**Puglia Winery**

Data Warehouse Initiatives

horizontal line

# 

# Table of Contents

|  |  |  |
| --- | --- | --- |
| [Introduction](#_ec8b6qnt17ur) |  | 2 |
| [Moving Forward](#_ltcwt9az8fgb) |  | 4 |
| [Data Models](#_weqn20cty90s) |  | 6 |
| [Data Analysis](#_ytyd64d4xsbp) |  | 7 |
| [ETL](#_dnzobuc0puga) |  | 8 |
| [Recommended Architecture](#_3ag2t7xz3y2l) |  | 9 |
| [Conclusion and Next Steps](#_xmf0431ydqo4) |  | 12 |

# Introduction

There comes a time in every business journey when the next steps forward require the implementation of new technologies. Sustainability and scalability require modern systemization and will support efforts in better customer satisfaction, profitability, organization, and business intelligence. This is the case for Puglia Winery.

A medium sized boutique winery located in Temecula, California, Puglia is exploring larger market diversity. In any circumstance where a business is aiming to understand sales trends to manage growth and improve revenue, a data warehousing and business intelligence system is necessary. Our team will propose a data warehouse solution that will integrate the data seamlessly and enable clarity on seasonality, product profitability and trends, and key performance indicators.

## Understanding the Current State of Puglia Winery

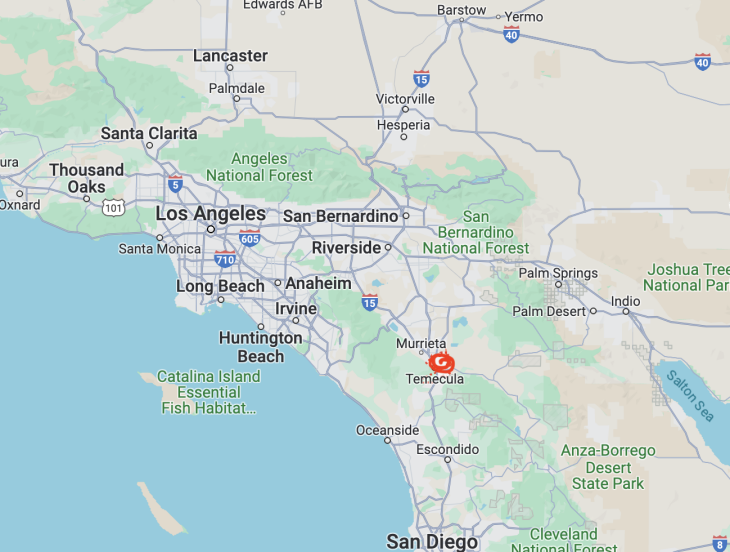
In order to justify the implementation of an advanced information technology infrastructure, we must first identify and develop an overview of Puglia’s business pre-established attributes and needs.

**Location**: 85 miles southeast of Los Angeles

**Managing Director:** Jack Gillespie

**Wine Maker:** Tommy Siragusa

**Sales Manager:** Stacie Giano



|  |  |
| --- | --- |
| **Operations** | **Sales** |
| Growing and maintaining vineyards  Total acres: 135   * 50 acres of pinot noir * 50 acres of merlot * 35 acres of pinot grigio   Producing and storing wine   * Focus: Quality * Total Products: 7   + 3 pinot noirs   + 2 pinot grigios   + 2 merlots   Selling and delivering wine   * Premium pricing | Wine Merchants (by cases of 12 bottles)   * California * Interstate * UK   Direct from cellar - very little of profit percentage |

The winery has branded itself as a quality wine and aims to continue with adhering to their policies in place to ensure the highest product standard. They are self-producing for the large majority of their wines but do purchase grapes from outside vendors which will be an added component of cost-efficiency. They have been producing the seven total wine products based on the three grapes they farm. A streamlined data warehouse will aid Puglia in developing proposals to answer questions concerning production value and efficiency and new market spaces.

# Moving Forward

The current analytical capabilities of Puglia are limited by the winery’s use of separate systems for production and sales. This leads to fragmented data and an inability to analyze the cost benefit of the seven products on an individual basis. The data also lies in varying formats, making aggregating data difficult and problematic. Below is the ensuing bus matrix that will guide the wineries data ventures and ensure the data warehouse is set up in the Kimball style to fit the needs of the business.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Business Process** | **Time** | **Raw Material** | **Supplier** | **Product** | **Market** | **Customer** |
| Organic Material Purchasing | X | X | X |  |  |  |
| Organic Inventory | X | X | X |  |  |  |
| Bill of Organic Materials | X | X | X |  |  |  |
| Production | X | X |  | X |  |  |
| Product Inventory | X |  |  | X |  |  |
| Orders | X |  |  | X | X | X |
| Shipping Orders | X |  |  | X | X | X |
| Invoicing | X |  |  | X | X | X |
| Payments | X |  |  | X |  | X |

## 

Fig 1: Bus Matrix

## Key Performance Indicators

In order to outline the best data warehouse structure for Puglia Winery, we must decide and define the key performance indicators. These will drive important business decisions and provide a baseline metric for sustainability.

