

Corporate Information model Catalog for Self Service using (Online Analytical Processing) OLAP methodology

Farhana Sethi¹

¹Schlumberger Ltd

September 14, 2020

Abstract

Corporate Information Models(CIM) are Online Analytical Processing (OLAP) Cubes – a data structure that enables simple and rapid analysis of data by the multiple dimensions that define a business challenge. A multidimensional cube for reporting composed of multiple Dimension e.g. Customers, Product Lines, business units and Date. By storing data using OLAP cubes, data is pre-calculated and ready-to-use, eliminating long-running data queries and enabling self-service business intelligence. A portal has been developed to provide and display the CIM detail. The CIM catalog links to over 60 CIMs categorized by the seven main categories called Value streams.

Corporate Information model Catalog for Self Service using (Online Analytical Processing) OLAP methodology

Farhana Sethi-Global Data & Analytics Business Intelligence - Quality & Governance Manager with Schlumberger Oilfield, Texas, Houston since 2006-Farhana sethi: FSethi@slb.com

Abstract — Corporate Information Models(CIM) are Online Analytical Processing (OLAP) Cubes – a data structure that enables simple and rapid analysis of data by the multiple dimensions that define a business challenge. A multidimensional cube for reporting composed of multiple Dimension e.g. Customers, Product Lines, business units and Date. By storing data using OLAP cubes, data is pre-calculated and ready-to-use, eliminating long-running data queries and enabling self-service business intelligence. A portal has been developed to provide and display the CIM detail. The CIM catalog links to over 60 CIMs categorized by the seven main categories called Value streams.

Index Terms —Application packages, Design Concept, Turnkey Solution, Intranet Portal

1 Introduction

OLINE

oAnalytical Processing (OLAP) allows users to analyze database information from multiple database systems at one time. An OLAP Cube is a data structure that allows fast analysis of data according to the multiple Dimensions that define a business problem. OLAP in a data warehouse enables companies to organize information in multiple dimensions, which makes it easy for businesses to understand and use data. Since OLAP contains multidimensional data usually obtained from different and unrelated sources, it requires a special

method of storing that data. OLAP tools enable users to analyze different dimensions of multidimensional data using self-service.

In our research and implementation scope, we have used OLAP method to provide corporate level multidimensional data to the business users. Corporate Information Models (CIMs) provide trusted and qualified data from multiple sources for everyone to build their own business intelligence reports and dashboards.

Corporate Information Models are Online Analytical Processing (OLAP) Cubes – a data structure that enables simple and rapid analysis of data by the multiple dimensions that define a business problem. A multidimensional cube for reporting composed of multiple Dimension e.g. Customers, Product Lines, business units and Date. By storing data using OLAP cubes, data is pre-calculated and ready-to-use, eliminating long-running data queries and enabling self-service business intelligence. A portal has been developed to provide and display the CIM detail. The CIM catalog links to over 60 CIMs categorized by the seven main categories called Value streams.

Application Feature

There are some key features for the application that make it unique and useful, as mentioned below.

CIM Catalog

CIM Catalog is a central repository for all the self-service OLAP models in the organization. It contained over 60 models with their description, facts and dimension detail. OLAP models detail synchronize with Azure Analytics Services using Azure Pipelines with daily refresh schedule. Catalog provides insight about every column definition approved by Data Owners, source system detail, connection and access related information. Every model has a hierarchy and map with main category called value Streams.

Social

Business users and analytics team members can easily communicate issues or feedback using focused discussions per analytics on CIM catalog such as Feedback, Reviews, and Support and feature requests.

Integrated with the ecosystem

CIM portal set up as Single-Sign-On with viewer integration. It uses Domain Name System (DNS) and CNAME (Canonical Name record) that is easy to remember and searchable from central internal hub .it also provides authentication on the mobile integrated with Microsoft Intune which is a cloud-based service that focuses on mobile device management (MDM) for the organization.

Structure the Catalog using Tags

To make all analytics on CIM catalog discoverable, it provides a taxonomy to the catalogs using tags. This allows business users to explore what is available in the organization.

Drill down to any combination

Users can easily drill down to the exact combination of what they want to explore. They can also mix and match with other filters that are available, such as a filter to only select a particular data source.

Supporting Content

Add supporting content to each CIM model, like: source data, a user guide, utilization metrics or a video walkthrough. This makes it substantially easier for new users to know their way around the CIM model, especially if it requires more context to understand the analytic assets.

Mobile View

Application supports mobile view using RWD. Responsive web design (RWD) is a web development approach that creates dynamic changes to the appearance of a website, depending on the screen size and orientation of the device being used to view it.

Integrated Viewer

It provides a view of all CIM model right from within Central Catalog. With central catalog users can request for access on the OLAP models, read the columns definitions, find the data owners and download pre-built Power BI report to explore and customize their own analytics.

Application Architecture

CIM Catalog has been designed and developed in the Microsoft Azure Cloud using SharePoint Online as described in high level architecture diagram Figure 1

Microsoft Azure is a cloud computing service created by Microsoft that can be used for deploying and managing applications as well as services through a global network of Microsoft-managed data centers.

In our research, corporate data source modeling being developed using Microsoft Azure Analysis Services which is linked with CIM catalog using Microsoft Azure Pipeline.

Azure Analysis Services is a fully managed platform as a service (PaaS) that provides enterprise-grade data models in the Microsoft Azure cloud. **Azure Data Factory** is Azure’s cloud ETL (Extract, Transform, Load) service for scale-out server-less data integration and data transformation.

