

---

# **Software Requirements Specification**

**for**

**How LinkedIn Scaled to 930 Million Users**

**Version 1.0 approved**

**Prepared by**

**Mrunmayee Kurhade**

**Dhanshree Dharpure**

**Suhani Taran**

**Pranav Darak**

**Guided by Dr. Heena Agrawal Mam**

**Shri Ramdeobaba College of Engineering & Management**

**03-04-2024**

# Table of Contents

<b>Table of Contents</b>	<b>ii</b>
<b>Revision History</b>	<b>ii</b>
<b>1. Introduction</b>	<b>1</b>
1.1 Purpose	1
1.2 Document Conventions	1
1.3 Intended Audience and Reading Suggestions	1
1.4 Product Scope	1
1.5 References	1
<b>2. Overall Description</b>	<b>2</b>
2.1 Product Perspective	2
2.2 Product Functions	2
2.3 User Classes and Characteristics	2
2.4 Operating Environment	2
2.5 Design and Implementation Constraints	2
2.6 User Documentation	2
2.7 Assumptions and Dependencies	3
<b>3. External Interface Requirements</b>	<b>3</b>
3.1 User Interfaces	3
3.2 Hardware Interfaces	3
3.3 Software Interfaces	3
3.4 Communications Interfaces	3
<b>4. System Features</b>	<b>4</b>
4.1 Connection Management	<a href="#">4</a>
4.2 Service-Oriented Architecture (SOA)	
4.3 Caching	
4.4 Multi Data Center	4
<b>5. Other Nonfunctional Requirements</b>	<b>4</b>
5.1 Performance Requirements	4
5.2 Safety Requirements	5
5.3 Security Requirements	5
5.4 Software Quality Attributes	5
5.5 Business Rules	5
<b>6. Other Requirements</b>	<b>5</b>
<b>Appendix A: Glossary</b>	<b>5</b>
<b>Appendix B: Analysis Models</b>	<b>5</b>
<b>Appendix C: To Be Determined List</b>	<b>6</b>

## Revision History

Name	Date	Reason For Changes	Version
<b>Mrunmayee Kurhade</b>	03-04-2024	Initial Version	0.1
<b>Dhanshree Dharpure</b>			
<b>Suhani Taran</b>			
<b>Pranav Darak</b>			

# **1. Introduction**

## **1.1 Purpose**

The purpose of this document is to provide a comprehensive overview of the software requirements necessary for scaling LinkedIn's infrastructure to accommodate 930 million users. It serves as a detailed guide for the development team to understand the functionality, constraints, and objectives of the project.

## **1.2 Document Conventions**

This Software Requirements Specification (SRS) adheres to established conventions to ensure clarity, consistency, and readability. It utilizes clear and concise language to describe requirements, prioritizes them effectively, and maintains a consistent formatting style throughout the document.

## **1.3 Intended Audience and Reading Suggestions**

This document is intended for various stakeholders involved in the LinkedIn scaling project, including:

Developers: To understand the technical requirements and design considerations for implementing the scalable infrastructure.

Project Managers: To oversee the project's progress, manage resources, and ensure alignment with business objectives.

Testers: To develop test cases and validate the functionality and performance of the system.

Stakeholders: To gain insight into the project scope, requirements, and expected outcomes.

Readers are encouraged to start with the overview sections, such as the purpose and product scope, to gain a broad understanding of the project. They can then proceed to sections that are most relevant to their roles and responsibilities.

## **1.4 Product Scope**

This SRS outlines the software requirements for enhancing LinkedIn's infrastructure to accommodate a significantly increased user base of 930 million users. The scope of the project includes:

**Scalability**: Implementing solutions to handle the anticipated growth in user traffic and data volume.

**Performance**: Ensuring the platform maintains optimal performance levels under increased load and user activity.

**Reliability**: Enhancing system reliability and fault tolerance to minimize downtime and disruptions for users.

The document focuses on specifying the necessary features, functionalities, and technical considerations to achieve these objectives.

## **1.5 References**

The following documents provide additional context and background information relevant to the LinkedIn scaling project:

LinkedIn Vision and Scope Document (Version 2.0): Outlines the overarching goals, objectives, and vision for the LinkedIn platform, including anticipated growth projections and strategic priorities.

LinkedIn System Requirements Specification (Version 1.5): Describes the initial set of system requirements and functionalities for the LinkedIn platform, serving as a foundational reference for the scaling project.

These references serve as valuable sources of information and guidance for understanding the project's context, goals, and existing system requirements. They provide a framework for aligning the scaling efforts with the overall vision and objectives of LinkedIn.

## **2. Overall Description**

### **2.1 Product Perspective**

The software specified in this document is an integral part of LinkedIn's existing infrastructure. It serves as a critical component responsible for enhancing the scalability and performance of the platform to accommodate the anticipated growth of 930 million users. While building upon the existing foundation of LinkedIn's ecosystem, the software introduces new functionalities and optimizations to meet the evolving needs and demands of the user base.