The winery must consider what information is necessary across the board and take note of growth prospects and retention efforts as well. A general overview of KPI’s will include:

* **Sales Volume**
* **Revenue**
* **Production Cost**
* **Market Growth**
* **Profit Margin**

These general KPI’s will aid Puglia in realizing the particulars within their business. Across the product types, varieties, and geographical markets, along with recognizing year-on-year peaks and troughs, Puglia can query these reports to drive their next steps.

If Puglia Winery is looking to expand their business and cut back on inefficiency, there are important questions to be considered. What season best supports the production of red and white varieties? What market? What are the trends of the wine-producing and sales season? What is the sunk cost and cost benefit of each variety? Where can money be saved on efficiency and where can it be spent to enhance growth? All of these questions and more begin by building historical documentation of KPI’s through reports that are easily accessed in data marts through the warehouse.

### Reports

Necessary and interesting reports need to be granular down to the day of purchase and the item purchased.

The Sales Performance Report will identify the most profitable products on a monthly, seasonal, and yearly basis.

* Product Name
* Product Type
* Quantity
* Dollar Sales
* Cost
* Margin
* *Viewers: Sales Manager, Director*

The Customer Profitability report will help to determine the most profitable customers in total and by product. It will also aid in noting churn and loss over the years by customer type as well as document growth on a monthly, seasonal and yearly basis.

* Customer Name
* Customer Type
* Product Name
* Product Type
* Unit Sales
* Dollar Sales
* Cost
* Margin
* *Viewers: Sales Manager, Director*

The Market Analysis will drive insights on the most profitable markets, for which products, and at what time of year. This report should cover the markets on a monthly, seasonal, and yearly basis so that Puglia can determine the direction of their expansion.

* Market Name
* Unit Sale
* Dollar Sales
* *Viewers: Director, Sales Manager*

These reports will guide important insights, answering essential questions in a timely and pertinent manner.

# Data Models

To address the business requirements and the reports that Puglia Winery has addressed, the data warehouse will employ a star schema model. This model will feature a centralized sales fact table containing key sales data, including unique identifiers for sales orders, customers, products, and dates, along with metrics such as quantity sold, sales amount, cost, and margin (where margin is calculated as sales\_amount - cost). Dimension tables will be created for dates, products, and customers. The date dimension table will provide detailed information like day, month, quarter, and year, while the product dimension table will include data on the wines, such as name, type, variety, year produced, and unit price. The customer dimension table will store customer names, addresses, and market segments. This dimensional model will support the desired reports concerning product and customer profitability.