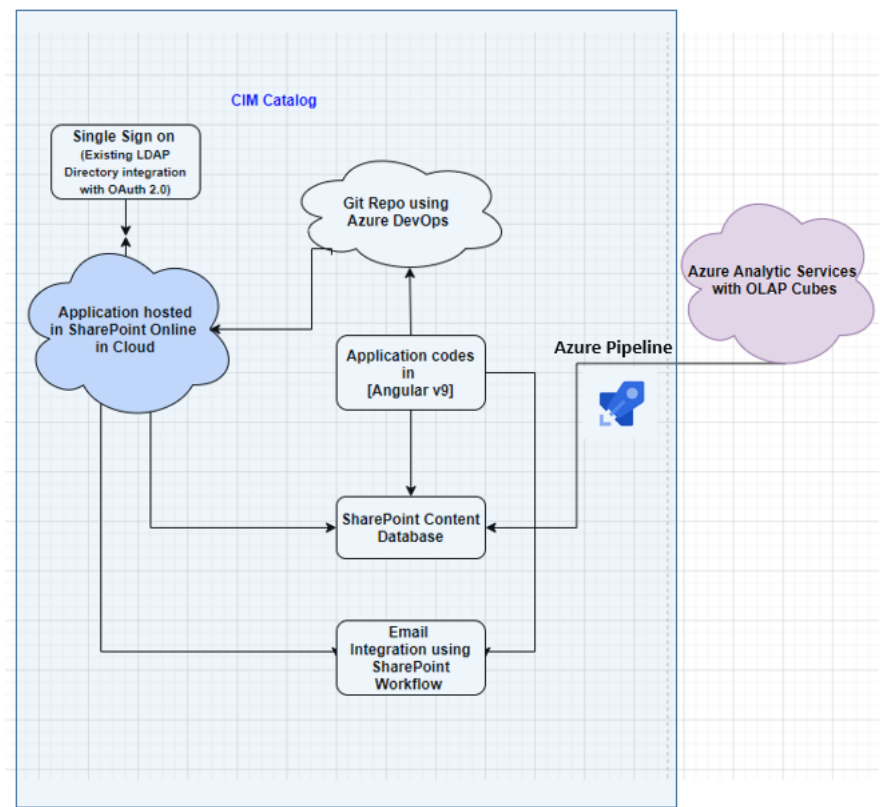


Figure 1: High-level Architecture diagram – CIM Catalog

Key technology and components

Key technology and components of Digital store are as follows:

Web Framework

In our research we have considered and compare Angular, Angular JS and React for the selection of web framework to develop the application.

There are several other web framework available such as Vue, Ember, Meteor, Polymer, Backbone, Aurelia and Mithril. In our research, we have not included the comparison of all other frameworks.

Angular: Angular is a web framework and a platform for building mobile and desktop web applications using HTML and Typescript. Angular is written in Typescript (Typescript is an open-source programming language developed and maintained by Microsoft). Angular is used in many public-facing application like Google and YouTube. It is a completely free framework helps to use HTML as a template language, creates RICH Internet Application and offers its developers a client-side application. We have used Version 9 in our application development, Angular 9 was released on February 6, 2020. This latest version moves all applications to use the Ivy compiler and runtime by default. Ivy is the code name for Angular's next-generation compilation and rendering pipeline. Also it has been updated to work with Typescript 3.6 and 3.7.

Angular VS. Angular JS

If we compare the performance on Angular with AngularJS then we have better reasons to use Angular. Angular is at least five times faster than AngularJS due to a much better algorithm for data binding and a component-based architecture. The components of an Angular application are quite independent and self-sufficient, which makes them reusable and test friendly.

Angular used Typescript and Angular JS uses JavaScript.

The main advantage of Typescript over JavaScript is that Typescript is a superset of JavaScript. So Typescript designed for the development of a large program that Trans compile to JavaScript. But Typescript may be used to develop a JavaScript application for both client-side and server-side.

Angular VS. React

There is a detail comparison which has been done between Angular and React as described in Table 1.

Features	Angular
Brief Overview	Full-fledged open-sourced JS Framework managed by Google and the Angular developer community Launched in Sep 2016
Universality	Both Web and Mobile For mobile development, it needs to be incorporated with Cordova. Moreover, for mobile development, it supports single and multiple-page apps
Self-Sufficiency	Angular is a full-fledged framework for software development Hence no additional libraries needed
Learning Curve	Angular itself is a huge library. Angular is more complex to understand, there is a lot of unnecessary code Long and difficult
Community	Large Google provides the long-term support of the framework and constantly improves it. However, the Angular community is also very active and provides a lot of support. Angular is used by such companies as McDonald's, AT&T, HBO, Apple, Forbes, Adobe, Nike, and many others.
Performance	Competitive Optimized with change detection
Language	JS, TypeScript
App Structure	Fixed and Complex Component-based Model, Controller, and view
UI Components	Built-in Material Design toolset
Directives	The greatest advantage of Angular rests in the fact that, unlike React, it supports dependency injection. Without it, the application becomes incomprehensible without knowledge of Angular.

Features	Angular
Dependency Injection	Supported
State Management	Effective improved with NgRx or RxJS
Data Binding	Bidirectional Data is mutable
Change Rendering	Real DOM, Change detection
Tools	Aptana, Sublime Text, Visual studio, Angular CLI, Angular Universal, Jasmine, protector, Karma

Table1: Comparison between Angular and React Framework

To conclude our comparison, Angular is a full-fledged mobile and web development framework. React is a framework only for UI development, which can be turned into a full-fledged solution with the help of additional libraries. In the end, React vs Angular is all a matter of personal preference, a matter of skills and habits. We have used Angular v9 for the application development.

Microsoft SharePoint online

SharePoint is a cloud-based service offered from Microsoft that helps organizations share and manage content, knowledge, and applications to empower teamwork, quickly find information and seamlessly collaborate across the organization.

In our research and deployment we have privileged enterprise infrastructure setup of SharePoint online in cloud to host the application that provides single sign on authentication using O365 with directory services, out of the box workflow development for email send and receive and host the database using SharePoint online Document library and SharePoint lists.

Modern Authentication

Modern Authentication is the term Microsoft uses to refer to their implementation of the OAuth 2.0 authorization framework for client/server authentication. By default, modern authentication is enabled for SharePoint online.

