Group 3’s Project

**Puglia Winery Case**



ISDS 556, Spring 2023, CSUF

Instructor: Dr. Peng Liu

Hy Luong

Eric Kingery

Queeny Mathayas

Miguel Cruz

Adrian Vasquez

**Table of contents**

[**1. Overview 2**](#_5lhw29d5n1j3)

[**2. Functionality and Features of BI system. 3**](#_l5dv06llaw64)

[**3. Data models 5**](#_e9dkefk51bzi)

[**4. Data Analysis 7**](#_an4d86m3qe68)

[**5. Extraction, Transformation, and Loading 8**](#_wxp2w9mnsoti)

[**6. The recommended architecture 9**](#_reamnrjtbf2u)

[**7. Summary, conclusion & recommendation for action 11**](#_wgoaenfb97u)

[**References 13**](#_sn6nwe7ddjaq)

# 

# Overview

The Puglia Winery is a medium-sized boutique winery located in Temecula, California. They specialize in hot climate wines and produce three famous varieties: pinot noir, merlot (red), and pinot grigio (white). They distribute their offerings locally in California, as well as interstate in the US and internationally in the United Kingdom. Over the last three years, the winery has seen dramatic growth. They are currently working with a traditional database and using mainly “Gut feel” when making decisions. The managing director, Jack Gillespie, believes that implementing data warehousing and business intelligence solutions can help the winery make more informed decisions to manage its growth effectively. We aim to provide a Dataware solution for Jack to make better-informed decisions with data on any given day or time.

The winery has identified three main business challenges: determining which products are most profitable, identifying key customers, and identifying the most lucrative markets. To address these issues, the winery has shared data from its production and sales systems.

**Gaps**

1. Production and Sales to merchants need to be integrated systems, thus making it difficult to analyze data for informed decision-making.
2. The winery needs a real-time dashboard to gain valuable insights and improve decision-making for business growth.
3. Time-based analysis can be complex since historical data is not stored in a centralized repository for easy access.

**Organization readiness**

1. Mr. Jack Gillespie recognizes that a data warehousing solution is necessary to maintain the exclusivity of the product.
2. Management is ready to invest.
3. Winery may have to hire skilled staff to support the proposed warehouse solution.

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

The columns represent the major organizations or workgroups involved in the requirements process. They are groups interested in the metrics associated with the business process rows. The rows correspond to business processes that will be prioritized within the architecture of the data warehouse system.

**Exhibit 1 - Organization/Workgroup**

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

# 

# 2. Functionality and Features of BI system.

**Focused KPI’s**

The foundation of the functionalities and features revolves around the business problems prioritized by Jack Gillespie, the managing director. He hopes to base the organization’s goals and strategies around these three problems:

1. Which products are most profitable?

2. Who are the key customers?

3. Which market is most profitable?

With those goals and strategies in mind, the mission is to measure the organization’s level of performance across multiple processes in order to determine its success. Key performance indicators are highly recommended as they are defined metrics that can be visualized via scorecards for monitoring performance and are impacted by everyone in the business. The team should focus on these KPIs: internal business process, learning and growth, customer, and financial.

**Generated Reports**

When it comes to reporting, there are three kinds of reports expected to be generated in this system.

The first report is the standard report. These are basic reports that provide users with key sets of information about a specific business area and can be generated on demand. Since they focus on a certain area, they are only derived from one data source. Everyone in the organization is expected to use these reports but they will mostly be used by business analysts and lower-level management. Examples of standard reports for the winery include YTD wine sales, production volume, and market locations.

The second report is generated from read/write applications. These are more back-end in nature and require the expertise of developers and niche mid-level managers such as sales managers and production managers. These reports are essentially used to assess performance by comparing actual performance to target goals. Examples and processes include planning, budgeting, forecasting, and what-if models. In nature, they are generated periodically on a quarterly or annual basis.

The third report is the dashboard and scorecard. This is the bread and butter of BI and it visually highlights metrics across the entire organization for information consolidation. They are more sophisticated in comparison to the standard report since they support high-level status reporting across multiple data sources. Like standard reports, everyone is expected to use these reports, but they are meant for more specific users, such as sales managers, finance managers, production managers, etc. Some of the many examples include reported sales vs forecasted sales, market sales, and seasonal sales. They are usually generated on demand and provided on a periodic basis.

