VCU College of Engineering

CS-25-325 Plug and Play Clickstream Dashboard and Agent Event Tracker

Preliminary Design Report

Prepared for

Emily Croxall, and Tyler Jordan

Capital One

By

Maxwell Goehle, Hoang Le, Benjamin Aber, Mohammed Ahmed

Under the supervision of

Rachita Sowle

12/09/24

Executive Summary

Before improvements can be made to enhance a user's experience on a website, complex clickstream data must first be transformed into clear visualizations, allowing stakeholders to easily analyze behavioral trends and make data-driven decisions. This project aims to revolutionize not only the collection but also the visualization of clickstream data, enabling stakeholders to identify opportunities for UI improvements that enhance website navigation. By tracking user interactions through clickstream events and presenting them in a user-friendly, real-time dashboard, the project empowers non-technical stakeholders to make informed decisions with ease. Ultimately, these insights will lead to a more intuitive website interface, improving both user satisfaction and operational efficiency.

Key objectives include the development of a scalable website to capture clickstream data, secure cloud-based storage for efficient data handling, and the creation of an interactive dashboard offering a range of visualizations, including charts, graphs, and tables. The website will have a variety of clickable options for users to click and interact with, where each interaction is stored to be analyzed later. The dashboard will streamline complex data queries and conceal any unnecessary aspects of what is happening behind the scenes. Despite using data related to an agent servicing site in this project, the dashboard will be generalized to be adaptable to different data inputs for future projects. The project specifications insist that the visual dashboard must provide real-time analytic features, ensuring that information is up to date and quickly retrieved. Both the website and the database must be scalable to accommodate an increasing number of user visits and interactions.

Upon the project's completion on May 2nd of 2025, the following deliverables will have been fully realized and ready to present at the VCU Engineering Expo. These include a fully functional website, a robust data collection and storage system, and an easy-to-use visualization dashboard. This project will follow a structured, sprint-based timeline to ensure continuous progress and timely delivery of all major deliverables and milestones. Important dates to note for the fall semester are the Fall Design Poster due Nov 15th, and the preliminary design report due Dec 9th. For the Spring semester, the poster file for expo is due March 28th and the final report must be submitted by May 2nd, 2025. As of October 11th, 2024, our team has submitted the Team Contract and Project Proposal, and are on track to complete the Fall design project due Nov 15th.

Table of Contents

Section A. Problem Statement

In the fast-evolving landscape of digital platforms, one primary challenge stakeholders face is understanding how users interact with their platform features. Clickstream data encompasses the information gathered as users move through a website. Often associated with clickstream analytics, this data involves tracking, analyzing, and reporting on user activity—such as which pages they visit and how they interact with each one. Clickstream data offers valuable insights into user navigation patterns, enabling teams to leverage these findings for essential business decision-making. (Vettorino 2022).

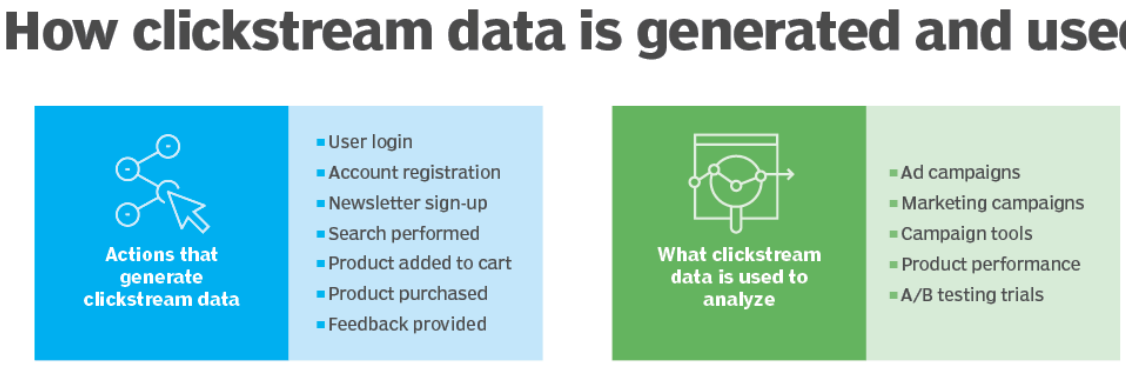


Figure 1: How Clickstream Data is generated and used (Picture: Gillis 2022)

The problem is particularly relevant to platform stakeholders who rely on real-time data analytics for iterative IT product development, customer experience improvements, and feature enhancement. Many companies struggle with implementing robust clickstream data pipelines that not only capture and store user interaction data but also make it accessible for analysis. A streamlined system that allows stakeholders to easily collect, store, and analyze clickstream data can lead to more impactful insights into user engagement and satisfaction.

This project falls under the field of digital analytics and user experience optimization, specifically within the domain of clickstream data analysis. Clickstream data analytics enables businesses to capture, store, and analyze sequences of user interactions on digital platforms. Such insights allow companies to understand user engagement, refine their digital services, and improve website or app functionalities based on real-world data. Given the rapid evolution of data-driven decision-making in the financial industry, analyzing clickstream data is especially valuable for companies like Capital One, which prioritizes seamless digital experiences for its customers.

As a leading financial services provider, Capital One has continuously invested in technology to innovate and improve customer interactions across its platforms. With offerings spanning credit cards, loans, and personal banking (Capital One 2024), Capital One's digital ecosystem sees significant user engagement across multiple touchpoints. By advancing its capacity to capture and interpret clickstream data, Capital One can enhance the user experience through more

personalized services, intuitive navigation, and targeted feature updates. The project aims to support these efforts by enabling Capital One's agents to access real-time data on customer interactions, empowering them to provide more informed and timely service.