SharePoint content database

A content database is a database file that stores content for one or more site collections for SharePoint web application. The content can be pages, files, documents, images and much more. In our case, we have used SharePoint document library, lists and images to save the Catalog information.

SharePoint Online Workflow

Workflows help people to collaborate on documents and to manage project tasks by implementing business processes on documents and items in a SharePoint site. There are five commonly used Workflows that can be set up within SharePoint: 1) Approval Workflows, 2) Status Workflows, 3) Notification Workflows, 4) Automation Workflows, and 5) Custom Workflows. In our research and deployment we have set up custom workflow to send emails for Feedback and support form.

Visual studio Code

We have used Visual studio Code for code development. Visual Studio Code is a lightweight but powerful free- source code editor made by Microsoft for Windows, Linux and macOS that runs on your desktop. It has built-in source code control including Git support. The Visual Studio (VS) Code product itself ships under a standard Microsoft product license, as it has a small percentage of Microsoft-specific customizations. It's free despite the commercial license. We have all the good reasons to use VS Code for our research and tool deployment.

Data Structure

There are 5 main list and 3 document libraries has been used to embrace the catalog information as described in Figure 2

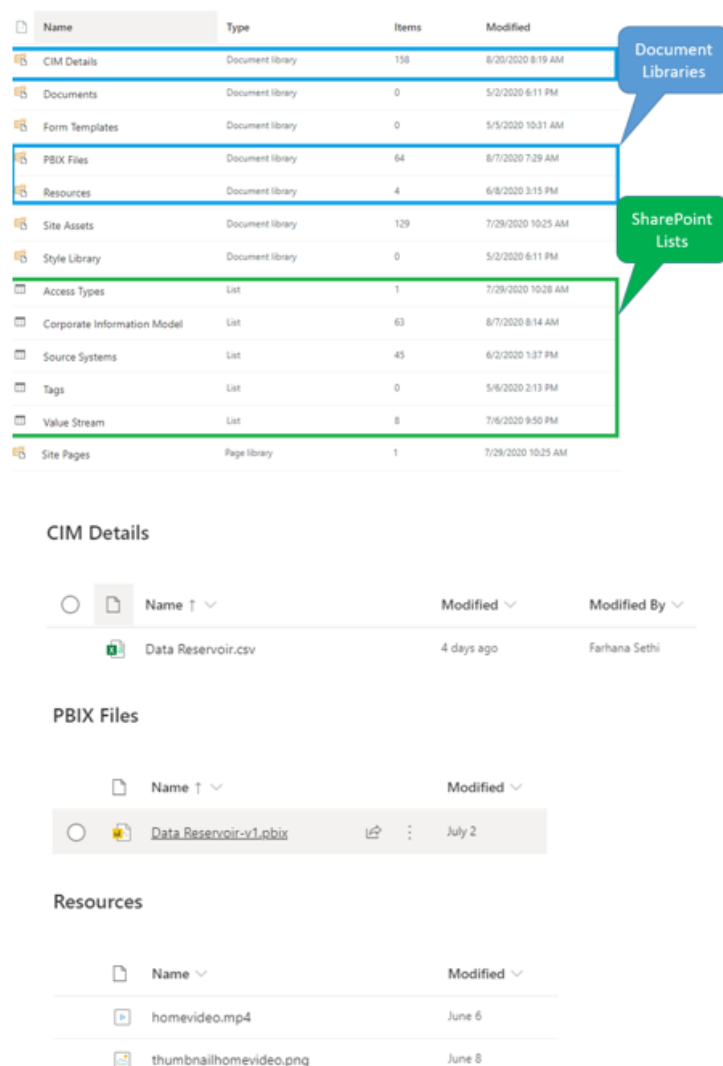


Figure2: SharePoint content database

Every SharePoint list has columns to keep the information for the catalog as described in Table 2.

List Name	Corporate Information Model
Column	Description
Corporate Information Model Name	Corporate Information Model Name
Value Stream	Value Stream (Main Category of the CIM)
Server Address	OLAP Connection Address
Source Systems	Source Systems Name
Description	Description
Refresh Frequency	Refresh Frequency

List Name	Corporate Information Model
History Available	Data retention information
Contact	CIM Contact (Data Owner)
Access	Access
Details Information	CIM Details Information
Available since	CIM Launch date
Tags	Tags
CSV File Path	CSV File Path for the CIM model data dictionary
PBIX File Path	PBIX File Path
Status	Status of the CIM Model (Production or under development)
List Name	Source Systems
Title	Enterprise data source system name
List Name	Tags
Tag Name	Tags name/ Taxonomy to the catalogs
TagColor	Display color for the tags
List Name	Value Streams
Value Stream Name	Main category mapping name
Value Stream Owner	Category Owner
Value Stream Data Owner	Category Data Owner
Value Stream Description	Category description
Value Stream Acronym	Category Acronym
List Name	Access Types
Access Name	Access application name
Access Link	Access application link
Help Link	

Table 2: SharePoint content database Column level detail

Mapping Techniques

A Smart mapping technique has been used to map OLAP catalog directory with CIM catalog and categories as described in Figure 3.

