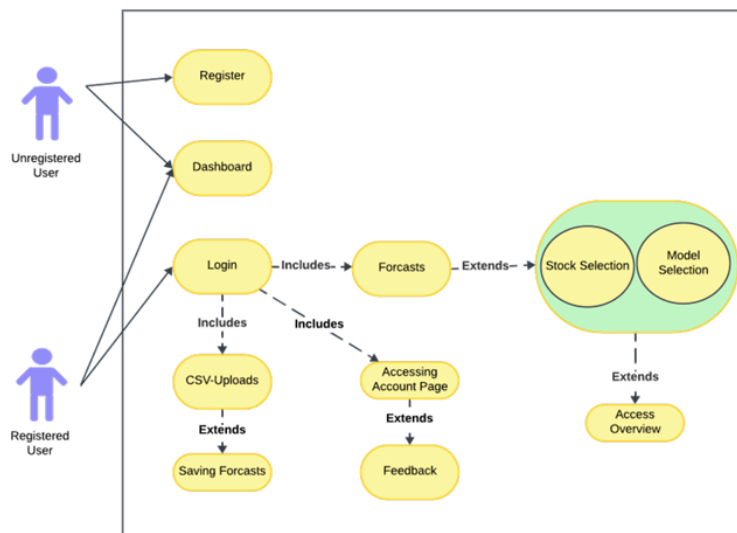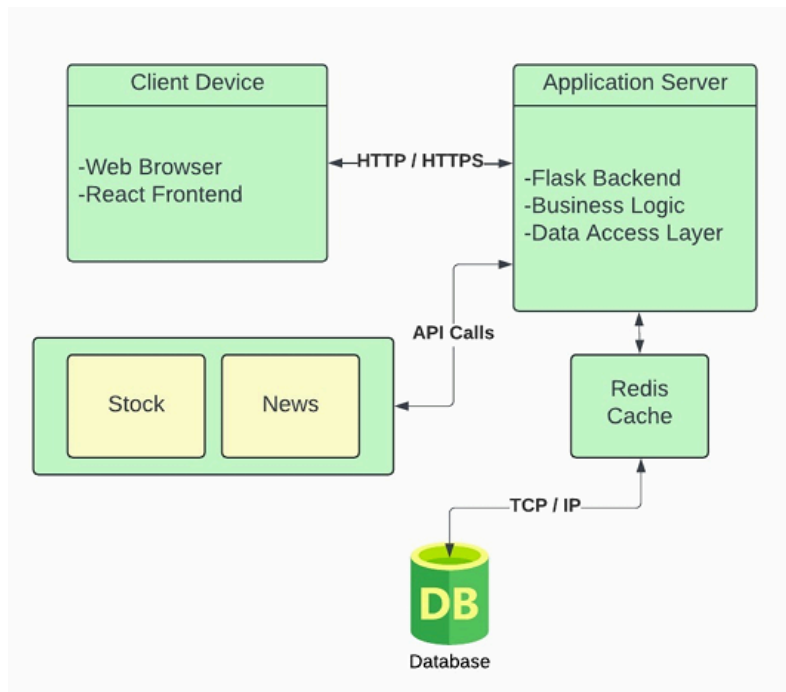# Stock Price Prediction System - End-to-End Design - Task 4

1. System Architecture Diagram

1.1 Use case diagram showing user interactions and functionalities

1.2 . Deployment View

2. Component Justification

Data Collection & Ingestion

Technology Choice:

- **Data Source:** Alpha Vantage API, Yahoo Finance, or IEX Cloud for real-time stock market data.
- **Ingestion Pipeline:** Apache Kafka (streaming) or AWS Glue (batch processing).
- **Storage:** Amazon S3 for raw data storage, PostgreSQL for structured data.

Justification:

- Reliable, real-time APIs provide high-frequency market data.
- Apache Kafka enables event-driven data ingestion for live updates.
- S3 offers scalable storage, while PostgreSQL supports structured querying.

Trade-offs:

- Real-time ingestion requires more computational resources compared to batch processing.
- Streaming systems introduce complexity in data handling.

Data Processing Pipeline

Technology Choice:

- **Preprocessing & Feature Engineering:** Apache Spark for large-scale processing.
- **Storage:** Feature Store using Amazon Redshift or Databricks.
- **Transformation:** Pandas for lightweight data processing tasks.

Justification:

- Spark processes large datasets efficiently.
- A Feature Store ensures consistency across training and inference.

Trade-offs:

- Spark requires cluster management and optimization.
- Maintaining a Feature Store adds operational overhead.

Model Operations

Technology Choice:

- **Training & Evaluation:** TensorFlow/Keras for deep learning models.
- **Model Versioning:** MLflow for tracking models.
- **Deployment:** AWS SageMaker, TensorFlow Serving, or FastAPI.
- **Monitoring:** Prometheus and Grafana for real-time model performance tracking.

Justification:

- TensorFlow is highly optimized for time series forecasting.
- MLflow enables experiment tracking and model lifecycle management.
- AWS SageMaker provides scalable deployment.

Trade-offs:

- SageMaker can be expensive for frequent retraining.
- TensorFlow Serving needs careful resource allocation.

Insight Delivery

Technology Choice:

- **Dashboarding:** Streamlit, Power BI, or Tableau.
- **Notification System:** WebSocket-based alerts.
- **User Interface:** Flask/Django backend with React.js frontend.

Justification:

- Streamlit simplifies interactive visualizations.
- React.js enables a dynamic and responsive UI.
- WebSocket ensures real-time alerts for brokers.

Trade-offs:

- Streamlit is limited in complex enterprise-level dashboards.
- WebSockets require persistent connections, adding server load.

System Considerations

Aspect	Solution	Justification

Scalability	AWS Lambda, Auto-scaling	Serverless computing scales based on demand
Reliability	Multi-AZ Database, Backup	Ensures fault tolerance
Latency	Edge computing for inference	Reduces response time.
Cost	Reserved Instances, Cost Optimization Strategies	Minimizes operational expenses

3. Data Flow Explanation

1. **Data Ingestion**: Market data is collected via APIs and stored in S3/PostgreSQL.
2. **Preprocessing**: Missing values handled, features engineered, and data standardized.
3. **Model Training**: LSTM/Transformer-based models trained on historical stock prices.
4. **Model Deployment**: The best model is deployed via TensorFlow Serving.
5. **Prediction & Insights**: Predictions sent to analysts via dashboards and alerts.

Stage	Batch or Streaming?	Reasoning
Ingestion	Streaming	Real-time price updates needed.
Processing	Batch	Feature engineering done periodically.
Prediction	Streaming	Immediate market insights required.

4. Challenge Analysis

Potential Challenges & Mitigation Strategies

1. **Data Latency & Inconsistencies**
   - Mitigation: Implementing a robust data pipeline with validation steps.

2. **Scalability Issues**

- Mitigation: Using auto-scaling cloud infrastructure to handle peak loads.

3. **Model Drift & Degradation**
   - Mitigation: Implementing continuous monitoring and periodic retraining.

4. **High Deployment Costs**
   - Mitigation: Using cost-optimized cloud services and scheduled retraining.

5. **Regulatory Compliance (GDPR, SEC, etc.)**

   - Mitigation: Ensuring data encryption and proper audit logs.

Conclusion

This end-to-end system ensures real-time stock predictions, integrates robust ML pipelines, and delivers valuable insights to financial analysts. With scalable infrastructure and continuous monitoring, it provides a production-ready solution that enhances trading strategies.