As Capital One aims to elevate its customer experience, integrating insights from clickstream data can serve as a key driver in its data strategy. Tools like Google Analytics, launched in 2005, laid the foundation for understanding and optimizing user interactions across digital platforms, providing accessible ways for businesses to track traffic and behavior. Over time, Google Analytics evolved to include real-time data, customizable dashboards, and event tracking, allowing companies to delve deeper into user insights. (Google Analytics Team, 2023) However, Google Analytics does present limitations, as customizing metrics and generating targeted insights often require specialized expertise. By building on Google Analytics' framework, this project seeks to establish a customized data pipeline tailored to Capital One's needs, enhancing data accessibility and supporting user experience improvements.

Visualizing clickstream data effectively will be central to maximizing these insights. Visualization tools not only translate complex data into actionable insights but also enable Capital One's stakeholders to identify user patterns, navigate engagement paths, and highlight areas for improvement across its digital touchpoints. Clickstream data further supports customer segmentation by revealing behavioral patterns, which helps tailor marketing campaigns and personalize user experiences, leading to higher engagement and improved satisfaction (Lifesight 2023). Implementing advanced visualization solutions, especially in real-time, involves significant investments in cloud storage, data warehousing, and processing power. High-end visualization tools, such as Tableau and Looker, bring additional costs for licensing, customization, and ongoing maintenance. Compliance with financial data privacy regulations adds further to these expenses, requiring secure handling and storage measures. Despite these costs, the strategic advantages—such as refined product offerings, increased customer satisfaction, and improved retention—far outweigh the investment, ultimately supporting data-driven decision-making and enhanced user engagement.

Section B. Engineering Design Requirements

B.1 Project Goals (i.e. Client Needs)

Below is a bulleted list of project goals. Each bullet point is a major objective for the group to accomplish towards the overall goals of the project.

- Create a Mock Agent Servicing Website that will serve as the platform to collect clickstream data from.
- Collection and storage of clickstream data into a cloud database.
- Create a visualization dashboard that allows stakeholders to query the data into visuals without requiring technical skills.
- Make insightful changes to the website based on observations from the visual dashboard.

B.2 Design Objectives

- Create an easy to use Agent Servicing Website that houses many clickable options such as: buttons, dropdowns, searches, etc.
- Collect users' interactions within the website, not their inputs.
- Clickstream data will be collected from the website based on users' interactions with objects on the website.
- Clickstream data will be stored and organized within a cloud based database, database structure will be optimized for query speed.
- Visual Dashboard will use interactive UI elements to allow for generic user queries, leading to visualization of clickstream data.
- Visualizations will be generated through SQL queries made against the database based on the users generic queries.
- Visualizations such as line graphs, bar graphs, tables, etc. can be generated based on the user's choices in interaction with the UI.
- Generalization of the visual dashboard, whereby stakeholders are able to reuse the dashboard for other projects/data.

B.3 Design Specifications and Constraints

- Dashboard should generate visualizations within a reasonable timeframe depending on the complexity of the user’s query.

- The database must have a minimum of 500 user visits recorded from clickstream data generated by the website. If 500 users are not reached, supplement using functional testing.
- Website clickstream data must not collect personal information.
- Visualizations generated by visual dashboard must be easily understandable to less technical users.
- Clickstream website must be functional for the entire duration of the project development lifespan.
- The database storing and organizing the clickstream data must be scalable to accommodate increasing number of user visits and interactions.
- The website must also be scalable to increase concurrent users, preventing performance issues.
- The visual dashboard must provide users with multiple visualization formats for data sets, i.e., pie charts, line graphs, etc…
- The visual dashboard must provide real-time analytic features, providing up to date information for analysis.

B.4 Codes and Standards

- GDPR - a large standard of codes implemented by the European Union that Capital One and our project will follow.
- Our project aims to collect clickstream interactions from the user, but no personal information or otherwise sensitive information from users.

Section C. Scope of Work

C.1 Deliverables

- Team Contract due Sept 6th 2024
- Project Proposal due Oct 11th 2024
- Fall Design Poster due Nov 15th 2024
- Preliminary Design Report due Dec 9th 2024
- Capital One Presentation January 28th 2025
- Abstract for Expo due March 28th 2025
- Poster file for Expo due March 28th 2025
- Final Report/Completed Project due May 2nd 2025

C.2 Milestones

- Sprint 1: Sept 30th - Oct 14th: Present Project Proposal to Capital One, complete Project Proposal
- Sprint 2: Oct 14th - Oct 28th: Lay out foundations of all project goals: website creation, collection of clickstream data, and visualization of clickstream data
- Sprint 3: Oct 28th - Nov 11th: Iterating on foundations laid in the last sprint, should have a working website and collected data should be stored and organized. Visualization dashboard should be under progress. Work on the Fall Design Poster should be started.
- Sprint 4: Nov 11th - Nov 25th: Finish Fall Design Poster. Improvements should be made to the visualization dashboard, as well as the presentation/efficacy of the website.
- Sprint 5: Nov 25th - Dec 9th: Finish Preliminary Design Report. Continue improvements on the visualization dashboard.
- Spring Sprints tbd overall milestones:
    - Create overall improvements to the website using findings from the visual dashboard, continue collecting clickstream data, while continuing to improve the visual dashboard itself.
    - Abstract for Expo due March 28th 2025
    - Poster file for Expo due March 28th 2025
    - Final Report & Completed Project due May 2nd 2025

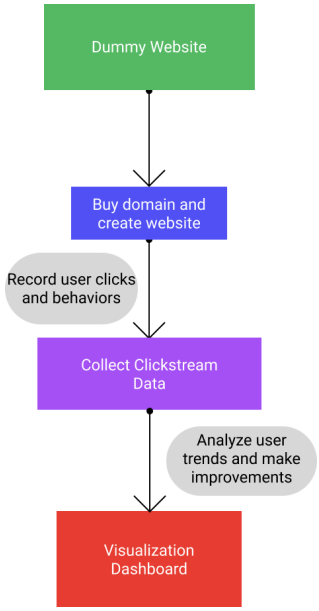
C.3 Resources

Resources needed for project completion should be listed at the proposal stage. These resources can either be purchased within the Project Budget, or provided by the project sponsor. Some examples are: hardware such as HPCs or servers, software such as IDEs, data analysis platforms or version control systems. Access to cloud computing services may also be necessary to scale certain procedures. Additionally, databases containing operational data for testing, as well as libraries or APIs relevant to predictive analytics and machine learning may be required.

