TESLA

DATA WAREHOUSE INITIATIVE

Robert Kaszubski

IS549 – Data Warehousing

Final Project

Company Background

Tesla, Inc. is an American manufacturer of electric vehicles, solar panels, and other energy-based goods. It was founded in 2003 and named after inventor Nikola Tesla. The company's founding mission was to prove that not only could automobiles run on electricity but could be faster and all around better than traditional gas-powered cars. By 2008, Elon Musk took over as the Chief Executive Officer, and in the same year, Tesla released their first production vehicle, the Roadster. The Roadster was a luxury all electric sports/super car, the first of its kind. The plan for Tesla was to release an expensive high-end vehicle, then use the revenue generated to fund lower cost vehicles that were accessible to a broader market. The revenue from the roadster was used to develop the Model S, Tesla's luxury, and record-breaking sedan. The company followed up by releasing their first electric SUV, the Model X. Finally, in 2017, Tesla was able to achieve their goal of releasing a lower priced, high volume electric car, the Model 3 sedan. The Model 3 quickly became Tesla's best-selling vehicle and led to Tesla unveiling the Model Y, two years later. The Model Y came in at a similar price range as the Model 3 and was Tesla's affordable mid-size SUV. The Model Y lived up to the expectations set by the Model 3 and began rivalling the Model 3's sales. Likewise, Tesla has an entirely separate energy division. They started manufacturing their battery packs in-house both for their vehicles and for residential use, as well as selling solar panels and state of the art solar roofs Tesla's success, however, has not been without any problems, some dealt with, others continuing to hinder the company's growth.

Currently Tesla manufactures their vehicles in two factories, Tesla Fremont in California is responsible for all vehicles made in North America, while Gigafactory Shanghai in China primarily produces vehicles for Asia and Europe. More Tesla factories are on the way, Gigafactory Texas and Gigafactory Berlin are close to opening and beginning production. Tesla also has several new products lined up. The company's futuristic pick-up truck, the Cybertruck, has already acquired well over a million preorders since its unveiling. The Cybertruck is being paired with a new ATV bike. Another exciting development is the release of the first electric truck, the Tesla Semi. With so much in the pipeline, the company will be under more pressure than ever to execute. During the initial design and production of the Model 3, the company had numerous logistical, and operational issues. The assembly processed was cited to be a mess, the cars had several defects and customer complaints, and the delivery of the car experienced numerous delays. Luckily the team was able to sort most of this out, but another similar situation could arise at any moment. The Covid-19 pandemic and subsequent supply shortage recently put additional strain on the company. Tesla needs to always be ready to adapt and make quick but informed decisions. Currently, customers are waiting upwards of six months for their vehicles to arrive. The demand for the vehicles is there but the production isn't, it is up to the business executives and leadership to navigate the company through troubled waters and continue their success. A crucial step in moving forward is the development and implementation of a data warehouse to store key information from their numerous sources that can then be used for business intelligence and analysis.

Business Objectives and Imperatives

Tesla's primary objective has always been to accelerate the transition of the world to sustainable energy. Their official goal statement is to "Create the most compelling car company of the 21st century while driving the world's transition to electric vehicles". However, no matter how well meaning they may be, the bottom line is that they are a business, and a business's main objective is to generate revenue. Many have noted Tesla's overvalued stock price in relation to comparatively low revenue and small profit margins on their vehicles. To generate more revenue, Tesla wants to drive up production to meet demand, expand to new locations, and importantly lower the cost of manufacturing. The long-term objective is selling affordable cars that almost everyone could purchase. To do so requires continued research and development particularly in battery and energy technology. Customer satisfaction is another key business objective. Many customers report long wait times to receive their vehicles, and vehicles arriving with several defects that require repair. Tesla wants to maintain a positive public image that will encourage potential customers to order their vehicles.

