**College of Professional Studies, Northeastern University**

**Professor Samuel Li**

**By**

**Rajan Rajeshrao Deshpande**

**Database Architecture Requirements for Loan Prediction and Risk Analysis System**

**1. Solution Client/Server Architecture**  
The Loan Prediction and Risk Analysis System will adopt a **three-tier client/server architecture** to optimize performance, scalability, and security.

* The **presentation layer** will consist of a user-friendly interface accessible to loan officers, system administrators, and data analysts through web or desktop applications.
* The **application layer** will handle business logic, connecting the user interface with the database via REST APIs or a middleware service. This layer will process user requests and manage interactions with the database.
* The **data layer** will host a relational database (e.g., MySQL, PostgreSQL) that stores customer information, loan applications, predictions, and system logs.

This architecture ensures efficient data access, modularity, and better management of concurrent operations.

**2. Cloud Hosting**  
Both the application and the database will be **hosted in the cloud** to ensure scalability, high availability, and cost-effectiveness. A cloud platform like **AWS (Amazon Web Services)** or **Microsoft Azure** will be used, leveraging services such as RDS (Relational Database Service) for the database and EC2 for the application. Hosting in the cloud provides flexibility to scale resources based on traffic and ensures disaster recovery through automated backups.

**3. Storage Requirements**  
The database will primarily handle structured data related to customers, loan applications, and predictions.

* **Data Volume**: Initial data volume is estimated to be moderate, with approximately 1–5 GB of storage for customer records, loan application data, and model predictions.
* **Growth**: The storage requirements are expected to grow over time as more customers and applications are processed. A scalable storage solution, such as Amazon S3 for archival data and RDS for active transactional data, will be implemented.
* **Backup and Recovery**: Regular automated backups will be scheduled to ensure data integrity and availability, with a recovery point objective (RPO) of less than 24 hours.

This architecture supports the project's objective of predicting loan risks while ensuring performance, reliability, and adaptability to future needs.

1. Snowflake and AWS Cloud to be used