

CASE STUDY

Business Intelligence Case Study: Sylt Fish Specialties

Aaron K. Althausen

Correspondence:
aaron.althausen@iubh.de
IU International University of
Applied Sciences, Berlin, DE
Full list of author information is
available at the end of the article

Abstract

The article details a business case for Sylt Fish Specialties GmbH as a practical application of Business Intelligence tools and procedures. Acting as an official report from a management consultancy firm, the document will provide technical advice on the digital optimization of the company's procedures. A reporting system will be made available to the company that utilizes a BI dashboard using Apache open-source technology. A Data Warehouse and associated data models will be developed and presented to the company as a digital solution. Finally, test analyses will be completed to present the functionality of the proposed system.

Keywords: Business Intelligence; Data Warehouse; Data Analysis; ETL; SQL

Executive Summary

The following report outlines the propositional design and development of a prototype BI (Business Intelligence) application reporting system for Sylt Fish Specialties. The report incorporates a proposed data model and DWH (data warehouse) design, tools for ETL (Extract- Transform-Load) processes and provides test analyses of hypothetical business questions using the suggested frameworks.

Sylt Fish Specialties is a restaurant chain with the business model of fresh fish provided via select inland routes. To maximize profit for the company, logistical operations need to be refined and evaluated completely. A prototypical BI application needs to be constructed in a way that allows all operational data to be queried and visualized in an ad hoc manner and by all user groups of the company. This will serve as the backend for a reporting system that will provide daily updated data accessible to all operational areas of the business. A set of test analyses need to be made to showcase the functionality of the proposed system and to ensure it is appropriate to query the company's operational data. All raw data used to exhibit the functionality of the design, along with SQL extractions and test data source code, is fully available online at the provided link.

The report will introduce a solution based on open-source technology previously demonstrated in various use cases for successful companies. Data warehousing, dashboard development and task scheduling will be provided by Apache Druid, Superset and Nifi, respectively; with PostgreSQL being utilized as a transactional processing database, and multiple ETL tools that provide a GUI (graphical user interface) for querying data. Test data is generated in Python to demonstrate various features of the working prototype, which is loaded into PostgreSQL to simulate a functioning OLTP (online transactional processing) system. The DWH design will make available crucial data to the company's business analysts and stakeholders alike,

providing immense value to the company. The power of Druid as a DWH lies in its unique segmented style of data storage that allows processing of transactional data to be indexed in time chunks, where the data stored in deep storage can be recalled easily by means of user-defined intervals of time. The Superset framework being used as a dashboard development tool alongside Druid makes the pair a valuable combination when analyzing and presenting data for stakeholders.

Findings

Analysis 1: Black Cod in Berlin

Berlin has the potential to be named the ‘cultural capital’ of Germany. It is a metropolitan region that will be crucial to the company’s successes when considered properly. There is no better place than Berlin to monitor KPIs (key performance indicator) as they pertain to one of the sought-after new specialties, Black Cod. A seafood known as a delicacy in markets like Japan, Black Cod is native to Alaska and requires efficient resources to distribute to Europe. Sylt Fish Specialties have been able secure a supplier of the delicacy, and therefore need to observe its performance as a menu item. A particular interest would be to understand which precise time of the day customers will prefer to eat Black Cod. Using test data as a replacement for what will occupy the company’s databases, we created a dashboard [1] that monitors the key metrics involved with this problem.