**Applications for Presentation and Analyses**

Potential applications that will extract data from the data warehouse for presentation and analysis depend on the business and technical requirements of the organization, the purchasing process, and the types of users and reports.

Based on the information provided by management on the winery, the company should decide on either Power BI or Tableau. Each of these applications has its own pros and cons and we recommend that a holistic analysis is necessary from the level of expertise of analysts to the overall size of the company to anticipations for future growth and to the data warehousing capabilities (size of data sets). An investment in either application would be most beneficial for the organization as the implementation of data visualization will allow for timely and accurate analysis of complex data and provide insight for managers to make business decisions like targeting specific countries for sales, the production of wine, and seasonal marketing approaches.

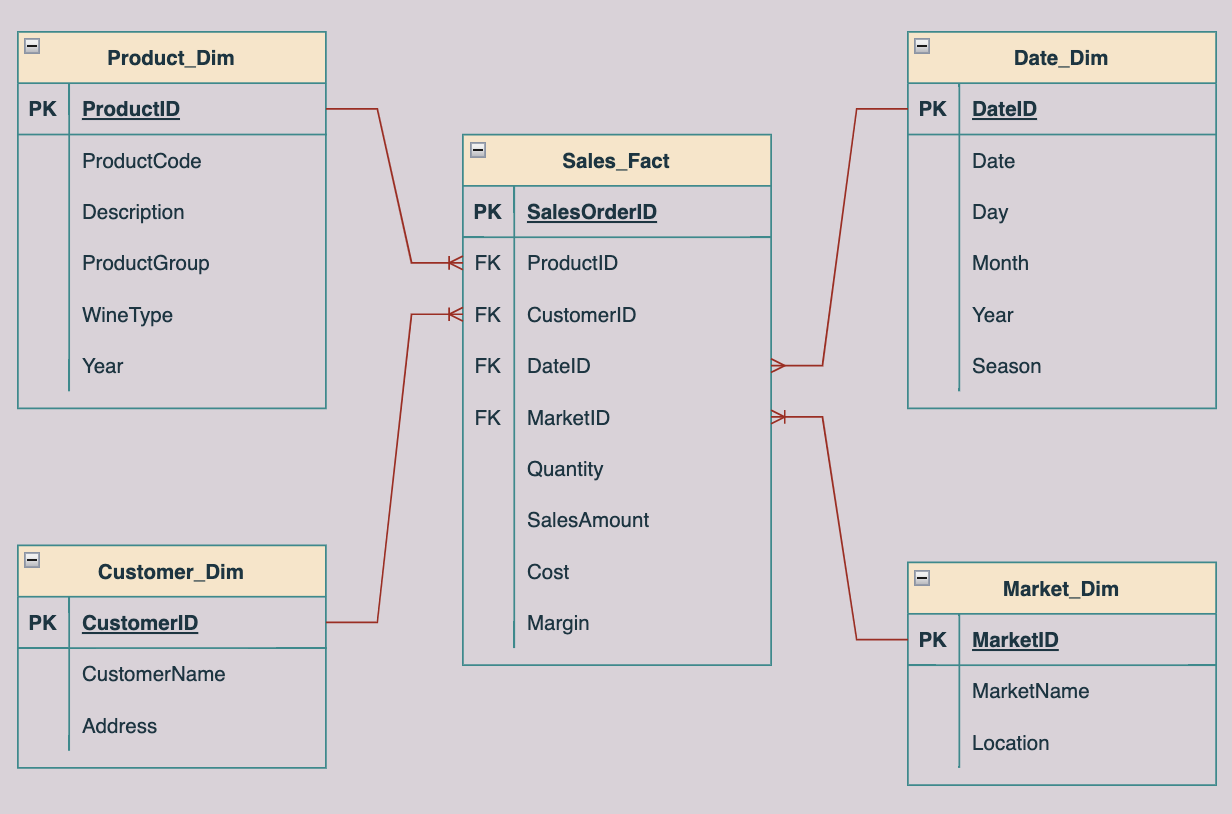
# 3. Data models

After conducting a competitive analysis, we concluded that creating logical and physical entity relationship data models is essential to aid system development. We will create physical models to complete our design and meet all requirements. These models will help us implement our system and inform new employees about the overall design.

