ARCHITECTURE
CROP PRODUCTION DATA ANALYSIS

Revision Number: 1.0

Last date of Revision: 25/09/2021

Submitted by:

Dattatreya Thunuguntla

Contents:

1. INTRODUCTION:

1.1 WHAT IS ARCHITECTURE DESIGN DOCUMENT

Any software needs the architectural design to represents the design of software. IEEE defines architectural design as "the process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system." The software that is built for computer-based systems can exhibit one of these many architectures.

Each style will describe a system category that consists of:

• A set of components (Ex: a database, computational modules) that will perform a function required by the system.

• The set of connectors will help in coordination, communication, and cooperation between the components.

• Conditions that how components can be integrated to form the system.

• Semantic models that help the designer to understand the overall properties of the system.

1.2 SCOPE

Architecture Design Document (ADD) is an architecture design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the design principles may be defined during requirement analysis and then refined during architectural design work.
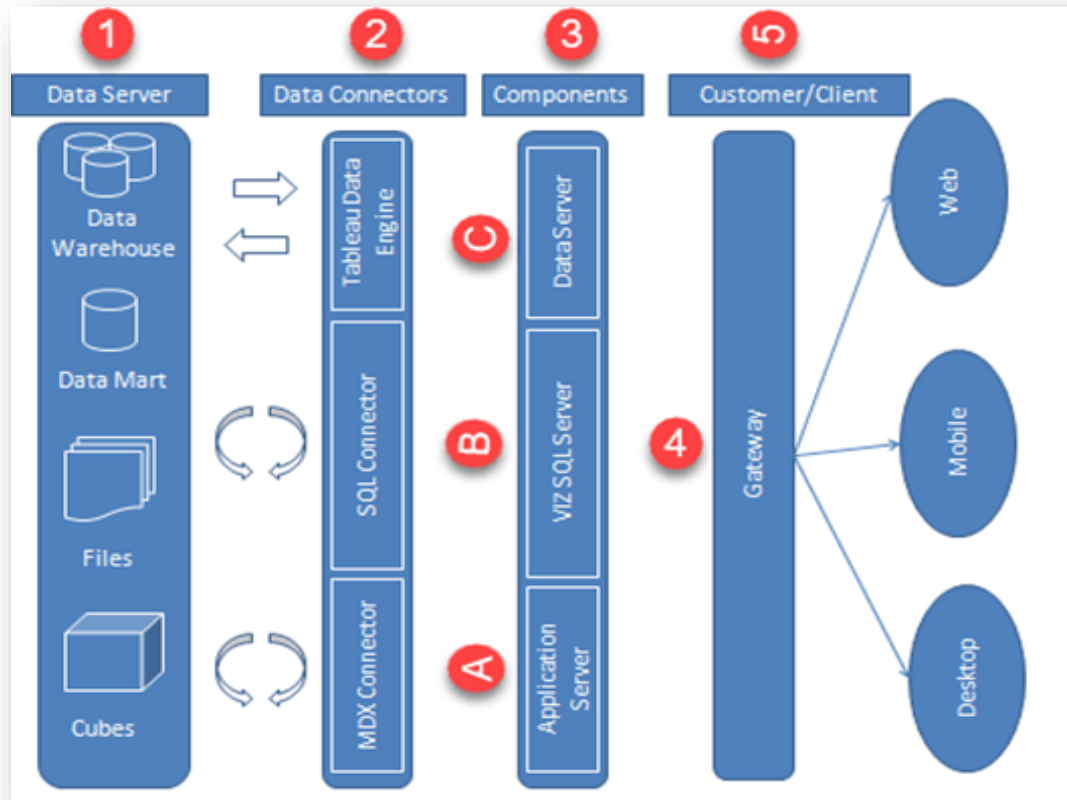
2. ARCHITECTURE:
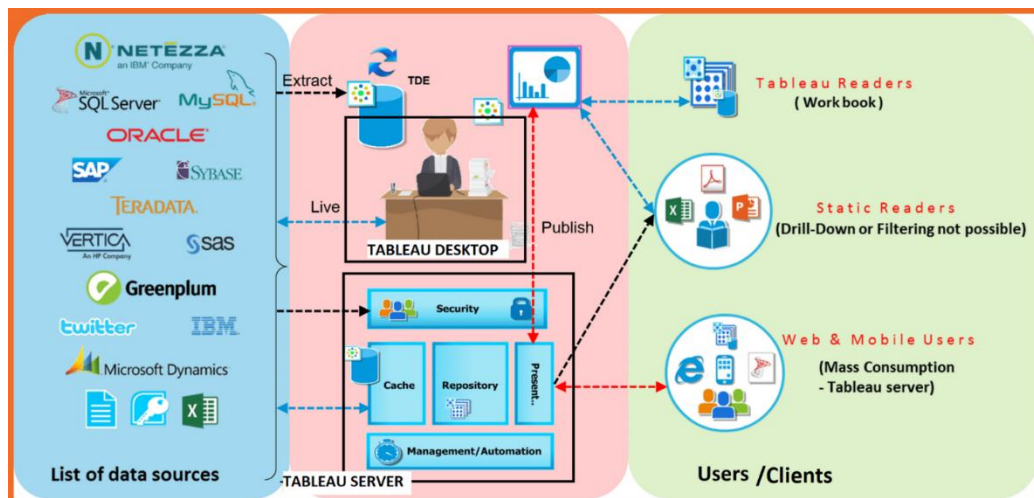