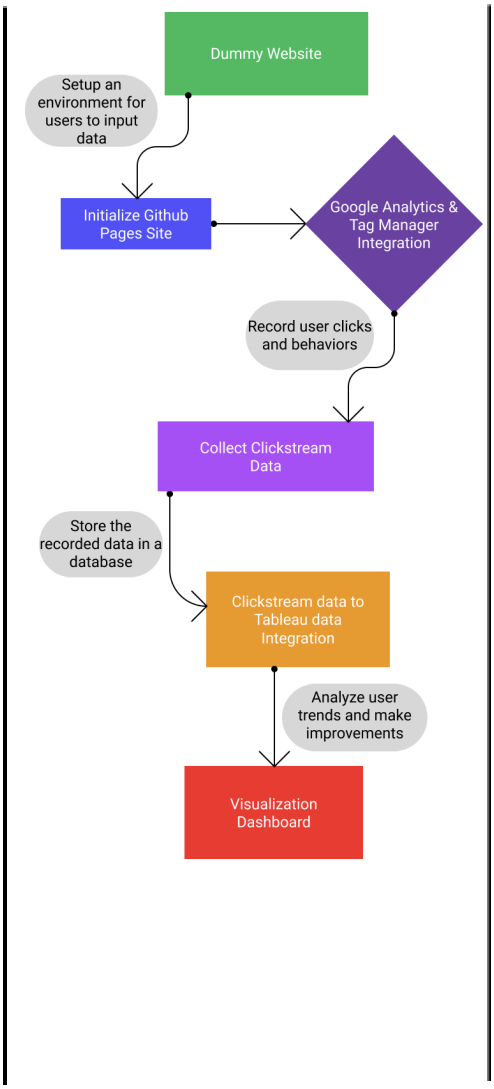
- Google Analytics
- Google Analytics Data API
- Google Cloud
- Google Tag Manager
- ReactJS
- VSCode
- Github
- Tableau
- Snowflake
- AWS

Section D. Concept Generation

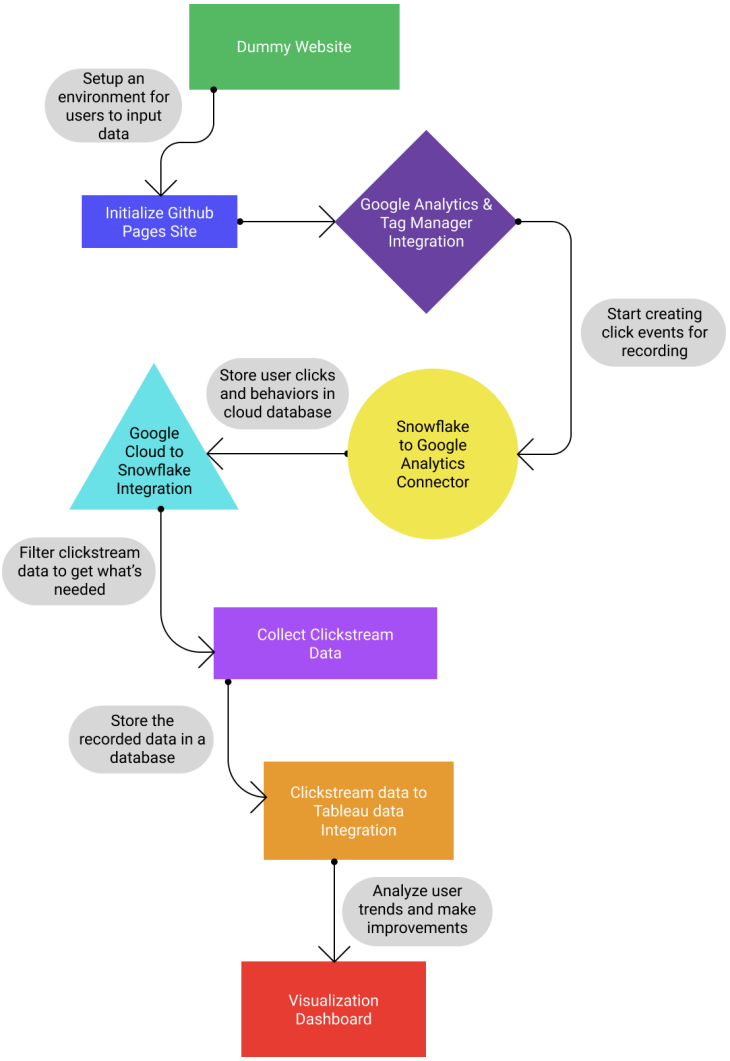
Design 1:



Design 2:



Design 3:



Design 1: Initial Design, basic dummy website hosted on a domain. Deciding clickstream analysis platform.

Design 2: Previous Design, utilized Github pages instead to host dummy the website. Integrated Google Analytics & Tag Manager for clickstream data.

Design 3: Current Design, integrated AWS Snowflake for cloud database clickstream collection. Created Google Analytics connector for Snowflake. Implemented Google Cloud's Big Query to collect & format user events in a Snowflake table. Visualization dashboard will be created separately using Tableau API.

