Streaming Data Analytics

In Consumer Services Data

Submitted By – **NAVISHA TANDON**
Roll No – **064091**

Project Use Case Overview

Introduction to the System

The project implements a Kafka-based streaming system designed to handle customer service interactions in real-time. It is structured into two primary components: the Producer and the Consumer.

The Producer reads customer service interaction logs from a CSV file and streams the data to a Kafka topic, `consumerdata`. Simultaneously, the data is stored in a MongoDB database. The Consumer processes the streamed data, triggers alerts based on predefined thresholds, and stores the processed data in MongoDB for long-term analysis and reporting.

This system is designed to manage and analyze large-scale data in real-time, ensuring better decision-making and operational efficiency.

Key Features

The project's key features include real-time data streaming, anomaly detection through alerts, secure data storage, and scalability. These features collectively enhance the system's utility in managing customer service operations.

Methodology

Producer Workflow

The Producer is designed to read customer service logs from a CSV file and stream the data to Kafka while simultaneously inserting it into MongoDB.

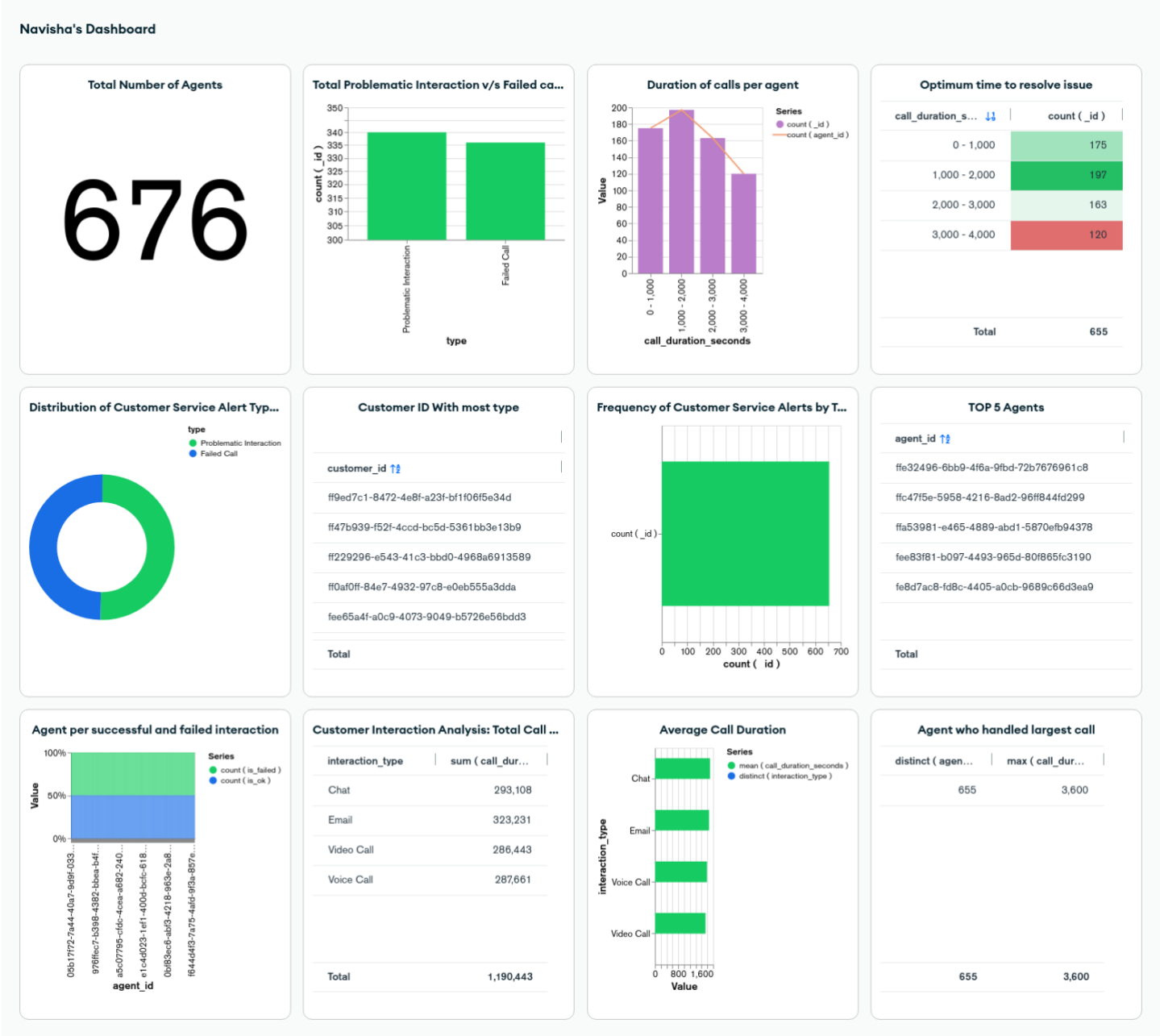
- **Step 1:** Initialize a Kafka producer to connect with the Kafka broker.
- **Step 2:** Read the CSV file containing customer service interaction data.
- **Step 3:** Iterate through each record, construct a JSON message, and send it to the Kafka topic `consumerdata`.
- **Step 4:** Insert the same data into MongoDB for persistent storage.
- **Intent:** The Producer ensures that data is both streamed to Kafka for real-time processing and stored in MongoDB for future analysis and audits.