Tesla's biggest imperative must be meeting demand and delivering cars to customers promptly soon after they order them. This is key to success and if Tesla is unable to do so much longer, the company will face more backlash and customers will seek out other electric vehicle manufactures. A six-month or longer wait is not an option for many nor are many willing to wait that long. Customers report minimal communication from Tesla along with constantly shifting delivery days leading to frustration and a bad reputation for the company. Recently, more and more car companies are entering the electric vehicle field. Tesla must solve their supply and manufacturing problem to continue their growth and success. It is imperative that Tesla scales meaningfully and quickly to hold onto their majority in the electric vehicle market share. It is also key that Tesla continues to address their production issues for greater speed and efficiency at each of their factories. This should raise the overall customer satisfaction score, another key imperative if Tesla wants to survive.

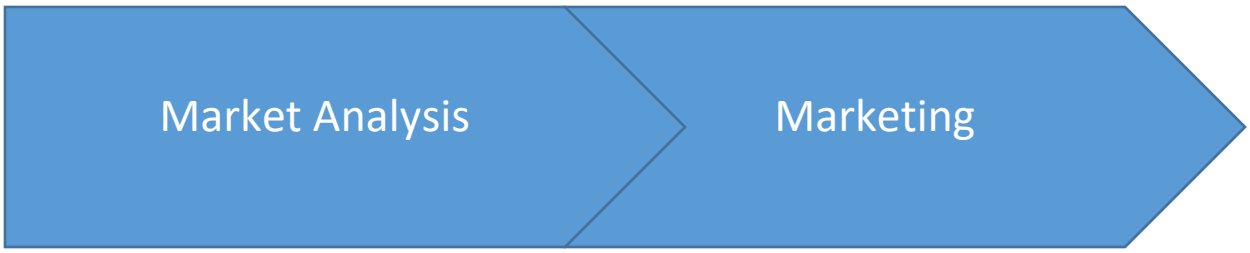
Several areas of the company should be monitored and improved. Information needs to be gathered showcasing the company's key performance indicators. Activities along the manufacturing process need to be observed to show where production is stalling and when parts are running out. Data needs to be gathered on deliveries and inventory. Daily figures should show the average cost and production time for each Tesla model. The company needs to be looking at speed and efficiency at each of their factories on a per model basis. The company also needs to gather sales information including total profits, total vehicles sold, and the total backup in orders left unfulfilled. Monitoring customer feedback and satisfaction scores will also be key at improving product quality.

Problems / Gaps Addressed by Data Warehouse

1. Ability to keep up with demand
2. More efficient manufacturing and production
3. Addressing shortage of key parts and equipment
4. Constant necessity for rapid growth and expansion
5. Improving customer satisfaction scores
6. Reducing car defects and improving product quality
7. Keeping track of employee wellness
8. Improved efficiency filling company vacancies
9. Increased effectiveness of marketing campaigns
10. Ability to maintain larger market share and top competitors
11. Ease of access to company information and key performance indicators
12. Ability to perform business intelligence analysis
13. Ability to make informed business decisions
14. Identifying areas of improvement
15. Tracking financial performance
16. Risk Management

Business Process Maps

Order/Sales	Manufacturing	Delivery
Order Agreement / Contract signed Product configuration confirmed Estimated Delivery Date generated Handle customer financing / billing	Inventory and Supply Availability Check Product Assembly Product Testing and Quality Control Time for Assembly	Transport Product to Delivery location Customer accepts product and signs paperwork Collect customer survey
Order numbers exceeded expectation x% of the time Total orders per store above x Total revenue above $x Total order backup under x%	Parts available for vehicles as needed x% of the time Vehicles assembled on time x% of the time Under x defects x% of the time Total manufacturing cost under x Total scrapped cost under x	Vehicle delivered on time x% of the time Perfect customer satisfaction score x% of the time 0 customer found defects x% of the time