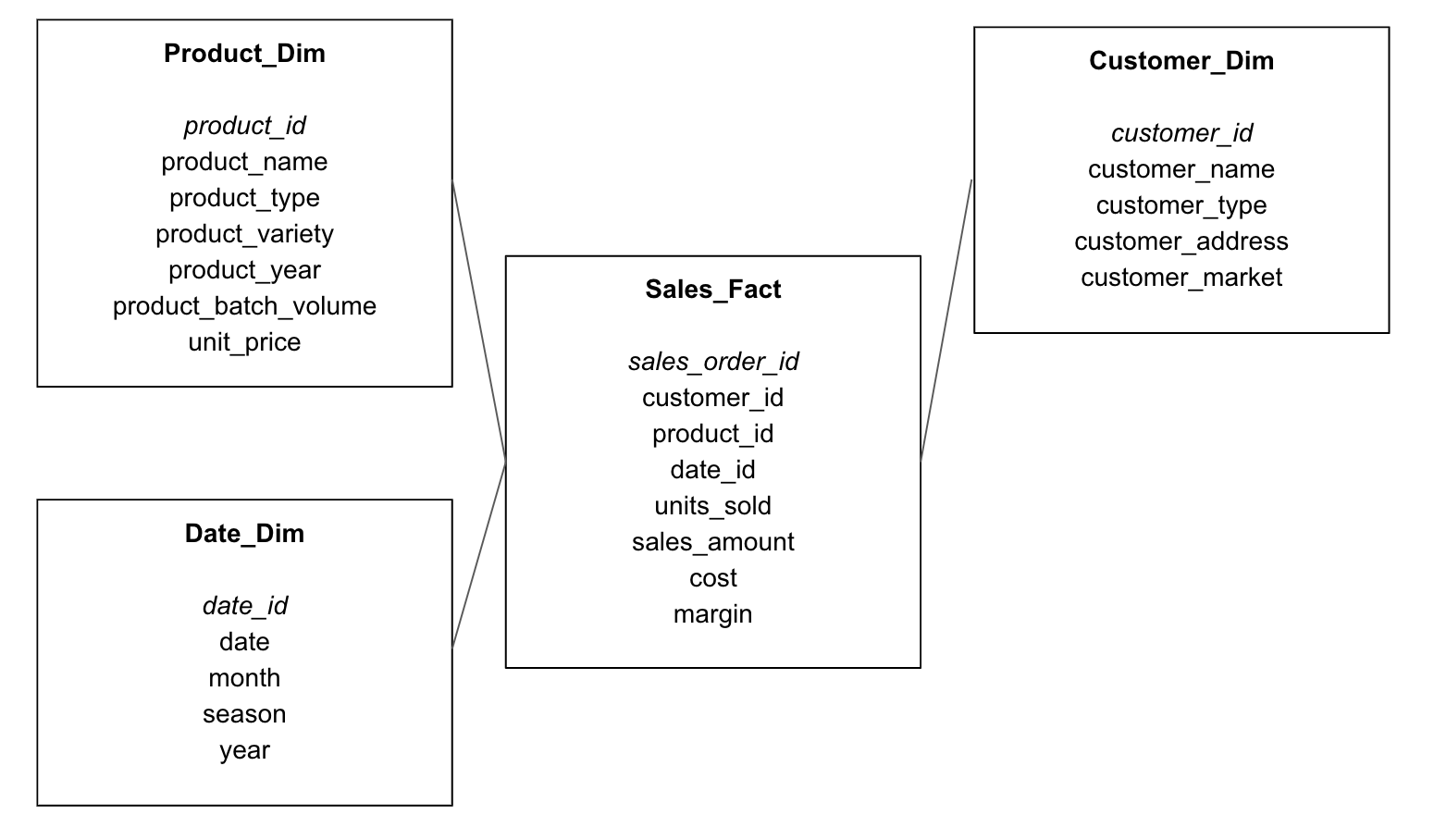
We select the star schema model because of the simplicity of the model, easy comprehension, and easy use for business directors and analysts. This star schema model will streamline data organization, improve query performance, and simplify reporting and analysis. The quantitative data, located in the fact table, is centralized and linked to a select few detailed dimensions. This model is optimized to make querying large data sets efficient and simplifies adding and changing new dimensions. 

Fig 2: Star Schema model for new Puglia Winery warehouse

Moreover, the star schema model is recommended over the entity-relationship (ER) model due to its simplicity and intuitive design for business users. While the star schema consists of a central fact table connected to several dimension tables, the ER model's normalized structure, with multiple related tables, is more complex and can be challenging for non-technical users. The ER model requires special attention if new dimensions are added to maintain a normalized structure. Although this minimizes redundancy in the tables and preserves data integrity, the queries become long and complicated in order to account for the necessary join statements.

On that note, for our data model, we will not be using snow-flaking since this would increase the complexity of our data model and make it more challenging for business users to navigate and comprehend. Because Puglia’s focus is in BI, the simplicity of the data model is key in aiding the success of the company.

# Data Analysis

Puglia Winery provided two operational systems that contain data for their production and sales processes. These systems are crucial for creating a comprehensive data warehouse and business intelligence (DW/BI) system to drive informed decision-making.

### The Production System Data

1. Product
   1. Code
   2. Description
   3. Group
2. Production History
   1. Code
   2. Product
   3. Year
   4. Production Volume
   5. Cost (Doz)

### The Merchant Sales System Data

1. Customer
   1. Customer ID
   2. Name
   3. Address
   4. Market
2. Product
   1. Product ID
   2. Description
   3. Group
   4. Year
   5. Unit Price (Doz)
3. Sales Order
   1. Sales Order
   2. Customer
   3. Date
4. Sales Order Line
   1. Sales Order
   2. Line
   3. Product
   4. Quantity (Doz)
   5. Price (Doz)

The data provided from both operational systems exhibit redundancy and inconsistency in naming conventions and formats. This issue further promotes the necessity for a DW/BI system to consolidate and standardize the data to ensure reliability and consistency. It is essential the data undergo cleansing and integration to establish a robust foundation for the DW/BI system and avoid inconsistencies. Puglia Winery must focus on standardizing product information, consolidating customer data, and aligning sales and production data.

