**RPC Data Architecture**

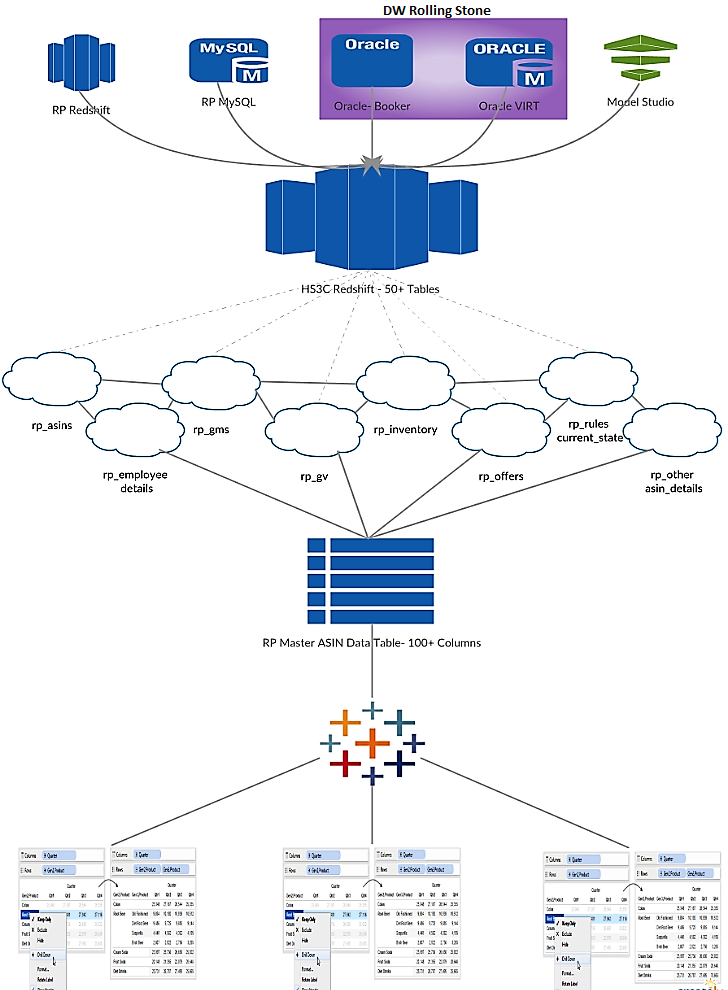
# Data Architecture

The HS3C BIA team has on-boarded 50+ required tables into HS3C redshift from multiple sources such as Selection Classification data, DW Booker and DW VIRT (the table details are in the [RP wiki](https://w.amazon.com/bin/view/HS3C/BIA/Restricted_Products_Compliance_%28RPC%29_BI_&_A/#HFactTablesandSimplifiedDataTable)). In order to simplify the querying and compiling part for the end user, the team created the fact tables as outlined in below table. By joining the fact tables and making some additional changes to it, the team created a De-Normalized table with 100+ columns, with most of the information RPC team requires. By having all the required data in one table in simplified form, anyone with basic SQL knowledge should be able to query RP related data. In addition, once the model studio tables are on board it will be easy to plug model studio ASINs data. Below are the fact tables created (additional details are in the [RP wiki](https://w.amazon.com/bin/view/HS3C/BIA/Restricted_Products_Compliance_%28RPC%29_BI_&_A/#HFactTablesandSimplifiedDataTable)).

|  |  |
| --- | --- |
| Table Name | What does table contain? |
| rp\_asins | Simplified version of RPS ASIN DATA with all the necessary computations such as open counts, false negatives etc. |
| rp\_employee\_details | This table contains RP users and their respective department details. |
| rp\_gms | This table contains calculated GMS, for RP ASINs, which have GMS. |
| rp\_gv | This table contains calculated Glance Views (GV), for RP ASINs, which have GVs. |
| rp\_inventory | This table contains latest inventory counts (Retail, FBA warehouse deals and others), for RP ASINs, which have inventory in last 2 days. |
| rp\_offers | This table contains latest offer counts (Retail, FBA, 3P), for RP ASINs, which have offers in last 2 days. |
| rp\_rules\_current\_state | This table contains current state of Rules detail for RP Rules instead of all the rule ids with additional info as review enabled, automation enabled. |
| rp\_asin\_details | This table contains other asin attributes from booker table eg : is\_adult\_product |

The tables have been created with necessary columns and logic. User Acceptance Testing is pending on this table.

# RPC Data Flow Diagram



# Business Use Case

The Restricted Products (RPC) Team i.e. Program/Product Managers request RP ASIN, Rules, User & Business related data on a daily basis to Business analyst, which they present to senior leadership or scope the projects and initiatives in hand. Also, RPC team only have automated excel dashboard with minimal functionalities i.e. won't be able to drill down and find answers to the question without the help of Business Analyst.

Currently the process is manual and consists of the following steps.