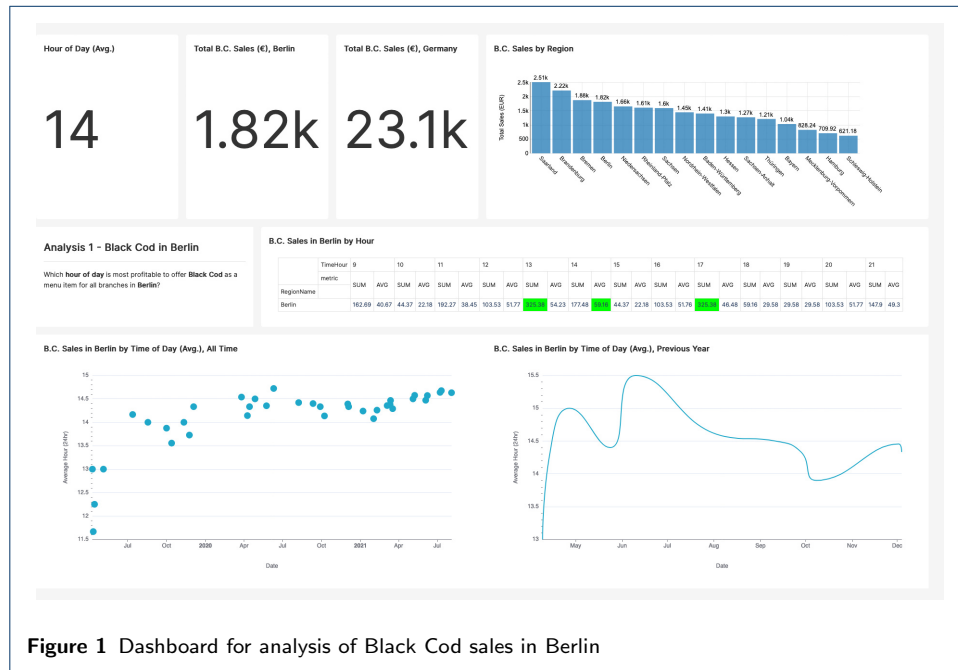


Figure 1 Dashboard for analysis of Black Cod sales in Berlin

We found the hour of 14 (2 pm) as the average order time for this menu item. Sales for this item were at their total maximum in the afternoon hour of 13 (1 pm), while having the highest average order total of €59.16 between 14:00 and 15:00. As seen on the dashboard’s scatter chart, the average hour of consumption fluctuates from 11.5 to 15 over the timespan of two years, with the average hour increasingly trending upwards. In the interval of just the previous year, the average hour starts around 15 in the start of the year, then drops to 14 in the autumn months. We

might also like to track which branches or regions are selling Black Cod the most, which our bar graph on the dashboard shows as the Saarland region, with €2.51k total sales of the item. Berlin is not far behind with €1.82k total sales, while the whole country has seen €23.1k total Black Cod sales.

Analysis 2: Saarland's Best Sellers

As a region with the smallest population, comparatively, in Germany, Saarland is worthy of tracking activity coming from its branches. Also being a landlocked region, which borders the rich cuisine culture of France, any company that provides fresh seafood should make calculated business decisions if expecting profit from the region. Especially in the summer months in Southwestern Germany, where the logistics of seafood travel become more complex. The test data powering our dashboard [2] has shown that this type of analysis is quite simple.

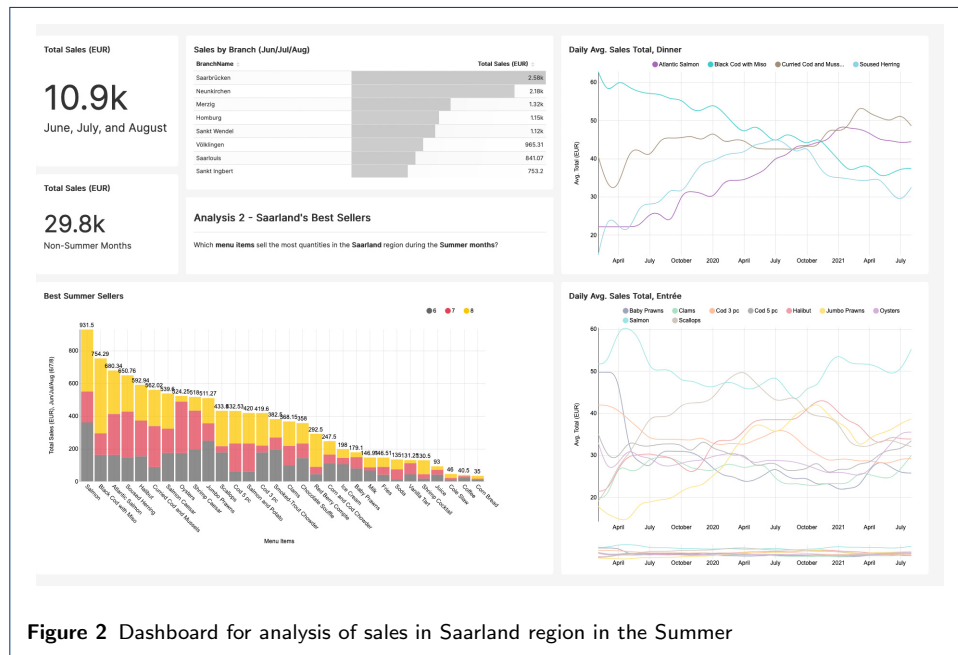


Figure 2 Dashboard for analysis of sales in Saarland region in the Summer

In the months of June, July, and August, Saarland saw €10.9k total sales, for a somewhat expected 26.78% of its total sales. The highest grossing branch, Saarbrücken, managed €2.58k total sales in the summer months. The best-selling menu item in the summer, Salmon, contributed €931.5 in sales, with the most sales coming in August. For items in the Dinner category in Saarland, the dashboard's line chart shows a daily average sales total just above €40 for all items in the category; for Entrees, the line chart shows a more diverse average across all items throughout the two year timespan.

Analysis 3: Breakfast Regions

When creating a menu that will fit with the business model and satisfy customers, it is difficult to understand how varied the menu should be; especially when the business hours of operation span the whole day. Initially, Sylt Fish Specialties decided not to include a specific breakfast category in the menu plan. Making a sample analysis as such shows how the business would be able to analyze which items have the

best performance in the hours of breakfast (before noon), along with which regions and branches are performing well in these hours; therefore, if there is a decision to one day include breakfast as a menu offering, there would be an upper hand in knowing which items to include and where to focus the marketing efforts.