Significant data cleansing is required to avoid issues with the date data. This process involves breaking down date attributes into a granular level such as month, season, and year. Puglia Winery will acquire the ability to uncover new business opportunities by achieving this level of granularity. Specifically, the date attributes will facilitate more thorough analyses for various time periods.

# ETL

Puglia Winery's rapid growth has created a need for more sophisticated data management and analysis to support informed decision-making. The current challenge involves integrating data from disparate production and sales systems that have different formats and use various database management systems. The crucial ETL issues to address include data integration, transformation, validation, scheduling, and aggregation.

The primary issue is integrating data from the winery's two separate systems, which will require transformations to unify the data formats. This can be mitigated by designing a comprehensive ETL process. Data extraction should focus on capturing relevant information from both systems, ensuring that all necessary data points are included.

Transformations will involve converting data into a unified format, applying validity checks to ensure data accuracy, and performing necessary aggregations to adjust granularity. For instance, sales data from different regions need to be aggregated to provide a consolidated view of market performance. Additionally, partitioning the data based on time or region can improve query performance and manageability.

Denormalization is another crucial transformation, especially for performance optimization in analytical queries. By denormalizing certain tables, redundant data is stored to reduce the complexity of joins, which speeds up query processing.

Scheduling the ETL process is critical to ensure timely data availability. Given the dynamic nature of sales and production, a near real-time data loading approach is recommended. This can be achieved by scheduling frequent data extractions, perhaps hourly or even more frequently, depending on the transaction volume. Loading data during off-peak hours can minimize the impact on production systems while ensuring that the data warehouse is updated regularly.

Amazon Redshift, recommended by our team as the ETL tool, will facilitate efficient data extraction from production and sales systems, transformation into a unified format, and loading into the data warehouse. This integrated approach will enable Puglia Winery to leverage a BI tool, such as Tableau, for visualization, providing Jack Gillespie and his team with actionable insights to drive business growth based on solid data evidence rather than intuition.

# Recommended Architecture

Puglia Winery data warehousing architecture solution meets the main business goals: control of production, knowledge of sales patterns, and data-based decision making. The proposed data warehousing solution consolidates data from several sources, guaranteeing uniform, precise, and punctual information. By using this architectural framework, Puglia Winery can acquire vital intelligence regarding product profitability, key clientele, and market performance. This will facilitate data-based decision-making and bolster future expansion.

### High Level BI Architecture Model

By utilizing Amazon Redshift and Tableau, Puglia Winery may construct a data warehousing solution that enables the integration of data from many sources, guaranteeing the provision of consistent, accurate, and timely information. Implementing this will facilitate evidence-based decision-making, assisting the winery in efficiently managing expansion and gaining a deeper comprehension of sales patterns. BI technologies will augment data analysis and visualization, offering significant insights into product profitability, key customers, and market performance.

