Merquery

Requirements

Team Green

Amy Nguyen
Jackeline Maldonado
Jacob Darensbourg
Rachita Gehi
Tina Bui

March 2016

Table of Contents

1. Introduction

This document specifies the requirements for the data exploration tool.

1.1. Background

The Viant Advertising Cloud is a people-based platform that infuses accuracy, transparency, and accountability into advertising executions. The company specializes in housing billions of individual data and selling it to key advertisers. If advertisers want to know how many individuals purchase a certain type of cereal or mouthwash, Viant helps answer these questions with their data. Based on the data, advertisers know how to target individuals and sell their products.

1.2 System Overview

The data exploration tool will allow employees at Viant Inc. to explore and understand their consumer data. Using the tool, the user can search through the data without any prior knowledge of programming languages (primarily SQL). It will then display the information in a format that is easy to comprehend. The tool will allow consumer data to be displayed to potential customers (while maintaining privacy of the data) and will allow employees to utilize the data without relying upon developers. The tool will be an advanced version of a previously developed system that only filtered through Viant's consumer data. The data exploration tool will connect the different data points together to show the relations between individual consumers and the broad spectrum of consumers.

1.3 Assumptions

1.3.1 Need for the System

It is assumed there is a need for this product at Viant to allow greater access to the large amounts of data available to the company and its employees. Currently, accessing the data is inefficient as only a handful of employees can access the data. Those with access are commonly software engineers who have experience with SQL, querying data, and handling databases. Viant has requested a system that would present the data in a relevant, understandable way. This system will increase efficiency and allow select employees access to relevant data.

1.3.2 Software Assumptions

The team is assumed to have access to necessary programming environments, to collaborate through GitHub, and to have sufficient knowledge of SQL as well as any other languages necessary.

1.3.3 Hardware Assumptions

The team will have access to computers with the necessary memory and efficiency through which they can make progress.

1.3.4 Assumptions about Environments/Data

Viant will provide a subset of anonymized data for prototyping. The subset is assumed to be similarly structured to allow easy integration of the system to the actual data. Viant will be able to answer/guide about the meaning of different areas of the data provided (for example: different labels on every section) through meetings/interviews.

1.4 Scope

There is currently a data-exploration tool that was developed by Viant over one year ago. This tool only works on a small-subset of data and does not provide a wide range of possible queries. We have been shown this tool as an example of something similar to what we are expected to produce. Instead of building on top of this tool, we will start from scratch. We have agreed with our sponsor, the SVP of data products, to build a tool that covers a wider range of data. The tool will allow users to traverse through the data in a more efficient manner. Our aim is to have this tool encompass a larger amount of data with more useful filters and presentation than the previous tool.

1.5 Statement of Work

Although Viant Inc. is an internet marketing service with massive amounts of data, they do not have cohesive way to organize or access it. Our task is to create a data exploration tool that makes it easier for a select number of employees to access the data. Currently, data is accessible only by developers who know SQL. Our goal is to organize the data and present a user-friendly interface that allows personnel to access and sort the data as they wish. This tool will increase the productivity of every team member and therefore minimize time spent on this task. This tool will also provide a more appealing and straight-forward way to find information needed to create advertisements. To accomplish these goals, we will create an easy-to-use interface that allows users the freedom to search for any information they may need.

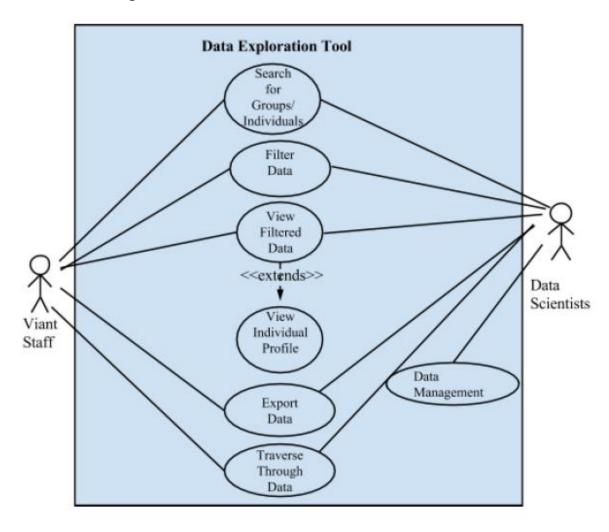### 1.5.1 Implementation Process

The tool will be developed by understanding the needs of different members of the Viant team. The tool will be initially developed and prototyped using a small, encrypted subset of the consumer data that will later be scaled to incorporate all of the data that Viant possesses.

2. System Requirements

2.1 Functional Requirements

    2.1.1 Data Exploration Requirements

        2.1.1.1 Display all of the consumer information

        2.1.1.2 Should be extensible so that later changes will dynamically alter search and table fields without having to alter code.