TABLEAU DESKTOP ARCHITECTURE:

Tableau Server is designed to connect many data tiers. It can connect clients from Mobile, Web, and Desktop. Tableau Desktop is a powerful data visualization tool. It is very secure and highly available. It can run on both the physical machines and virtual machines. It is a multi-process, multi-user, and multi-threaded system. Providing such powerful features requires unique architecture. The different layers used in Tableau server are given in the above architecture diagram.

Let's get to know about the different component of the Tableau architecture:

2.1 Data server: - The primary component of Tableau Architecture is the Data sources which can connect to it. Tableau can connect with multiple data sources. It can blend the data from various data sources. It can connect to an excel file, database, and a web application at the same time. It can also make the relationship between different types of data sources.

2.2 Data connector: - The Data Connectors provide an interface to connect external data sources with the Tableau Data Server. Tableau has in-built SQL/ODBC connector. This ODBC Connector can be connected with any databases without using their native connector. Tableau desktop has an option to select both extract and live data. On the use's basis, one can be easily switched between live and extracted data.

- Real-time data or live connection: Tableau can be connected with real data by linking to the external database directly. It uses the infrastructure existing database by sending dynamic multidimensional expressions (MDX) and SQL statements. This feature can be used as a linking between the live data and Tableau rather than importing the data. It makes optimized and a fast database system. Mostly in other enterprises, the size of the database is large, and it is updated periodically. In these cases, Tableau works as a front-end visualization tool by connecting with the live data.
- Extracted or in-memory data: Tableau is an option to extract the data from external data sources. We make a local copy in the form of Tableau extract file. It can remove millions of records in the Tableau data engine with a single click. Tableau's data engine uses storage such as ROM, RAM, and cache memory to process and store data. Using filters, Tableau can extract a few records from a large dataset. This improves performance, especially when we are working on massive datasets. Extracted data allows the users to visualize the data offline, without connecting to the data source.

3. Components of Tableau server: Different types of components of the Tableau server are:

1. Application server
2. VizQL server
3. Data server

3.1 Application server: The application server is used to provide the authorizations and authentications. It handles the permission and administration for mobile and web interfaces. It gives a guarantee of security by recording each session id on Tableau Server. The administrator is configuring the default timeout of the session in the server.

3.2 VizQL server: VizQL server is used to convert the queries from the data source into visualizations. Once the client request is forwarded to the VizQL process, it sends the query directly to the data source retrieves information in the form of images. This visualization or image is presented for the users. Tableau server creates a cache of visualization to reduce the load time. The cache can be shared between many users who have permission to view the visualization.

