CROP PRODUCTION ANALYSIS

High Level Design

Domain: Business Intelligence

Date:07/08/2024

Contents

Abstract

Knowing about cropping patterns or the cropping behavior gives you an extra advantage over the others. We are way ahead of our traditional times and we need to leave all the traditional methodologies while going ahead.

Agriculture is a sector which has changed a lot, it is a dynamic sector. New varieties, diseases, technologies, trends are some of the changes we generally see happening. Now, we are more inclined towards precision agriculture rather than the traditional methods. Making agriculture a pro table business needs to analysis areas, crops, production and future trends.

Introduction

What is High-Level Design Document?

The goal of this HLD or a high-level design 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 design aspects and define them in detail
- Describe all user interfaces being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and architecture of the project
- List and describe the non-functional attributes such as security, reliability, maintainability, portability, reusability, application compatibility. resource utilization, serviceability

Scope

The HLD documentation presents the structure of the system, such as 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.

General Description

Problem Statement

The agriculture business domain, as a vital part of the overall supply chain, is expected to highly evolve 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 the crop production.

Make views and dashboards first.

Make a story out of it.

Proposed Solution

The proposed solution leverages Exploratory Data Analysis (EDA) techniques to uncover key insights in crop production data. By using Power BI for interactive dashboards, the solution visualizes trends in crop production over years, regional and seasonal variations, and crop-specific performance.

Dataset Details

The dataset contains information about crop production in different districts across various states of India. Here is a summary of the dataset:

General Information

- **Number of Entries:** 246,091
- **Number of Columns:** 7

Column Details

1. **State_Name:** Name of the state (33 unique states).
2. **District_Name:** Name of the district (646 unique districts).
3. **Crop_Year:** Year of crop production (from 1997 to 2015).
4. **Season:** Season of the crop (6 unique seasons: Kharif, Rabi, Whole Year, Summer, Winter, Autumn).
5. **Crop:** Type of crop (124 unique crops).
6. **Area:** Area in which the crop is produced (in hectares).
7. **Production:** Production of the crop (in tonnes).

Statistical Summary

- **State_Name:** Most frequent state is Uttar Pradesh (33,306 entries).
- **District_Name:** Most frequent district is Bijapur (945 entries).
- **Crop_Year:** Average year is approximately 2005.64.
- **Season:** Most frequent season is Kharif (95,951 entries).
- **Crop:** Most frequent crop is Rice (15,104 entries).
- **Area:**
  - Mean: 12,002.82 hectares
  - Standard Deviation: 50,523.40 hectares
  - Minimum: 0.04 hectares
  - Maximum: 8,580,100 hectares

- **Production:**
  - Mean: 582,503.40 tonnes
  - Standard Deviation: 17,065,810 tonnes
  - Minimum: 0 tonnes
  - Maximum: 1,250,800,000 tonnes

Tools used

Business Intelligence tools and libraries works such as NumPy, Pandas, Power BI , are used to build the whole framework.

1. Python: Python is a popular programming language for data science. It offers a wide range of libraries and frameworks, such as NumPy, pandas, etc..
2. Jupyter Notebook: Jupyter Notebook is an interactive development environment that allows you to combine code, visualizations, and explanatory text in a single document. It's commonly used for exploratory data analysis and model prototyping.
3. Matplotlib: Matplotlib is a widely-used data visualization library in Python. It provides a flexible and comprehensive set of functions to create various types of static, interactive, and animated plots.
4. Seaborn: Seaborn is a Python data visualization library built on top of Matplotlib. It provides a high-level interface for creating aesthetically pleasing statistical graphics.
5. Microsoft Power BI: Power BI is an interactive data visualization software product developed by Microsoft with a primary focus on business intelligence.It is part of the Microsoft Power Platform. Power BI is a collection of software services, apps, and connectors

that work together to turn various sources of data into static and interactive data visualizations.