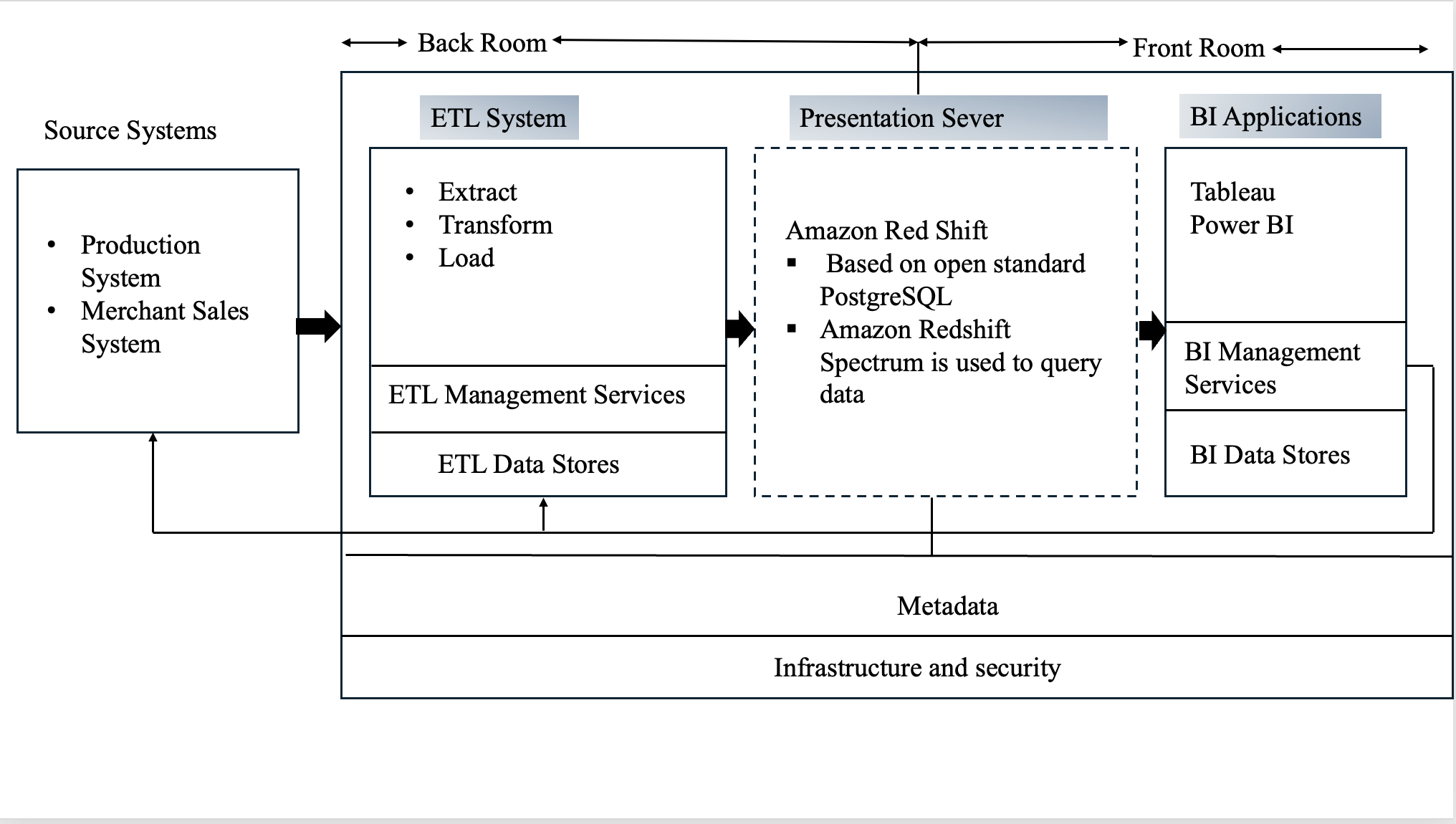


Fig 3: High level BI architecture model

### Backroom System Architecture Model

The implementation of an ETL (Extract, Transform, Load) system for Puglia Winery entails various processes, such as configuring the ETL tools, establishing the ETL process, overseeing data sources, and maintaining metadata.

#### Implementation of ETL Process

1. Extract Data

Identify Data Sources: Specify the origin systems from which data will be retrieved.

Establish Data Connections: Utilize the Extract, Transform, Load (ETL) tool to establish connections with the manufacturing and sales systems.

1. Transform Data

Data cleaning: Involves the elimination of duplicate entries, addressing missing values, and ensuring that data formats are standardized.

Data Mapping: Transfer data from the source schema to the target schema.

Data Aggregation: Condense data as needed

1. Load Data

Target Data Warehouse: Transfer the modified data into Amazon Redshift, which serves as the target data warehouse.

Data Marts: Import data into designated data marts such as Sales, Production, and Customer.