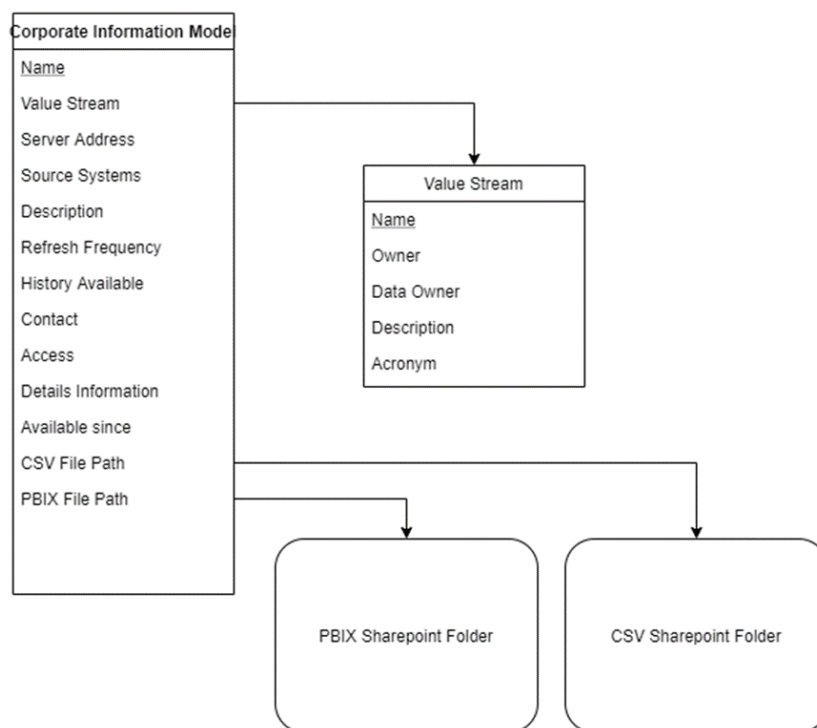


Figure 3: Column mapping for Data association

Folder Structure

The key guidelines for creating folder structure for an Angular Project is to organize the app such that you can locate code quickly and have a near-term view of implementation and a long-term vision. We have built the structure which focuses on a multiple-module architecture which in turn are divided into pages and a shared set of components with a large focus on scaling as described in Figure 4.


```
> .vscode
> e2e
v src
  v app
    > configs
    > core
    v modules
      > error
      v home
        > pages/home
          TS home-routing.module.ts
          TS home.module.ts
      v value-stream
        > pages/value-stream-detail
          TS value-stream-routing.module.ts
          TS value-stream.module.ts
          TS modules.module.ts
    v shared
      v components
        > avatar
        > brand
        > corporate-information-card
        > lateral-menu
        > layout
        > loader
        > outdated-browser
        > page-footer
```

Figure 4: High level Overview of Folder Structure

Code architecture

We have adopted the MVCS code architecture for the application development. The Model-View-Controller (MVC) is an architectural pattern that separates an application into three main logical components: the model, the view, and the controller. MVC is one of the most frequently used industry-standard web development framework to create scalable and extensible projects

Models - To manage the data of an application. The models will be anemic (they will lack functionalities) since they will be referred to the services.

Views/Templates - The page/GUI where the user interacts with the app.

Controllers - Links between services and views.

Services - Manages the SharePoint request data.

Angular Components

Components are the most basic building block of an UI in an Angular application. An Angular application is a tree of Angular components. Every page has multiple components and one component can belong to many pages. Below is an example of components breakdown and association with the Storefront landing Page as described in Figure 5.

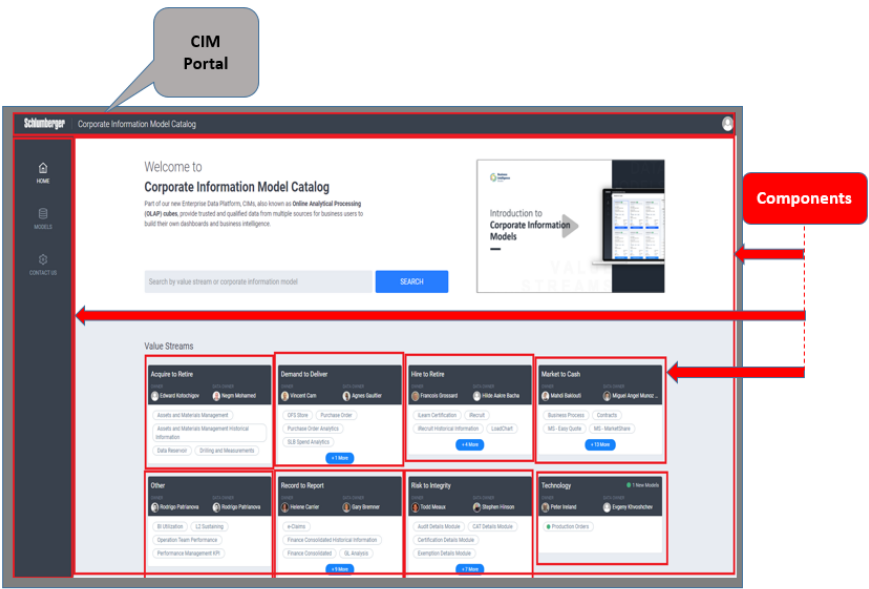


Figure 5: Components Association with the Web page

There are no more or less important components, all are important, but there are more complex or more used components.

Key Components

There are over 20 components being used to develop the application with all the required features in the scope. Top 3 key components has been identified and mentioned below in Figure 6, 7 and 8.

Value Stream Card

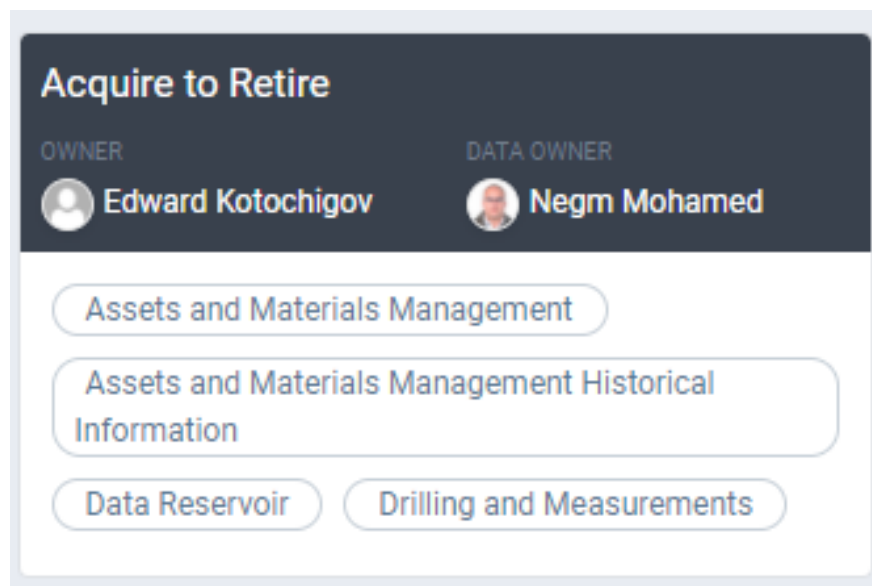


Figure 6: Value Stream Card
Corporate Information Card

Assets and Materials Manage...