### **2.2 Product Functions**

The software encompasses a range of essential functions essential for the seamless operation of LinkedIn's platform:

- User Authentication: Enabling users to securely log in to their accounts and access platform features.
- Profile Management: Allowing users to create, edit, and manage their profiles, including personal information, professional experience, and skills.
- Content Delivery: Facilitating the delivery of diverse content types, including posts, articles, videos, and job listings, to users' feeds and search results.
- Messaging: Supporting real-time messaging capabilities for users to communicate with each other privately.
- Recommendation Systems: Leveraging algorithms to provide personalized content recommendations, job suggestions, and connection recommendations based on users' interests and interactions.

These functions collectively contribute to delivering a seamless and engaging user experience on the LinkedIn platform.

## **2.3 User Classes and Characteristics**

LinkedIn caters to various user classes, each with distinct characteristics and needs:

- Regular Users: Individuals using the platform for networking, job searching, content consumption, and professional development.
- Premium Users: Subscribers of premium LinkedIn services, such as LinkedIn Premium, Sales Navigator, and LinkedIn Learning, who access advanced features and insights.
- Administrators: Platform administrators responsible for managing user accounts, content moderation, and enforcing platform policies.
- Third-party Integrations: External applications and services integrated with LinkedIn's platform to extend its functionality and provide value-added services to users.

These user classes exhibit differences in usage frequency, technical proficiency, and access privileges, which influence their interactions with the platform.

## **2.4 Operating Environment**

The software operates within a distributed environment comprising various components:

- Servers: Hosting the application logic, databases, and caching mechanisms to handle user requests and data processing.
- Databases: Storing user data, content metadata, activity logs, and other relevant information in a scalable and efficient manner.
- Networking Infrastructure: Facilitating communication between different system components and ensuring reliable data transmission.
- Third-party Services: Integrating with external services, APIs, and cloud platforms to leverage additional functionalities and resources.

Compatibility with multiple operating systems, web browsers, and devices is essential to ensure a seamless user experience across diverse user environments.

## **2.5 Design and Implementation Constraints**

Several constraints shape the design and implementation of the software:

- Regulatory Compliance: Adherence to data protection regulations, privacy laws, and industry standards governing user data handling and security.

- **Hardware Limitations:** Consideration of hardware resources, such as processing power, memory, and storage capacity, available for deploying and running the software.
- **Software Compatibility:** Integration with existing software components, frameworks, libraries, and APIs used within LinkedIn's ecosystem.
- **Security Considerations:** Implementation of robust security measures, including encryption, access controls, and threat detection mechanisms, to protect user data and prevent unauthorized access.

Addressing these constraints is critical to ensuring the software's effectiveness, reliability, and compliance with regulatory requirements.

## **2.6 User Documentation**

User documentation plays a vital role in guiding users on how to effectively utilize the platform's features and functionalities. It includes:

- **User Manuals:** Comprehensive guides outlining platform features, navigation instructions, and best practices for maximizing user engagement.
- **Online Help:** Interactive help resources accessible within the platform, providing context-sensitive assistance and troubleshooting guidance.
- **Tutorials:** Step-by-step tutorials and video demonstrations covering specific tasks, workflows, and use cases to help users get started and achieve their goals on the platform.

Clear and accessible user documentation enhances user satisfaction, reduces support overhead, and promotes user adoption and retention.

## **2.7 Assumptions and Dependencies**

Certain assumptions and dependencies underpin the software development and deployment process:

- **Stable Network Connectivity:** Reliance on stable internet connectivity to ensure uninterrupted access to the platform and seamless data transmission.
- **Availability of Hardware Resources:** Availability of sufficient hardware resources, including servers, storage, and networking infrastructure, to support the software's operational requirements.
- **Adherence to Regulatory Requirements:** Compliance with applicable regulations, standards, and policies governing data privacy, security, and user rights.
- **Dependencies on Third-party APIs:** Integration with third-party APIs, services, and platforms to leverage additional functionalities and enhance the platform's capabilities.

Addressing these assumptions and dependencies is essential for ensuring the successful development, deployment, and operation of the software within the specified constraints and requirements.

## **3. External Interface Requirements**

### **3.1 User Interfaces**

The user interface (UI) of the software plays a pivotal role in shaping the user experience (UX) and driving user engagement on the LinkedIn platform

User interfaces should be intuitive, responsive, and accessible across various devices and screen sizes.

### **3.2 Hardware Interfaces**

The software must seamlessly interface with underlying hardware components, including servers, storage devices, and networking equipment, to ensure optimal performance, reliability, and scalability of the LinkedIn platform. Hardware interfaces should be designed to leverage the capabilities and resources of the underlying infrastructure efficiently, maximizing system throughput, processing speed, and data storage capacity.

