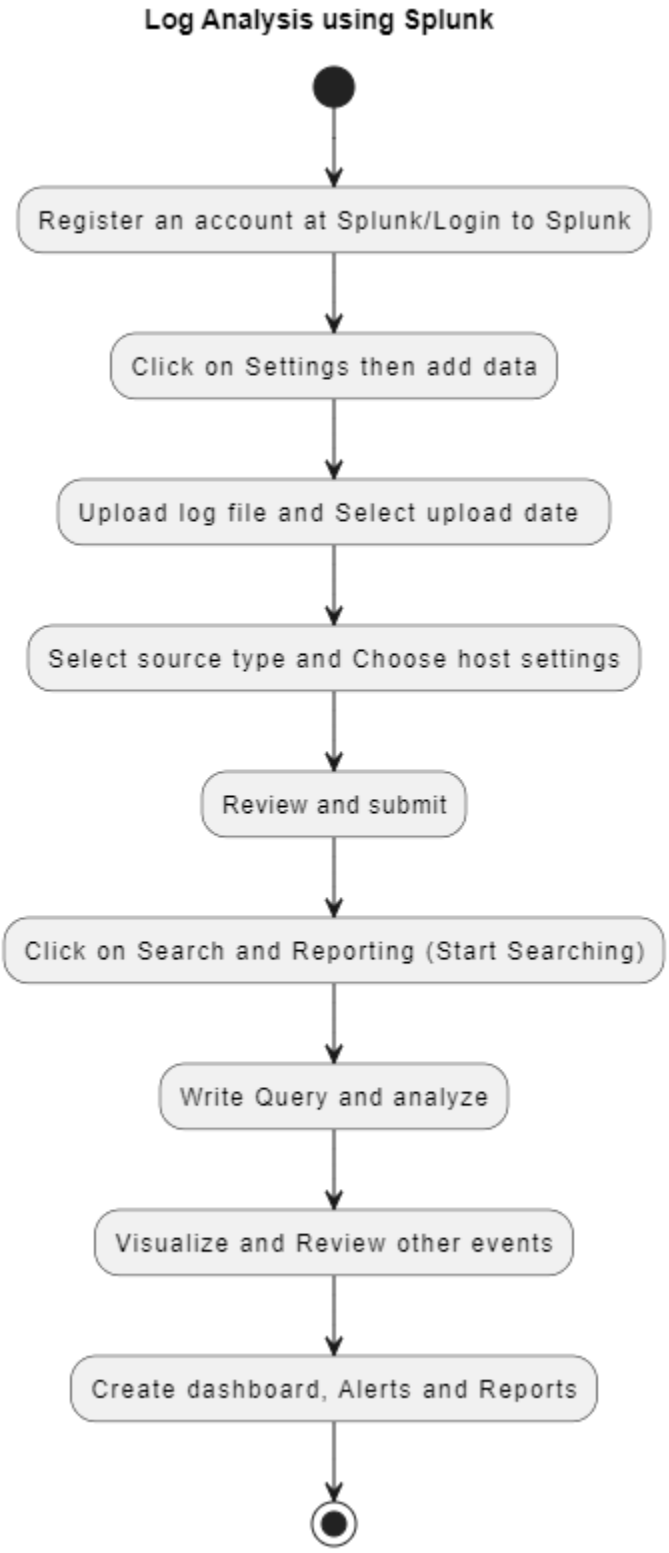




Class diagram: It define the classes, attributes, and relationships within the system.