Indicators on Assets across the company such as Asset Availability (aka. Capacity) and Asset Utilization.

VALUE STREAM
Acquire to Retire

SOURCE SYSTEMS
FTL / QTRAC / RITE

TAGS

ADDRESS (Click to copy)
[link://slbbidatamodelsOperations.azurew...](#)

ACCESS SAMBA ?	REFRESH RATE Weekly
CONTACT	AVAILABLE SINCE 01-Oct-18

Get Started in Power BI

Figure 7: Corporate Information Card

Search Bar

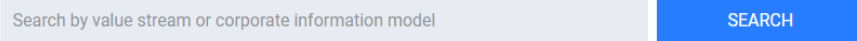


Figure 8: Search Bar

Codes for the top 3 key components has been describe in Appendix 1.

Inside the Components

Every Component has to be written in three languages.

HTML - Hypertext Markup Language (HTML) is the standard markup language for documents designed to be displayed in a web browser.

Scss - Scss stands for Sassy Cascading Style Sheets and it's an extension of CSS which adds nested rules Variables, mixin, selector, inheritance and a lot more features.

TS - TypeScript is an open-source programming language developed and maintained by Microsoft.

SharePoint Connection

SharePoint ship with an updated REST API that can be used to access and control nearly every aspect of a site collection: Lists, Libraries, Permissions, and Users are just some of the objects that have defined endpoints. Integrating Angular, a powerful client-side framework, gives developers and content managers another option for customizing the user experience. The most important library used in the project is to create and manage the connection between the application and the SharePoint structure.

Deployment technique

We used Minimum Viable Product (MVP) approach for product development strategy. A minimum viable product (MVP) is a version of a product with just enough features to satisfy early customers and provide feedback for future product development.

Every MVP includes below key elements:

Functionality - the set of features deliver clear value to the user.

Design - the design of the MVP was up to the highest industry standard.

Reliability - production quality standard has been achieved by rigorous testing.

Usability - the MVP was easy to use and intuitive.

We have used Azure DevOps for code deployment. Azure DevOPS is a Software as a service (SaaS) platform from Microsoft that provides an end-to-end DevOps toolchain for developing and deploying software.

Azure DevOps being used for code versioning and release pipeline using Git repository with Visual Studio code tool.

Git is a program that tracks changes made to files. Once installed, Git can be initialized on a project to create a Git repository. A Git repository is the .git/ folder inside a project. This repository tracks all changes made to files in the project, building a history over time.

A release pipeline is a conceptual process by which we take committed code into production. A release pipeline takes a build artifact, a result of the build process and deploys that to one or more environments. We have used three environment for the product deployment i.e. Development, Quality Assurance and Production as described in Figure 9 below.

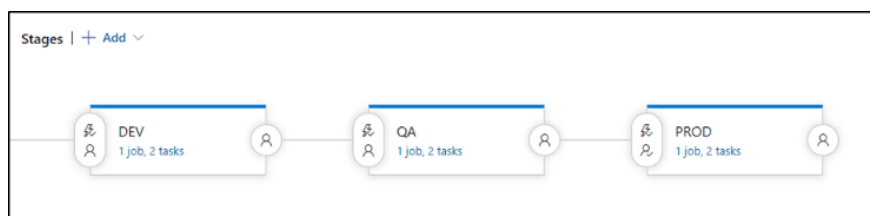


Figure 9: Release pipeline in Azure DevOps Repository

Code Development environment integrates with Azure DevOps Git repository to enable source control management. With source control, developers can collaborate, track and save their changes to a branch and these changes will be merged into the main branch and deployed to the higher-level environments (From development to Quality Assurance/User Acceptance Test environment to Production), where it will also be tested and validated. Once the above mentioned steps are complete the developed codes be published into the development environment as described in Figure 10

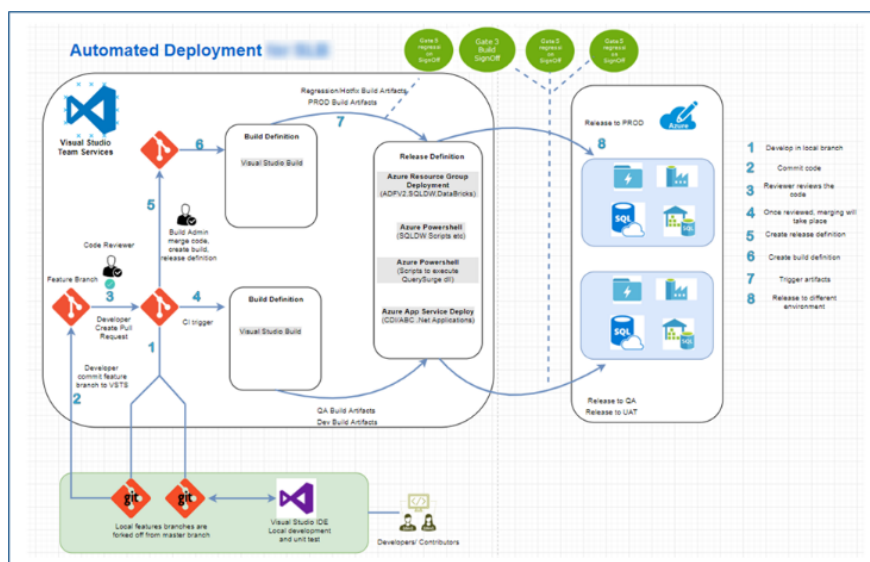


Figure 10: Automated deployment using Azure DevOps (Git) Repository

Azure Release Pipeline Approval/Pre Deployment Conditions

Approvals and gates gives further control over the execution and success of the deployment pipeline. Each step in a release pipeline can be built with pre-deployment and post-deployment settings that can embrace waiting for authority users to manually approve or reject deployments, and testing and validating the quality checks with any computerized systems until definite requirements are complete. Also, we can constitute a manual interference to pause the deployment pipeline and prompt developers to carry out manual tasks, then resume or reject the deployment.