Our proposed data model for meeting business requirements is a star schema consisting of a single central fact table and several dimension tables. We chose the star schema due to its simplicity, ease of implementation, and quick query performance. After analyzing the competition, we found that the star schema is a widely used data model due to its many benefits.

**Star Schema**

Our proposed data model for meeting business requirements is a star schema consisting of a single central fact table and several dimension tables. We chose the star schema due to its simplicity, ease of implementation, and quick query performance.



There is no need to complicate the schema by creating multiple levels for the dimension tables, as they are not overly complex. Denormalizing them into a single level would not result in significant performance benefits.

**Slowly Changing Dimensions (SCD)**

It is recommended to treat Product\_Dim and Customer\_Dim as Type 2 SCDs since they may experience changes over time, and it is important to preserve historical data. An instance of this is when a product's description changes, a new record should be created with the updated information and a new ProductID while still keeping the old record.

**Fast-Changing Dimensions**

The Date\_Dim is considered a fast-changing dimension because dates change frequently. However, the dimension table will not become excessively large as the granularity is set at the day level.

**Degenerate Dimensions**

The SalesOrderID in the Sales\_Fact table acts as a degenerate dimension. Although it lacks its own dimension table, it provides essential context for the fact table. By implementing the suggested data model, the winery can tackle their primary business issues. They can thoroughly analyze the profitability of their products, identify key customers, and evaluate market profitability. Users can generate several reports and perform various analyses, including descriptive statistics, PCA, and data visualization. This allows for informed decision-making and optimization of sales, marketing, and production strategies.

# 

# Data Analysis

**Available Data**

The data available for the Puglia Winery Case Study comes from the existing operational systems, such as production and merchant sales. This data covers different aspects of the winery's operations, including product information, customer details, sales transactions, production data, and market information.

**Data Source**

The main sources of data are the operational systems, which include production and sales systems. These sources are reliable since they are directly connected to the winery's activities and offer the latest and most precise information for analysis. While secondary data sources, like census data or market research, may offer more context and insights, they are not necessary to solve the business issues presented in the case study.

**Data Cleaning**

Before loading data into the data warehouse, it's crucial to perform data cleansing. This involves standardizing the data by ensuring consistent date formats and units of measurement, filling in or erasing missing or incomplete data, detecting outliers, and removing duplicate records. This process guarantees that the data is accurate, consistent, and reliable, which is essential for effective analysis and decision-making. If necessary, create a matrix of the data source mapped to the measures of interest.

| **Data Source** | **Measures of Interest** |
| --- | --- |
| Production System | Product information, production data, product costs |
| Merchant Sales System | Sales transactions, customer details, market information |
| (Optional) Market Research/Census Data | Additional market insights, demographic data |

To solve the business problems in the case study, we've created a matrix that shows which data sources and measures are relevant. The production system provides information about product details, production data, and product costs. Meanwhile, the merchant sales system offers data on sales transactions, customer details, and market trends. For a more complete understanding of the winery's business landscape, you may also consider using market research or census data to gain additional insights into the market and demographics.

# Extraction, Transformation, and Loading

**ETL issues**

Standardizing or streamlining all of the data from two separate data sources into the same format. Once those two information systems are formatted under the same dimensions or configurations, then that data must be implemented into a single database system using solutions such as Snowflake or Microsoft SQL Server. This would also allow for much greater scalability and elasticity.

**Transformation**

Standardizing the data from multiple sources into the same formatting. Aggregating those two different information systems into one database. This process will help ensure the improvement of data integrity. It will also remove any potential duplicates and ensure that the initial raw data is transferred in the appropriate form and ready to use.

Data will be denormalized into a simpler and less complex form since we are using a Star Schema for our database design. Therefore, the granularity will be at a relatively high level. More aggregated and summarized forms of the data. This will allow for streamlined use, ease of implementation, and faster queries and running of reports and ad hocs.

