**NTSBE-31**

**RTT Transition – Create Druid Loader for Hourly Score Data**

**Task**: https://onejira.verizon.com/browse/VZWHD-1379

**Analyzed By**: Akhilesh Varma Guduri

**Designed By**: Akhilesh Varma Guduri

**1. Overview:**

The Hourly loader table is getting ingested into the druid on daily basis with the records taken after correlating both the same area device score hourly and hourly Vmas score data

**2. Internal/External Impacts:**

Input tables to be considered “**stp\_rtt\_subscriber\_area\_device\_scores\_hourly**” and “**stp\_rtt\_vmas\_kpi\_scores\_subscriber\_agg\_hourly**”

**3. Assumptions:**

Assuming data is available in the “**stp\_rtt\_subscriber\_area\_device\_scores\_hourly**” and “**stp\_rtt\_vmas\_kpi\_scores\_subscriber\_agg\_hourly**” hourly basis.

**4. Detailed Design:**

**4.1 Input changes:**

Below is the input table used in this process for loading into druid and the structure details as below

* **stp\_rtt\_vmas\_kpi\_scores\_subscriber\_agg\_hourly**

|  |  |
| --- | --- |
| Column Name | Data Type |
| score\_date\_hr | STRING |
| subscriber\_id | BIGINT |
| mdn | STRING |
| imsi | STRING |
| imei | STRING |
| make | STRING |
| model | STRING |
| attach\_failure\_pct | DOUBLE |
| attach\_failure\_cnt | BIGINT |
| attach\_attempts\_cnt | BIGINT |
| rrc\_setup\_failure\_pct | DOUBLE |
| rrc\_setup\_failure\_cnt | BIGINT |
| rrc\_setup\_attempts\_cnt | BIGINT |
| srf\_pct | DOUBLE |
| service\_request\_failures\_cnt | BIGINT |
| service\_request\_attempts\_cnt | BIGINT |
| pcf\_pct | DOUBLE |
| session\_setup\_failures\_cnt | BIGINT |
| session\_setup\_attempts\_cnt | BIGINT |
| cd\_pct | DOUBLE |
| context\_drops\_cnt | BIGINT |
| context\_events\_cnt | BIGINT |
| sip\_dropped\_calls\_pct | DOUBLE |
| volte\_voice\_calls\_dropped\_cnt | BIGINT |
| volte\_voice\_setup\_incomplete\_calls\_cnt | BIGINT |
| rrc\_radio\_drop\_pct | DOUBLE |
| radio\_bearer\_drops\_cnt | BIGINT |
| radio\_bearer\_setup\_attempts\_cnt | BIGINT |
| downlink\_throughput\_kbps | DOUBLE |
| uplink\_thoughput\_kbps | DOUBLE |
| downlink\_data\_volume | BIGINT |
| downlink\_active\_time\_ms | BIGINT |
| uplink\_data\_volume | BIGINT |
| uplink\_active\_time\_ms | BIGINT |
| travelling\_indicator | BIGINT |
| volume\_weighted\_uplink\_thpt\_kbps | DOUBLE |
| volume\_weighted\_downlink\_thpt\_kbps | DOUBLE |
| attach\_failure\_pct\_weight | DOUBLE |
| rrc\_setup\_failure\_pct\_weight | DOUBLE |
| srf\_pct\_weight | DOUBLE |
| pcf\_pct\_weight | DOUBLE |
| cd\_pct\_weight | DOUBLE |
| sip\_dropped\_calls\_pct\_weight | DOUBLE |
| rrc\_radio\_drop\_pct\_weight | DOUBLE |
| downlink\_throughput\_kbps\_weight | DOUBLE |
| uplink\_thoughput\_kbps\_weight | DOUBLE |
| UES\_Accessibility | DOUBLE |
| UES\_Availability | DOUBLE |
| UES\_Retainability | DOUBLE |
| UES\_ServiceIntegrity | DOUBLE |
| UES\_Mobility | DOUBLE |
| USE\_Voice\_Retainability | DOUBLE |
| UES\_Voice\_Reliability | DOUBLE |
| UES\_Data | DOUBLE |
| UES\_SMS | DOUBLE |
| UES\_3GPP | DOUBLE |
| UES\_AllSERV | DOUBLE |
| preference\_ratio | DOUBLE |
| load\_time | STRING |
| trans\_dt | STRING |
| trans\_hr | STRING |
|  |  |