### **3.3 Software Interfaces**

Integration with external software components, databases, and APIs is crucial for extending the functionality and capabilities of the LinkedIn platform and enabling seamless interoperability with third-party systems and services. Software interfaces should be designed to facilitate data exchange, interoperability, and communication between different software modules, subsystems, and external applications

### **3.4 Communications Interfaces**

Effective communication interfaces are essential for enabling seamless data exchange, collaboration, and interaction among users, systems, and services within the LinkedIn ecosystem. Support for a variety of communication protocols, such as TCP/IP, UDP, SMTP, and HTTP, is required to facilitate different types of communication channels, including messaging, email, notifications, and real-time data streams

## **4. System Features**

### **4.1 User Authentication**

#### **4.1.1 Description and Priority**

User authentication is a critical system feature aimed at ensuring the security and integrity of the LinkedIn platform. With a high priority designation, this feature encompasses the implementation of robust and secure mechanisms for verifying the identity of users accessing the platform.

#### **4.1.2 Stimulus/Response Sequences**

Upon initiating the authentication process, users provide their credentials, typically comprising a username/email and password combination. The system then verifies these credentials against stored user data to determine authentication success or failure. Upon successful authentication, users are granted access to their accounts and the platform's features, while access is denied in case of authentication failure.



### 4.1.3 Functional Requirements

#### Connection Management

- Description: With LinkedIn's growth, managing connections between users efficiently became crucial for the platform's functionality and user experience.
- Importance: Connection management allows users to establish professional relationships, network, and access relevant content and opportunities on the platform.
- Implementation: LinkedIn introduced a separate in-memory graph service to manage connections between users, enabling efficient handling of user interactions and network growth.

#### Service-Oriented Architecture (SOA)

- Description: Transitioning from a monolithic architecture to a Service-Oriented Architecture (SOA) was pivotal for LinkedIn to ensure high availability and handle high traffic.
- Importance: SOA enables LinkedIn to break down the monolithic application into smaller, stateless services, facilitating scalability, and improving engineering agility.
- Implementation: LinkedIn divided the monolith into smaller services such as frontend, mid-tier, and backend data services, allowing them to scale out by replicating services and enhancing the platform's resilience to traffic spikes.

#### Caching

- Description: LinkedIn employed caching extensively to meet scalability needs and improve performance by reducing the load on backend systems.
- Importance: Caching helps to store frequently accessed data closer to users, reducing latency and improving overall system responsiveness.
- Implementation: LinkedIn utilized Content Delivery Networks (CDN), browser cache, and stored precomputed results in the database to optimize content delivery and minimize the strain on backend resources, ensuring a smooth user experience.

#### Multi Data Center

- Description: To ensure high availability and reliability, LinkedIn replicated data across multiple data centers, minimizing the risk of single points of failure.
- Importance: Multi-data center replication enhances system resilience by distributing data across geographically dispersed locations, reducing the impact of downtime and improving performance for users worldwide.
- Implementation: LinkedIn implemented mechanisms to redirect user requests to nearby data centers, ensuring that users access data from the nearest location, thereby improving performance and reliability of the platform.

## 4.2 Profile Management

### 4.1.1 Description and Priority

Profile management is a core functionality of the LinkedIn platform, allowing users to maintain and customize their profiles to showcase their professional identity effectively. Given its high priority designation, this feature focuses on providing users with intuitive tools and functionalities for managing their profile information efficiently.

#### 4.1.2 Stimulus/Response Sequences

When users initiate profile updates or modifications, they interact with the system through various input methods, such as form fields or interactive UI elements. Upon submitting profile changes, the system validates the input data and updates the corresponding user profile records in the database. Once the updates are successfully processed, the system generates a confirmation message to inform the user of the successful completion of the profile management operation.

## 5. Other Nonfunctional Requirements

### 5.1 Performance Requirements

The system should maintain an average response time of under 500 milliseconds for user interactions: Ensuring optimal system performance is crucial for delivering a responsive and engaging user experience on the LinkedIn platform. To achieve this, the system should be designed and optimized to handle user interactions swiftly, with an average response time of under 500 milliseconds. This includes processing user requests, fetching and rendering content, and executing backend operations efficiently. Performance monitoring and profiling tools should be employed to identify and address bottlenecks or performance issues proactively, ensuring that the platform consistently meets the specified response time targets.

The system should support a peak concurrency of 10 million simultaneous users: Scalability is paramount for accommodating the anticipated growth in user traffic and ensuring that the LinkedIn platform can handle peak loads without degradation in performance or responsiveness. The system should be architected and provisioned to support a peak concurrency of 10 million simultaneous users, allowing for seamless user interactions, content delivery, and system operations even during periods of high demand. Horizontal and vertical scaling strategies, along with load balancing and caching mechanisms, should be employed to distribute and manage user traffic effectively across the system infrastructure, mitigating the risk of overload or downtime during peak usage scenarios.

