|  |
| --- |
| C:\Users\bhavinkumarp\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.Word\medseek.png |
| Reporting 2.0 |
| Influence Health Engineering |
| Scope, Process, Infrastructure, Approach, Data Model, ETL & Visuals |
|  |
| **Reporting 2.0 - Influence Health India** |
| **11/4/2016** |
|  |
| Version 2.0 |

|  |
| --- |
| **InfluenceHealth** 3000 Riverchase Galleria, Suite 1500, Birmingham, AL. 35244 <http://www.influencehealth.com> |

# Contents

[1 Contents 2](#_Toc466036030)

[2 OBJECTIVE 4](#_Toc466036031)

[2.1 INTRODUCTION 4](#_Toc466036032)

[3 PROCESS & CHALLENGES 4](#_Toc466036033)

[3.1 Discussions /Team Meetings - PROCESS 4](#_Toc466036034)

[3.1.1 Roles & Responsibilities 4](#_Toc466036035)

[3.1.2 JIRA process- To be updated if there is a change in future 5](#_Toc466036036)

[3.1.3 Step by Step Report Creation Process 6](#_Toc466036037)

[3.1.4 Working with IHI 6](#_Toc466036038)

[3.2 Discussions /Team Meetings - CHALLENGES 6](#_Toc466036039)

[3.2.1 Data Movement 6](#_Toc466036040)

[3.2.2 Custom Reporting 7](#_Toc466036041)

[3.2.3 1 report per client v/s 1 report for all clients 7](#_Toc466036042)

[3.2.4 Raw Data access for clients 7](#_Toc466036043)

[3.2.5 UX consistency 7](#_Toc466036044)

[4 INFRASTRUCTURE 8](#_Toc466036045)

[4.1 ARCHITECTURE 8](#_Toc466036046)

[4.2 Infrastructure Planning 8](#_Toc466036047)

[5 SCOPE 9](#_Toc466036048)

[**5.1** Short Term Objective 9](#_Toc466036049)

[**5.2** Long Term Objective 9](#_Toc466036050)

[6 Data Model 9](#_Toc466036051)

[7 ETL 10](#_Toc466036052)

[8 Visuals 10](#_Toc466036053)

[9 QA - IHI 10](#_Toc466036054)

[9.1 ETL testing 10](#_Toc466036055)

[9.2 Visualization testing 11](#_Toc466036056)

[9.3 Process and Documentation 11](#_Toc466036057)

[9.4 Future Scope 11](#_Toc466036058)

# OBJECTIVE

To set up process and project plan around Reporting 2.0 Project.

## INTRODUCTION

This section has the Summary / Agenda around the Reporting 2.0 discussions occurred between [05/31/2016 – 06/014/2016], first week to discuss higher-level items, with the second week focused on designing actual reports.

* Process discussions
* Challenges

# PROCESS & CHALLENGES

## Discussions /Team Meetings - PROCESS

Following are the details of various discussions held to determine process around Reporting 2.0 project.

### Roles & Responsibilities

**Product Owners:** To build Product requirements that caters client needs and approves the final deliverables.

1. Phil & Mary: Client interaction, communication, product requirement and approvals.

**Project Management:** To plan and manage the deliverables by coordinating with different teams.

1. Dave – IHUS
2. Asha - IHI

**Development Team:** This team is responsible to build data model for Reporting 2.0 project along with the Tableau development.

1. IHI – Srihari , Venkat
   1. Team Lead – Srihari (Team management / Task Distribution / Development)
2. IH-US – Chue
   1. Team Lead – Chue (Team management / Task Distribution / Development)

**ETL Team:** This team is responsible to build and Test ETL process for Reporting 2.0 project that bring data from Cassandra (EDH) to Redshift (Reporting 2.0 Data Warehouse)

1. ETL Development:
   1. IH-US – Nathan (Tentative) , Jane
2. ETL QA:
   1. Chris Wendt / Jane
   2. Supriya – Primary IHI resource