Research and collect data on electric vehicle and energy competitors

Analyze Market Share and top performing regions

Create targeted photo and video advertising

Track page views, traffic, and order conversion

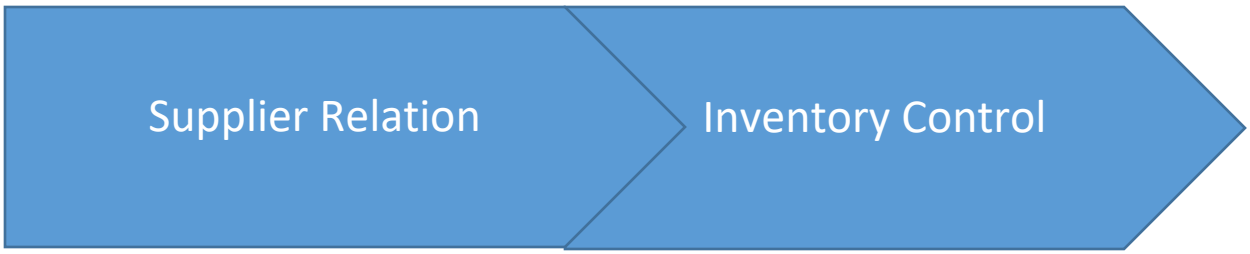
% of market share in state/country exceeds x%

% of market share increases over x time

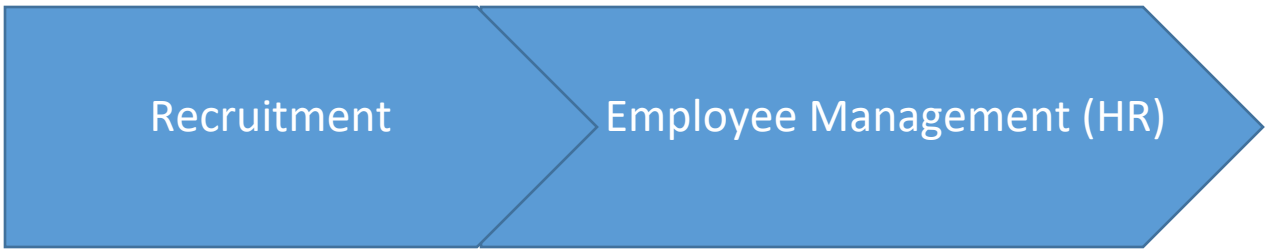
Total $ spend on advertising

Advertising traffic exceeds x%, x% of the time

Advertising converts to orders x% of the time



Part orders placed Supplier parts availability Part delivery frequency and scheduling	Inventory supply check and record management Warehouse to Factory delivery frequency and scheduling Part Space/Capacity designation
Parts orders fulfilled on time x% of the time Order quantity available as needed x% of the time Total $ spent on parts	Inventory supply checked x times over x time Number of parts available exceeds x% of total capacity



