Team 2’s Project

**Puglia Winery Case**

**A logo of a university

Description automatically generated**

ISDS 556, Summer 2023, CSUF

Instructor: Dr. Peng Liu

Sairaj Jadhav

Pragati Ingole

Mine Salman

Vishnu Menon

Ray Lien

**Table of contents**

[1. Overview 2](#_Toc142042362)

[2. Functionalities and features of the BI system: 3](#_Toc142042363)

[3. Data Models 6](#_Toc142042364)

[4. Data Analysis 8](#_Toc142042365)

[5. Extraction, Transformation, and Loading 9](#_Toc142042366)

[6. Recommended Architecture: 11](#_Toc142042367)

[7. Summary, conclusion & recommendation for action 13](#_Toc142042368)

[References 16](#_Toc142042369)

# 

# 1. Overview

The Puglia Winery, situated in Temecula, California, is a medium-sized boutique winery specializing in producing hot climate wines. Their three renowned varieties include pinot noir, merlot (red), and pinot grigio (white). Their distribution covers both local areas within California, as well as interstate in the US and internationally in the United Kingdom. In recent years, the winery has experienced substantial growth. However, their decision-making process primarily relies on "gut feel" due to their use of a traditional database. To address this limitation and manage their growth more effectively, the managing director, Jack Gillespie, believes that implementing data warehousing and business intelligence solutions is essential. Our goal is to provide a data warehousing solution to enable Jack to make better-informed decisions based on real-time data.

The winery has identified three main business challenges that need to be tackled. Firstly, they need to determine which products are the most profitable. Secondly, they seek to identify key customers to better cater to their preferences and needs. Lastly, they aim to pinpoint the most lucrative markets to further expand their reach. To address these challenges, the winery has shared data from its production and sales systems.

**Gaps**

1. The winery requires a real-time dashboard to obtain valuable insights and enhance decision-making processes to foster business growth.

2. The lack of integration between Production and Sales systems with merchants creates challenges in analyzing data for informed decision-making.

3. Conducting time-based analysis becomes complicated due to the absence of a centralized repository for storing historical data, making access difficult.

**Organization readiness**

1.The management is prepared and willing to make investments in implementing the solution.

2. Mr. Jack Gillespie acknowledges the need for a data warehousing solution to uphold the

product's exclusivity.

3. The winery may need to recruit skilled personnel to provide support for the proposed warehouse solution

**High-level enterprise data warehouse bus matrix**

The columns in the matrix represent the key organizations or workgroups engaged in the requirements process. These groups are interested in the metrics linked to the business process rows. On the other hand, the rows indicate the business processes that will be given priority in the design of the data warehouse system.

**Exhibit 1 - Organization/Workgroup**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Business Process/Event** | Production (Tommy Siragusa) | Marketing & Sales (Stacie Giano) | Finance | Customer |
| Market transactions |  | X |  |  |
| Line unit production | X |  | X |  |
| Product transactions |  | X |  | X |
| Wine merchant transactions |  | X |  | X |
| Finished goods inventory | X |  |  | X |

# 

# 2. Functionalities and features of the BI system:

**1. Product Profitability Analysis:**

**Functionality:**

**- Data Integration:** The BI system should extract relevant data from the production system's Oracle database and integrate it into the data warehouse. This process involves data transformation and cleansing to ensure consistency and accuracy.

**- Sales and Revenue Analysis:** The system should perform calculations to determine the total unit sales, dollar sales, cost, and margin for each product (e.g., Bellarine Pinot Grigio, Downunder Pinot Noir) and wine type (red and white).

**- Time Period Analysis:** The BI system should allow users to analyze product profitability over different time periods, such as yearly, seasonally (e.g., Autumn, Winter, Spring, Summer), and monthly.

**Features:**

**- Data Integration:** From the Oracle database of the production system, the BI system should extract relevant data and integrate it into the data warehouse. For consistency and accuracy, this procedure incorporates data transformation and cleansing.

**- Sales and Revenue Analysis:** The system must compute the total units sold, dollars sold, costs, and margins for each product (for instance, Bellarine Pinot Grigio and Downunder Pinot Noir) and wine variety (red and white).

**- Time Period Analysis:** The BI system should enable users to examine the profitability of a product over a variety of time frames, including annually, seasonally (for example, during the Autumn, Winter, Spring, and Summer), and monthly.

**2. Key Customer Identification:**

**Functionality:**

**- Data Integration:** The BI system should extract relevant details from the SQL-Server database of the merchant sales system and integrate it into the data warehouse. Customer data, sales orders, and sales order line items fall under this category.

**- Customer Segmentation:** The system should divide customers into groups depending on how they make purchases and reveal information about the most valued and devoted clients. Users should be able to assess customer performance over time, considering annual and seasonal sales data.