**QA Team:** This team is responsible for validation of Tableau visuals for Reporting 2.0 project.

1. Data Validation: Supriya.
2. UXP : Liz
3. UAT: Mary , Phil / Dave

### JIRA process- To be updated if there is a change in future

There will be three boards that tracks every involved activity of Reporting 2.0

1. **Request board** : This board is used between the Client Services and Product for new feature/report requests , following are its work buckets
   1. **New requirement :** CS team member / user raises new feature/ report request
   2. **Need more info :** To capture more details on the requirement if things are not clear
   3. **Under review :** Initial analysis of user request to decide the product scope / approve the user request
   4. **Rejected :** Feature will not be implemented
2. **Elaboration board ( Product-Technical Team ) :** This board is used to elaborate the approved requests coming from Request board, following are its work buckets
   1. **To Do :** Any approved request from the Request board will be in this queue
   2. **In Progress :** User request is elaborated to cater more details for development team
   3. **Ready for Grooming :** Feature overview & Feasibility check to understand the technical implications
3. **Construction board (Technical Team – Project Management ) :** This board is used to track the development, validation, user acceptance and deployment activities of elaborated tickets, following are the current work buckets (To be updated by Dave)
   1. **To Do :** Development team picks up work items from this bucket based on First In First Out order
   2. **In Progress :** Any work that is in progress
   3. **In Progress Review :** To showcase the progress and to accommodate any change requests / cosmetic changes to avoid last minute changes
   4. **QA :** Validation phase
   5. **UAT :**  Final approval phase
   6. **Done :** Feature Integration to application

### Step by Step Report Creation Process

Following are the steps involved in this process

1. Requirement Elicitation and Analysis – Product Team
2. Feasibility Check – Technical Team
3. Data Model - Technical Team
4. ETL – Technical Team
5. Visualizations – Technical Team
6. User Acceptance - Product
7. Integration – Technical Team

### Working with IHI

Following points describe the communication and Interaction between two teams (IHI-US & IHI)

1. Work Item prioritization – Ideally First In First Out order is followed to pick up the items from construction board unless specified by Project management team
2. ETL – IHI can pitch in on need basis if there is need to perform any new ETL tasks for Reporting 2.0 future requests
3. QA – IHI will contribute depending on the area of specialization
4. Status Updates - JIRA and Daily Stand Up Calls [6.30 PM – 7.00 PM IST or 8.00 AM – 8.30 AM CST]

## Discussions /Team Meetings - CHALLENGES

Following are the challenges that team may encounter during the project development.

### Data Movement

This task is to bring the EDH data from Cassandra database to Redshift data warehouse through ETL and has following challenges

1. Data availability for different clients on EDH impacts Reporting 2.0 timelines
2. Any change on EDH structure may impact ETL to redshift
3. Number of updates on EDH = Number of data refresh on Redshift

### Custom Reporting

Custom reporting is dependent on client level agreements and may involve different reporting strategies

1. Pre-defined reporting hours per client
2. Categorization of client report requests to determine the efforts
   1. Small – No changes to Data model
   2. Medium – Minimal changes to Data model but no ETL level changes
   3. Large – New ETL and Data model efforts

### 1 report per client v/s 1 report for all clients

The deployment strategy for hosting Tableau visuals on the server and following are the implications of each approach:

1. 1 Report Template / Client
   1. More Deployments
   2. Role based user mapping is less complex
2. 1 Report Template for all clients
   1. One time deployment
   2. Role based mapping is complex

### Raw Data access for clients

This can be done and will be accomplished if client is willing to pay extra money for the new Relational Database set up to access the raw data on redshift.

Currently this is out of scope in the Reporting 2.0 objective.

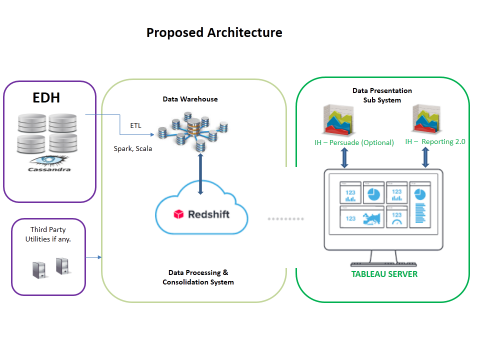
### UX consistency

As several people are involved in Tableau visual development maintaining the UX consistency is difficult and hence UX specialist (Liz) will be involved in the In-progress review of every report development and guide team whenever necessary on the cosmetic related changes.

IH standard style guide is also available for the team to refer and implement the best practices of visualizations.