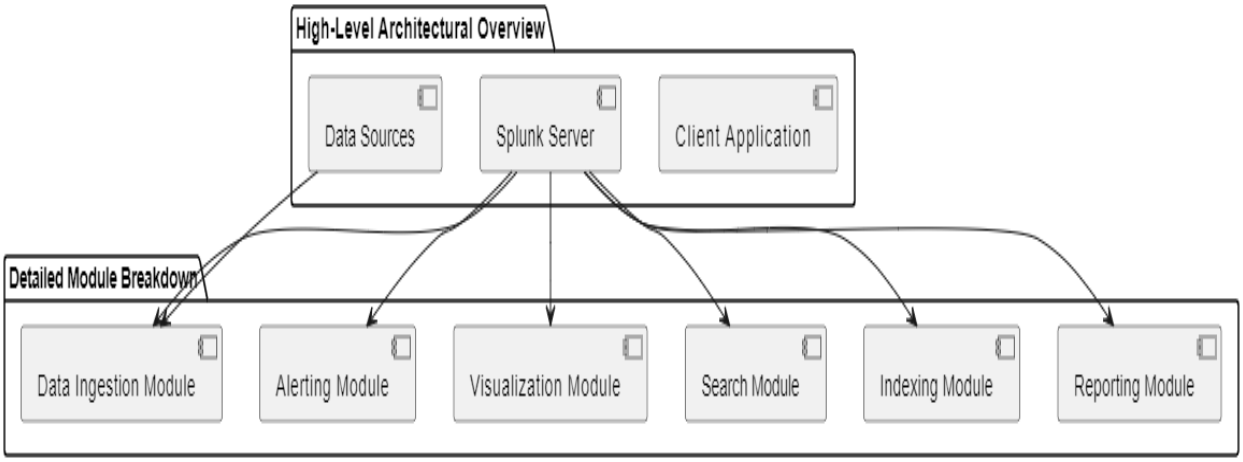


Activity diagram: It Show the flow of activities within the system.