Once the approvals are in place the Release would be promoted to higher environments and finally deployed in Production. Assigned approvers would receive email notification along with the link to approve. Unless approved deployment will not take place. Once it is approve and release, pipeline shows with success as described in Figure 11

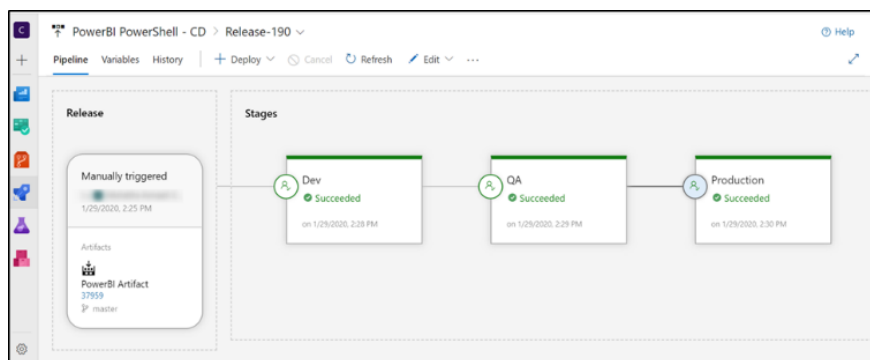


Figure11: successful Azure Release Pipeline

Output

Application in production has below key displays that will be interesting to show as described in Figure 12, 13, 14 for **CIM Catalog Landing page** , **Inside the CIM** and **Report Selection** respectively.

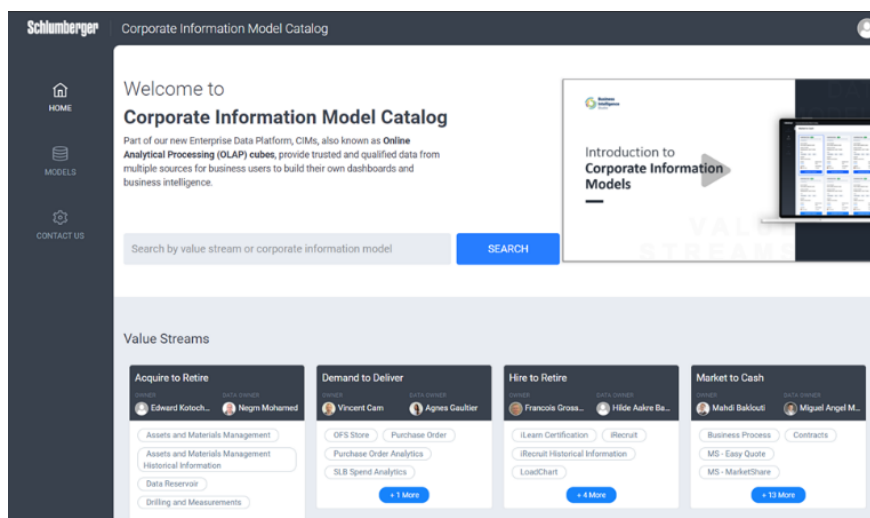


Figure 12: CIM Catalog Landing page

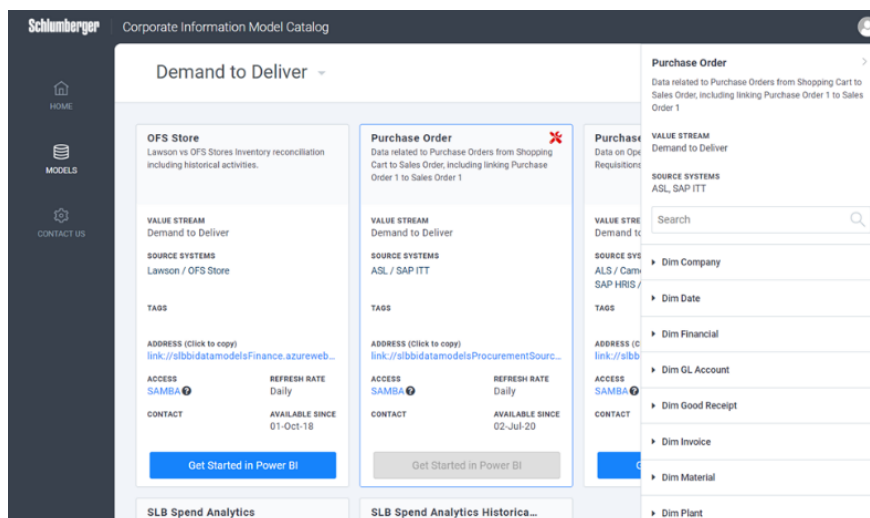


Figure 13: Inside the CIM



Figure 14: Report Selection

Utilization tracking

Utilization tracking for the Web application is one of the key strategy to get the adoption over the period of time. If you're not tracking enough data, you won't see how trends change over time or why some users behave one way versus another. You won't understand user engagement. Ultimately, you just won't have the knowledge to make smart product or marketing decisions. We have enabled and configured Utilization of the CIM Catalog application due to the precise reasons as follows.

- *To find out which storefront bring the most traffic and conversions.
- *Determine where the best visitors are located in the worldwide.
- *Learn what people are searching for on the application.
- *Visualize what people click on the most.
- *Uncover the top content. Which pages keep the visitors on the site the longest.