# INFRASTRUCTURE

## ARCHITECTURE



## Infrastructure Planning

1. **Redshift Cluster Set Up**:
   1. Currently a redshift server (10.128.64.41) has been set up to start work on the Reporting 2.0 project
   2. Performance Testing is yet to be performed over larger set of data
2. **Tableau Server / User Licencing:** 
   1. **Tableau server hosting URL: (**[**www.tableau.influencehealth.com**](http://www.tableau.influencehealth.com)**)**
   2. **User licencing mode is yet to be decided**
      1. Named user licencing -
      2. Core user licencing – Expensive
   3. **Tableau / Redshift Server Monitoring** – Max will be looking into these tasks

# SCOPE

Reporting 2.0 has short term and long term goals and corresponding details are as follows:

## Short Term Objective

1. **Leads report – Chue , IHI** (Post transition)
   1. Source – EDH (Persuade data on Redshift for Piedmont client)
   2. Business Rules – (JIRA Ticket BI- 46)
2. **Performance Report** (Marketing Effectiveness report without ROMI calculations) **- IHI**
   1. Source – EDH (Predict / EDH data on Redshift)
   2. Business Rules – (BI-23)
3. **Demographic profile report - IHI**
   1. Source – EDH (Predict / EDH data on Redshift)
   2. Business Rules – (BI – 47)
4. **Patient Acquisition report - IHI**
   1. Source – EDH (Predict / EDH data on Redshift)
   2. Business Rules – (BI – 48)
5. **Persuade set of reports – Chue & IHI** (Need Basis)

## Long Term Objective

Long term objective of this project is to create a common reporting platform for AI and Persuade users so that user can assess performance on any type of list and link the two worlds.

# Data Model

Star Schema has been proposed to build the data model for Reporting 2.0 project lead by Chue accompanied with IHI.

Benefits are as follows:

1. High analytical / complex query performance
2. Model can evolve over time without changing existing structures
3. Efficient navigation through data
4. Load performance and Administration

**Business Logic Implementation Strategies:**

1. **Implementation during data load** – Certain business rules has to be implemented during the ETL to gain more performance (Fast Rendering of reports) or to avoid custom query writing while building visualizations.
2. **Implementation after data load –** Certain business rules can be achieved using Tableau constructs to optimize ETL performance

# ETL

1. ETL from Cassandra to Redshift is through Spark framework using Scala programming language.
2. Initial ETL implementation is led by Nathan and will be carried out by Chris in the later stages
3. IHI can contribute to ETL tasks depending on work load at IH-US or on need basis

# Visuals

1. Mock Up screen has been developed using existing data on redshift
2. This helps to finalize the Visuals for report requests
3. Once actual data is available then the data sources are to be replaced to get the actual data

# QA - IHI

QA tasks are mainly categorized into two streams

1. ETL
2. Visualizations

## ETL testing

Following process will be followed to ensure the correctness of data in the redshift server

1. Data validation between Cassandra and Redshift staging
   1. Data in redshift staging database will be validated against source data (Cassandra)
   2. A meeting will be held with Mary to understand the current data validation process
   3. Necessary tool set up (Data Stax Console)
2. ETL testing within redshift ( Raw tables in Staging v/s Star schema model)
   1. ETL scripts that are used to load the star schema will be validated and tested with new client data
   2. Scripts that are used to load data from Staging (Redshift) to Production (Redshift) is validated and tested
3. Above process is carried out once during every new client set up

## Visualization testing

1. Every report request/ticket on construction board has to go through QA at various stages
   1. First build
   2. Data refresh
   3. Change requests
2. A set of scenarios / test cases will be created/ updated for every report request for above mentioned stages
3. Necessary tool set up (My SQL workbench)
4. Report data will be validated by writing backend queries

## Process and Documentation

**PROCESS:**

1. Test cases has to be created using the defined QA template and is to be attached to the JIRA ticket before unit testing
2. Scenarios for any change request on the existing ticket has to be updated on the same document where the initial scenarios are written
3. No separate ticket is created for QA tasks, once the development ticket from ‘In Review’ to ‘QA’ one has to start working on the validation based on priority
4. In case of multiple tickets in the QA queue Dave / Phil will set the priority to address the tickets

**DOCUMENTS:**

1. QA Test case template
2. QA tracker for internal use

## Future Scope

Once the data refresh and ETL development is stabilized then we can plan to automate some of the manual testing process with the help of dashboards specific to data loads. Currently this is in the ideation phase and this section will be updated soon with concrete details.