Hardware Requirements

- A computer system capable of running python. The chosen IDE (such as Visual Studio Code). The system should meet the minimum system requirements for the selected software.
- Minimum 1.10 GHz processor or equivalent.
- Between 1-2 GB of free storage
- Minimum 512 MB of RAM
- 3 GB of hard-disk space

Constraints

The System should be user-friendly, the user should get all proper messages while using the web app. He/she also should get a proper error message if he/she has done something wrong on the web-app page. All the errors and results should be delivered in the easiest possible way and all the buttons are going to insert on the webpage should be labeled properly, so the user did not get confused to use the system.

Design Details

Functional Architecture

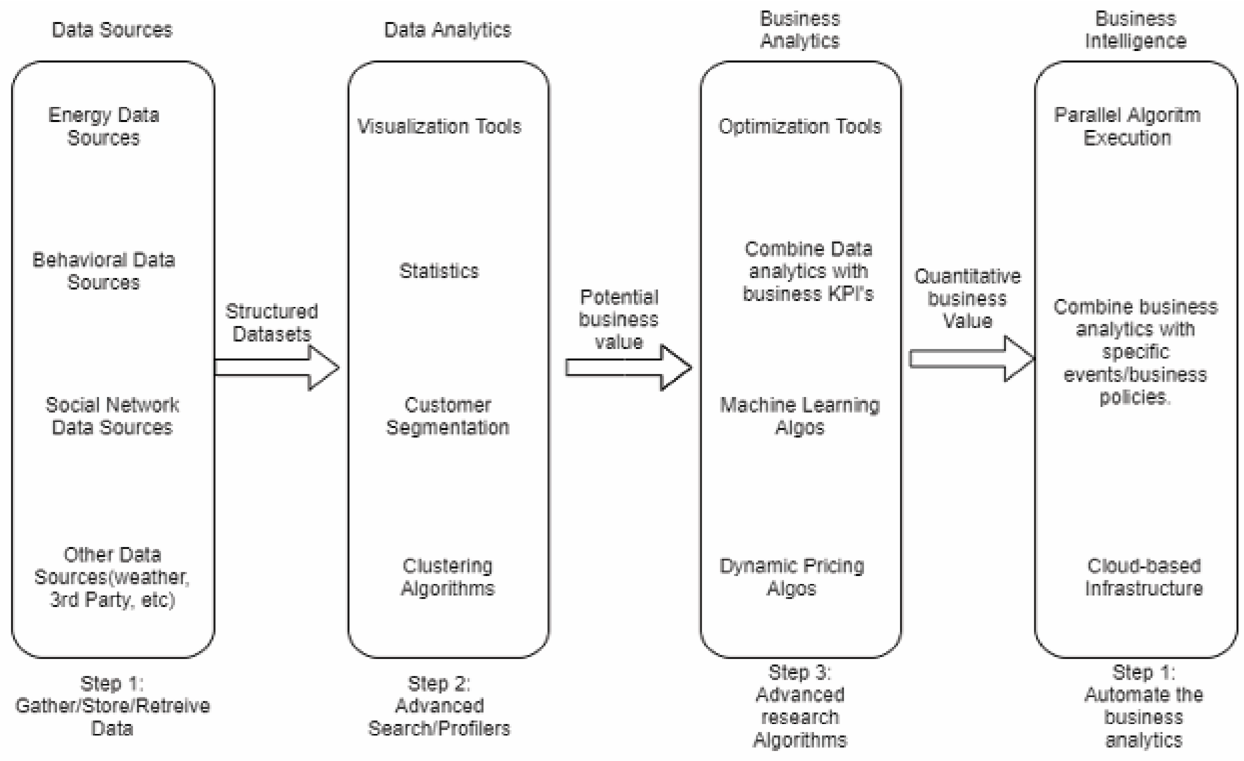
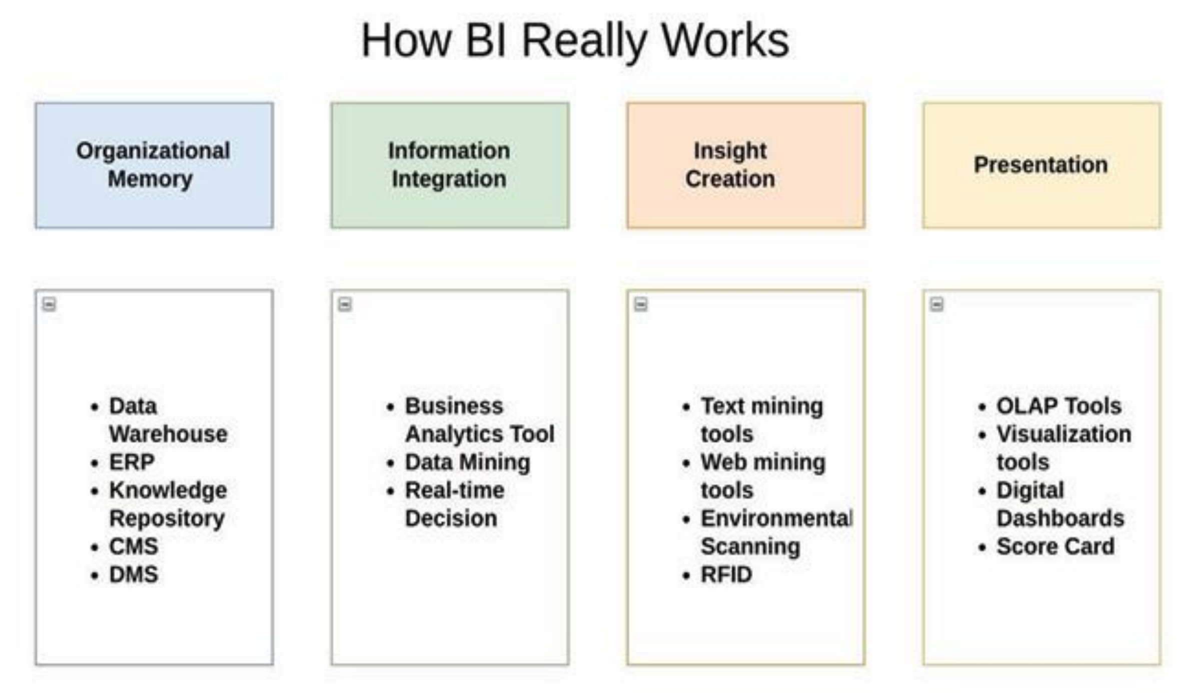