* **stp\_rtt\_subscriber\_area\_device\_scores\_hourly**

|  |
| --- |
| **stp\_rtt\_subscriber\_area\_device\_scores\_hourly** |
| subscriber\_id |
| enb\_id\_list |
| enb\_weighted\_area\_agg\_score |
| enb\_simple\_area\_agg\_score |
| enb\_area\_subscriber\_count |
| manufacturer |
| model |
| enb\_weighted\_device\_agg\_score |
| enb\_simple\_device\_agg\_score |

**4.2 Logic**

1. Script has to verify whether both the tables are loaded with hourly records while running for that hour.
2. If any or both the tables are not loaded, then exit the script and process that hour in the next hour.

Ex: while looking for 4th hour if aggregation table is not available with data, then the script exits for 4th hour, then on 5th hour it processes both 4th and 5th hours once the data is loaded

1. Load the tables **stp\_rtt\_subscriber\_area\_device\_scores\_hourly** and **stp\_rtt\_vmas\_kpi\_scores\_subscriber\_agg\_hourly** into variables
2. Correlate the tables **stp\_rtt\_vmas\_kpi\_scores\_subscriber\_agg\_hourly** and **stp\_rtt\_subscriber\_area\_device\_scores\_hourly** based on subscriber\_id
3. Add the partition into the hive table using msck repair command
4. Configure the Json file with all the specifications required for druid ingestion (Ex: column details, input path etc.)
5. Call the Pig script to store the records into hdfs with partitions score date.
6. Call the druid ingestion loader through the curl command by passing json which is configured in the shell script.
7. Exit the script according to the status of the druid ingestion.

**Output Path:**

**Output Table:** **stp\_rtt\_vmas\_scores\_druid\_summary\_hourly**

|  |  |
| --- | --- |
| Column Name | Data Type |
| score\_date\_hr | STRING |
| subscriber\_id | BIGINT |
| mdn | STRING |
| imsi | STRING |
| imei | STRING |
| manufacturer | STRING |
| model | STRING |
| attach\_failure\_pct | DOUBLE |
| attach\_failure\_cnt | BIGINT |
| attach\_attempts\_cnt | BIGINT |
| rrc\_setup\_failure\_pct | DOUBLE |
| rrc\_setup\_failure\_cnt | BIGINT |
| rrc\_setup\_attempts\_cnt | BIGINT |
| srf\_pct | DOUBLE |
| service\_request\_failures\_cnt | BIGINT |
| service\_request\_attempts\_cnt | BIGINT |
| pcf\_pct | DOUBLE |
| session\_setup\_failures\_cnt | BIGINT |
| session\_setup\_attempts\_cnt | BIGINT |
| cd\_pct | DOUBLE |
| context\_drops\_cnt | BIGINT |
| context\_events\_cnt | BIGINT |
| sip\_dropped\_calls\_pct | DOUBLE |
| volte\_voice\_calls\_dropped\_cnt | BIGINT |
| volte\_voice\_setup\_incomplete\_calls\_cnt | BIGINT |
| rrc\_radio\_drop\_pct | DOUBLE |
| radio\_bearer\_drops\_cnt | BIGINT |
| radio\_bearer\_setup\_attempts\_cnt | BIGINT |
| downlink\_throughput\_kbps | DOUBLE |
| uplink\_thoughput\_kbps | DOUBLE |
| downlink\_data\_volume | BIGINT |
| downlink\_active\_time\_ms | BIGINT |
| uplink\_data\_volume | BIGINT |
| uplink\_active\_time\_ms | BIGINT |
| travelling\_indicator | BIGINT |
| volume\_weighted\_uplink\_thpt\_kbps | DOUBLE |
| volume\_weighted\_downlink\_thpt\_kbps | DOUBLE |
| attach\_failure\_pct\_weight | DOUBLE |
| rrc\_setup\_failure\_pct\_weight | DOUBLE |
| srf\_pct\_weight | DOUBLE |
| pcf\_pct\_weight | DOUBLE |
| cd\_pct\_weight | DOUBLE |
| sip\_dropped\_calls\_pct\_weight | DOUBLE |
| rrc\_radio\_drop\_pct\_weight | DOUBLE |
| downlink\_throughput\_kbps\_weight | DOUBLE |
| uplink\_thoughput\_kbps\_weight | DOUBLE |
| USE\_Voice\_Retainability | DOUBLE |
| UES\_Voice\_Reliability | DOUBLE |
| UES\_Data | DOUBLE |
| UES\_AllSERV | DOUBLE |
| preference\_ratio | DOUBLE |
| load\_time | STRING |
| score\_dt | STRING |
| Score\_hr | STRING |
| enb\_id\_list | STRING |
| enb\_simple\_area\_agg\_score | DOUBLE |
| enb\_area\_subscriber\_count | INT |
| enb\_weighted\_area\_agg\_score | DOUBLE |
| enb\_simple\_device\_agg\_score | DOUBLE |
| enb\_device\_subscriber\_count | INT |
| enb\_weighted\_device\_agg\_score | DOUBLE |