### 5.2 Safety Requirements

User data should be encrypted both at rest and in transit to ensure data safety: Protecting the confidentiality and integrity of user data is paramount to maintaining user trust and compliance with regulatory requirements. Therefore, all sensitive user data, including personal information, communications, and authentication credentials, should be encrypted using strong encryption algorithms both at rest (stored in databases or storage systems) and in transit (during data transmission over networks). Transport Layer Security (TLS) protocols should be implemented to secure communication channels and prevent unauthorized access or tampering with data during transmission. Additionally, encryption keys should be managed securely, with strict access controls and regular key rotation to minimize the risk of data breaches or unauthorized access.

Regular data backups should be performed to prevent data loss in case of system failures:

Data integrity and resilience are critical for ensuring business continuity and mitigating the impact of system failures or disasters. Therefore, the system should implement robust data backup and recovery mechanisms to safeguard against data loss or corruption. Regular backups of user data, configuration settings, and system logs should be performed at predetermined intervals, with multiple redundant copies stored in geographically distributed locations to mitigate the risk of data loss due to hardware failures, human errors, or malicious attacks. Backup processes should be automated, monitored, and tested regularly to ensure reliability and effectiveness in restoring data in the event of a system outage or data loss incident.

### 5.3 Security Requirements

The system should adhere to industry-standard security practices such as OWASP Top 10: Adherence to industry-standard security practices and guidelines, such as those outlined by the Open Web Application Security Project (OWASP), is essential for mitigating security risks and vulnerabilities in the LinkedIn platform. The system should undergo regular security assessments, code reviews, and vulnerability scans to identify and remediate common security threats and weaknesses highlighted in the OWASP Top 10 list, including injection attacks, cross-site scripting (XSS), broken authentication, and sensitive data exposure. Secure coding practices, such as input validation, output encoding, parameterized queries, and proper authentication and authorization mechanisms, should be followed throughout the software development lifecycle to minimize the risk of security breaches and unauthorized access to sensitive data.

Access control mechanisms should be implemented to restrict unauthorized access to sensitive data: Granular access control mechanisms should be enforced to regulate user permissions and privileges based on roles, responsibilities, and data sensitivity levels. Role-based access control (RBAC) should be implemented to define and manage user roles, access rights, and permissions within the system, ensuring that users can only access and manipulate data and functionalities that are relevant to their roles and responsibilities. Additionally, fine-grained access control policies should be applied to sensitive data and resources, with strict enforcement of authentication, authorization, and audit logging mechanisms to prevent unauthorized access, privilege escalation, and data breaches.

### 5.4 Software Quality Attributes

The software should be highly maintainable, with well-documented code and modular design: Ensuring the maintainability of the software is essential for facilitating ongoing development, troubleshooting, and enhancements to the LinkedIn platform. The codebase should be well-structured, modular, and extensively documented to facilitate code comprehension, modification, and collaboration among development teams. Best practices such as code commenting, code refactoring, version control, and code reviews should be followed rigorously to maintain code quality and readability. Additionally, adherence to design principles such as separation of concerns, loose coupling, and high cohesion should be prioritized to enhance code maintainability and scalability over time.

Usability testing should be conducted to ensure an intuitive user experience: Usability testing is crucial for evaluating the effectiveness, efficiency, and satisfaction of users interacting with the LinkedIn platform. User-centric design principles should be applied throughout the software development lifecycle, with a focus on understanding user needs, preferences, and pain points through user research, personas, and user journey mapping. Usability testing sessions should be conducted with

representative user groups to gather feedback, identify usability issues, and validate design assumptions. Iterative design and testing cycles should be employed to refine the user interface, navigation flows, and interaction patterns, ensuring that the platform delivers an intuitive and user-friendly experience that meets the needs and expectations of its diverse user base.