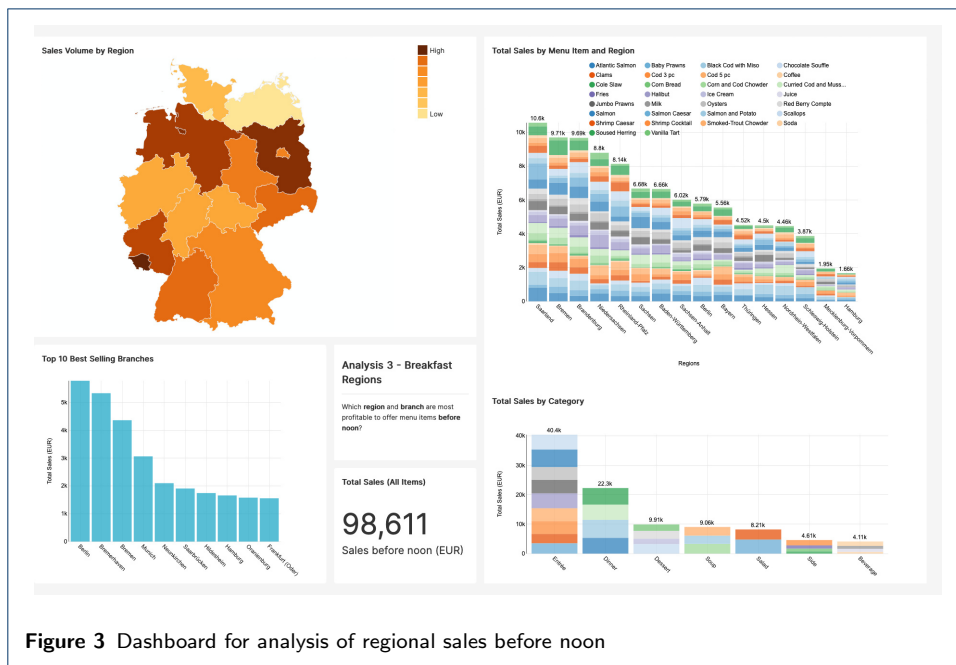


Figure 3 Dashboard for analysis of regional sales before noon

As our dashboard shows [3], the Saarland region is outperforming the others in terms of total sales before noon, with €10.6k total; Salmon and Halibut are a majority in the morning sales in Saarland, but as our stacked bar chart shows, the menu item preferences vary from region to region. The best-selling breakfast branch is Berlin (Berlin region) with Bremerhaven (Lower Saxony) not far behind, each with above €5k sold. Another stacked bar chart on the dashboard shows the category of Entrees as the highest seller before noon. The company has seen €98,611 total sales before noon.

Analysis 4: Bavaria Logistics

When analyzing inland routes to regions in Germany, the state of Bavaria is the most logistically challenging. Being in the more extreme climate of the southeast, the southernmost and furthest region from the ports of the Atlantic Ocean, Baltic and North seas, Bavaria has its challenges when fresh seafood needs delivered. The proposed dashboard [4] will display a sample analysis showing which time of the year is most affected, along with showing supplier performance in relation to the products being shipped and the time of year.

The average lead time for orders to Bavaria is minimized in Q1 (15 days), with Q2 Q3 and Q4 having average times of 19, 16, and 18 days, respectively. The heat map of average process time for orders shows a heavy lead time in the months of May-June and October. The tree map showing average lead time by product shows mussels as the quickest to be shipped, and scallops and cod as the slowest. In terms of supplier, Chord Seafood from Sweden have the lowest minimum and

