**DATASET EXPLANATION**

A **centralised RDS** has been set up where the data will be hosted. You will be required to import the data from this RDS and then use this data to create a Kafka topic for the Patient's vital information. We will also provide the dataset for Patient’s Contact information and for the Reference Threshold values. You will have to use Sqoop to consume the data from the contact information table whereas, for the latter, you will have to create the table yourself.

**Hostname**: [upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com](http://upgradtest1.cyaielc9bmnf.us-east-1.rds.amazonaws.com/)  
**username**: student  
**password**: STUDENT123  
**dbname**: testdatabase  
**table-name**: patients\_vital\_info

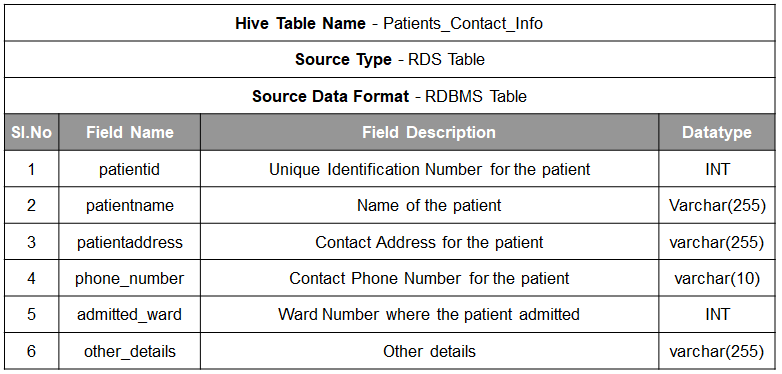
The data is based on daily health monitoring mechanisms which take the heart rate and blood pressure of the patient into consideration to check for any abnormalities.

The Datasets that you will be working with, in this Capstone Project are as follows -

Patient’s Contact Info Dataset

This data set will be used to access the contact information of the patient so that it can be used in case of any emergencies.

The schema for this dataset is given below



Patient's Contact Info Dataset Schema

Here is a snippet of how the Patient's Contact Info dataset looks like:

| **patientid** | **patientname** | **patientaddress** | **phone\_number** | **admitted\_ward** | **age** | **other\_details** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Alex S | XDC test Address | 8982739282 | 1 | 23 | NULL |
| 2 | Sammy A | New Building Address | 2382739282 | 2 | 45 | NULL |
| 3 | Karan C | Aws Address | 8923739282 | 3 | 56 | NULL |
| 4 | Dara M | India Address | 2182739282 | 4 | 67 | NULL |
| 5 | Pam | ABC test Address | 4982739282 | 5 | 72 | NULL |

The details for this table in RDS are as follows:

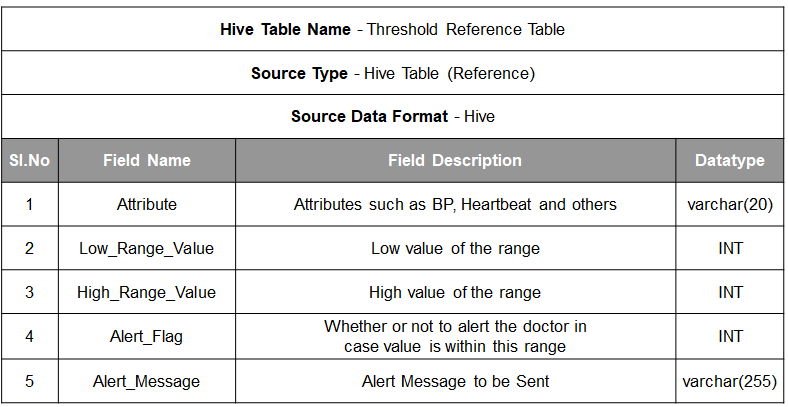
**Hostname**: [upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com](http://upgradtest1.cyaielc9bmnf.us-east-1.rds.amazonaws.com/)  
**username**: student  
**password**: STUDENT123  
**dbname**: testdatabase  
**table-name**: patients\_information