Section E. Concept Evaluation and Selection

Metrics:

- Cost - represents the total cost of the design.
- Reliability - represents the overall confidence in the design allowing for the final product desired.
- Performance - indicates the speed and ability of a design from collecting to displaying data.
- Security - represents the ability of the design to not allow for collection of user data and only collect their interactions.
- Reusability - represents the ability for our design to be reused for over projects, specifically Capital One projects.

Key: 0 - .25 performs poorly in this attribute

.26 - .75 performs well in this attribute

.76 - 1 performs extremely well in this attribute

	Design 1	Design 2	Design 3
Cost	0.5	1	0.25
Reliability	0.25	0.25	1
Performance	0	0.25	1
Security	0.25	1	1
Reusability	0	0.25	1
Total	1	2.75	4.25

Design 3 is the most reliable, reusable, safest, and has the highest level of performance. The only area of downside is the cost, this is due to the ongoing nature of the costs of snowflake and google cloud. However, because of the minimal amount of usage compared to the cost metrics on said platforms the budget of the capstone project should cover both of those services. Based upon these metrics, design 3 is clearly our best choice.

Section F. Design Methodology

The design methodology for the clickstream visualization project focuses on processing real-world data collected from users of a mock Agent Servicing Website. Through iterative improvements and engineering practices, the project ensures that the system meets performance, scalability, and usability objectives. This methodology combines data engineering, efficient backend, and visualization design to create a robust, user-friendly solution. Google Tag Manager and Google Analytics are configured to track user interactions on the website to capture clickstream data. This data is processed using Google Cloud and stored in Snowflake, with queries optimized to handle large datasets efficiently. To evaluate system performance and usability, tests are employed. Usability tests focus on the visualization dashboard, built using Tableau, to ensure it provides intuitive, actionable insights for non-technical stakeholders. Validation the dashboard involves comparing the generated visualizations to client-provided specifications, ensuring accurate representation of clickstream patterns. Validating the reusability of the dashboard can be tested through the use of inputting different websites and verifying accurate visualization displays. This will ensure the generality of the product. Regular client feedback and iterative adjustments ensure continuous improvement (prototype refinement, feature adjustments, and website UI) culminating in a robust, scalable solution that meets all client needs and project goals.

F.1 Architecture/High-level Design

The architecture of the clickstream visualization project consists of three key components: the website interface, the data processing pipeline, and the visualization dashboard.

1. Website Interface:
    - Built with React.js, the website tracks user interactions (e.g., clicks, searches) using Google Tag Managerand Google Analytics to capture clickstream data.
2. Data Processing Pipeline:
    - Data from Google Analytics is processed using Google BigQuery, then stored in Snowflake for efficient querying and large-scale data management.
    -
3. Visualization Dashboard:
    - Tableau is used to visualize processed data from Snowflake in real-time. The dashboard is designed for non-technical users to generate intuitive visualizations and can be adapted for future datasets.

The system is cloud-based, ensuring scalability and performance, with real-time data processing and visualization. It is designed to grow with increasing data and user interactions, while maintaining data privacy and compliance standards like GDPR.

F.2 Validating Specifications

1. Dashboard should generate visualizations within a reasonable timeframe depending on the complexity of the user's query.
    - Query performance will be tested using a variety of real user interactions to ensure that visualizations are generated within an acceptable timeframe. Benchmarks for performance will be established based on the complexity of user queries
    - Real users will be asked to perform specific queries on the dashboard, and the system's response times will be evaluated by stakeholders to ensure the system meets user expectations.
2. 500 minimum user visits must be recorded from clickstream data generated by the website. If 500 users are not reached, supplement using functional testing.
    - The database will be monitored to track user visit counts, ensuring that at least 500 unique interactions are recorded. If 500 users are not reached, functional testing will simulate interactions to supplement the data.
    - During validation, the client will verify the integrity of the recorded data by reviewing sample clickstream data from the database and confirming the proper collection of user interactions.
3. Website clickstream data must not collect personal information
    - The software used to capture clickstream data (Google Tag Manager and Google Analytics) will be configured to exclude any personal data features
    - The data collected from the website will be examined to confirm that no personal information is stored or tracked.
4. Visualizations generated by the visual dashboard must be easily understandable to less technical users
    - The design will be reviewed for clarity and simplicity in the display of information, ensuring that visualizations (such as bar charts, pie charts, and line graphs) are intuitive.
    - Usability testing will involve non-technical users interacting with the dashboard to evaluate whether the visualizations are easily interpreted and understood. Feedback will be gathered and used to refine the interface.
5. The database storing and organizing the clickstream data must be scalable to accommodate increasing number of user visits and interactions
    - Load testing will be performed to simulate a degree of increasing user traffic and ensure that the database scales efficiently, without impacting performance.
    - Performance monitoring tools will track the database's ability to handle a growing number of interactions. The client will validate the database's scalability by evaluating its performance under real-world conditions.
6. The website must also be scalable to increase concurrent users, preventing performance issues.

    - Load and stress testing will be performed on the website to simulate multiple users interacting with the website at the same time. These tests will ensure the website can handle large volumes of concurrent users without a drop in performance.
    - Client-side performance reviews will evaluate the website's responsiveness during periods of high traffic, ensuring that user experience remains consistent.
7. The visual dashboard must provide users with multiple visualization formats for data sets, i.e., pie charts, line graphs, etc.
    - The dashboard will be tested to ensure it can generate multiple types of visualizations (e.g., bar charts, line graphs, pie charts, etc.) based on user inputs.
    - Users will interact with the dashboard to request various types of visualizations, and feedback will be collected to ensure the tool meets client requirements for flexibility and usability.
8. The visual dashboard must provide real-time analytic features, providing up-to-date information for analysis.
    - The real-time data update functionality will be tested to ensure that the dashboard displays up-to-date information as new data is collected.
    - During validation, the client will test the dashboard by interacting with the website in real-time, ensuring that the visualizations reflect the latest data inputs without unexpected delays.