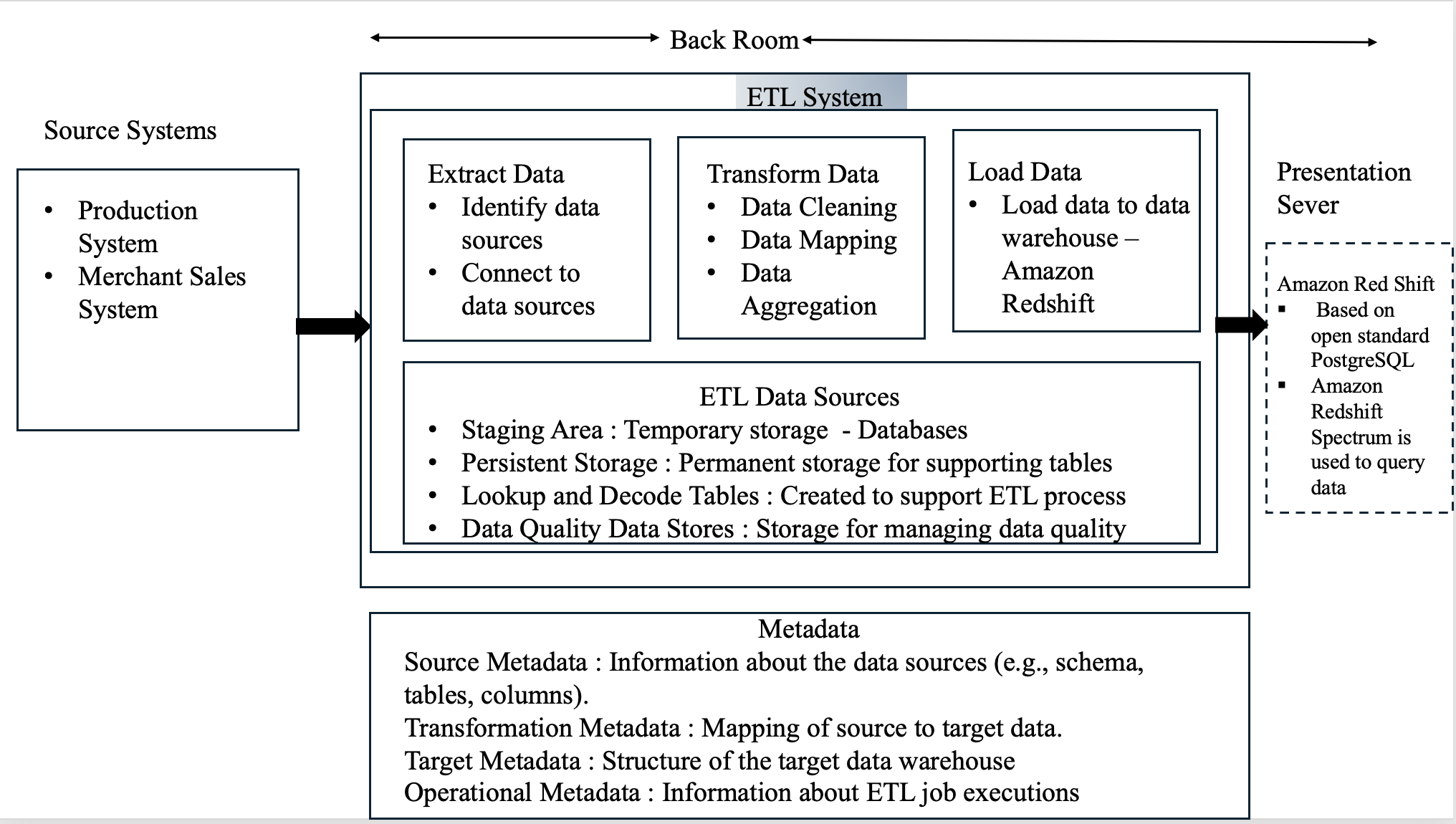


Fig 4: Backroom system architecture model

To implement an Extract, Transform, Load (ETL) system for Puglia Winery, the necessary steps include configuring the ETL tools (such as Talend or Apache Nifi), establishing the ETL process, ensuring data quality and integration, and managing metadata to enhance data governance. Puglia Winery can attain a comprehensive data warehousing solution that facilitates data-driven decision-making by adhering to these measures.

### Presentation Server Architecture Model

Deploying a presentation server with Amazon Redshift entails configuring the necessary infrastructure to enable end-users to access, analyze, and visualize data. This configuration enables business intelligence (BI) tools and other reporting applications to establish a connection with the Redshift data warehouse in order to retrieve and display data.