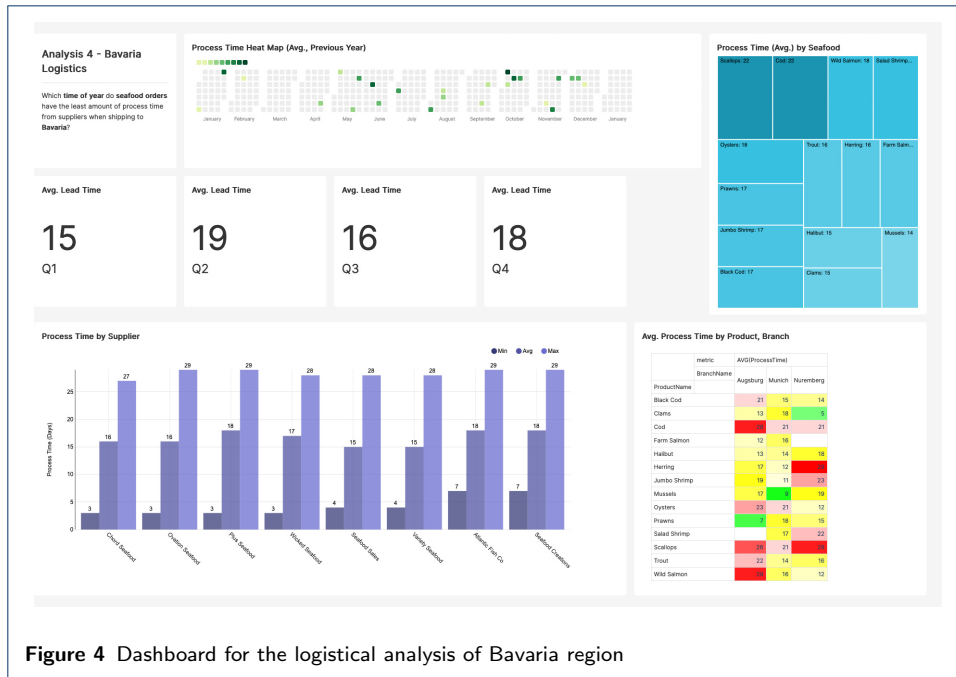


Figure 4 Dashboard for the logistical analysis of Bavaria region

maximum process times for orders, with average time hovering around 16-18 days for all suppliers. The branch in Augsburg, Bavaria, has the longest process times for orders, while the Munich branch seems to be more accessible for inland routes.

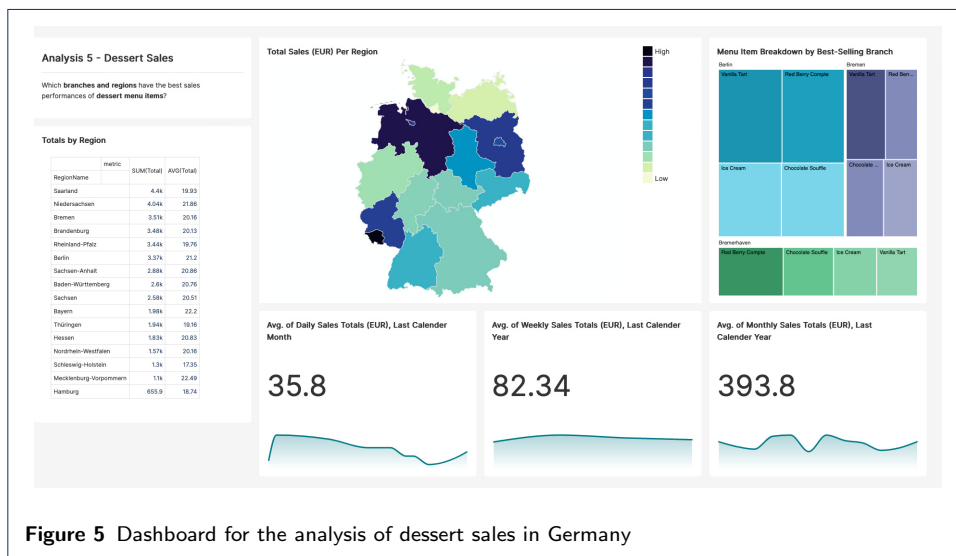


Figure 5 Dashboard for the analysis of dessert sales in Germany

Analysis 5: Dessert Sales

Apart from being a seafood distributor, Sylt Fish Specialties takes pride in their delicious desserts. While the options are slimmer for the sweet variety, plans to introduce a monthly special could be a positive trend for the restaurant. One would need to consider which regions are going to offer the most profitability for a dessert special, and which branch would be smartest to introduce the campaign at. The dashboard created for this purpose [5] is useful for such analysis.

Saarland seems to be the ideal region to introduce a new dessert: the state sold €4.4k worth of desserts in a two-year span. Other notable regions are Lower Saxony and Bremen, both with over €3.5k sales. Berlin is the best-selling dessert branch, with Bremen and Bremerhaven as the next highest grossing branches. As shown on the tree map, Vanilla Tart is the most coveted dessert at both Berlin and Bremen branches, with Red Berry Compote winning in Bremerhaven. Across all Germany, there was an average daily sales total of €35.8 for the last calendar month; a weekly sales total of €82.34 for the last calendar year; and a monthly average of €393.8 for the same year.

Discussion

Our sample analyses show that a blossoming company like Sylt Fish Specialties can scale profits using clever analysis of daily business. Being able to monitor company metrics in an efficient and cost-effective manner such as this is the most advantageous way to increase profits. When the company wants to know which hour to offer a discount on Black Cod, they can rely on the developed dashboard to understand the trends as they are happening; when the concern is about supplier performance for seafood logistics, it is also important to have quick dashboard access to understand the data. Apache Druid and Superset were used here in tandem to reflect the accessibility and power of the designs; but there are many options available that are up to the discretion of the client and what they are comfortable investing in. The recommendation to use open-source technology as a reporting system for company data is strong due to the vast open-source development support given to Apache products. The versatility of using custom Python scripts to develop reports and design dashboards in a framework written completely in Python (Superset) adds a gained advantage to the designed system because Python knowledge is common in the education path of data analysts.