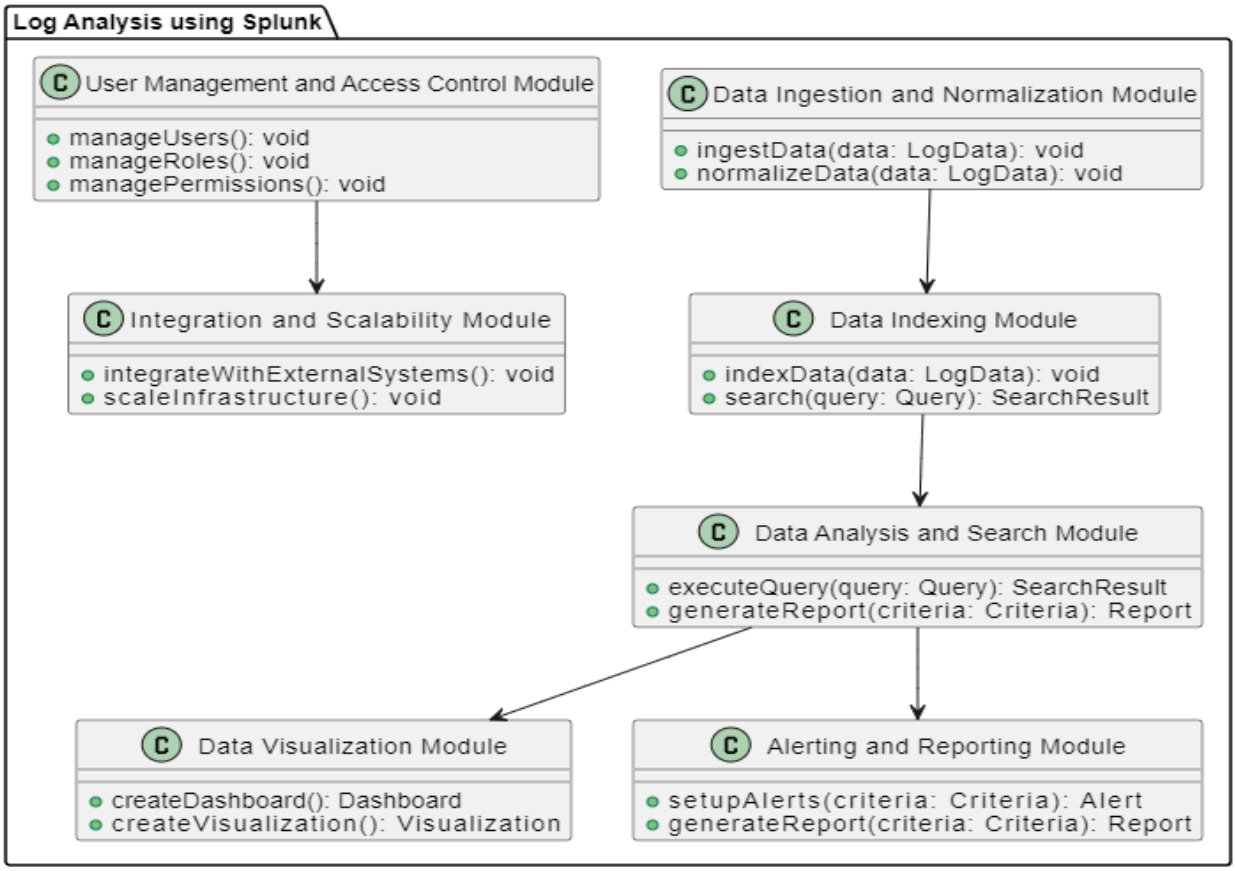


Architecture

High-level architecture: It show the components and their interactions.

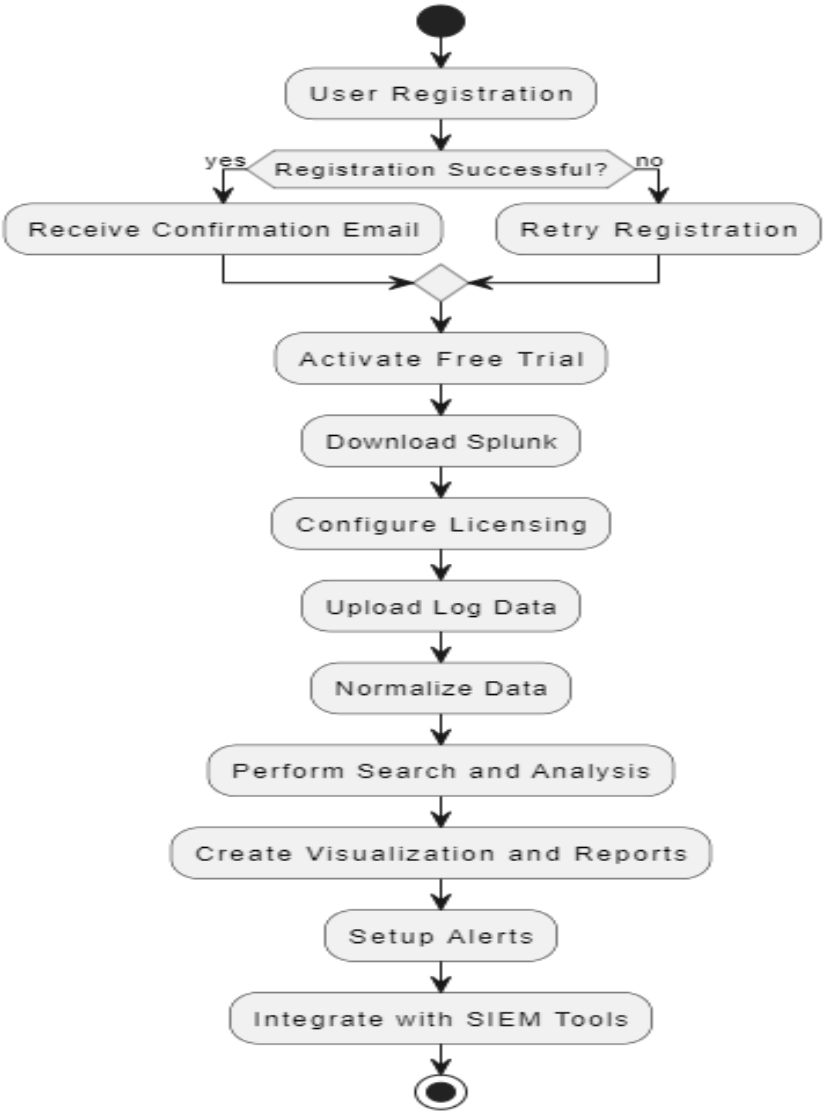


Detailed architecture: It shows the technical details of each component.