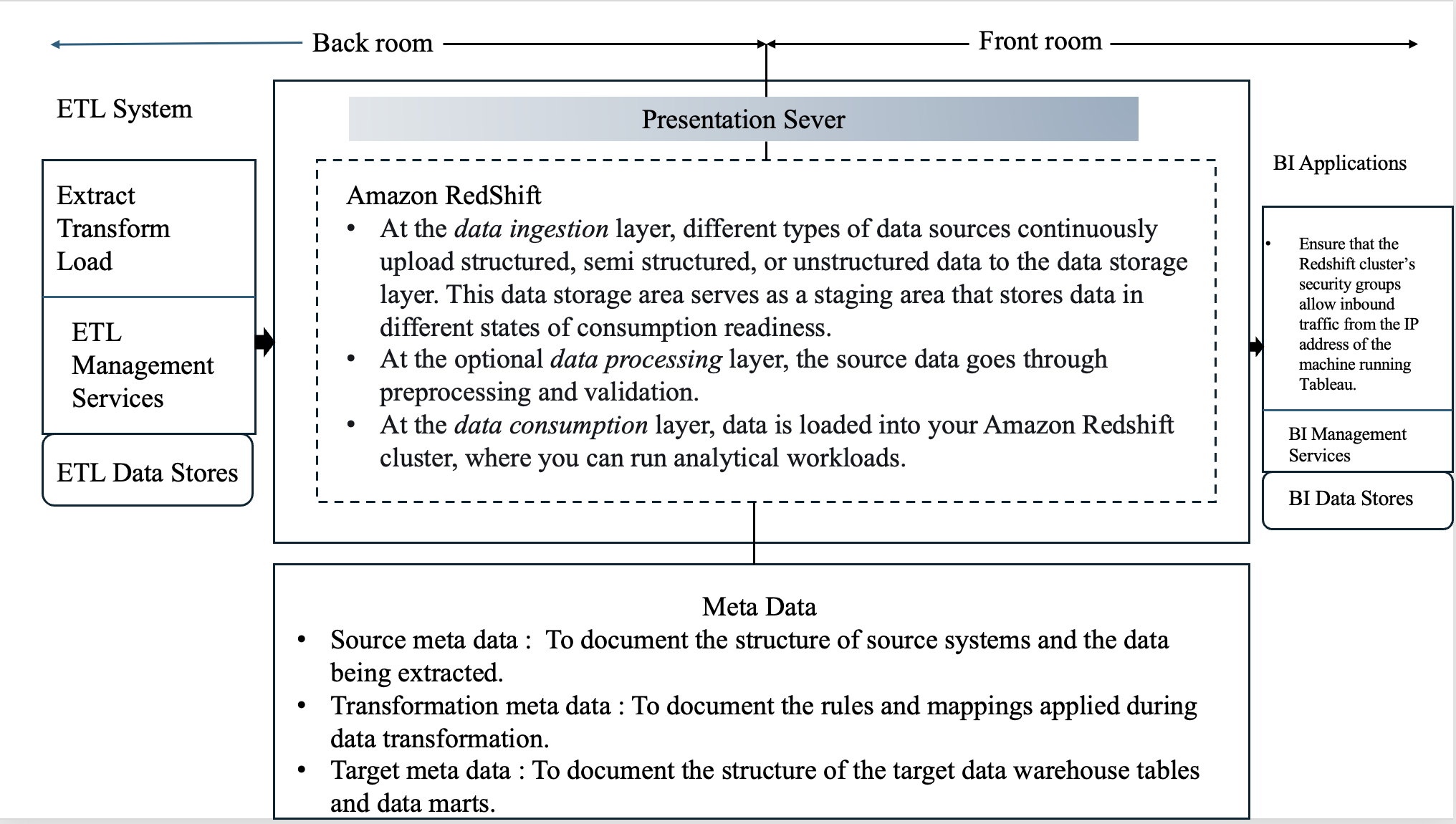


Fig 5: Presentation server system architecture model

This configuration will provide effortless data retrieval, analysis, and presentation, allowing for data-informed decision-making and comprehensive business insights. Utilizing business intelligence (BI) technologies such as Tableau, Power BI, and Looker will enable end-users to generate interactive dashboards and reports, offering vital insights into the winery's operations and performance.

### Front Room Architecture Model

Puglia Winery Project establishes a strong connection between Amazon Redshift and Tableau and Power BI, allowing for robust data visualization and reporting functionalities. Effective metadata management guarantees that users possess a comprehensive comprehension of the data, hence improving the quality of the data and fostering trust in the generated reports and dashboards. Tableau and Power BI are robust business intelligence tools that serve a wide range of application types. These apps can be classified according to their functionality, deployment, and utilization within an organization.