*To identify the worst performing pages.

*Determine where people abandon due to having multiple clicks

We have used and compared SharePoint activity tracker and Google analytics to enable application tracking.

Microsoft SharePoint Analytics

We get a view into SharePoint activity using Microsoft SharePoint Analytics by looking at the Files and Users views. It is out of the box services available within Microsoft SharePoint Online in Cloud.

This can simply be enabled by the Site admin from the admin center, Go to the Reports > Usage page.

From the Select a report drop-down, select SharePoint > Activity.

The Pages view shows the number of unique pages visited by users as mentioned in the figure 15.

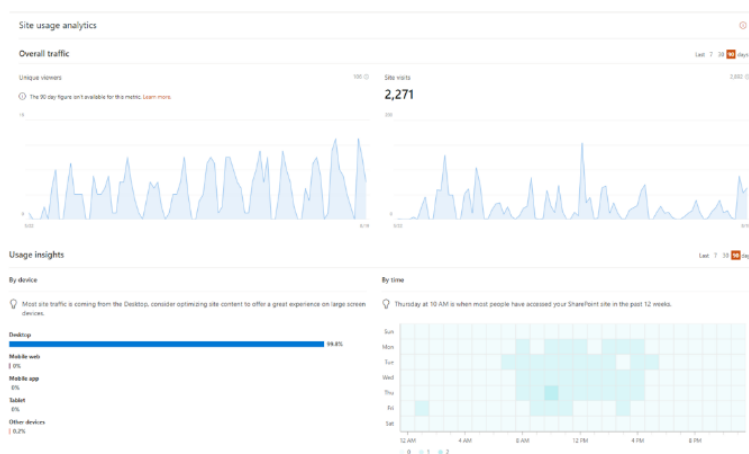


Figure 15: Utilization analysis from SharePoint Online

Google Analytics

1 “Site owners using Google Analytics have the capacity to keep track of views over time, follow site visitors’ move-ments in real time, pinpoint where site visitors leave, view demographics and maps, and understand the search terms that initially brought visitors to the website”¹.

Google Analytics provide detail analysis and is used to track website activity such as session duration, pages per session, bounce rate etc. of individuals using the site, along with the information on the source of the traffic. In order for Google Analytics to function, we were required to put JavaScript tracking code on each page of our application. When a visitor comes to the application, the tracking code runs and sends tracking data to Google’s data collection servers. Hence, we have enabled monitoring for usage analysis over the period of time by adding below line of tracking code in the html as described in Table 3.

```
<!-- Global site tag (gtag.js) - Google Analytics --> <script async src="https://www.googletagmanager.com/gtag/js?id=UA-
```

Table 3: Tracking code for Google Analytic

In our analysis we decided to use Google analytics as it provides detail and comprehensive active user’s trend over the period of time as described in Figure 16 and Figure 17 below.