As a DWH, Oracle Data Warehouse would provide sufficient storage capabilities and speed, but will be a costly expense and will require experts with specialized skills and certifications to build and develop; Amazon Redshift is built off of PostgreSQL with modifications such as a column-oriented concepts in how it processes data, but comes with systems in place called “vacuum” that automatically reclaim deleted tables as disk space; Apache Hive offers more traditional SQL querying on top of distributed systems, but often they are too complex to be setup in a timely manner.

For creating data dashboards, Tableau is a popular choice in the sector because it allows for easy integration with Google Analytics, but the cost of subscription per user is high; while Microsoft’s PowerBI is quite powerful also as an ETL tool, it has a steeper learning curve when compared to other options. What is ultimately chosen is dependent on the operational systems the company has in place and how it plans to extract the data from each system.

Recommendation

The idea that OLAP (online analytical processing) is an extremely advantageous system to use for business intelligence is at the heart of our design decision for this business case. Apache Druid as an OLAP database and DWH is currently in use by profitable companies such as eBay, Netflix, and Nielsen; companies that require

high- level batch and real-time processing of data. AirBnB uses Druid and Superset to power analytics and visualize data for the company and are developing ways to push these tools to their limits (Muthiah & Li, 2018).

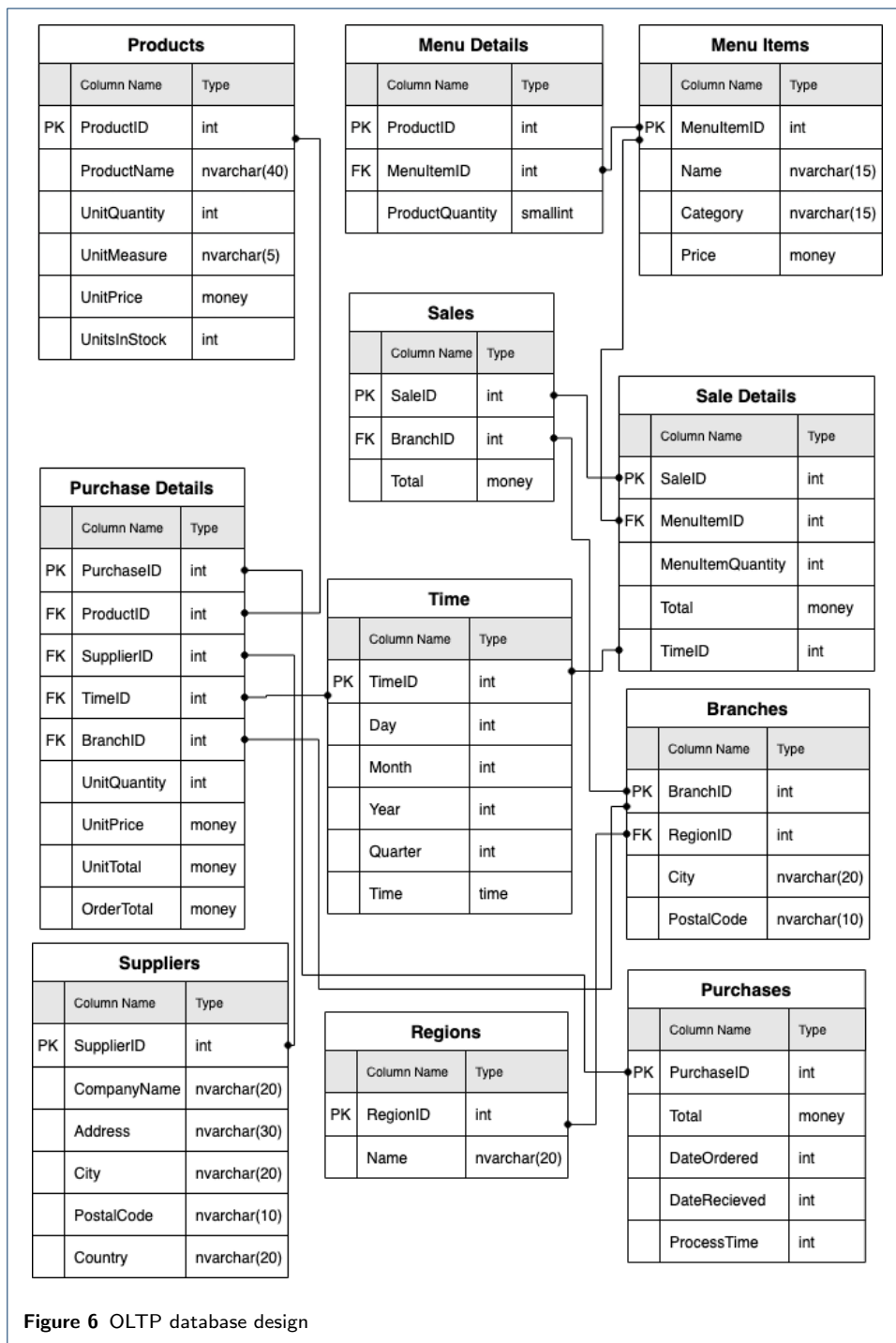


Figure 6 OLTP database design

Design

OLTP

The OLTP system [6] is designed to cover all aspects of the business operational data. For the sole purpose of brevity, some processes have been left out of the design and test data. As this follows the structure of a simple analysis, the main dimensions of the business' data covered in the report are Sales, Geography, Product, and Time.