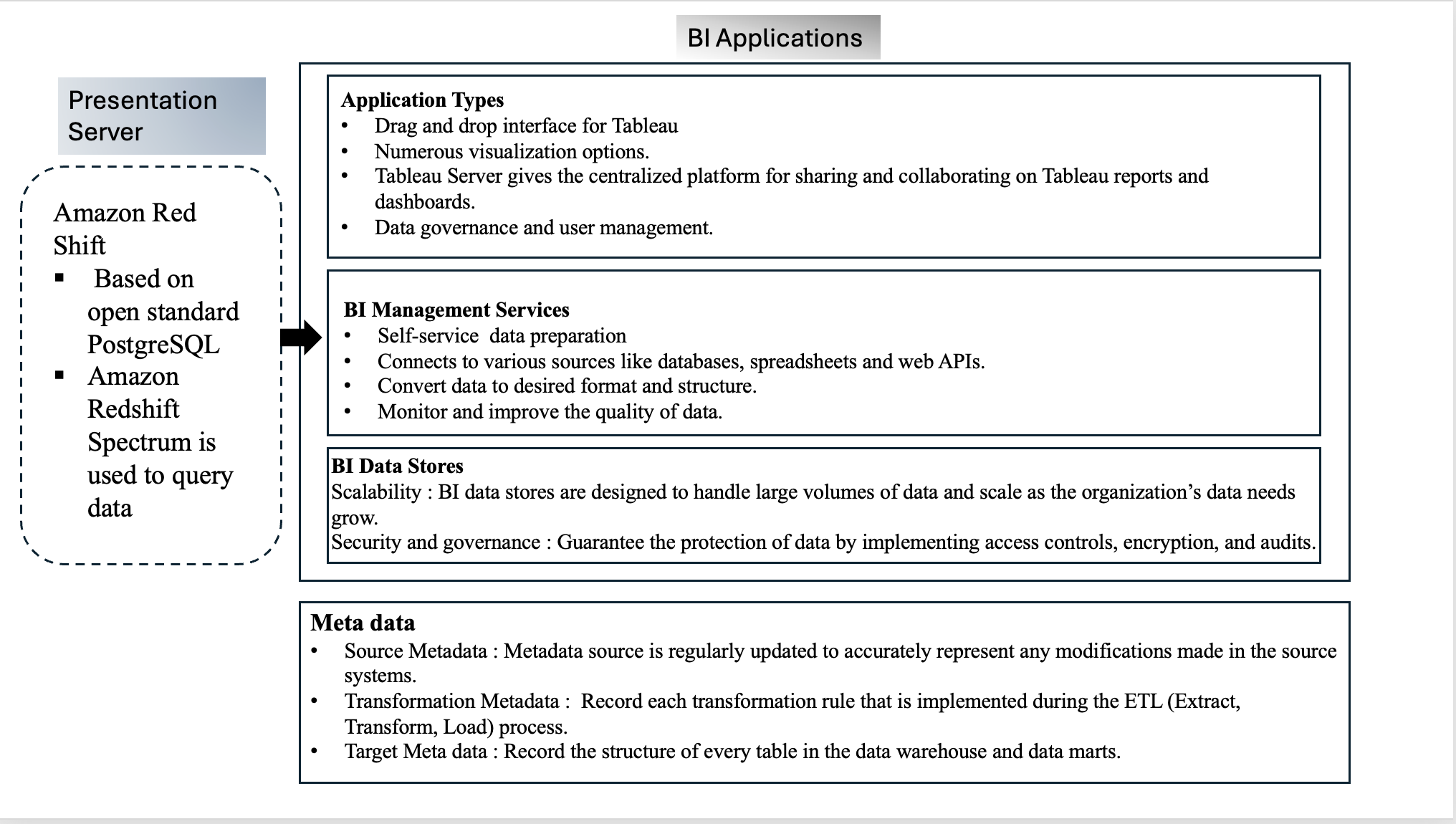


Fig 6: Front room technical architecture model

Tableau and Power BI provide a variety of specialized apps to meet the specific requirements of different departments in a business. These applications facilitate the access, visualization, and sharing of data insights for data analysts, decision-makers, and developers. Tableau's portfolio comprises Desktop, Server, Online, Public, Prep, and Mobile, whereas Power BI's suite consists of Desktop, Service, Report Server, Mobile, Embedded, and Gateway. Each application type has a distinct purpose, ranging from data preparation and visualization to sharing and embedding, to provide comprehensive support for business intelligence and analytics.

# Conclusion and Next Steps

For Puglia Winery, the implementation of a data warehouse and business intelligence tool represents a crucial step toward enhancing the winery’s operational efficiency, scalability, and strategic decision-making. As Puglia Winery ventures into larger markets and seeks to diversify its wine offerings, it is essential to their success that they adopt a data management solution that can provide the reports and statistical analysis necessary for directives.

In conclusion, by leveraging a star schema data model within Amazon Redshift, the proposed data warehouse solution will streamline data organization and improve query performance. The star schema’s simplicity, containing only the necessary dimension tables, will facilitate clearer insights into sales trends, product profitability, and customer behavior. Dimension tables for dates, products, and customers, along with a centralized sales fact table, will enable comprehensive analysis while maintaining ease of use for the director.

The ETL process will play a pivotal role in addressing the challenges of data fragmentation and format inconsistency. Amazon Redshift will serve as the core data warehouse, consolidating data from production and sales systems. The ETL process will involve extracting relevant data, transforming it into a unified format, and loading it into the data warehouse. Data cleaning, standardization, aggregation, and denormalization to enhance query performance and data accuracy will compile the key transformations. Near real-time data loading will ensure that decision-makers have access to the most current information.

As for next steps, we would recommend hiring a contracted data architect or a full time business intelligence role to help implement the data warehouse, build reports, and guide initial analysis. Having an expert in the fields, especially one who is able to actively engage with the data and have relevant experience within the wine-market will be most influential.

With the integration of Tableau for visualization, Puglia Winery will be equipped to generate detailed and interactive reports. These reports will cover various key performance indicators (KPIs) such as sales volume, revenue, production costs, market growth, and profit margins. The insights derived from these reports will enable the company and all internal stakeholders to make data-driven decisions, optimize production and sales strategies, and effectively manage market expansion.