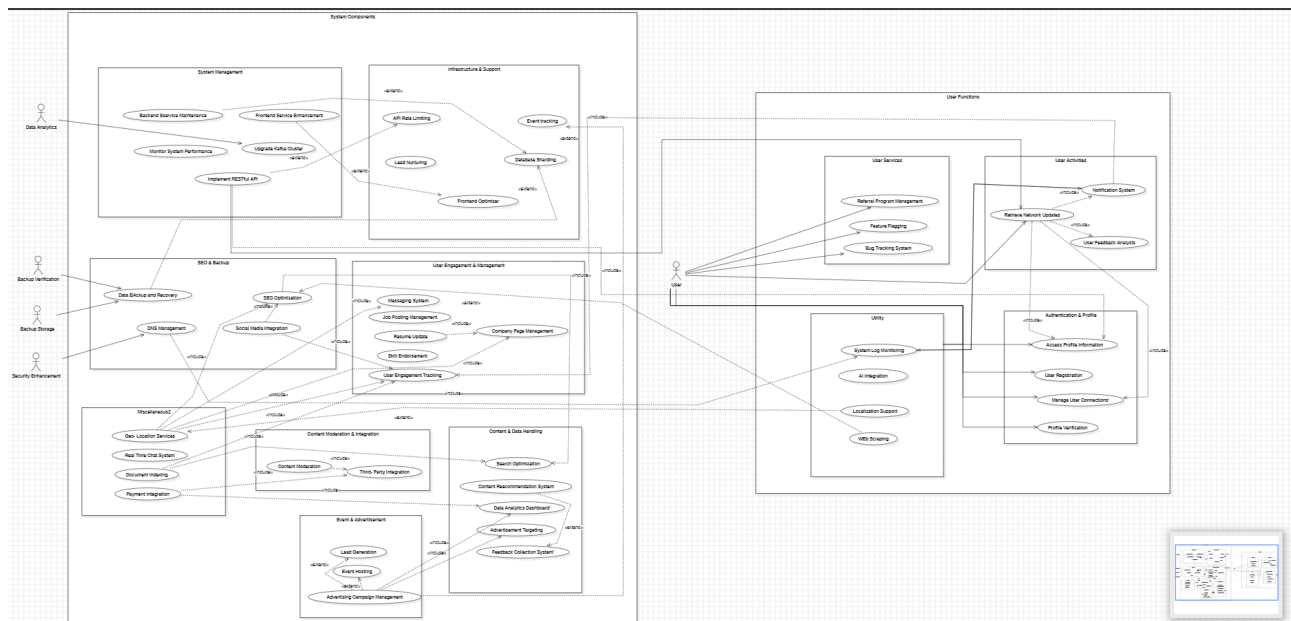
## 5.5 Business Rules

Premium users should have access to additional features such as advanced analytics and insights: Offering premium features and services to subscribed users is a key aspect of LinkedIn's business model. Premium users should be granted access to exclusive features such as advanced analytics tools, personalized insights, and enhanced networking capabilities. These premium features provide added value to subscribers and incentivize premium membership upgrades, contributing to revenue generation and customer retention.

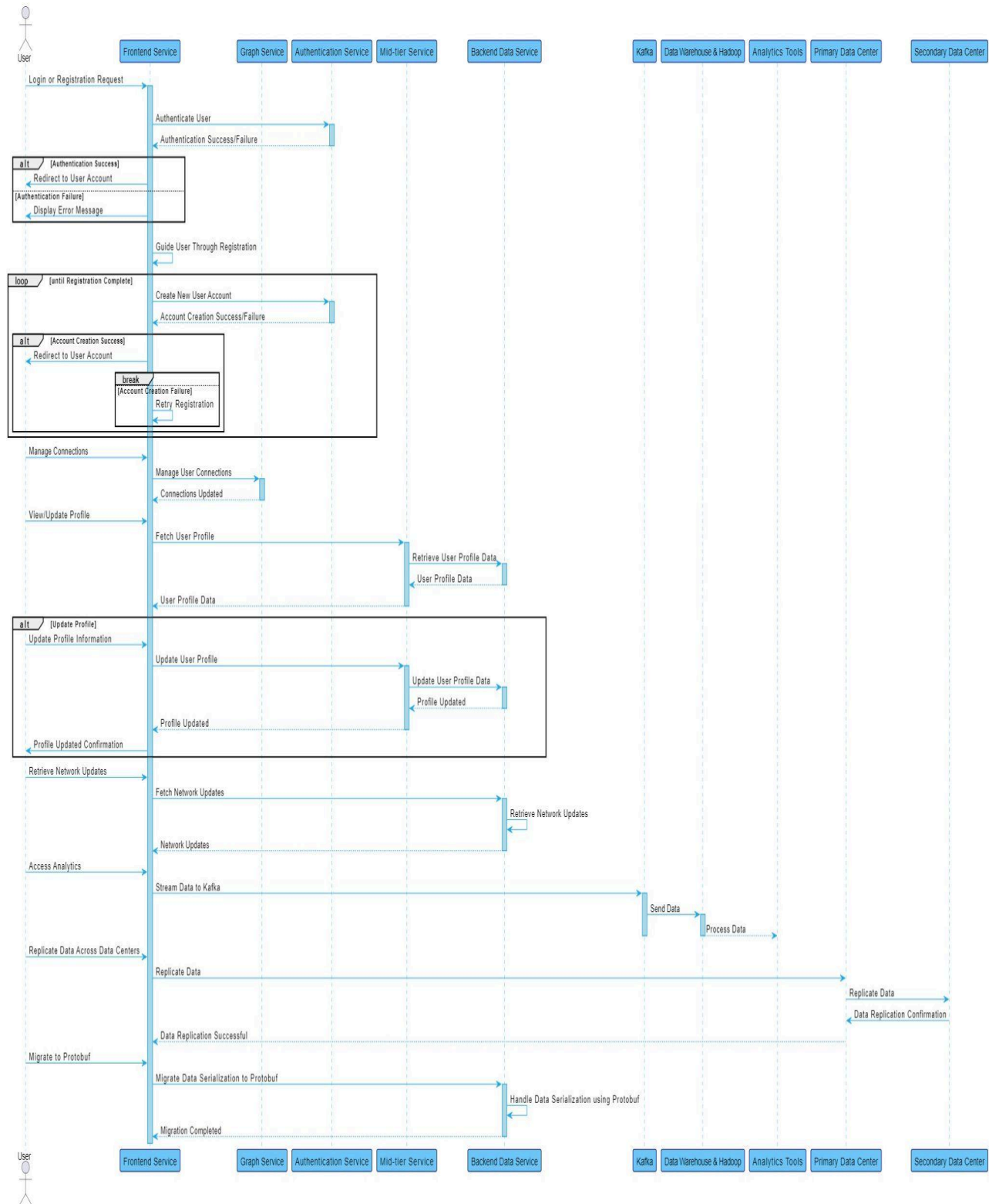
Certain actions, such as deleting an account, should require confirmation from the user: Implementing safeguards for critical user actions is essential for preventing accidental data loss and unauthorized account access. Certain actions, such as deleting an account or permanently removing data, should require confirmation from the user to ensure intentionality and prevent irreversible consequences. Confirmation dialogs or verification steps should be implemented to prompt users to confirm their actions before proceeding, reducing the risk of unintended actions and enhancing user control over their accounts and data.