Consumer Workflow

The Consumer processes the streamed data, performs analysis, and triggers alerts.

- **Step 1:** Connect to the Kafka topic `consumerdata` to consume messages.
- **Step 2:** Parse the incoming data to extract relevant fields such as customer ID, agent ID, call duration, and timestamps.
- **Step 3:** Store the processed data in MongoDB and evaluate predefined thresholds, such as long call durations or failed calls.
- **Step 4:** Trigger alerts and insert them into an alerts collection in MongoDB.
- **Intent:** The Consumer ensures real-time anomaly detection and stores data for reporting, enabling timely action and operational insights.

Insights from MongoDB Dashboard:



1. Total Number of Agents (676)

The total workforce consists of 676 agents, indicating a medium-scale operation.

2. Total Problematic Interactions vs. Failed Calls

Problematic interactions and failed calls are evenly distributed, with failed calls making up nearly 50% of the total. These issues are often concentrated among specific customers and agents, indicating potential systemic or skill-related problems.

Example: A customer (`ffe0b4`) recorded over 20 failed calls, highlighting unresolved grievances or recurring technical issues.
Actionable Insight: Identify top agents/customers with repeated failed calls and implement targeted troubleshooting, such as training agents or improving technical infrastructure.

3. Duration of Calls per Agent

The majority of calls fall within the 1000–2000 seconds range. Calls lasting over 3000 seconds significantly consume resources and often involve complex issues. Specific agents, such as `ffa53981`, are consistently handling extended calls, which may indicate a need for escalation protocols.

Example: Agent `ffa53981` managed 5 calls exceeding 3000 seconds, likely due to unresolved issues requiring additional attention.
Actionable Insight: Introduce a process to flag and escalate calls exceeding a predefined duration to senior agents or specialized teams to optimize resource utilization.

4. Optimum Time to Resolve Issues

Most issues are resolved within 1000–2000 seconds, with 197 issues successfully addressed in this range. Beyond 2000 seconds, resolution rates decline, indicating inefficiencies in extended interactions.

Actionable Insight: Focus on resolving 90% of issues within the 2000-second window to improve efficiency.

5. Distribution of Customer Service Alert Types

Problematic interactions and failed calls are evenly split. Customers raising the most alerts often use email and voice call channels, indicating challenges in these specific modes of communication.

Example: A customer (`ffe65a4`) frequently raised alerts through email and voice calls, suggesting potential channel-specific inefficiencies.

6. Customer ID with Most Alerts

Certain customers, such as `ff9ed7`, consistently trigger the highest number of alerts across multiple interaction types. These customers are at a higher risk of churn if their issues are not adequately addressed.

Example: Customer `ff9ed7` logged 12 failed voice calls within a short period, highlighting unresolved issues.

7. Frequency of Customer Service Alerts by Type

Alerts are most frequently associated with video calls, which tend to involve more complex issues and longer handling times.

Example: Customers using video calls contribute disproportionately to the alert volume due to technical or resolution challenges.