1. SIM will be filed to a Business Analyst (BA).
2. BA reviews the request and asks for additional information if needed
3. Based on the request, BA develops SQL queries in Redshift, Data Warehouse and downloads the data into Excel after the queries execute.
4. Once the user downloads the data into Excel, the requested information is compiled based on the request.
5. Finally the BA shares the final output with the requester.

This requires a Business analyst/Manager to have RP subject matter expertise, SQL & Excel expertise and spend time in manually compiling the data from multiple sources. This is a time consuming and error prone task. Due to varying expertise levels in SQL, queries written can slow down the HS3C Redshift cluster as well. The average SQL development and compilation time by a BA to resolve a SIM is about 1-2 days based on the availability and currently RP BA gets around average 33 SIMs per month.

For example, in this [SIM](https://issues.amazon.com/issues/RPBI-417) the legal team has requested for a list of ASINs for rule Auto\_Tuner\_Claims and its respective Retail/3P offers, keywords caught, ASIN title, GL and GMS data. To get this output, a BA needs to follow the steps below:

1. Run a redshift query in RP Redshift to get the list of ASINs.
2. Create a DW segment for a list of ASINs and run one query to get offer data and one query to GMS data for the respective ASINs.
3. After retrieving data into Excel, do necessary lookups to consolidate them.

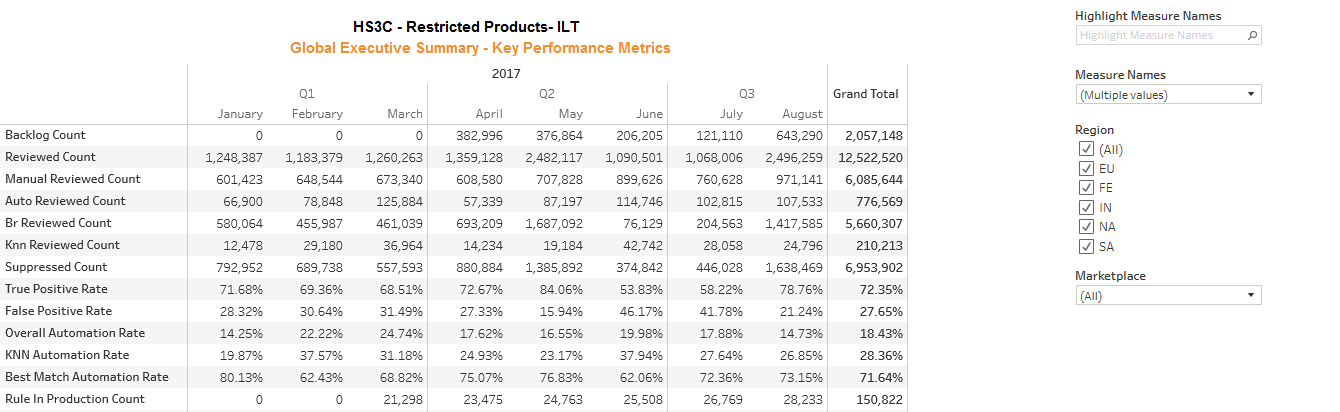
To make this an easy and fast process and to reduce operational risk to the cluster, the team on-boarded Data tables from multiple sources into one Redshift cluster [HS3C] and created a de-normalized table combining multiple facts. With the architecture outlined below a single query on the HS3C Cluster can achieve the same results. The team is targeting to expand the proposed architecture to address 90% of the regularly received requests, within the next 3-6 month timeframe.

# Tableau Dashboard

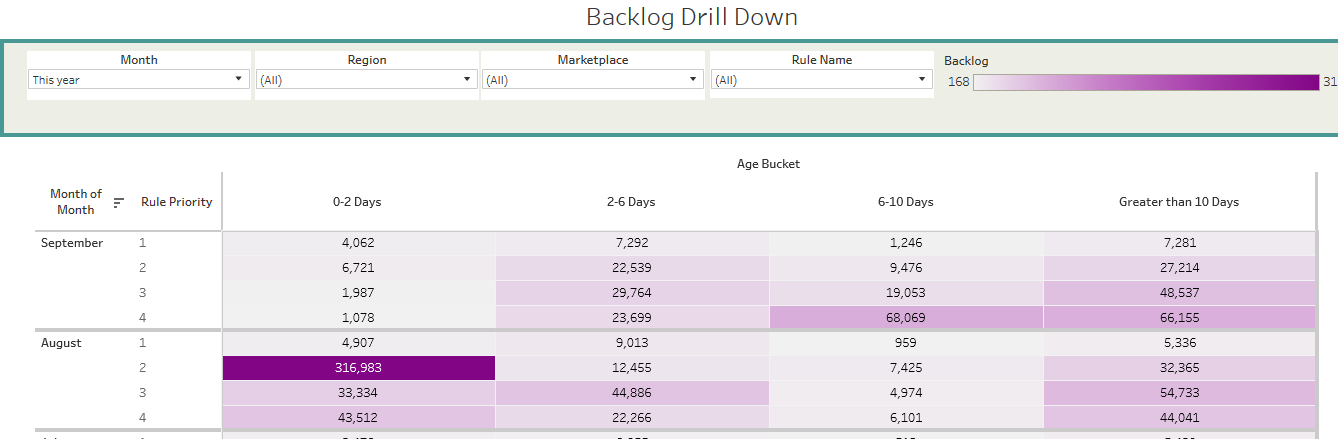
The BI team has also created the ILT Dashboard as requested by the program team. (Linked [here](https://hs3c-insights.aka.amazon.com/#/views/RPCITL-GlobalExecutiveSummary/Dashboard1)). The request was to replicate the current Excel dashboard into Tableau and automate it. BI team has automated the report in Tableau and added additional Tableau functionality for better user performance.