Products. Ingredients and supplies that contribute to a Menu Item, such as flour or oil. *Menu Items.* Each offering on the menu that Sylt Fish Specialties offers to customers, such as Black Cod with Miso or Vanilla Tart.

Menu Details. All ingredients from the Products table that correspond to each Menu Item. Each Menu Item may have many Products associated with it.

Purchases. Orders of products that Sylt Fish Specialties makes to suppliers.

Purchase Details. All data relating to the details of each purchase, including time the order was made and received, which Products were included and how much they cost, which Branch made the order, and the total order cost.

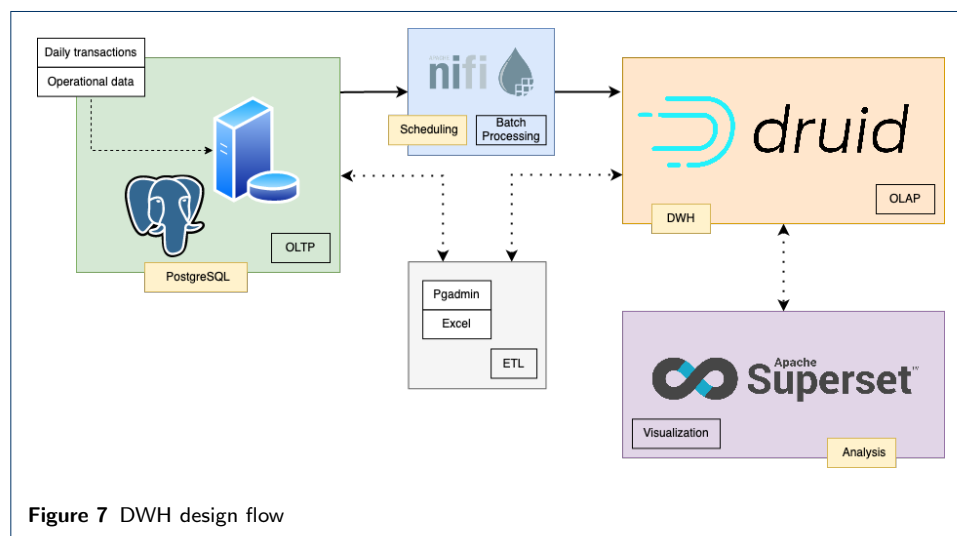
Branches. Geographical data about each of the 400 branches owned by the company.

Regions. Geographical data about each region within Germany.

Suppliers. Geographical data about suppliers used by the company.

Sales. Data about each Sale, including Branch ID where the sale was made and Sale ID.

Sale Details. Data about each Sale connected by a unique Sale ID. Includes Menu Items purchased, Time of order, quantity of Menu Items per Sale, and total Sale price.



Data Warehouse

PostgreSQL. Data from customer orders and product purchases at each branch are aggregated and sent to a central PostgreSQL database that serves as the OLTP system.

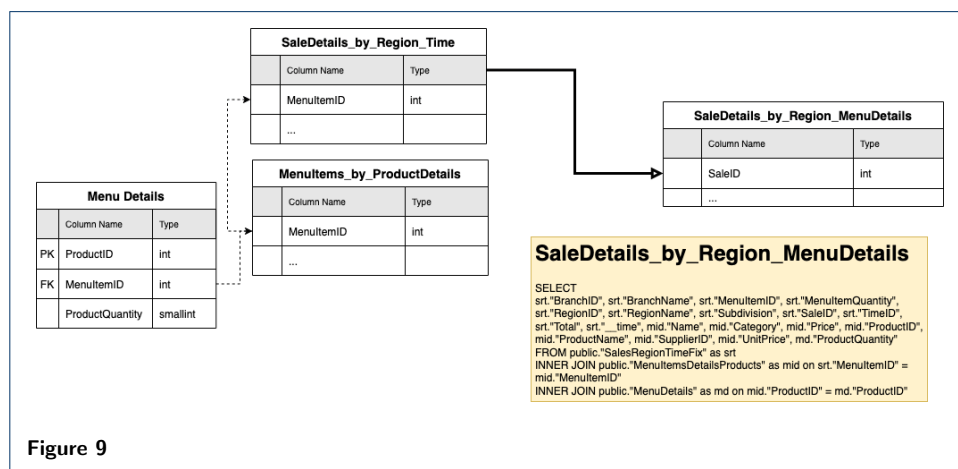
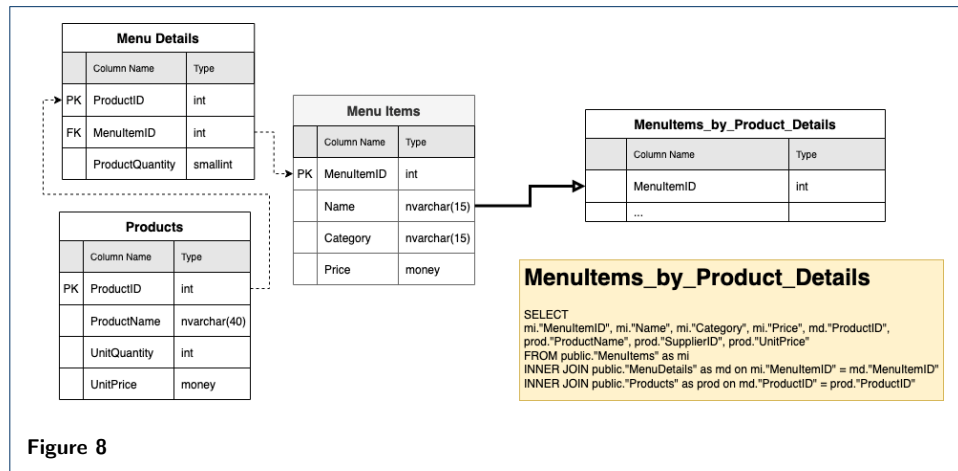
Data Warehouse. Data is ingested through batch deliveries and stored in segments in the Druid DWH. Incoming data from ETL processes is stored in deep storage