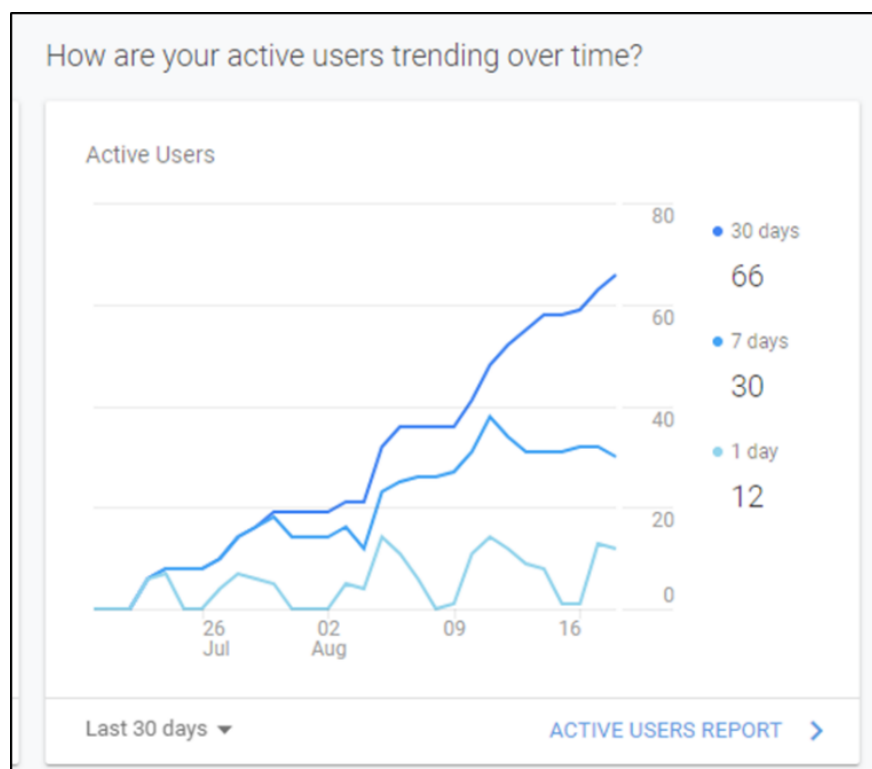


Figure 16: Active user's trend in Google Analytics

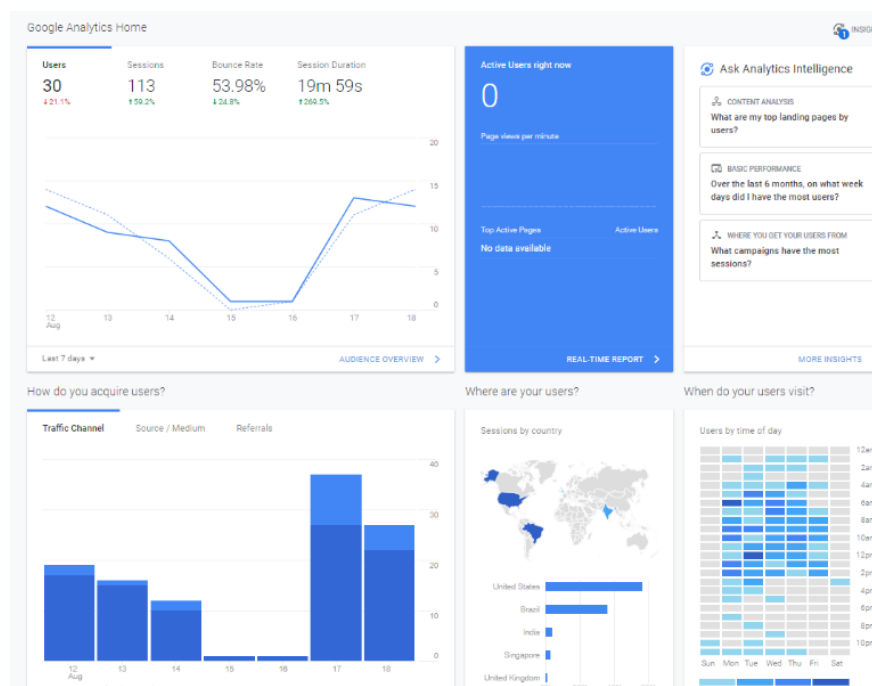


Figure 17: Overall traffic analysis from Google analytics

Security Overview

CIM Catalog is hosted in Microsoft SharePoint Online environment in Cloud, hence it is inherited the security framework what SharePoint offered. SharePoint has the default security groups which are the SharePoint objects that have “users” (Azure Active Directory Users and Azure Active Directory Groups by default) as members and come with their own settings.

There are 5 main layers of Security available in the CIM Catalog environment:

Read: Can view pages and list items and download documents.

Contribute: Can view, add, update, and delete list items and documents.

Edit: Can add, edit and delete lists; can view, add, update and delete list items and documents.

Restricted Edit: Can add, edit and delete specific rows from the list based on the defined role.

Design: Can view, add, update, delete, approve, and customize.

Full Control: Has full control.

There are three security groups. **Members, Owners and Visitors.**

Users required to be added into a security group. Each security group has unique Permission level as described in Figure 18.

Security Groups	Permission Level					
	Read	Contribute	Edit	Restricted Edit	Design	Full Control
Members	✓	✓	✓	✗	✗	✗
Owners	✓	✓	✓	✓	✓	✓
Visitors	✓	✗	✗	✗	✗	✗
site collection Admins	✓	✓	✓	✓	✓	✓
Category owner	✓	✓	✗	✓	✗	✗
Value Stream Owners	✓	✓	✗	✓	✗	✗

Figure 18: CIM Catalog Security Overview

Self-Service for Catalog maintenance

Self-service is the practice of serving oneself and empowered the clients. We have designed and built the application in a way that it required least support and maintenance from the CIM portal Admins and IT Staff. One of the key benefits of self-service to an organization is that they save on resources. Also Customers have better control and privilege on the Content and hence engagement and adoption will increase eventually. We have considered all those factors in the consideration and designed CIM Catalog application in such a manner where Category Owner, Product Owners and Co-Owners can maintain their assigned catalog and CIMs on their own.

Content is the essence of any digital application. The CIM catalog has to be current, consistent, and correct. There are three main components of CIM Catalog that require to be updated and accurate all the time.

- CIM catalog Models and their detail

- Data definitions
- Connection detail and Power BI report

Every Category Owner, Product Owners and Co-Owners can add/edit or update their assigned CIM models information from the administrator console from the front end of the CIM portal application.

Conclusions

In this paper, we demonstrate that how Corporate information model using OLAP approach can serve as self-service strategy to business users from corporate level systems. The proposed method can be implemented to any large organization. The developed scripts has been implemented in real time in the organization. Furthermore, if the identified approach is also suitable for different processes, the scope of the analysis can be extended using a similar approach.

Acknowledgement

The author would like to thank the technical support from the department of IT in the Schlumberger Oil industry to gain the information and data gathered for program development.

Citation

E. J. Petersen and B. M. Martin, "Misuse, play, and dis-use: Technical and professional communication's role in understanding and supporting website owners' en-gagement with Google Analytics," 2015 IEEE Interna-tional Professional Communication Conference (IPCC), Limerick, 2015, pp. 1-5, doi: 10.1109/IPCC.2015.7235786.

Reference for the tools

- Angular coding style guide <https://angular.io/guide/styleguide>
- Angular Workspace Configuration <https://angular.io/guide/workspace-config>
- How to define a highly scalable folder structure for your Angular project- <https://itnext.io/choosing-a-highly-scalable-folder-structure-in-angular-d987de65ec7>
- SharePoint Connection Library - <https://www.npmjs.com/package/@cwi/sharepoint-connector>
- Microsoft SharePoint Document <https://docs.microsoft.com/en-us/sharepoint/introduction>
- Visual Studio Code:Version: 1.48.0 (user setup)

Commit: db40434f562994116e5b21c24015a2e40b2504e6

Date: 2020-08-13T07:50:42.600Z

Electron: 7.3.2

Chrome: 78.0.3904.130

Node.js: 12.8.1

V8: 7.8.279.23-electron.0

OS: Windows_NT x64 10.0.17134

- Visual Studio Code Keyboard<https://www.arungudelli.com/microsoft/visual-studio-code-keyboard-shortcut-cheat-sheet-windows-mac-linux/>
- Microsoft Azure DevOps<https://azure.microsoft.com/en-us/services/devops/>

Appendix 1

Card name

Card name

FARHANA SETHI Global Data & Analytics Business Intelligence - Quality & Governance Manager with Schlumberger Oilfield, Texas, Houston since 2006. Bachelors in Engineering - Electrical & Computers science. Microsoft Certified Architect (Azure Solution Architect. Professional .Certification in Reservoir Geomechanics from Stanford University - United States, California. Professional Certification in Internet of Things from Stanford University - United States, California. Professional certification in Data Science: Machine Learning from Harvard University - United States, Massachusetts