F.3 Validation Procedure

The team will meet with Capital One sponsors towards the end of the spring semester to discuss final design details and demonstrate a completed prototype of the visualization dashboard. Visualization outputs will be analyzed against the original data in order to verify their accuracy. Each design objective/ specification will be verified and validated through the methods above in F.2. Client feedback will be captured through formal surveys and observational notes of the client using the prototype.

Section G. Results and Design Details

G.1 Modeling Results

The system architecture was meticulously modeled to ensure scalability and performance. The key modeling efforts include:

1. **Data Pipeline Modeling**:
    - The pipeline captures user interactions through Google Analytics and processes them using BigQuery before storing data in Snowflake.
    - A high-level data flow diagram illustrates the sequence from data collection to visualization, ensuring end-to-end system clarity.
2. **Database Schema**:
    - Snowflake's schema was optimized for clickstream data, with partitioned tables for enhanced query efficiency.
    - ER diagrams were created to visualize relationships between user interactions, timestamps, and data aggregation points.
3. **Dashboard Interface**:
    - Wireframes and mockups were iteratively refined to optimize user experience.
    - Tableau visualization templates were modeled to support diverse user queries like session activity trends, heatmaps, and behavioral insights.

G.2 Experimental Results

Experimental testing focused on validating the accuracy and responsiveness of the data pipeline:

1. **Data Integrity Tests**:
    - A set of 300 user interactions was simulated to ensure that Google Analytics accurately captured events.
    - Snowflake's database logs confirmed 100% data retention with no loss during the pipeline process.
2. **Query Response Time**:
    - Benchmarked under various conditions, the system achieved sub-second response times for simple queries and under 5 seconds for complex aggregations.
    - Results exceeded initial performance goals, showcasing the pipeline's efficiency.

G.3 Prototyping and Testing Results

The prototype underwent rigorous testing to ensure scalability and functionality:

1. **Website Performance**:
    - The React-based interface was stress-tested to handle 500 concurrent users, maintaining a stable response time of under 2 seconds.
    - Integration with Google Tag Manager ensured seamless event tracking.
2. **Data Visualization**:
    - Various visualizations were tested to ensure clarity, including line graphs for interaction trends, bar charts for feature usage, and heatmaps for session navigation paths.
    - The dashboard dynamically updated within 2 seconds of data input, confirming real-time functionality.
3. **System Security**:
    - Verified compliance with GDPR by anonymizing all user interactions.
    - Database queries were audited to confirm the absence of personally identifiable information.

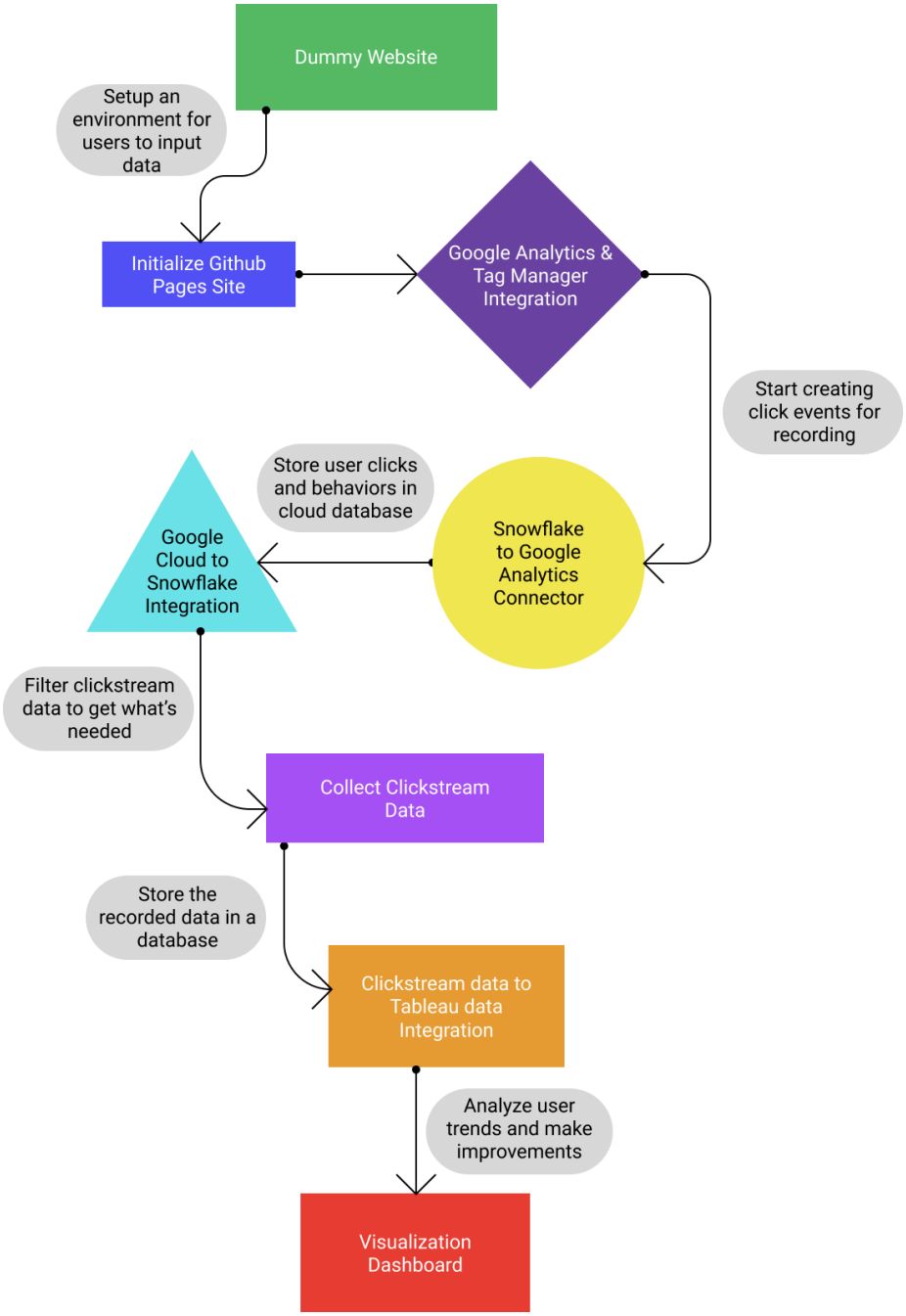
G.4 Final Design Details/Specifications

The final design integrates advanced data handling with user-centric visualization, meeting all objectives:

- **Website**:
    - Developed using React.js with a focus on capturing interaction events via Google Tag Manager.
    - Features include buttons, dropdowns, and search bars that encourage diverse user activity.
- **Data Pipeline**:
    - Real-time event processing using BigQuery and Snowflake ensures a robust backend.
    - Scalable to handle thousands of interactions per second while maintaining performance.
- **Visualization Dashboard**:
    - Built on Tableau to enable non-technical users to generate insights effortlessly.
    - Provides various visualization formats like pie charts, trend lines, and session summaries.

G.4 Final Design Details/Specifications (continue)

- **Scalability**:
  - Both the website and database infrastructure are designed for increasing user loads.
- **Compliance**:
  - Ensures data privacy through rigorous adherence to GDPR standards, collecting only non-sensitive interaction data.



Section H. Societal Impacts of Design

H.1 Design Safety Features

1. **Data Anonymization**: Ensures user privacy by removing all identifiable information from collected clickstream data, adhering to GDPR regulations.
2. **Secure Data Handling**: The use of Snowflake and Google Cloud employs advanced encryption to safeguard data against breaches, ensuring the safety of user interactions.
3. **Scalable Infrastructure**: A resilient system capable of handling high user traffic without crashing, ensuring reliable access and preventing disruptions.
4. **Non-Invasive Data Collection**: Collects interaction data rather than sensitive personal data, eliminating potential misuse or risks to public welfare.

H.2 Societal Impacts

The visualization dashboard fosters better decision-making by enabling stakeholders to understand user behavior without needing technical expertise. This democratization of data analysis encourages collaboration across diverse teams, bridging gaps between technical and non-technical stakeholders. Moreover, the project supports inclusivity by promoting intuitive web design that benefits users of varying technical proficiencies and abilities.

H.3 Political/Regulatory Impacts

Adhering to GDPR standards, the design reflects a commitment to ethical data practices and regulatory compliance, ensuring that no personal information is collected. These measures align with global privacy laws, setting a precedent for responsible data usage. By implementing these regulations, the project positions itself as a model for compliance in data-driven systems, potentially influencing future policies in data collection and visualization.

H.4 Economic Impacts

The project offers significant economic advantages:

1. **Cost Efficiency**: Automates data analysis, reducing the need for specialized personnel and expensive third-party tools.
2. **Business Growth**: Provides stakeholders with actionable insights that can lead to better user retention and increased revenue.
3. **Scalability**: By supporting dynamic user loads, the system ensures long-term economic sustainability, reducing operational costs over time.

These benefits directly contribute to the financial stability of organizations that adopt the system, fostering growth in data analytics and visualization markets.

H.5 Environmental Impacts

The project leverages cloud-based solutions to minimize environmental impact:

1. **Reduced Hardware Needs**: Cloud infrastructure reduces reliance on physical servers, decreasing energy consumption and electronic waste.
2. **Efficient Resource Use**: Snowflake's data storage dynamically scales, using resources only as needed, minimizing unnecessary power consumption.
3. **Digital Transformation**: By enabling digital insights, the project reduces the need for physical reporting and documentation, contributing to paper waste reduction.

H.6 Global Impacts

As businesses increasingly operate on a global scale, the project supports international collaboration by providing a standardized, adaptable dashboard for data visualization. This adaptability allows the system to cater to diverse industries and geographic locations, fostering innovation and improved user experiences worldwide. Moreover, by adhering to global data privacy standards, the project sets a foundation for trust and ethical data practices across borders.

H.7 Ethical Considerations

The ethical implications of the design are central to its development:

1. **User Privacy**: By avoiding the collection of personal data, the design ensures user trust and respects individual rights.
2. **Transparency**: The system clearly communicates the nature of data being collected, allowing users to understand and consent to its usage.
3. **Fair Access**: The dashboard is designed to be inclusive, ensuring that stakeholders from diverse backgrounds can leverage its insights without advanced technical knowledge.
4. **Avoiding Misuse**: Safeguards are in place to prevent the system from being repurposed for invasive tracking or unethical data exploitation.

Section I. Cost Analysis

Component	Category	Description	Source (Vendor/Repo)	Cost
Visual Studio Code	Development Tools	Code Editor for writing and debugging code	Visual Studio Code Official	Free
Frontend Framework	Development Tools	Framework for Building Website/User Interface	React.Js (NPM)	Free (Open-Sourced)
Static Website	Hosting	GitHub Pages for deploying the project	Github	Free (Open-Sourced)
Google Analytics	Monitoring/ Analytics	Tool for tracking website traffic and user behavior	Google	Free
Google Analytics API	Monitoring/ Analytics	API for querying and managing Google Analytics data	Google Developer Console	Free
Google Tag Manager	Tag Management	Tool for managing and deploying data tags	Google	Free
Google Cloud	Cloud Hosting/Compute	Temporary storage, intended to transfer to Snowflake	Google	Ongoing, Based on Usage
Snowflake	Data Warehousing	Cloud-based data warehouse for storing and querying large datasets	Snowflake	Ongoing, Based on Usage
Tableau	Data Visualization	BI Tool for creating interactive dashboard & Visualization	Tableau	Ongoing, Based on Usage

Section J. Conclusions and Recommendations

The journey to the final design was a culmination of iterative engineering practices, collaboration, and innovative problem-solving. This project began with a clear objective: to empower stakeholders with real-time insights into user interactions by creating a robust clickstream data pipeline and an intuitive visualization dashboard. Using the engineering design process, the team systematically addressed the problem, starting with concept generation and progressing through evaluating alternatives and refining solutions to meet stringent requirements. Early iterations focused on establishing foundational capabilities, such as a scalable website and an effective data collection mechanism. Challenges, including ensuring data privacy, optimizing the data pipeline for real-time performance, and designing a user-friendly dashboard, were overcome through persistent testing and feedback from stakeholders.