Along with ILT dashboard BI team has also built Backlog Dashboard and Ticketing Dashboard. The team is also building other dashboards based on the requirements of the program team. Link to Tableau Dashboard can be found [here](https://hs3c-insights.aka.amazon.com/#/projects/35/workbooks).

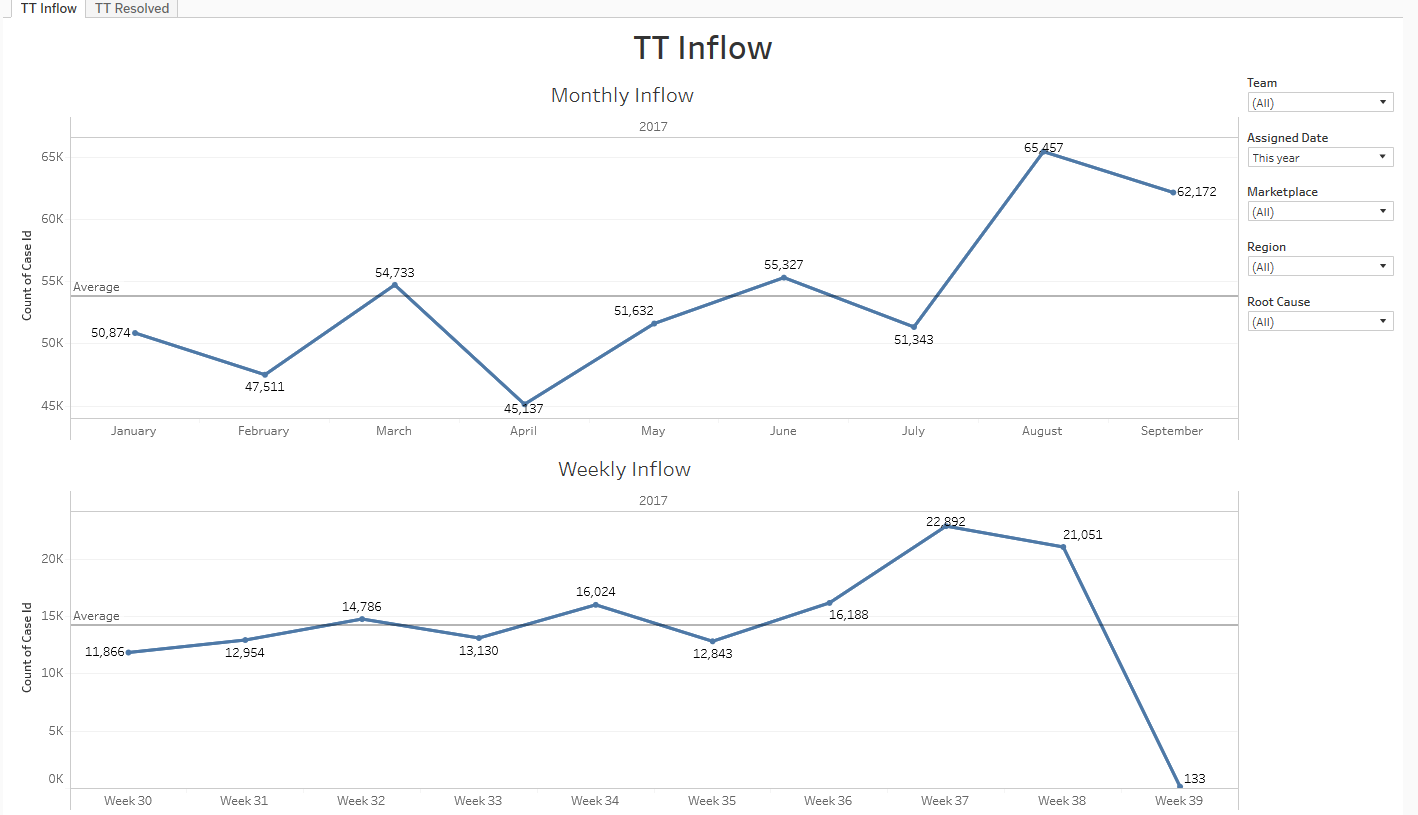
### ILT Dashboard



### Backlog Dashboard



### Ticketing Dashboard



# Road Ahead

1. Roll out the table & dashboard for UAT.
2. Gather additional requirements from the Program team if any.
3. Roll out the final tables & dashboard into production for everyone to consume.
4. Measure the success.
5. Plan for self-serving tools without SQL.