High-Level Design (HLD)
Crop Production Analysis in India

Revision Number: 1.0
Last Date of Revision: 31/01/2023

Madhav Seth

Contents

Document Version Control

Date Issued	Version	Description	Author
31st January 2023	1.0	First Version of Complete HLD	Madhav Seth

Abstract

The Agriculture business domain, as a vital part of the overall supply chain, is expected to highly evolved in the upcoming years via the developments, which are taking place on the side of the Future Internet. This paper presents a novel Business-to-Business collaboration platform from the agri-food sector perspective, which aims to facilitate the collaboration of numerous stakeholders belonging to associated business domains, in an effective and flexible manner.

This dataset provides a huge amount of information on crop production in India ranging from several years. Based on the Information the ultimate goal would be to predict crop production and find important insights highlighting key indicators and metrics that influence crop production.

1 Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

2. General Description

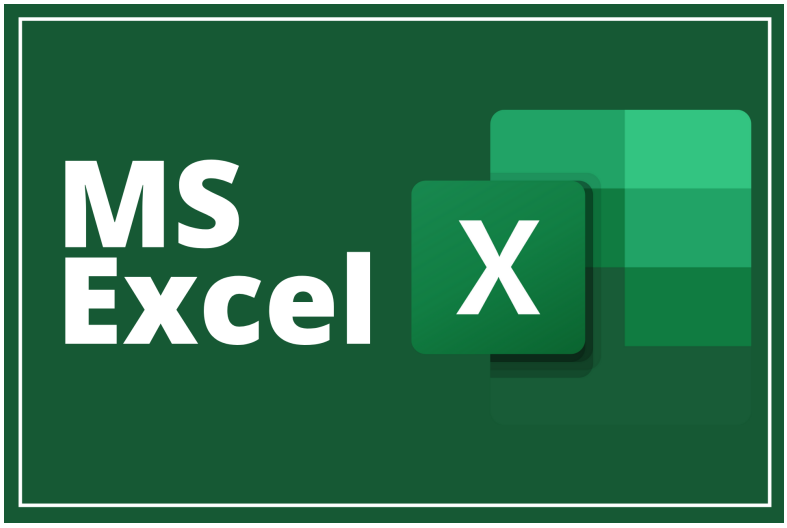
2.1 Product Perspective and Problem Statement

Agriculture is critical to the Indian economy. Agriculture supports more than 70% of rural households. Agriculture is an important sector of the Indian economy, accounting for approximately 17% of the total GDP and employing more than 60% of the workforce.
In this project, I analyzed crop production data in India to gain insights into crop production in various Indian states and districts. The project's goal is to gain insights and use data visualization techniques to learn more about the customers. The goal of this project is to use various Python libraries to gain insights and a visual understanding of the data.

2.2 Tools Used

2.2.1. MS Excel

MS Excel is spreadsheet software used to organize, analyze, and store data. It allows one to create tables and charts and performs calculations and in-depth analysis.