3.3 Data server: Data server is used to store and manage the data from external data sources. It is a central data management system. It provides data security, metadata management, data connection, driver requirements, and data storage. It stores the related details of data set like calculated fields, metadata, groups, sets, and parameters. The data source can extract the data as well as make live connections with external data sources.

3.4 Gateway: The gateway directed the requests from users to Tableau components. When the client sends a request, it is forwarded to the external load balancer for processing. The gateway works as a distributor of processes to different components. In case of absence of external load balancer, the gateway also works as a load balancer. For single server configuration, one gateway or primary server manages all the processes. For multiple server configurations, one physical system works as a primary server, and others are used as worker servers. Only one machine is used as a primary server in Tableau Server environment.

3.5 Clients: The visualizations and dashboards in Tableau server can be edited and viewed using different clients. Clients are a web browser, mobile applications, and Tableau Desktop.

Web Browser: Web browsers like Google Chrome, Safari, and Firefox support the Tableau server. The visualization and contents in the dashboard can be edited by using these web browsers.

Mobile Application: The dashboard from the server can be interactively visualized using mobile application and browser. It is used to edit and view the contents in the workbook.

Tableau Desktop: Tableau desktop is a business analytics tool. It is used to view, create, and publish the dashboard in Tableau server. Users can access the various data source and build visualization in Tableau desktop.

4. Deployment:

Prioritizing data and analytics couldn't come at a better time. Your company, no matter what size, is already collecting data and most likely analyzing just a portion of it to solve business problems, gain competitive advantages, and drive enterprise transformation. With the explosive growth of enterprise data, database technologies, and the high demand for analytical skills, today's most effective IT organizations have shifted their focus to enabling self-service by deploying and operating Tableau at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

Tableau prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Tableau Server and Tableau Online leverage your existing technology investments and integrate into your IT infrastructure to provide a self-service, modern analytics platform for your users. With on-premises, cloud, and hosted options, there is a version of Tableau to match your requirements. Below is a comparison of the three types:

1. Tableau Server - On Premises:

   - Full control of hardware and software.
   - Infrastructure and data remain behind your firewall.
   - Need dedicated administrators to manage hardware and software.
   - Additional infrastructure needed to access off-network (mobile, external)

2. Tableau Server - Public Cloud (IaaS):

   - Full control of software on managed hardware

- Puts infrastructure in same place as data (for migration to cloud)
- Flexibility to spin up/down hardware as needed
- Need dedicated administrators to manage software
- Additional infrastructure needed to access off-network (mobile, external)

3. Tableau Online (SaaS):

- Fully hosted solution (hardware, software upgrades)
- Fast to deploy
- Easy for external audience to access
- Single-site in multi-tenant environment
- Cubes are not supported
- No guest account access
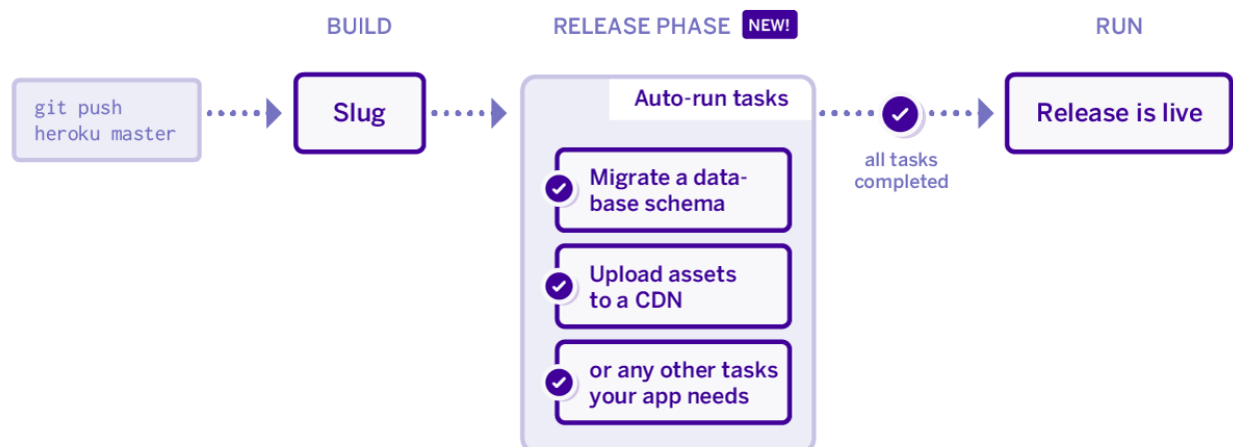
5.Deployment in Heroku:

The deployment in the Heroku consists of 3 types like

1. Heroku Git
2. Personal GitHub
3. Container Registry

In this project we had used Personal GitHub Deployment method. You can check the source files in the personal GitHub repository in the following link:
https://github.com/dattu-98/crops_india

- The GitHub repository consists of Procfile, app.py, requirements.txt, runtime.txt, setup.sh
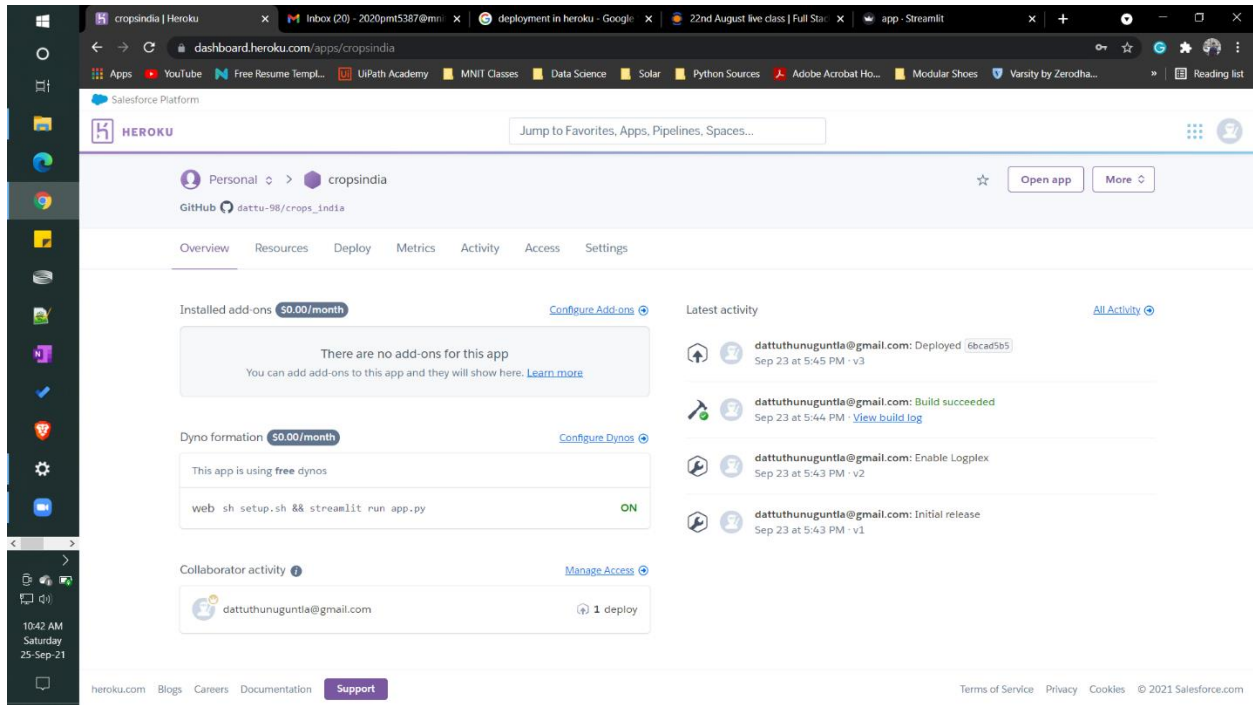
After deployment in Heroku the Webapp looks in the following fashion:

Heroku Webapp Link: https://cropsindia.herokuapp.com/