        2.1.1.3 Should track user searches and queries through the use of breadcrumbs for filtering purposes.

        2.1.1.4 Users will be able to traverse through data points by clicking on the relevant field value.

        2.1.1.5 Should display profile, device, and household cards when icons are clicked

        2.1.1.6 Should not display encrypted e-mail addresses (MD5)

2.2 Non-functional Requirements

    2.2.1 Hardware Requirements

        2.2.1.1 Tool shall be intended to be used as a web interface.

    2.2.2 Performance

        2.2.2.1 Tool should respond to queries within a reasonable time frame.

    2.2.3 Accessibility

        2.2.3.1 Tool shall be accessible via any web browser.

        2.2.3.2 Tool will run on Windows, Mac, and Linux.

    2.2.4 Licensing

        2.2.4.1 Tool should first be an internally used product.

    2.2.5 Training

        2.2.5.1 Tool should be intuitive to use.

        2.2.5.2 Tool should require little to no actual training to use.

        2.2.5.3 Tool should not require any inherent programming knowledge to use.

    2.2.6 Usability

        2.2.6.1 Should have consistent interface elements.

        2.2.6.2 Tool shall not display any code to the user.

        2.2.6.4 Tool should be capable of showing data in an easy to read format.

    2.2.7 Reliability and Robustness

        2.2.7.1 Users should be notified when there is an invalid query.

        2.2.7.2 Tool should link together all the data without crashing.

3. Appendix

3.1 Use Case Diagram

3.2 Individual Use Cases
3.2.1 Use Case 1

Section	Content/Explanation
Use Case Name	1.1 Viant Staff an Individual's profile
Priority	High
Criticality	This use case is high priority level feature because it enables staff to view an individual's profile and study how typical users behave.
Source	Client Meetings
Short Description	Staff accesses the tool to view an individual.
Goal(s)	- Visualize Data - Allow select staff to view user profiles and traverse through the data
Primary Actor	Viant Staff
Secondary Actors	N/A
Preconditions	- Staff must be authorized to access the tool
Success End Condition	- Staff have successfully viewed an individual's profile
Failed End Condition	- Staff cannot view an individual's profile
Trigger	- Viant staff submits a query to view an individual's profile
Results	- The system will a profile page in a easily readable and understandable format based on: age, ethnicity, interests, gender, education, income and location, devices.
Main Success Scenario	Description of the main scenario of the use case, successfully completing

	Step	Action
	1	Staff accesses the tool on the Viant network
	2	Front page is displayed with filters
	3	Staff selects appropriate filters and views results
	4	Staff selects an individual and views their profile
	5	Staff closes the tool

Alternative Scenarios	Description of alternative scenarios – e.g., extension scenarios specifying the action above at which each extension applies	
	1.a - Staff is not validated	1.b Tool displays error message
Exception Scenarios	- The system is down due to management or other technical factors	
Relationship to other use cases	- This use case can be directly linked to Use Case 2 since after the data has been filtered, the user can view an individual's profile.	
Supplementary Information	- Sales engineers will be the only ones who will have access to update, load and delete the anonymized data.	
Open Issues	N/A	

3.2.2 Use Case 2

Section	Content/Explanation	
Use Case Name	1.2 Data exploration tool filters through data	
Priority	High	
Criticality	This is highly critical because it allows the staff to filter through the data for a more specific statistics.	
Source	User interviews and client meetings	
Short Description	Staff uses the tool to select the parameters (e.g.gender, age, interests, etc.) desired to refine their search through the data.	
Goal(s)	Filter through the data	
Primary Actor	Viant staff	
Secondary Actors	N/A	
Preconditions	There must be data in the database to filter through	
Success End Condition	System filters through the data	
Failed End Condition	System is unable to bring up results based on parameters given	
Trigger	Staff selects different parameters to filter through data with	
Results	The system filters through the data based on the specified parameters	
Main Success Scenario	Description of the main scenario of the use case, successfully completing	
	Step	Action
	1	Staff selects parameters to filter through data (e.g. interests)
	2	Tool filters through data based on the selected parameters (i.e. interests)
	3	Staff traverses through data based on user profiles
Alternative Scenarios	Description of alternative scenarios – e.g., extension scenarios specifying the action above at which each extension applies	
	3a	There is no data in the system that matches up with the parameters	Tool displays an error message
Exception	There are technical issues or there is no data in the database.	

Scenarios	
Relationship to other use cases	Filtering through the data is associated with viewing and exporting the filtered data
Supplementary Information	Sales engineers have the access to update and delete the anonymized data
Open Issues	N/A

3.2.2 Use Case 3

Section	Content/Explanation		
Use Case Name	Traverse through data		
Priority	High		
Criticality	This is highly critical because this allows Viant employees to traverse through the filtered data		
Source	Client meeting		
Short Description	Viant employees are able to select a piece of data and be shown other items which are connected to that piece of data.		
Goal(s)	Connecting appropriate nodes based on the filtered data		
Primary Actor	Viant staff		
Secondary Actors	N/A		
Preconditions	There should be data which are connected/similar to each other.		
Success End Condition	A list/chart of people, who are connected to a particular data, are shown.		
Failed End Condition	There are technical issues or system is down for maintenance.		
Trigger	Staff selects a field description from a profile		
Results	The system generates data, which matches the field selected, and displays the information in an orderly format.		
Main Success Scenario	Description of the main scenario of the use case, successfully completing		
	Step	Action	
	1	Staff selects a field from the profile	
	2	Data which are connected to that data point are generated and displayed	
Alternative Scenarios	Description of alternative scenarios – e.g., extension scenarios specifying the action above at which each extension applies		
	2a	There is no data which are connected to that data point	Tool displays error message
Exception	There is no data in the database		