**Scheduling issues**

Extracting and loading data would be done on a daily basis. Extracting and loading would be tasked toward the end of every day. The loads would be completed in incremental portions, with the possibility of refreshes throughout the day and updates after hours. Incremental Loads allow for much faster data loads with individual batches of data instead of with one big load (full load). It is a more manageable but less comprehensive way to compare incoming data with what is already on hand. This type of data architecture allows smaller and less expensive operations and data warehouses like the one we are designing for Puglia to maintain and manage. However, each load transfer must be maintained and double-checked to ensure all dimensions and fact tables are properly aligned and joined.

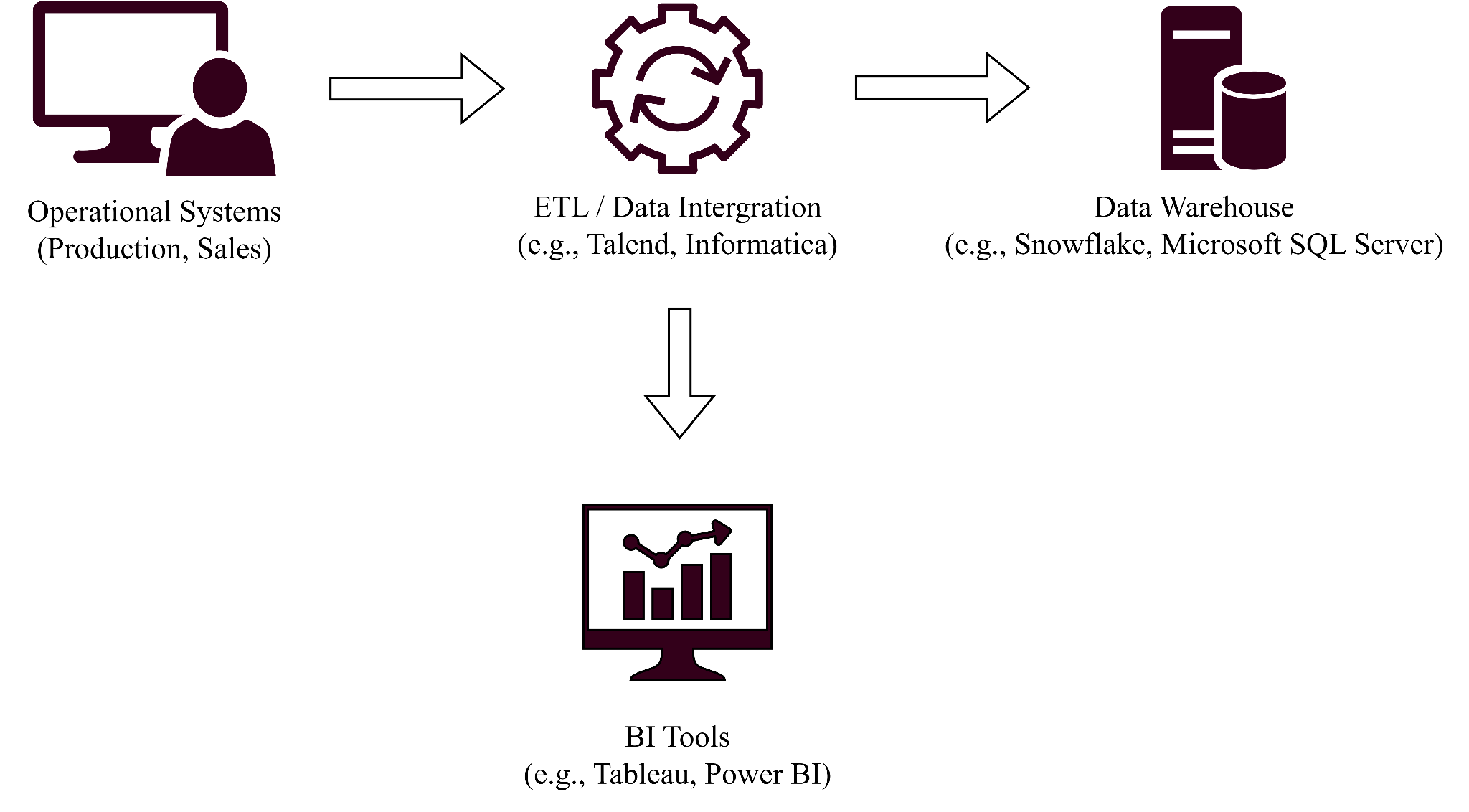
**Real-Time or Near Real-Time Loading**

In our case, near real-time loading will be ideal. Up-to-the-minute or second loading is not necessarily required, as extraction and loading (incremental) would be initiated and completed at the end of every day. In terms of frequency, incremental loads will be completed at least once a day, or every 12-24 hours, depending on needs.