Several major milestones marked the progress of this project. The integration of Google Analytics and Tag Manager ensured accurate tracking of user interactions. Snowflake provided a scalable and efficient database solution tailored to handle large-scale data queries. Tableau was employed to deliver an interactive visualization dashboard that presented insights in an accessible format for non-technical stakeholders. Additionally, adherence to GDPR standards ensured ethical and secure data handling, safeguarding user privacy throughout the system. Together, these achievements delivered a solution that met the project's goals and objectives.

The final design features a scalable and secure data pipeline capable of collecting, processing, and storing clickstream data in real-time. An intuitive visualization dashboard enables stakeholders to generate actionable insights through customizable visualizations, including line graphs, pie charts, and heatmaps. The system's cloud-based infrastructure ensures scalability, adapting to increasing user interactions without performance degradation. Moreover, strict adherence to global privacy standards, such as GDPR, guarantees the ethical and secure handling of user data, reinforcing user trust.

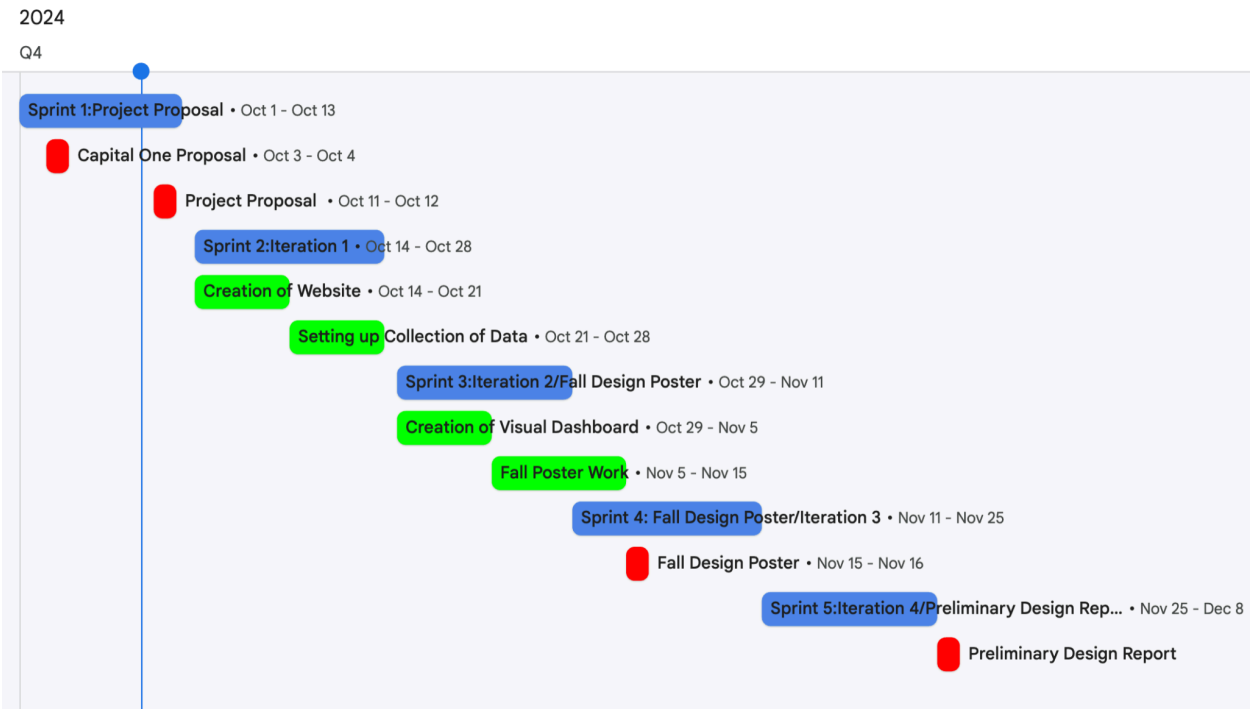
This project provided valuable lessons that shaped the final design. Regular feedback from stakeholders proved essential for improving visualization clarity and usability. Planning for scalability early in the design process was critical to avoiding performance bottlenecks as the system matured. Balancing performance with cost was another significant learning experience, requiring careful analysis of cloud services like Snowflake and Tableau to align with project constraints.

Although the final design meets its objectives, several recommendations can guide future development. Incorporating advanced analytics such as predictive modeling or AI-driven insights could enhance the depth of the dashboard's outputs. Expanding the dashboard to include live session tracking and alert systems for unusual activity would increase its functionality. Generalizing the system to support diverse platforms beyond the mock Agent Servicing Website could unlock broader adoption, while exploring alternative data warehousing and visualization tools may reduce long-term operational costs.

Despite its success, the project raises some unanswered questions. The scalability of real-time analytics in environments with exponentially increasing user interactions is an area for further exploration. Alternatives to Google Analytics and Tag Manager, could significantly reduce costs by providing a simpler pipeline. Coupling this change with a shift from Google Cloud would further optimize the project's budget without sacrificing system performance.