2.2.2. Jupyter Notebook

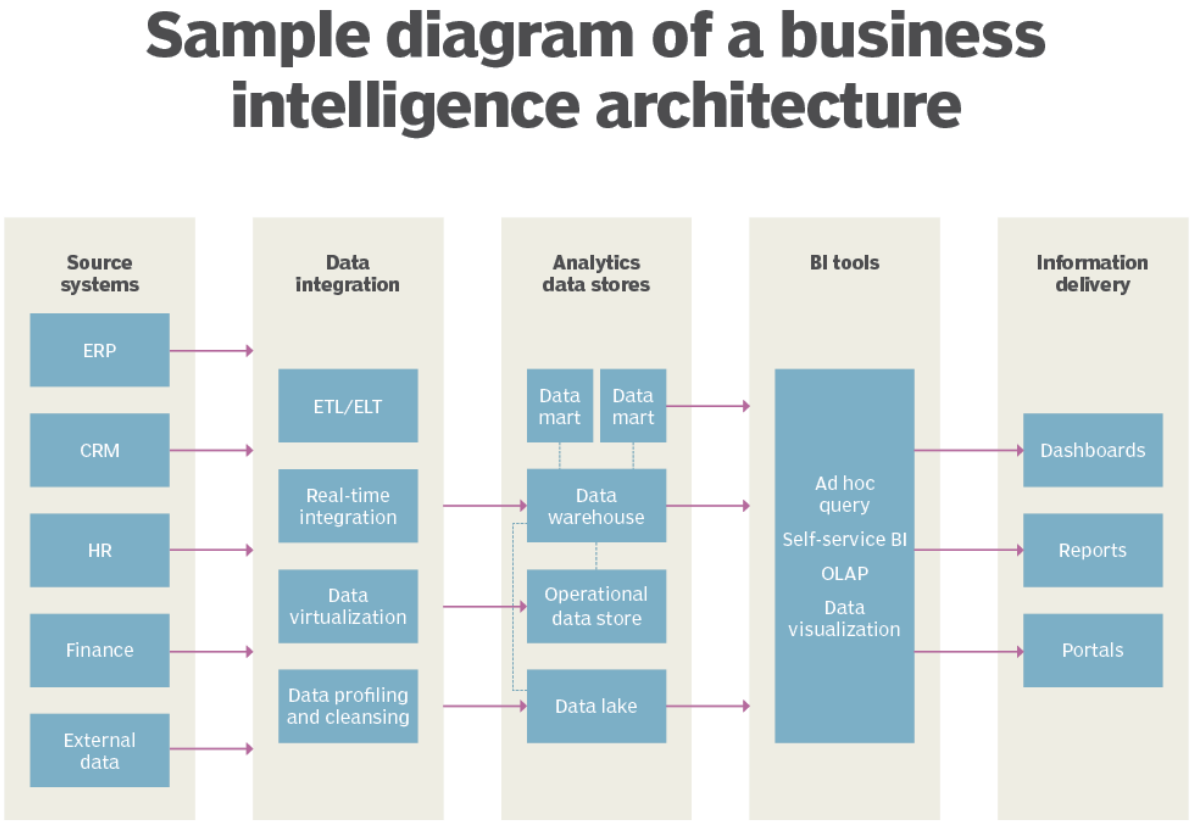
I uploaded data in a Jupyter notebook, imported necessary libraries like NumPy and Pandas, loaded the data into a DataFrame, cleaned the data by removing missing values, transformed data by normalizing, aggregating, and, exported the cleaned and transformed data to a new CSV file. Data is now ready for analysis and get better insights.

2.2.3. Tableau Public

It is a free data visualization tool that allows users to connect to various data sources, create interactive dashboards, and share their visualizations with the public. It is a powerful tool to communicate insights and trends from data in a visually compelling way. It's easy to use and has a wide range of options to customize and format the visualizations. It allows us to connect and analyze data from various sources and create interactive and shareable dashboards.

3. Design details

3.1 Functional Architecture



A business intelligence architecture is the framework that an organization uses to run business intelligence and analytics applications. It includes the IT systems and software tools used to collect, integrate, store, and analyze BI data before presenting it to corporate executives and other business users as information on business operations and trends.

The underlying BI architecture is a critical component in the successful implementation of a business intelligence program that uses data analysis and reporting to assist an organization in tracking business performance, optimizing business processes, identifying new revenue opportunities, improving strategic planning, and making more informed decisions overall.

3.1.1 Importance of a BI architecture

A Business Intelligence architecture articulates the technology standards, data management, analytics practices, and specific platforms and tools that will be deployed to support an organization's BI efforts. It serves as a technology blueprint for collecting, organizing, and managing business intelligence data before making it available for analysis, data visualization, and reporting. A strong BI architecture also includes policies that govern how the technology components are used.

Putting such a framework in place enables a BI team to work in a coordinated and disciplined manner to build an enterprise BI program that meets the data analytics needs of its organization. The BI architecture also assists BI and data managers in developing an efficient process for handling and managing data as it enters the environment.