**Features:**

**- Customer Segmentation Reports:** Based on sales and revenue data, the BI system should produce reports that categorize clients into various segments (such as high-value customers, repeat customers, and new customers).

**- Customer Loyalty Analysis:** The system should track repeat buyers for Puglia Winery by tracking customer retention and loyalty.

**- Geographical Insights:** The BI system could offer visuals that highlight the markets or regions that are most responsible for client sales.

**3. Market Profitability Assessment:**

**Functionality: -** Data Integration: The BI system should gather information from the merchant sales system, concentrating on sales information for various markets, including California, the rest of the US, and internationally.

**- Market Comparison:** The system should enable users to assess the effectiveness of sales and profitability in various markets.

**- Time Period Analysis:** Users should be able to look back on a year's worth of market performance monthly.

**Features:**

**- Market Performance Dashboards:** Comparative dashboards showing sales and revenue information for each market should be shown by the BI system.

**- Market Trend Analysis:** The system should make it possible for users to spot recurring patterns and trends in market sales.

Data visualizations that display prospective growth prospects in various markets could be a feature of the BI system.

**Additional Features of the BI System:**

**1. Data Visualization:** To aid users in intuitively understanding and interpreting data, the system should include a variety of data visualization options, such as bar charts, line graphs, pie charts, and heatmaps.

**2. Forecasting and Predictive Analytics:** To predict future sales patterns, consumer behavior, and market demand, the BI system should include forecasting models and predictive analytics.

**3. Data Drill-Down:** To get deeper understanding and provide ad hoc inquiries during analysis, the system should enable users to drill down to certain data points.

**4. Data Security and Access Control:** To guarantee that confidential corporate information is only available to authorized individuals, strong data security measures should be put in place.

**5. Mobile Compatibility:** By making the BI system mobile-friendly, users will be able to get crucial insights while on the go.

**6. Collaboration and Sharing:** The system should encourage teamwork by enabling members to exchange information with one another, such as reports and insights.

**7. Alerts and Notifications:** To inform stakeholders of important changes or trends in key indicators, the BI system could automatically produce alerts and notifications.

**8. Ad-hoc Querying:** Users should be able to conduct ad-hoc queries on the BI system to freely explore the data and acquire insights outside of the scope of normal reports.

The BI system for Puglia Winery will enable stakeholders to make data-driven decisions, optimize product offers, recognize key clients, and focus on profitable areas for long-term business growth by including these functionalities and features.

# 3. Data Models

In OLAP (online analytical processing), we will focus on high understandability and flexible data analysis. While redundancy or storage matters much on the OLTP side of the system, those are not the case for OLAP, but the efficiency of reporting.

For the data aggregation aspect of data modeling, the main concern for reporting is efficiency due to its time saving aspect of aggregating data. When we run the report, we already have it recalculated. Although, that makes reporting so much faster, in our model, we do not need to add such attribute in the fact table based on provided report requirements. Together these principles and business requirements lead us to "star schema" consisting of a single central fact table and several dimension tables.See ER diagram below:

**A screenshot of a diagram

Description automatically generated**

Sales\_Order\_Line\_Fact (fact table) uses foreign key links to dimension tables (MarketID,ProductionDateID, etc.) and does not contain a composite primary key. In our model, dimension tables are customer, product, vineyard, date, and market, with descriptive information for all related fields in the fact table records. It will be filled and grew quickly with large volume of sales and its details data. Snow-flaking is not necessary since the schema is relatively straightforward, and there is no hierarchical relationships in dimension tables that would benefit from normalization.

**Degenerate Dimensions**

By definition, degenerate dimensions are attributes (columns) that are not stored in separate tables but are directly attributes of the fact table. So, in our model, the “line” attribute of the fact table can be treated as a degenerate dimension.

**Slowly Changing Dimensions (SCD)**

SCD change over time as its name speaks itself. However, this update/change needs to be handled in the model, the way of managing these historical data lead slowly changing dimensions diffrentiate different types which are type1,type2 and type3. In type1, no track of historical data in the model on the contrary, type2 keeps existing record as well in case of any update. Customer\_dims and Product\_dims can be treated as type2. Whenever there is an update to a customer’s columns, a new record will be stored in the Customer\_dims. This way, we will be able to monitor the changes over time.

**Fast-Changing Dimensions**

These dimensions are called when frequent changes happens.In our model, the price of a product may change frequently especially in a fluctuated economy. There are two approach to address this, either storing a new records for each price change into Product\_dims table or keeping a start and end date for each price period to keep track of historical price data. We go with the first option.

# 4. Data Analysis

The data available for analysis in the Puglia Winery case study encompasses various aspects of the winery's operations, including product information, customer details, sales transactions, production data, and market information.