Testing

Test cases: It defines the test scenarios and expected results.



Test plan: It include various types of testing to ensure the reliability, functionality, and performance of the system. Here are some common test plans used :

- Unit Testing: Verify individual components.
- Integration Testing: Test module interactions.
- System Testing: Ensure end-to-end functionality.
- Functional Testing: Validate against requirements.
- Non-Functional Testing: Assess performance, security, and scalability.
- Regression Testing: Check for impacts after updates.
- User Acceptance Testing: Validate user satisfaction.

Deployment

Deployment strategy: It describe the steps to deploy the system

- Plan: Define objectives and scope, identify log sources, and allocate resources.
- Install: Download and install Splunk Enterprise or Splunk Cloud on designated hardware or cloud platform.
- Configure: Set up data inputs, define source-types and indexes, and configure Splunk forwarders if needed.
- Ingest: Start ingesting log data into Splunk using configured data inputs.
- Analyze: Utilize Splunk's search functionality to query and analyze log data.
- Visualize: Create dashboards and visualizations to present insights from analyzed data.
- Alert: Set up alerts to notify of critical events or anomalies detected in log data.
- Secure: Implement security measures and access controls to protect sensitive data.
- Train: Provide training to users and administrators on Splunk usage.
- Test: Conduct testing to ensure system meets performance requirements.
- Optimize: Monitor system performance and optimize as needed.
- Maintain: Establish processes for ongoing maintenance, updates, and support.

Environment requirements: It contains the list of hardware and software requirements for the system

Hardware Requirements:

Processor: Intel Core i5 or higher

RAM: 8GB minimum (16GB recommended for larger deployments)

Storage: Minimum 500GB HDD/SSD (1TB or higher recommended for data storage)

Network: Ethernet or Wi-Fi connectivity for data transfer

Display: Monitor with at least 1280x800 resolution

Other: Keyboard, mouse, and sufficient power supply

Software Requirements:

Operating System: Supported versions of Windows, Linux, or macOS

Database: Splunk Enterprise or Splunk Cloud

Web-Browser: Supported versions of Chrome, Firefox, Safari, or Edge for web-based-interface

Additional Software: Depending on specific use cases, additional software may be required for data preprocessing, analysis, or integration with other systems.

Maintenance

Maintenance plan

- Monitoring and Updates: Regularly monitor system health, performance, and data ingestion rates. Set up alerts for abnormal conditions. Stay updated with Splunk software updates and patches.
- Data and User Management: Implement data retention policies, optimize storage, and archive historical log data for compliance. Manages user accounts and permissions for data security. Provide user training and support.
- Performance Optimization: Optimize search queries, indexing configurations, and address performance bottlenecks.
- Backup & Recovery: Regularly backup Splunk configurations and develop a disaster recovery plan.
- Security: Review and update security configurations, and apply patches promptly.
- Documentation & Sharing: Maintain documentation and facilitate knowledge sharing among team members.
- Monitoring & Reporting: Set up performance monitoring dashboards and generate regular reports.
- Continuous Improvement: Gather feedback for improvement and stay updated on industry trends.

Support plan

- Basic Support Plan: Email support, online knowledge base access, software updates.
- Standard Support: Phone and email support, knowledge base access, software updates.
- Premium Support: Priority email and phone support, dedicated Technical Account Manager (TAM).
- Enterprise Support: 24/7 email and phone support, dedicated TAM, customized training.
- Custom Support Plans: Tailored support, additional services available as add-ons.