Scenarios	
Relationship to other use cases	Traversing through the data is associated with filtering and viewing the data
Supplementary Information	Anonymized data is updated and added by sales engineers.
Open Issues	N/A

3.3 Personas and Scenarios

3.3.1 Personas

3.3.1.1 Persona 1

Tim graduated a few years ago and decided to start a business along with his brother. He put much work into it and now is the CEO of the company Viant Inc. He is a very busy man that occasionally holds meetings with the most important clients to acquire their business. Tim didn't graduate with a computer science degree and therefore doesn't have a wide range of ability when it comes to programming. He has very limited knowledge of SQL and often times have to ask developers to assist him in writing SQL scripts. Tim is a businessman and therefore seeks to always have efficient meetings displaying the best aspects and data of Viant. He prides himself in the ability to acquire the best and biggest clients and appreciates the assistance of his team in doing so.

3.3.1.2 Persona 2

Jonathan is a team leader for Viant. He has been working for the company for seven years now. He understands the importance of data when trying to analyze sales patterns and convincing clients to use their services. Although he knows enough SQL to look through the data and search for simple things, he normally does not have access to the data and must go find the sales engineers who do have access. He also realizes that there is a lot of information available for him. However, his limited knowledge of SQL and difficulty in organizing the data, he cannot utilize the data to its fullest potential.

3.3.1.3 Persona 3

James is a sales engineer at Viant. He is very passionate about his job and has been with the company for four years. He has access to all the data and is very proficient at SQL. He is aware that Viant has a lot of data which can be organized in a way to further to benefit the company. While he is able to view different metrics for the data, he does not have a way to view the connections the data has with each other. Though he has always kept up with the information Viant currently provides, he hopes there will be a good solution to allow him to visually traverse through the data.

3.3.2 Scenarios

3.3.2.1 Scenario A (Persona 1)

Tim is preparing for a presentation with a big client this afternoon. Tim usually presents to the potential client and assures them of Viant's ability and the accuracy of the data it holds. Lately he has found it very difficult to have a persuasive and efficient presentation because of the way he has to access data. Tim does not know much SQL and therefore must always bring a developer with him for presentations. During presentations, when the potential client has a special request for a demo they would like to see, Tim must pass this information to the developer. The developer must then write the scripts to show the desired information. Tim is tired of this system as he feels it is unprofessional to have the potential client waiting during this process and feels the whole process isn't efficient. In preparation for his afternoon meeting he asks his team at Viant if there are any better ways in which the company as a whole could improve the way it explores its data.

3.3.2.2 Scenario B (Persona 2)

Jonathan is currently working on recruiting a large business to use Viant's services. Their potential client is skeptical of whether or not they should work with Viant to expand their marketing as they are already successful. To convince them that Viant has very valuable data for their company, Jonathan wants to display data and show how much information can be gathered for individuals. He also does some research and identifies some areas where the company is not targeting. The data exploration tool allows him to quickly filter through the information and display key individuals. With this information, Jonathan is confident they will gain a new client.

3.3.2.3 Scenario C (Persona 3)

Currently, James is looking at the different metrics on the dashboard. He wishes to view the data connections on different fields. Using the tool, James pulls up a data profile and views the different fields on there. There are lots of data on this person, but is currently interested in the different campaigns the individual is associated with. He finds this way of traversing through information useful because he can report the information to the other teams such as marketing.

3.4 Requirements Priority Level (MoSCoW)

The following is a list of requirements, listed by their level of priority using MoSCoW:

1. Must-Have
    a. Pull consumer data from GBQ
    b. Display consumer data in a table
    c. Left navigation search fields that change according to the information contained within GBQ
    d. Traverse through data table via clicking
    e. Display user profile, device, and household cards
2. Should-Have
    a. Show the table across multiple pages
    b. Run all within the same web page without opening new tabs
    c. A simple help popup to assist first-time users understand the tool
    d. Visual indicators to demonstrate that system is working, not broken
3. Could-Have
    a. Clickable fields in the user profile windows
    b. Images of user in the user profile windows
    c. Dropdown menus and radio buttons for left navigation search fields
    d. Visually-appealing search fields, tables, and user profiles
4. Won't-Have
    a. Visualizations of consumer data information
    b. Metrics for consumer data information

3.5 Definitions, Acronyms, and Abbreviations

- SQL - Structured Query Language
- IP Address - Internet Provider Address
- GBQ - Google BigQuery, database where Viant consumer data is stored
- Data Dictionary - listing of all data points in schema, what their values are, and an example of their values
- Data Schema - formal organization schema for data points
- GitHub - online code repository used for system version control

3.6 References

- IEEE Recommended Practice for Software Requirements Specifications
  - https://eee.uci.edu/15w/37010/assignments/IEEEStd830-1998.pdf
- Use Case Diagrams
  - http://www.uml-diagrams.org/use-case-diagrams.html