**Primary and Secondary Data Sources:**

The primary data sources are the operational systems, namely the production and merchant sales systems. These systems capture data from the winery's activities, making them primary sources. On the other hand, secondary data sources like census data or market research are optional to address the specific business problems outlined in the case study.

**Appropriateness of Data Sources:**

The primary data sources, the production and merchant sales systems, are appropriate for analysis as they provide up-to-date and relevant information about the winery's operations. These sources offer accurate insights into product details, production volumes, sales transactions, and customer information, enabling a comprehensive analysis of the business problems.

**Data Cleansing:**

Before loading the data into the data warehouse, it is essential to assess whether significant data cleansing is required. While the case study does not explicitly mention data quality issues, it is advisable to conduct standard data cleansing procedures to ensure accuracy, consistency, and reliability. This may involve tasks such as standardizing data formats, handling missing values, detecting outliers, and removing duplicate records.

To effectively address the business problems, it is crucial to create a matrix that maps the data sources to the specific measures of interest. Based on the provided case study, the following matrix can be generated:

|  |  |
| --- | --- |
| **Data Source** | **Measures of Interest** |
| Production System | * Product information: Includes product description, code, and group * Production data: Provides production volumes for each product * Product costs: Offers cost information for each product |
| Merchant Sales System | * Sales transactions: Captures unit sales and dollar sales for each product * Customer details: Provides customer information, such as ID, name, address, and market category |

This matrix serves as a guide to understanding which data sources to utilize and how they are mapped to the specific measures of interest for analysis purposes.

By considering these aspects of data analysis, we can ensure a thorough and effective approach to address the business challenges faced by Puglia Winery.

# 5. Extraction, Transformation, and Loading

**ETL issues**

The process involves making the data from two distinct data sources uniform and consistent. Once these information systems are aligned in terms of structure and format, the data is integrated into a unified database system using platforms like Snowflake or Microsoft SQL Server. This integration offers improved scalability and flexibility.

**Transformation**

The data from various sources will be standardized to have a consistent format. The next step involves merging the two information systems into a single database. This process aims to enhance data integrity by eliminating duplicates and ensuring that the raw data is transformed into a suitable and usable form.

To achieve this, the data will be denormalized into a simpler and less complex form by adopting a Star Schema for our database design. As a result, the granularity will be relatively high, and the data will be presented in more aggregated and summarized forms. This approach facilitates streamlined use, ease of implementation, and faster execution of queries, reports, and ad hoc analyses.

**Scheduling issues**

Data extraction and loading will be performed on a daily basis, specifically toward the end of each day. The loading process will be done incrementally, and there is also the option for refreshes throughout the day and updates after hours. Incremental loads offer faster data loading as they handle individual batches of data rather than a single large load (full load). This approach allows for more manageable and cost-effective operations and is suitable for smaller data warehouses like the one we are designing for Puglia Winery. However, it's crucial to carefully manage each load transfer to ensure proper alignment and joining of all dimensions and fact tables. This maintenance and verification process are essential to maintain data accuracy and consistency.

**Real-Time or Near Real-Time Loading**

For our situation, a near real-time loading approach would be most suitable. It is not necessary to achieve up-to-the-minute or second-by-second loading since our extraction and incremental loading processes will occur daily, specifically at the end of each day. The frequency of incremental loads will be at least once a day or every 12-24 hours, depending on the specific requirements.

# 6. Recommended Architecture:

Puglia Winery's recommended data warehousing architecture will include both backend and front-room technologies, ensuring effective data processing and giving end users illuminating visualizations and reports. These elements will be included in the architecture:

**Backroom Technology:**

1. Data Integration and ETL Tools:

The winery can use robust ETL tools for the extraction, transformation, and loading (ETL) process, such as Apache Spark or Microsoft SQL Server Integration Services (SSIS). These technologies will make it easier to retrieve data from production and merchant sales systems, clean it up, change it, and load it into the data warehouse. Prior to being made available for reporting and analysis, the ETL process will verify that the data is of high quality and consistency.

2. Data Warehouse:

The architecture's central element will be the data warehouse itself. A scalable and high-performance database platform should be selected to accommodate the winery's current data volume and anticipated future growth. For processing massive datasets and complicated queries, options like Amazon Redshift or Microsoft Azure SQL Data Warehouse are excellent choices. To improve query performance, the data warehouse will store the combined and processed data in a star or snowflake schema.

3. Metadata Repository:

To hold details about data sources, transformations, and mappings, a metadata repository will be set up. Users will find it simpler to comprehend the data and its context thanks to this repository's documentation of the data warehouse's structure and meaning.