Create new job listings

Send out recruitment emails to potential hires

Process Resumes

Conduct Interviews

Hire new employees

Handle and resolve employee conflicts

Keeps records of salary and process payroll

Handle employee benefits

Send employee surveys

x number of vacancies x% of the time

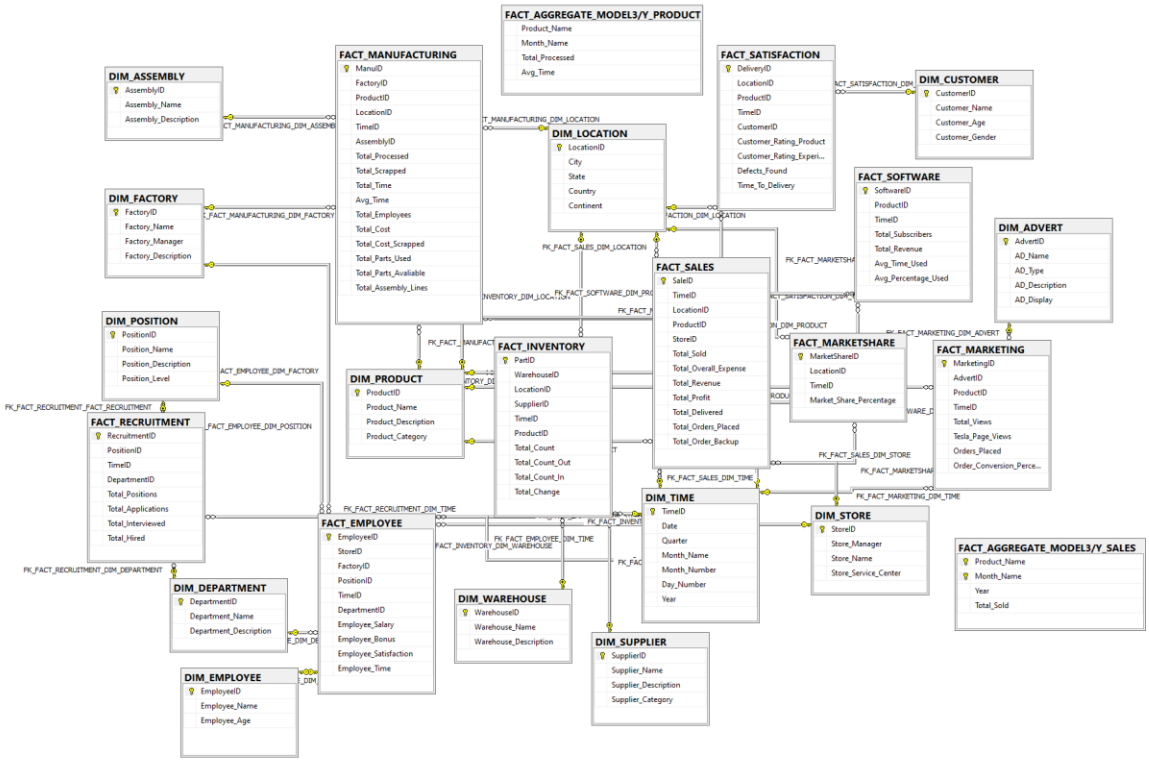
number of new hires over x period of time

number of resumes to job openings

Average salary increases over x% during x period of time

Employee satisfaction over x for x% of the time

Dimensional Model



Data Quality

Data Quality and ETL is done using SSIS packages created in SQL Server Import Export Wizard. These extract, transform and load the data into a specific table while validating it and checking that all constraints are met.

