A picture containing text

Description automatically generated

Capstone Project

In this document, all rules and requirements on how to proceed with defining and implementing the Revature capstone project are defined.

**Fields (Schema)**

**Screeners**:

one who evaluates a Qualified Lead during a Technical Screening

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| id | Screener Id |
| first\_name | Screener First Name |
| last\_name | Screener Last Name |

**Recruiters**:

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| id | Recruiter Id |
| first\_name | Recruiter First Name |
| last\_name | Recruiter Last Name |

**Qualified Lead**:

individual who has registered through Revature and expressed interest in joining

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| id | Qualified Lead Id |
| first\_name | Qualified Lead First Name |
| last\_name | Qualified Lead Last Name |
| university | University attended by Qualified Lead |
| major | Major of Degree of Qualified Lead |
| email | Qualified Lead Email |
| home\_state | Qualified Lead State |

**Contact Attempts**:

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| recruiter\_id | Recruiter Id |
| ql\_id | Qualified Lead Id |
| contact\_date | Date of Contact |
| start\_time | Start Time |
| end\_time | End Time |
| contact\_method | Method of Communication (Phone, Email, SMS) |

**Screening**:

The Technical Screening you took before entry into Revature

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| screener\_id | Screener Id |
| ql\_id | Qualified Lead Id |
| screening\_date | Date of the Screening |
| start\_time | Start Time |
| end\_time | End Time |
| screening\_type | Type of Screening (Spark, Standard, Business Analyst) |
| question\_number | Number of questions asked durring the Screening |
| question\_accepted | Number of accepted answers from the Screening |

**Offers**:

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| screener\_id | Screener Id |
| recruiter\_id | Recruiter Id |
| ql\_id | Qualified Lead Id |
| offer\_extended\_date | Date the offer was extended |
| offer\_action\_time | Date the offer was addressed by the QL |
| contact\_method | Method of Communication (Phone, Email, SMS) |
| offer\_action | Response to the offer (Accept, Reject, Delay) |

**Sample Data (CSV)**

**Screeners**:

id, first\_name, last\_name

1, Alisa, Figgures

2, Oralle, Druhan

3, Onida, Radin

**Recruiters**:

id, first\_name, last\_name

1, Brantley, Pickance

2, Joyous, Macconachy

3, Robbin, Ferdinand

**Qualified Lead**:

id,first\_name, last\_name, university, major, email, home\_state

1, Maddy, Blees, Fayetteville State University, Business, mblees0@wiley.com, NM

2, Valencia, Guillet, Al Turath University College, Hospitality, vguillet1@dedecms.com, LA

3, Milty, Tracey, Chatham College, Business, mtracey2@sbwire.com, NY

**Contact Attempts**:

recruiter\_id, ql\_id, contact\_date, start\_time, end\_time, contact\_method

3, 1, 5/6/2018, 9:00:00 AM, 10:25:00 AM, Phone

1, 2, 2/17/2019, 10:00:00 AM, 11:25:00 AM, Email

2, 2, 11/24/2019, 12:00:00 PM, 1:09:00 PM, Phone

2, 3, 6/2/2019, 12:00:00 PM, 1:00:00 PM, Email

**Screening**:

screener\_id, ql\_id, screening\_date, start\_time, end\_time, screening\_type, question\_number, question\_accepted

3, 1, 6/17/2018, 10:00:00 AM, 11:26:00 AM, Spark, 34, 14

1, 2, 3/24/2019, 9:00:00 AM, 10:04:00 AM, Spark, 32, 8

2, 2, 12/29/2019, 10:00:00 AM, 11:06:00 AM, Standard, 31, 9

3, 3, 6/30/2019, 2:00:00 PM, 3:04:00 PM, Business Analyst, 33, 20

**Offers**:

screener\_id, recruiter\_id, ql\_id, offer\_extended\_date, offer\_action\_date, contact\_method, offer\_action

3, 1, 1, 7/8/2018, 7/15/2018, Email, Accept

3, 2, 3, 7/14/2019, 7/21/2019, SMS, Delay

**Data Flow for Mocking:**

Screeners, Recruiters, and Qualified leads are all individuals and can be generated independently.

Contact Attempts will be made first, where qualified leads can have multiple Contact Attempts.

Some Contact Attempts, but not all, will lead to Screenings where qualified leads can have multiple Screenings.

Finally, some Screenings will lead to Offers, where qualified leads can have multiple Offers.

**Tasks:**

1. Create a producer program that will ingest data to a Kafka Topic.
   1. Data will have to be generated in the program.
   2. Ingest the data every 2 seconds into the Kafka Topic.
2. Display the data from the input Kafka Topic in a console consumer (CLI).
3. Create a consumer program in Spark that will read the data stream from the input Kafka Topic and will process the data further.
   1. Read the data into DataFrame objects.
   2. Print the schema of the input data stream
   3. Apply the [above-mentioned schema](https://usc-word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en%2DUS&rs=en%2DUS&wopisrc=https%3A%2F%2Frevature0-my.sharepoint.com%2Fpersonal%2Fkenneth_davis_revature_com%2F_vti_bin%2Fwopi.ashx%2Ffiles%2Fe01fd0c103d44eb9aab888967d941d53&wdenableroaming=1&mscc=1&wdodb=1&hid=2D1104A0-5020-C000-9604-AA267C2E75BC&wdorigin=ItemsView&wdhostclicktime=1637169363407&jsapi=1&jsapiver=v1&newsession=1&corrid=9876280e-7c32-66ef-03ee-a964f097f973&usid=9876280e-7c32-66ef-03ee-a964f097f973&sftc=1&mtf=1&sfp=1&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush&preseededsessionkey=b1a4c05f-eb23-0772-1dd5-67395f8b035c&preseededwacsessionid=9876280e-7c32-66ef-03ee-a964f097f973&rct=Medium&ctp=LeastProtected#_Fields_(Schema)) to the dataframes and print the schema.
   4. From the consumer program:
      1. Determine and display on the console the total number of Qualified Leads
      2. Determine and display on the console the number of contact attempts and total number per recruiter
      3. Determine and display on the console the number of screenings and total number per screener
      4. Determine and display on the console the number of offers and totals by offer action
   5. (Optional) Display all above outputs in Tableau or some other data visualization tool