Figure 1: Functional Architecture of Business Intelligence



Performance

Optimization

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.

Reduce the marks (data points) in your view

- Practice guided analytics. There's no need to t 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.

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).

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.
  - LODs - Look at the number of unique dimension members in the calculation.
  - 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.

KPIs

Dashboards will be implemented to display and indicate certain KPIs.

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.

KPIs (Key Performance Indicators)

Key indicators displaying a summary of the Housing Price and its relationship with different metrics

1. Trend and pattern identification.
2. Comparing crop yields.
3. Identifying areas with growth potential.
4. Regional production influence.
5. How production does not depend on area.

Conclusion

The Power BI dashboard for crop production in India provides a comprehensive view of agricultural data, facilitating informed decision-making for various stakeholders in the agri-food sector. Key insights derived from the data include:

- **Yearly Trends**: Analysis of crop production over the years reveals trends and helps in understanding the impact of policies and climate changes.
- **Regional Analysis**: Comparing production across states and districts highlights the regions with high and low productivity, enabling targeted interventions.
- **Seasonal Trends**: Understanding how crop production varies with seasons helps in planning and resource allocation.
- **Crop-wise Analysis**: Identifying the most produced crops and their distribution aids in strategic planning for crop diversification and resource management.

The deployment of this dashboard ensures that stakeholders have access to real-time, actionable insights, driving efficiency and innovation in the agriculture business domain. With regular updates and continuous improvements, the dashboard will remain a valuable tool for decision-making and collaboration in the agri-food sector.