**Front-room Technologies:**

1. Data Visualization and Reporting Tools:

Data visualization and reporting tools like Tableau, Microsoft Power BI, or QlikSense can be used to present data in a clear and aesthetically pleasing way. Users will be able to produce interactive dashboards, reports, and ad hoc studies using the data in the data warehouse using these tools, including Jack, Stacie, and Tommy. Drilling down into particular data points and examining trends will improve decision-making abilities.

2. Query and Analysis Tools:

Front-room technologies should also contain tools that let expert users and data analysts run complex queries and do in-depth data analysis. OLAP (Online Analytical Processing) tools or SQL-based query tools can do multidimensional analyses and deliver insightful information about the data.

3. User Access and Security:

To maintain data security and privacy, user access to the data warehouse and related technologies should be tightly controlled. To prevent unauthorized access to sensitive data and to guarantee that users can only access the information necessary for their responsibilities, role-based access restrictions and authentication procedures will be put in place.

**Integration and deployment:**

Close cooperation between IT teams, business stakeholders, and data analysts is crucial for the successful implementation of the suggested design. The data models created for the project should be in line with the business needs, and appropriate testing techniques should be used to confirm the reliability and correctness of the data in the data warehouse.

Deployment of the data warehousing solution may involve setting up on-premises infrastructure or leveraging cloud-based services. Cloud solutions, such as Amazon Web Services (AWS) or Microsoft Azure, offer scalability and flexibility, allowing the winery to adapt to changing business needs and data volume.

By adopting the recommended architecture, Puglia Winery will establish a robust data warehousing solution that empowers its decision-makers with data-driven insights. The integration of backroom technologies for data processing and storage, combined with front-room tools for visualization and analysis, will provide a comprehensive business intelligence platform. Ultimately, the data warehousing solution will enhance the winery's ability to make informed decisions related to products, customers, and markets, driving continued growth and success.

# 7. Summary, conclusion & recommendation for action

**Summary**

The Puglia Winery, a medium-sized boutique winery that specializes in wines from hot climates, encounters a number of business obstacles that limit their ability to expand. The managing director, Jack Gillespie, understands the need for a data warehousing and business intelligence solution to handle these issues and make better informed decisions. The objective is to offer in-the-moment insights regarding profitable markets, important clients, and profitable products.

The study describes the features and functionalities of the suggested BI system, such as market profitability analysis, identification of the most important clients, and product profitability analysis. In order to provide efficient data integration and display, it also discusses data models, data analysis, and the extraction, transformation, and loading (ETL) procedure.

The suggested design contains both front- and back-end technologies, as well as tools for data integration, reporting, and data visualization. It also has a data warehouse for storing information. To safeguard sensitive data, user access and security are also underlined.

**Conclusion:**

For Puglia Winery to continue expanding and succeeding, it is crucial to implement a data warehousing and business intelligence (BI) solution. The winery may expand into lucrative markets, cater to key clients, and maximize product offers by utilizing real-time data analytics. Decision-makers will have access to precise and fast information thanks to the proposed architecture and features, enabling data-driven decision-making processes.

**Recommendation for Action:**

It is advised that Puglia Winery adopt the data warehousing and BI solution right away in light of the report's conclusions. The actions listed below can be taken:

1. Assess and Select BI technologies: The winery should assess various BI technologies, such as Tableau and Microsoft Power BI, to determine which is best for their demands, financial situation, and technical requirements.

2. Create Data Models: To enable effective data integration and analysis, design and build data models, including dimension tables and the primary fact table.

3. Establish the ETL process to extract, transform, and load data from production and sales systems into the data warehouse. This can be done by using tools like Apache Spark or Microsoft SQL Server Integration Services.

4. Deploy Data Warehouse: When deciding where to host the data warehouse, take scalability and cost-effectiveness into account.

5. Create Interactive Dashboards and Reports: Employ data visualization and reporting technologies to create interactive dashboards and reports that offer insightful information on the profitability of products, the performance of key clients, and the state of the market.

6. User Training and Security: Educate the necessary parties on how to utilize the BI system, including management, sales, and production teams. Use role-based access control to protect the privacy and security of your data.

7. Monitor and Improve: Keep track of how well the BI system is working and get user feedback to pinpoint problem areas and make the required corrections.

These actions will help Puglia Winery create a data-driven culture of decision-making, improve operational performance, and accelerate business growth through real-time analytics and insights.

# References

Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (3rd ed.). Wiley.

Inmon, W. H., & Hackathorn, R. D. (2015). Using the Data Warehouse. Wiley.

Khan, A. R., & Negi, P. S. (2016). ETL Testing - A Comprehensive Guide. Packt Publishing.

Microsoft Docs - SQL Server Integration Services (SSIS). (n.d.). <https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services>

Amazon Web Services - Amazon Redshift. (n.d.). <https://aws.amazon.com/redshift/>