8. Top 5 Agents

The top-performing agents include `ffe32496` and `ffc4715e`, who handle a high volume of calls efficiently while maintaining resolution quality. These agents can serve as role models for training underperforming staff.

Example: Agent `ffe32496` successfully resolved 25 complex calls exceeding 2000 seconds, showcasing exceptional adaptability.

9. Agent Performance in Successful and Failed Interactions

Certain agents, such as `ffa53981`, show a higher failure rate due to disproportionate workloads or skill gaps. Redistribution of calls or targeted training can help address this.

Example: Agent `ffa53981` recorded 10 failed calls, with an average call duration exceeding 2000 seconds.
Actionable Insight: Reallocate calls for underperforming agents or provide additional resources (e.g., scripts, tools).

10. Customer Interaction Analysis by Total Call Volume

Email and chat interactions dominate the total call volume, indicating a preference for asynchronous communication among customers. Voice and video calls, while fewer, tend to involve longer durations and complex issues.

Example: Emails account for 293,231 seconds of interaction time, highlighting their importance as a preferred communication channel.

11. Average Call Duration by Interaction Type

Video calls have the highest average duration (~900 seconds), reflecting their complexity. Chats have the lowest duration, suggesting efficiency in asynchronous interactions.

Example:
Reducing video call duration by 10% (e.g., through pre-call forms) can save significant resources.
Actionable Insight: Create FAQs and pre-call guides to reduce call duration while maintaining satisfaction.

12. Agent Who Handled the Largest Call

The longest call (3600 seconds) was managed by agent `ffa53981`, involving a video interaction. Such instances are rare but require efficient handling strategies to avoid resource drain.

Example: This prolonged call highlights the need for escalation protocols for lengthy or complex interactions.

Application in Business and Customer Service Industry

- **Customer Service Optimization:** The system improves customer service operations by identifying and addressing inefficiencies. For instance, prolonged call durations may indicate complex issues requiring additional agent training.
  Example: In call centres for e-commerce platforms, prolonged calls during peak seasons can indicate systemic issues such as product unavailability or technical difficulties, prompting immediate operational adjustments.

- **Fraud Detection:** The system detects fraud attempts through monitoring of suspicious activities. For example, multiple failed calls or problematic interactions might highlight phishing attempts or fraudulent bots.
  Example: In the financial sector, repeated failed calls could indicate unauthorized attempts to access customer accounts, enabling early intervention to prevent breaches.

    - Fraud email detection can also be integrated by analyzing email metadata, identifying mismatched sender addresses, and tracking repeated invalid login attempts.
      Example: A bank identifying phishing scams targeting customer accounts through metadata analysis can act promptly to secure user data.

- **Strategic Decision-Making:** The system provides actionable insights from the data, aiding strategic decisions. Analyzing customer behavior trends helps businesses refine their services.
  Example: Telecom companies using the system can identify regions with high call volumes, reallocating resources effectively to manage demand.

Future Enhancements

- **Advanced Alert Mechanisms-** Incorporating machine learning algorithms will enhance pattern detection and enable sentiment analysis to identify dissatisfied customers.
  Example: Telecom companies can use automated sentiment analysis to identify frustrated customers, ensuring timely escalations.

- **Multi-Channel Integration-** Expanding the system to monitor email, chat, and social media interactions will provide a unified customer service view.
  Example: Retail chains can integrate all customer interaction channels, delivering seamless support across platforms.

- **Dashboard and Visualization-** Interactive dashboards can visualize KPIs and alerts, empowering businesses to adapt thresholds and configurations.
  Example: Healthcare providers monitoring patient inquiries can prioritize urgent cases using real-time interactive dashboards.

- **Expanded Fraud Detection-** Voice data analysis and external database integration can improve fraud detection capabilities.
  Example: Financial institutions analyzing voice interactions can prevent synthetic voice phishing attempts, protecting user assets.

Conclusion

This project provides a framework for handling customer service data in real-time. By integrating advanced alerting and fraud detection mechanisms, the system enhances operational efficiency and protects businesses from potential threats. The system also delivers actionable insights that drive strategic decision-making and elevate customer service standards.

With future enhancements, this solution can cater to broader industry needs, ensuring scalability and adaptability in a dynamic business environment.