for the purpose of OLAP. Data from ad hoc queries is recalled using segments in the form of time intervals. Dimension tables are created allowing quick access to various business areas.

ETL. Data is extracted and transformed by analysts through SQL queries from the PostgreSQL database, from native queries in Druid, from SqlLab queries in Superset, or in the form of flat files from Excel. Data is loaded back into Druid after transformation from any of the ETL tools.

Scheduling. Daily batch jobs are scheduled that move replicated batches of transactional data into the DWH, making data readily available for analysts the following morning.

Analysis. After the creation of dimension tables in the DWH, data is queried into datasets within Superset, forming the basis for dashboard development. Once dashboards are published, data is connected from the DWH to the dashboard and receives real-time updates.



Implementation

Multiple different queries need to be performed by employees and users of the system. Below will be a demonstration and explanation of some example queries that may be necessary for daily monitoring and processing of data.

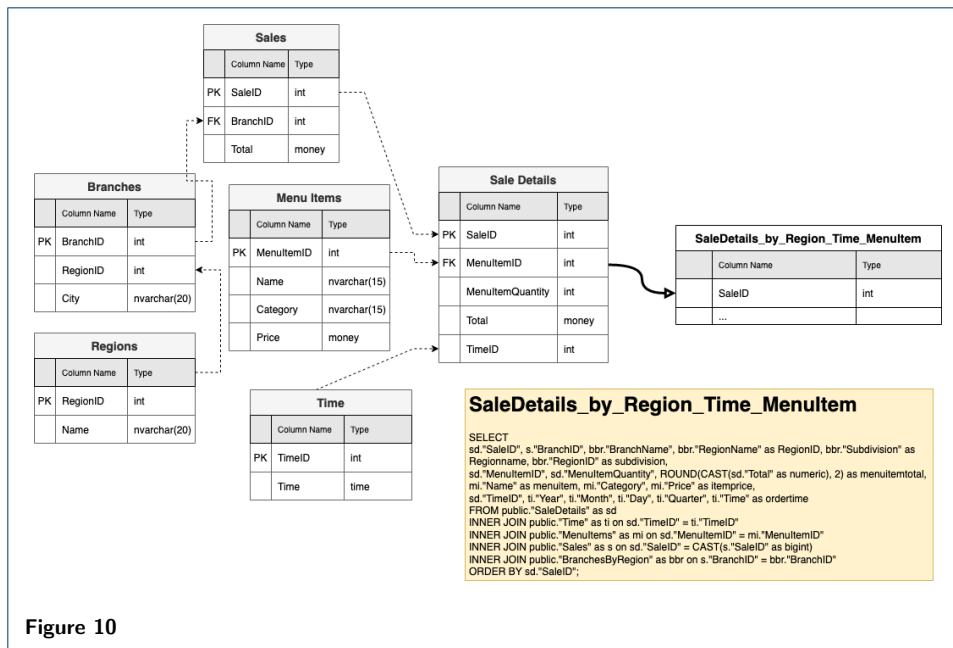


Figure 10

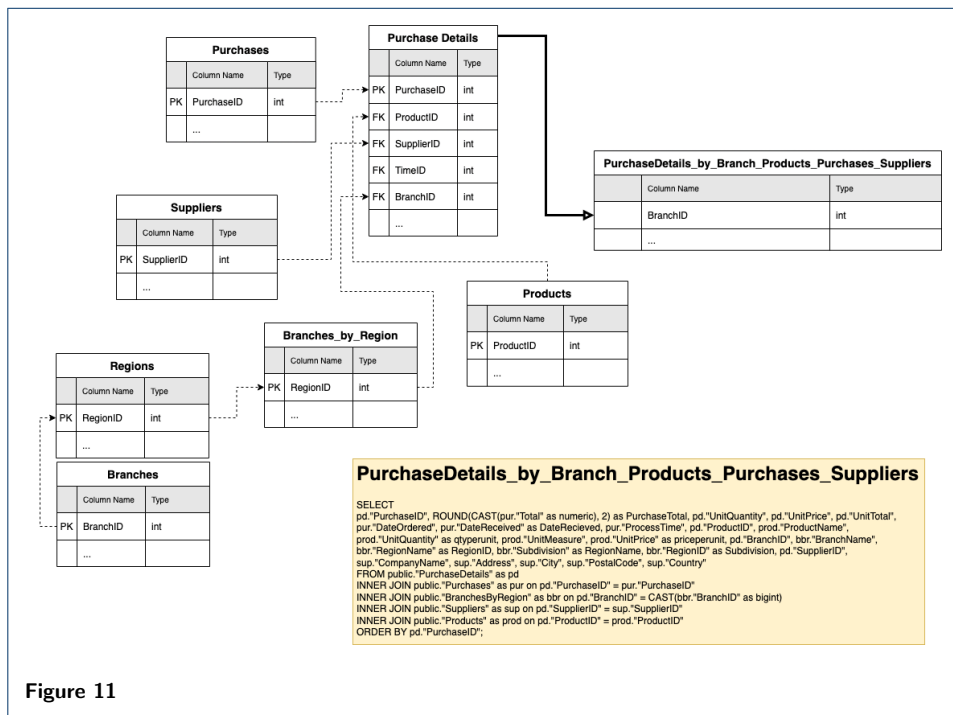


Figure 11

Menu Items by Product and Menu Details [8]. Used in multiple test analyses. Especially useful when querying for menu items, as the details of each item can also be extracted.

Sale Details by Region and Menu Details [9]. A useful query to analyze the cost of menu items per region. Utilizes a prior join to extend and reindex into a new data source in Druid.

Sale Details by Region, Time, and Menu Item [10]. To make analyses about specific items, which region they sold in, times they were ordered, and specific details about each menu item. After transformation, data is loaded into DWH using a flat file generated from Pgadmin, as a new data source entitled “SaleDetails.by.Region.Time.MenuItem”. Used in Analysis 1, 2, 3, and 5.

Purchase Details by Purchase, Branch, Product, and Supplier [11]. Used specifically for Analysis 4: Bavaria Logistics. It was necessary to join Products, Suppliers, and Purchases tables to monitor supplier performance in Bavaria region.

Conclusion