If this project continues as a future senior design effort, future teams can build on the established foundations of data collection, processing, and visualization. Key focus areas should include enhancing real-time capabilities, expanding dashboard functionality, and conducting further load testing for peak user traffic. Supporting documentation, including detailed system architecture diagrams, ER schemas, and annotated code, is available in the appendices to aid in the transition. All relevant digital resources, such as source code and dashboard templates, have been stored in Github for seamless access.

Appendix 1: Project Timeline



Key:

- Red: Deadlines
- Blue: Major Sprints
- Green: Tasks to accomplish during sprints

Appendix 2: Team Contract (i.e. Team Organization)

Team Member Name	*Strengths each member bring to the group*	*Other Info*	*Contact Info*
Hoang Le	Communicates well, can be flexible with workload responsibilities	Experience with Python, Java, Javascript, and C.	leht8@vcu.edu
Ben Aber	Creative thinker, adaptable and reliable in a team.	Proficient in Java, Python and C. Always open to new ideas and constructive criticism.	aberb@vcu.edu
Maxwell Goehle	Problem solver, works well in a group environment by being patient and flexible.	Experience with Java, C, and Python, and always looking to learn.	goehlemc@vcu.edu
Mohammed S. Ahmed	Time-management, Attention detail, Collaboration	AWS Certified, Experienced in java, python, and C. Always looking for ways to improve.	ahmedm12@vcu.edu

Other Stakeholders	*Notes*	*Contact Info*
Mahesh Nair (Capital One)	*Capital One Team Lead*	mahesh.bahulleyannair@capitalone.com
Rachita Sowle(Mentor/ Advisor)	*VCU Faculty Advisor*	sowler@vcu.edu
Tyler Jordan (Capital One)	*Capital One Advisor*	tyler.jordan@capitalone.com
Emily Croxall (Capital One	*Capital One Advisor*	emily.croxall@capitalone.com

Culture Goals	***Actions***	***Warning Signs***
Punctual for both meetings and delivering tasks	- Arriving on time for meeting - Submitting work and completing work by the agreed date - Set reasonable timeline and schedules for project tasks	- Normally arriving late for team meetings - Constantly missing deadlines or needing more time - Lack of urgency or respect for other people time in scheduling or communication
Clear and concise communication at each step of the process	- Text/email each other when issues or delays occur - Hold each other accountable regarding their respective work - Individual tasks are specified so everyone understands their specific role	- Teammate is ghosting the groupchat for an extended period of time - Not asking questions when confused about task - Teammate takes feedback but then accidentally overcomplicates the assignment
Put teammates in positions to work well together	- Utilize each other's strengths to avoid relying on others weaknesses. - Be understanding of issues that may come up over the course of the project, be accommodating and understanding of teammates.	- Teammate(s) are struggling to meet deadlines or has sloppy work. - Individuals struggle to complete tasks that another individual or as a group could be done more efficiently.

Meeting Participants	***Frequency*** ***Dates and Times / Locations***	***Meeting Goals*** ***Responsible Party***
Students Only	Meet Wednesday 10 am to 11 am, discord call	-Update group on week-to-week challenges and accomplishments -Communicate what needs to be done and plan ahead -Update Github
Students Only	Available for discord calls/chats in discord server as necessary.	-Solve problems that arise during the project -Understand others' statuses on tasks and help each other along the way.
Students + Faculty advisor	*Weekly 9am Fridays*	-Provide faculty advisor with weekly status reports -Check in weekly and report project progress
Project Sponsor(s)	*Weekly 9am Fridays*	-Update stakeholders on progress and current roadblocks -Work together to find ways to improve project and or work flow

Team Member	*Role(s)*	*Responsibilities*
Maxwell Goehle	Project Manager	• Be a contact person for Capstone staff if a problem arises • Leads and records important information discussed in meetings throughout the duration of the project • Mediate any potential problems that occur outside or within the group pertaining to the completion of the project
Hoang Le	Logistics Manager	• Coordinates group communication and meetings throughout the year • Obtaining necessary information/specifications to help complete tasks at each step • Keeps track of what is being done by each group member • Sets up communication between sponsor advisor and VCU advisor
Ben Aber	Systems Engineer	• Understand client's initial design requirements and specifications to ensure the product matches the expectations • Clarify any ambiguities that may arise in the design requirements • Coordinate, recommend and manage any development and system architecture design improvements • Manage product interfaces
Mohammed Ahmed	Manufacturing Engineer	• Identify and integrate relevant data sources, and ensure accuracy and clean data collection • Design and develop data process for meaningful insights • Continuously optimize and maintain the dashboard for performance

Step 5: Agree to the above team contract

Team Member: Maxwell Goehle *Signature: Maxwell Goehle*

Team Member: Hoang Le *Signature: Hoang Le*

Team Member: Ben Aber *Signature: Ben Aber*

Team Member: Mohammed Ahmed *Signature: Mohammed Ahmed*

References

Citations:

[1] Madison Zoey Vettorino, “What Is Clickstream Data? Everything You Need to Know,” *HubSpot Blog*, September 21, 2022, accessed October 10, 2024, https://blog.hubspot.com/website/clickstream-data

[2] Capital One. “About Us.” Capital One, 2024. Accessed October 10, 2024. https://www.capitalone.com/about/

[3] Google Analytics Team. “A Brief History of Google Analytics.” Google, 2023. Accessed October 10, 2024. https://blog.google/products/analytics/

[4] Gillis, Alexander S. “Clickstream Analysis (Clickstream Analytics).” *SearchCustomerExperience*, TechTarget, May 2022. Accessed October 10, 2024. https://www.techtarget.com/searchcustomerexperience/definition/clickstream-analysis-clickstream-analytics

[5] Lifesight. “How Clickstream Data Can Benefit Your Business.” *Lifesight Blog*, September 6, 2024. Accessed October 10, 2024. https://www.lifesight.io/blog/clickstream-data-for-business.