**Stp\_ druid \_ingestion\_hourly.json.variable**

**"columns" :** ["score\_date\_hr","mdn","imsi","imei","make","model","attach\_failure\_pct","attach\_failure\_cnt","attach\_attempts\_cnt","rrc\_setup\_failure\_pct","rrc\_setup\_failure\_cnt","rrc\_setup\_attempts\_cnt","srf\_pct","service\_request\_failures\_cnt","service\_request\_attempts\_cnt","pcf\_pct","session\_setup\_failures\_cnt","session\_setup\_attempts\_cnt","cd\_pct","context\_drops\_cnt","context\_events\_cnt","sip\_dropped\_calls\_pct","volte\_voice\_calls\_dropped\_cnt","volte\_voice\_setup\_incomplete\_calls\_cnt","rrc\_radio\_drop\_pct","radio\_bearer\_drops\_cnt","radio\_bearer\_setup\_attempts\_cnt","downlink\_throughput\_kbps","uplink\_throughput\_kbps","downlink\_data\_volume","downlink\_active\_time\_ms","uplink\_data\_volume","uplink\_active\_time\_ms","travelling\_indicator","travelling\_indicator\_cnt","volume\_weighted\_uplink\_thpt\_kbps","volume\_weighted\_downlink\_thpt\_kbps","load\_time","subscriber\_id","attach\_failure\_pct\_weight","rrc\_setup\_failure\_pct\_weight","srf\_pct\_weight","pcf\_pct\_weight","cd\_pct\_weight","sip\_dropped\_calls\_pct\_weight","rrc\_radio\_drop\_pct\_weight","downlink\_throughput\_kbps\_weight","uplink\_throughput\_kbps\_weight","ues\_voice\_retainability","ues\_voice\_reliability","ues\_data","ues\_allserv","preference\_ratio","enb\_id\_list","enb\_simple\_area\_agg\_score","enb\_area\_subscriber\_count","manufacturer","enb\_simple\_device\_agg\_score","enb\_device\_subscriber\_count","score\_dt","score\_hr"]

**Timestamp spec**

Score\_dt

**DimensionsSpec : [** "subscriber\_id", "manufacturer","model","make","mdn","imei","imsi" ]

**Metric Specs**

**Scripts Details:**

**stp\_druid \_ingestion\_hourly\_loader.pig**

**stp\_druid \_ingestion\_hourly\_loader.sh**

**Stp\_ druid \_ingestion\_hourly.json.variable**

**Unit Test Cases:**

|  |  |  |  |
| --- | --- | --- | --- |
| SL.No | Test Case Description | Expected Results | Test Case Status |
| 1 | Run the druid ingestion script to verify whether the druid ingestion is working or not. | Druid script has to run properly without any errors and the druid loader has to return success |  |
| 2 |  |  |  |