In the modern context, data is extremely powerful to own. Being able to make reports from different areas of business in a timely manner relies on that data being pre-processed and accessible to all users. Because not all users carry the same knowledge as a data engineer, the data also needs to be structured in a manner that is relevant to the task at hand. For a restaurant chain with close to 400 branches, the ability to monitor data closely and accurately is critical to the company’s success.

Our proposed solution of open-source tools to store data in a DWH will guarantee accessibility of data to all users and ensure that essential queries are being fed with clean and accurate data; all without requiring expert technicians as in-house staff. As this report has established, a data model based on open-source tools such as Druid and Superset leads the way to an elegant solution to a business case that requires a reporting system across all operational functions of the business.

Availability of data and materials

Source code, raw extracts, and images can be found online at: <https://github.com/aaronalt/BICaseStudy>

Competing interests

The authors declare that they have no competing interests.

Author details

IU International University of Applied Sciences, Berlin, DE. Frankfurter Allee 73a 10247, Berlin, Germany.

References

1. Garadi, M. (2019). Monitoring at eBay with Druid. <https://tech.ebayinc.com/engineering/monitoring-at-ebay-with-druid/>
2. Muthiah, P., & Li, J. (2018). How Druid enables analytics at Airbnb. The AirBnb Tech Blog [Medium]. <https://medium.com/airbnb-engineering/druid-airbnb-data-platform-601c312f2a4c>
3. Natarajan, K., & Khera, D. (2018). PayPal merchant ecosystem using Apache Spark, Hive, Druid, and HBase. <https://www.youtube.com/watch?v=HJvuU0CQS44>
4. Sykes, B. (2020). How Netflix uses Druid for Real-time Insights to Ensure a High- Quality Experience. Netflix Technology Blog. <https://netflixtechblog.com/how-netflix-uses-druid-for-real-time-insights-to-ensure-a-high-quality-experience-19e1e8568d06>
5. Yaffe, I. (2019). How Nielsen Marketing Cloud Uses Druid for Audience and Marketing Performance Analysis. User Stories. <https://imply.io/post/nielsen-marketing-audience-analysis>