Error Checking

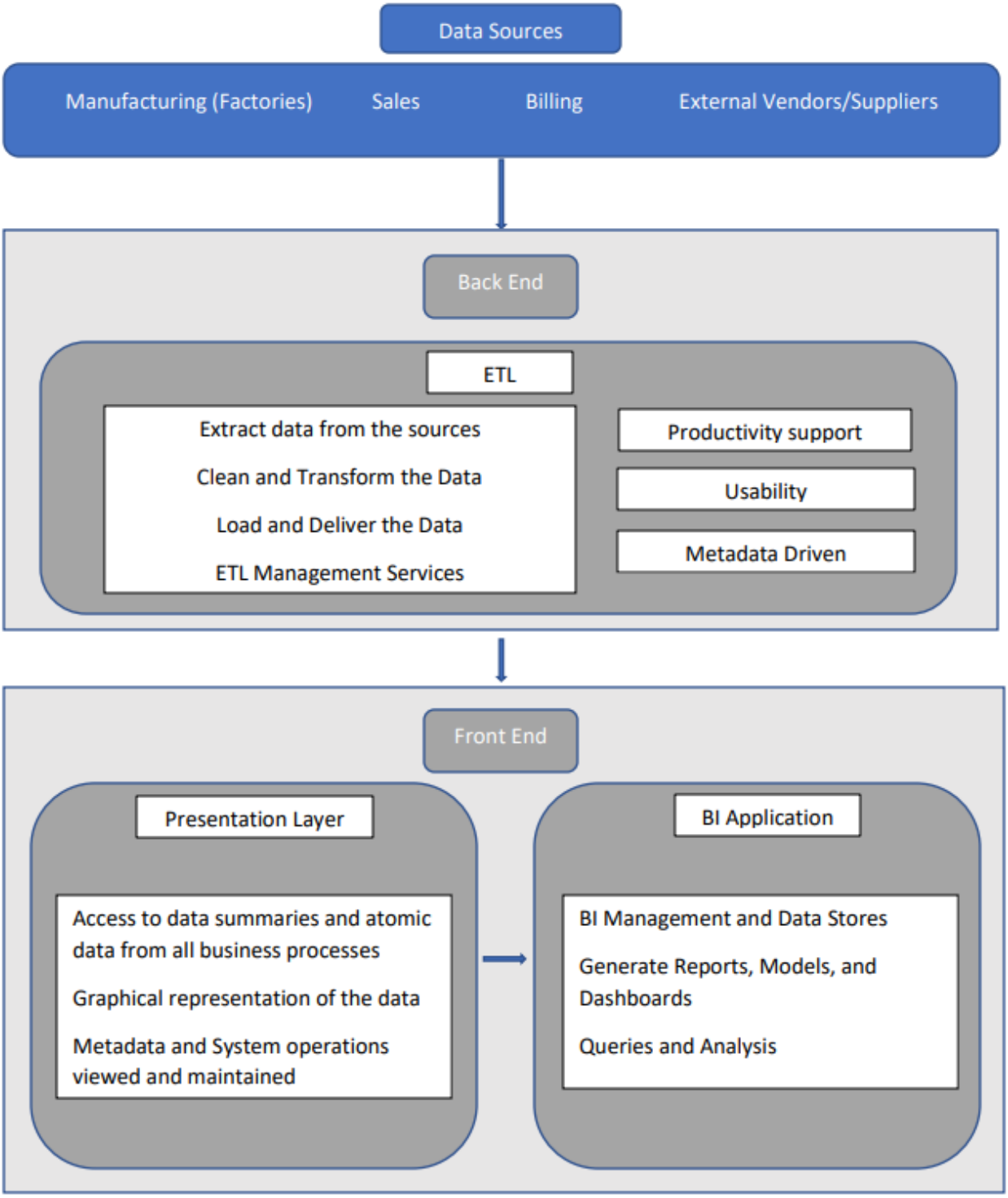
Area of Applicability	Error	Definition	Example
Fact tables	Foreign Key Constraints	Each foreign key in a fact table row must already be in a dimension table	ProductID must exist in the DIM_PRODUCTS table prior to importing fact table
Table Columns	Column Count	Number of columns imported must match columns at destination	DIM_POSITION has 4 columns, the data imported must also have 4
Data Values	Data Type Check	Data type must match destination column,	int to int, string to string, date object to date object
Data Length	Data Length Too Long	Each column has a specific number of bytes	Varchar(10) – maximum of ten characters
Table Columns	Mapping	Each column must be mapped to the correct destination column	FirstName -> Name Salary -> Salary
Data Values	Truncation	Inserting more characters than possible to a column	Text file has 10 characters, ssis creates buffer of 50, column should be 10 – this will work, but if the column is 9, it'll fail
Data Values	Data Conversion	Converting an inappropriate data type to another	Alphanumeric string to an integer won't work

Standards

Area of Applicability	Standard	Example
TimeID – Time Dimension	Letter T followed by abbreviated date mmddyy	T010121 = 01/01/2021
FactoryID – Factory Dimension	Country Code + Number	US01 – Fremont Factory
ProductID – Product Dimension	Model Code + Year Released	MS21 – Model S 2021
CustomerID – Customer Dimension	Country Code + Number (max 8 digits)	US78797414
AdvertID – Advert Dimension	AD + Number	AD1
AssemblyID – Assembly Dimension	A + Number	A009 – Defect Checking
DepartmentID – Department Dimension	DP + Number	DP001 - Assembly
PositionID – Position Dimension	P + Number	P0001
StoreID – Store Dimension	Country Code + Number (max 5 digits)	US04311
SupplierID – Supplier Dimension	S + Number	S0123
LocationID – Location Dimension	Country Code + Number (max 8 digits)	US413412311
PartID – Part Dimension	PRT + Number	PRT1
RecruitmentID – Recruitment Dimension	R + Number	R141342
MarketingID – Marketing Dimension	MK + Number	MK1
SoftwareID – Software Dimension	SF + Number	SF1
ManuID – Manufacturing Dimension	M + Number	M41231
SaleID – Sale Dimension	S + Number	S1
Sales Fact, Manufacturing Fact, Employee Fact, Software Fact	Any cost/profit or revenue figures are all reported in $	$1,000,000
Employee Name, Customer Name, Manager Name	Are always reported First (space) Last	Elon Musk
Average Time, Total Time	Any measures of KPIs that are time based are reported in minutes	10 minutes
Market Share Percentage, Percentage used	Any percentages are reported as an integer from 0-100	50 = 50%