# The recommended architecture

The recommended architecture for the Puglia Winery data warehouse solution consists of several components, including data integration, data storage, data processing, and data presentation.

**Diagram**



**Backroom technologies**

**ETL / Data Integration**

Tools such as Talend or Informatica can be used to extract, transform, and load data from the operational systems into the data warehouse. These tools enable efficient data cleansing, transformation, and integration to ensure that the data is accurate and consistent.

**Data Warehouse**

A cloud-based data warehouse solution like Snowflake or an on-premise solution like Microsoft SQL Server can be used to store the integrated data. These solutions provide scalable and high-performance data storage, which is crucial for supporting the winery's growth.

**Front-room technologies**

**BI Tools**

Business intelligence tools such as Tableau or Power BI can be used to create reports, dashboards, and visualizations to present the insights derived from the data warehouse. These tools enable users to easily explore and analyze the data to support informed decision-making.

**Justification for specific solutions**

**ETL / Data Integration**

Talend and Informatica are widely recognized for their robust data integration capabilities, making them suitable choices for handling the data extraction, transformation, and loading requirements of the Puglia Winery case study.

Data Warehouse: Snowflake offers a highly scalable, cloud-native data warehouse solution that can grow with the winery's needs, while Microsoft SQL Server is a well-established on-premise data warehouse solution that offers comprehensive features and performance.

**BI Tools**

Tableau and Power BI are popular business intelligence tools known for their intuitive interfaces and powerful data visualization capabilities, making them ideal for presenting the insights derived from the data warehouse to support decision-making at Puglia Winery.

In conclusion, the recommended architecture leverages a combination of backroom and front-room technologies to address the business needs of the Puglia Winery. By integrating data from operational systems, storing it in a data warehouse, and presenting insights using BI tools, the solution enables the winery to make data-driven decisions to effectively manage its growth.

# 

# Summary, conclusion & recommendation for action

In this report, we described a data warehousing and business intelligence solution for the Puglia Winery case study to address their business problems and support informed decision-making. We assessed the organization's readiness, identified gaps in its current analytical capabilities, and proposed a high-level enterprise data warehouse bus matrix. We also discussed the data models required for the system, the data analysis process, and the recommended architecture for the data warehouse solution.

We described a combination of backroom and front-room technologies, including ETL/data integration tools (Talend or Informatica), data warehouse solutions (Snowflake or Microsoft SQL Server), and BI tools (Tableau or Power BI) to create a comprehensive solution that enables the winery to make data-driven decisions.

**Next steps**

1. Obtain stakeholder buy-in: Present the proposed solution to the management team, highlighting the benefits of implementing a data warehouse and business intelligence solution for Puglia Winery.
2. Select vendors and technologies: Evaluate and choose the most suitable vendors and technologies for each component of the architecture based on factors such as cost, scalability, and ease of use.
3. Develop a project plan: Create a detailed project plan outlining the implementation timeline, budget, and resource requirements.
4. Implement the data warehouse solution: Set up the chosen data integration, data storage, and BI tools, and begin the process of integrating data from the operational systems into the data warehouse.
5. Train staff: Provide training to the staff who will be responsible for maintaining and using the data warehouse solution, ensuring they have the necessary skills to effectively utilize the new tools and insights.
6. Monitor and evaluate: Regularly monitor the performance of the data warehouse solution, assess its impact on the organization's decision-making and operations, and make adjustments as necessary to optimize its effectiveness.

By following these steps, Puglia Winery can implement a data warehousing and business intelligence solution that will enable the organization to make better-informed decisions, identify growth opportunities, and manage its operations more effectively.

# 

# References

Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., &amp; Becker, B. (2011). The Data

Warehouse Lifecycle Toolkit (2nd ed.). Wiley Publishing.

Liu, P. (n.d.). Puglia Winery Case Study. Fullerton; Cal State Fullerton.