3.1.1.2 The Advantages of Business Intelligence

The following are the advantages of business intelligence:

- Business intelligence is a faster and more accurate process of reporting critical information and enables better and more efficient decision-making.
- Business intelligence delivers timely information to improve customer relationship management.
- Business intelligence enables organizations to assess their readiness to meet new business challenges.
- Business intelligence facilitates the use of best practices and identifies all hidden costs.
- Business intelligence usage can be optimized by identifying key projects on which the company wishes to concentrate its efforts.
- Business intelligence governance refers to the process of highlighting key projects.

3.2 Optimization

3.2.1. Your data strategy drives performance

• Minimize the number of fields
• Minimize the number of records
• Optimize extracts to speed up future queries by materializing calculations, removing columns, and the use of accelerated views

3.2.2. Reduce the marks (data points) in your view

• Practice guided analytics. There's no need to fit everything you plan to show in a single view. Compile related views and connect them with action filters to travel from overview
to highly-granular views at the speed of thought.
• Remove unneeded dimensions from the detail shelf.
• Explore. Try displaying your data in different types of views.

3.2.3. Limit your filters by number and type

• Reduce the number of filters in use. Excessive filters on a view will create a more complex query, which takes longer to return results. Double-check your filters and remove any that aren't necessary.
• Use an include filter. Exclude filters load the entire domain of a dimension, while include filters do not. An include filter runs much faster than an exclude filter, especially for dimensions with many members.
• Use a continuous date filter. Continuous date filters (relative and range-of-date filters) can take advantage of the indexing properties in your database and are faster than discrete date filters.
• Use Boolean or numeric filters. Computers process integers and Booleans (t/f) much faster than strings.
• Use parameters and action filters. These reduce the query load (and work across data sources).

3.2.4. Optimize and materialize your calculations

• Perform calculations in the database
• Reduce the number of nested calculations.
• Reduce the granularity of LOD or table calculations in the view. The more granular the calculation, the longer it takes.

o LODs - Look at the number of unique dimension members in the calculation.
o Table Calculations - the more marks in the view, the longer it will take to calculate.

• Where possible, use MIN or MAX instead of AVG. AVG requires more processing than MIN or MAX. Often rows will be duplicated and display the same result with MIN, MAX, or AVG.
• Make groups with calculations. Like including filters, calculated groups load only named members of the domain, whereas Tableau's group function loads the entire domain.
• Use Booleans or numeric calculations instead of string calculations. Computers can process integers and Booleans (t/f) much faster than strings. Boolean>Int>Float>Date>DateTime>String.

4 KPIs

KPI (Key Performance Indicator) is a metric used in data analysis to measure the performance and success of an organization or specific business process. As and when the system starts to capture the historical/periodic data for a user, the dashboards will be included to display charts over time with progress on various indicators or factors

4.1 KPIs (Key Performance Indicators)

Key indicators displaying a summary of the Crop Production area and the production quantity information based on various parameters –

- Crops Producing States of India.
- Crop Producing districts of India.
- The year of crop production.
- Season under which crops fall.
- Names of the crop produced in India.
- Area cultivated under each crop (in Hectares).
- Total production quantity of a crop (in Tons).

5 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 them 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:

5.1.1. 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)

5.1.2. Tableau Server - Public Cloud (IaaS)

• Full control of software on managed hardware
• Puts infrastructure in the same place as data (for migration to the cloud)

- Flexibility to spin up/down hardware as needed
- Need dedicated administrators to manage software
- Additional infrastructure needed to access off-network (mobile, external)

5.1.3 Tableau Online (SaaS)

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

Depending on your organizational roles and responsibilities, Tableau Server should be installed by a systems administrator and the designated Tableau Server Administrator in coordination with the appropriate IT roles. For Tableau Online, you will integrate with your existing technology and configure the site settings. The Data & Analytics Survey, completed by business teams, identifies and prioritizes data use cases, audience size, and users. You will use the information collected in both surveys to plan your deployment strategy, including sizing, installation, and configuration of your Tableau Server or integration and configuration of Tableau Online. In addition to installing Tableau Server or configuring Tableau Online, administrators will also need to plan for the client software installation of Tableau Prep Builder, Tableau Desktop, Tableau Mobile, and Tableau Bridge for Tableau Online where applicable.