## Diagrammatic Analysis

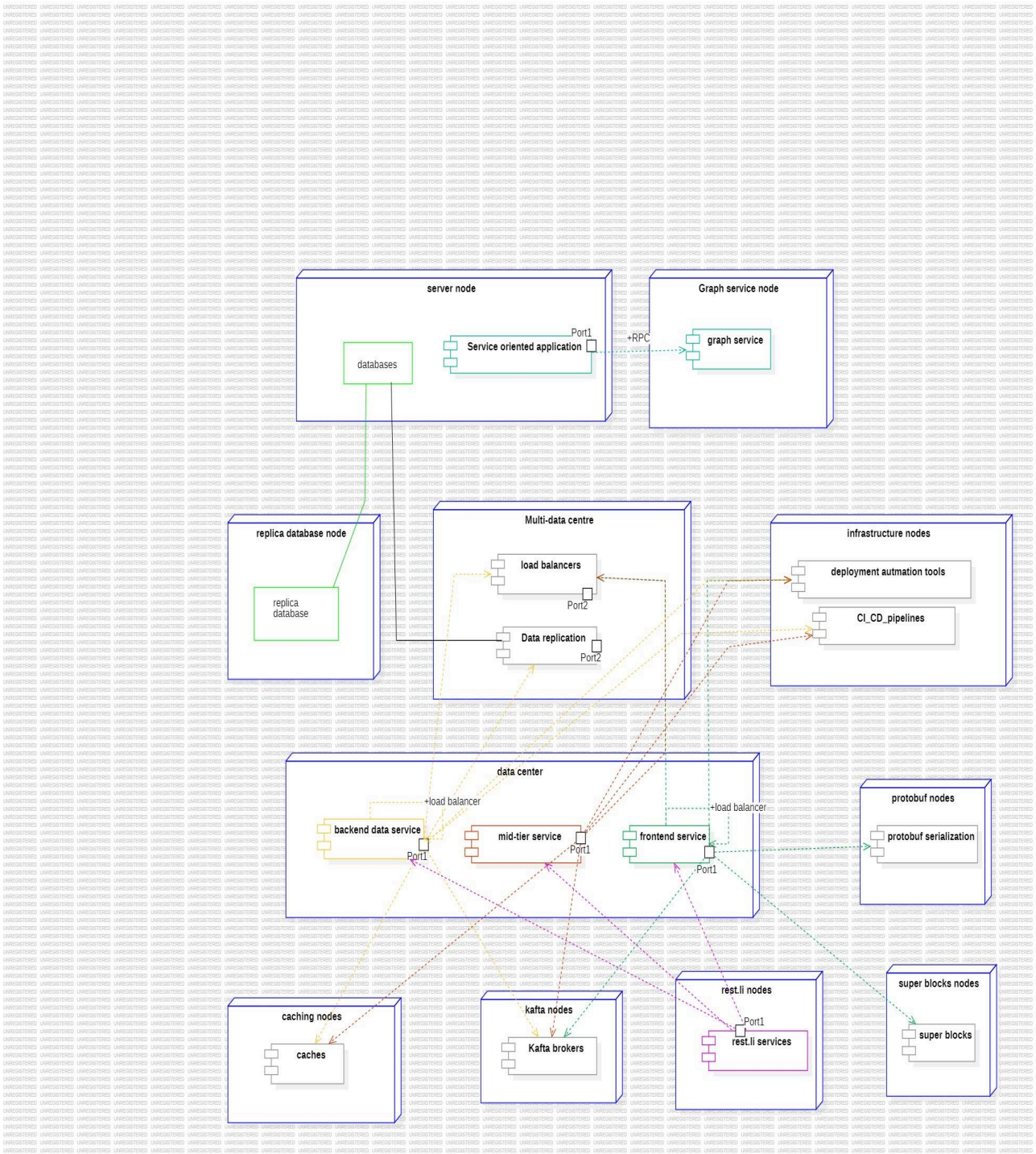
### Use Case Diagram:



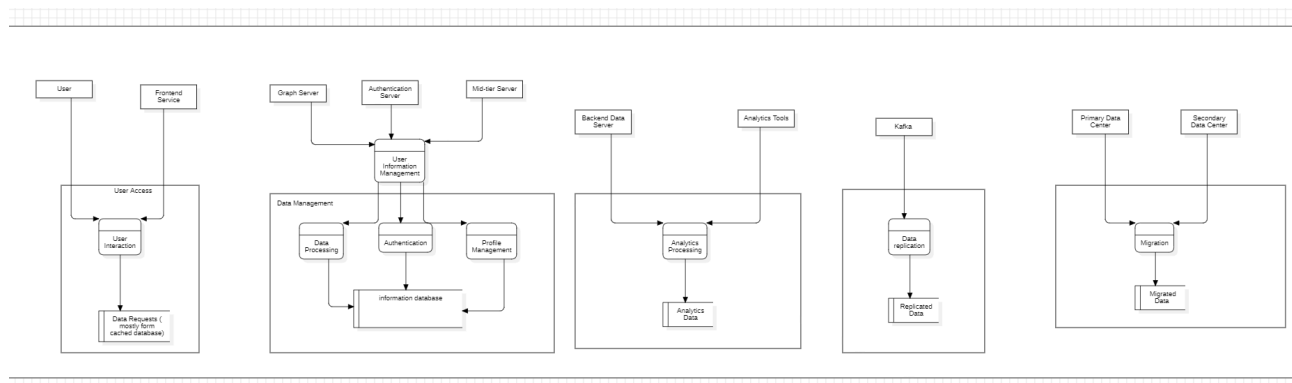
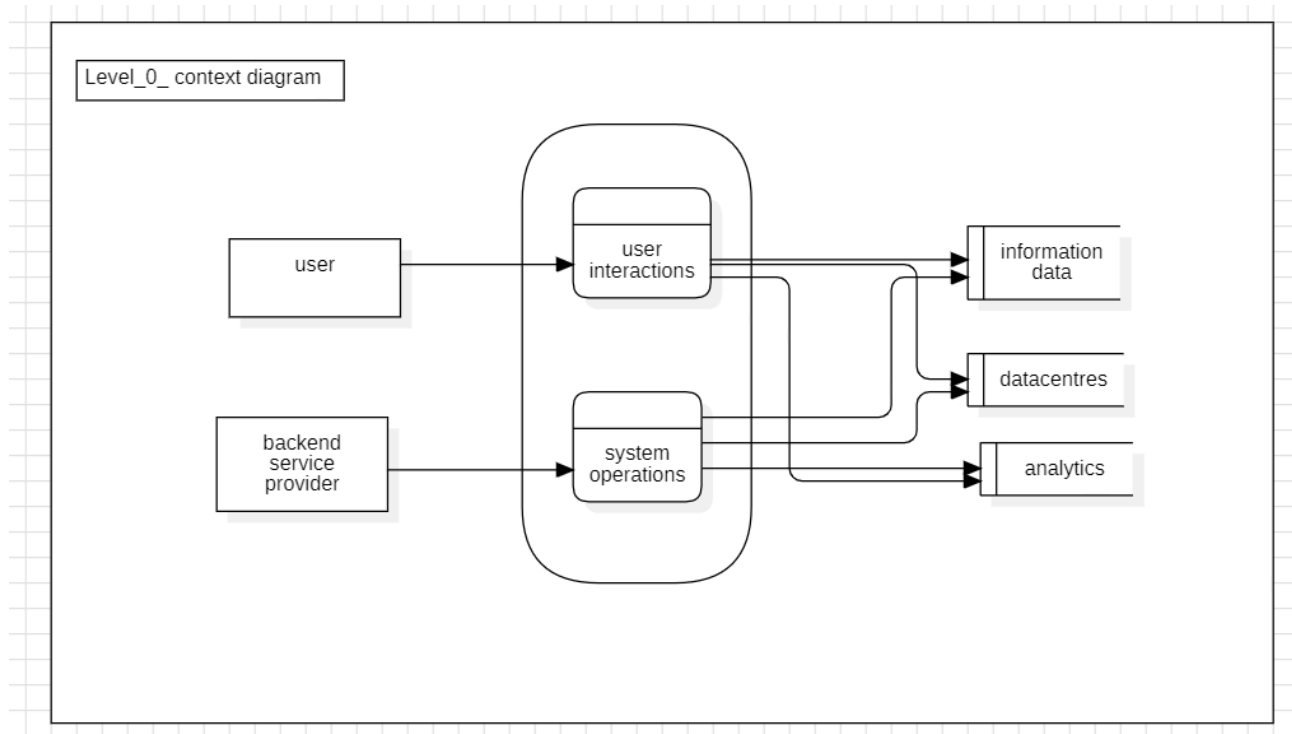
## Sequence Diagram