Threshold Reference Table

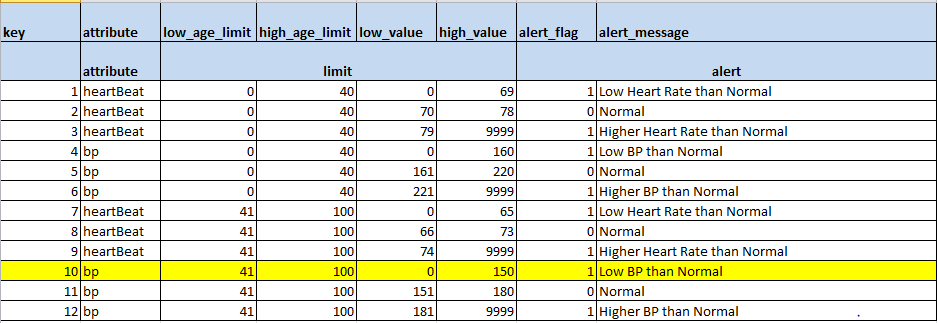
This data set will be used for comparing the vital incoming data from the patient with standard threshold information to check for anomalies.

The schema for this dataset is given below

**Note**: There are 2 columns which are not mentioned in the schema - **low\_age\_limit and high\_age\_limit**, both with the data type INT which are also supposed to be there in the table as mentioned in the Threshold Reference Table Image.



Threshold Reference Table Schema

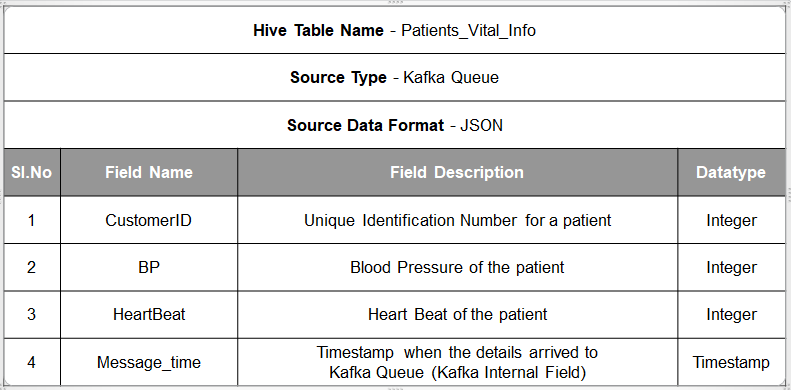


Threshold Reference Table

Patient’s Vital Information Dataset

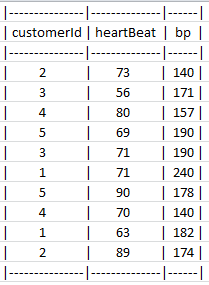
This data set will be used to monitor the patient on a real-time basis and check their overall well-being.

The schema for this dataset is given below:



Patient's Vital Info Dataset Schema

The vital information snapshot for 10 seconds would like:



10 sec Vitals Snapshot

**Output:**

Once it is decided that the alert has to be sent, the notification message will have both, the vital information and the contact information of the concerned patient which will be sent in JSON format. Here is a sample output format of how the alert message will be sent to the registered email address:

{"patientname":"Sammy A",

"age":**45**,

"patientaddress":"New Building Address",

"phone\_number":"2382739282",

"admitted\_ward":**2**,

"bp":**140**,

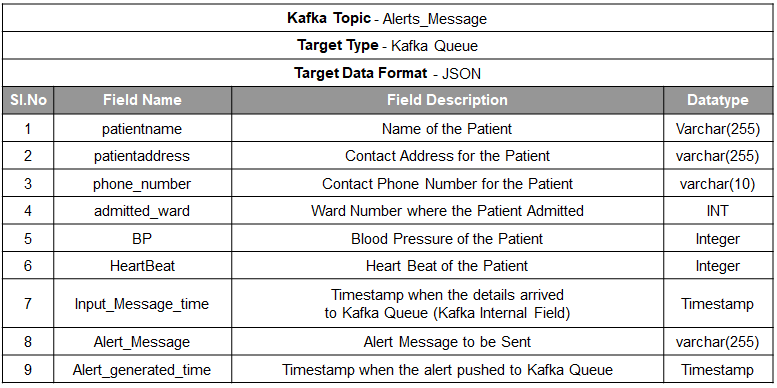
"heartBeat":**73**,

"input\_message\_time":"2020-11-02T05:52:01.885Z",

"alert\_generated\_time":"2020-11-02T07:00:20.942Z",

"alert\_message":"Low BP than Normal"}

The schema for the output will look like this:



Output Schema