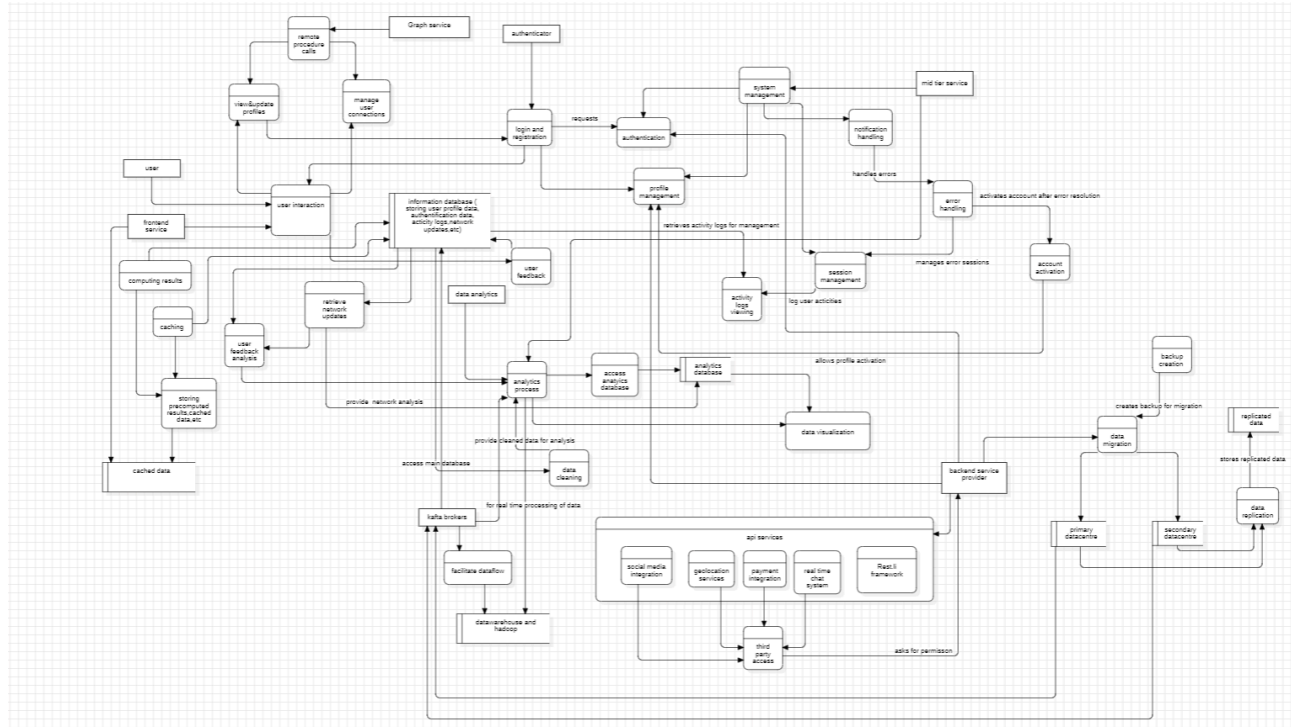


Deployment Diagram:





**DFD:**



## 6. Other Requirements

- Database requirements: Support for both relational and NoSQL databases.
- Internationalization requirements: Support for multiple languages and localization options.
- Legal requirements: Compliance with data protection regulations such as GDPR.

## Appendix A: Glossary

- LDAP: Lightweight Directory Access Protocol
- API: Application Programming Interface
- OWASP: Open Web Application Security Project

## Appendix B: Analysis Models

- Data Flow Diagram: Illustrating the flow of data within the system.
- Sequence Diagram: Describing the relationships between system components.



## **Appendix C: To Be Determined List**

- *TBD items: Performance benchmarks, specific hardware requirements.*