Architecture

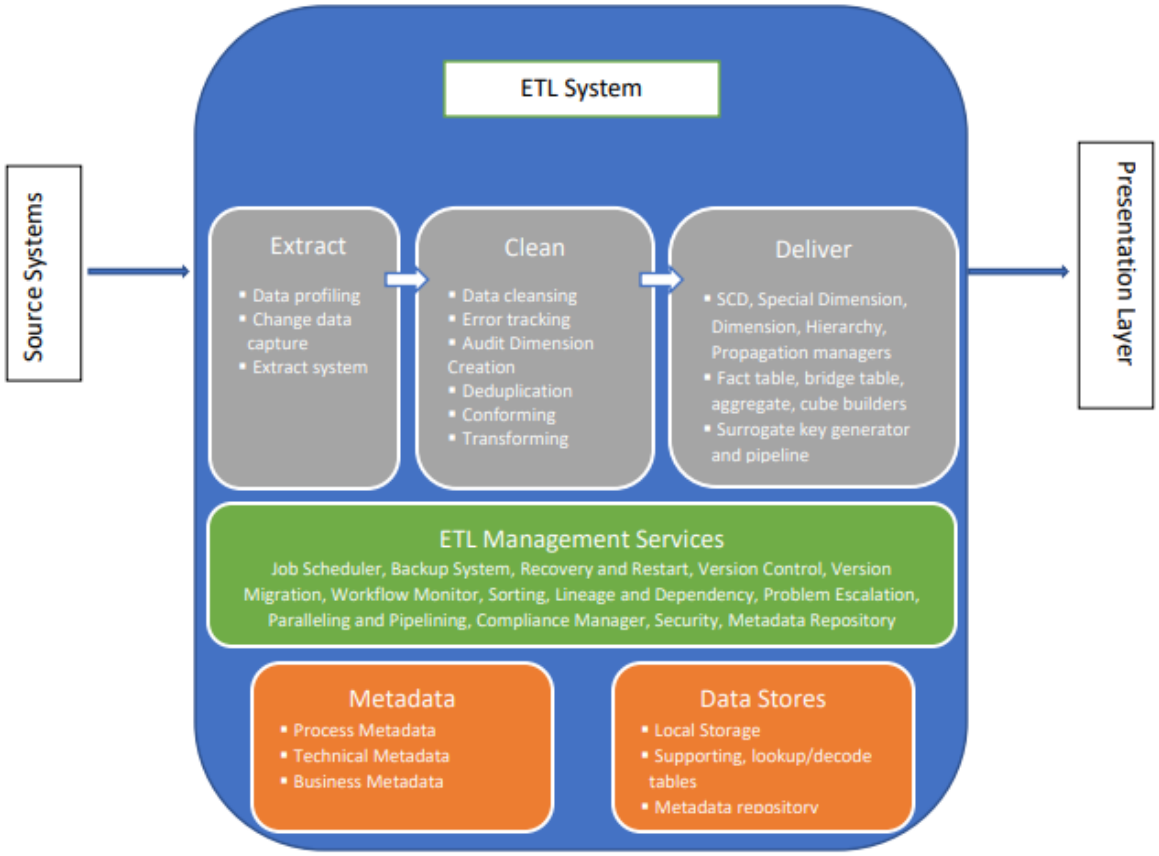
a) End to End Architecture



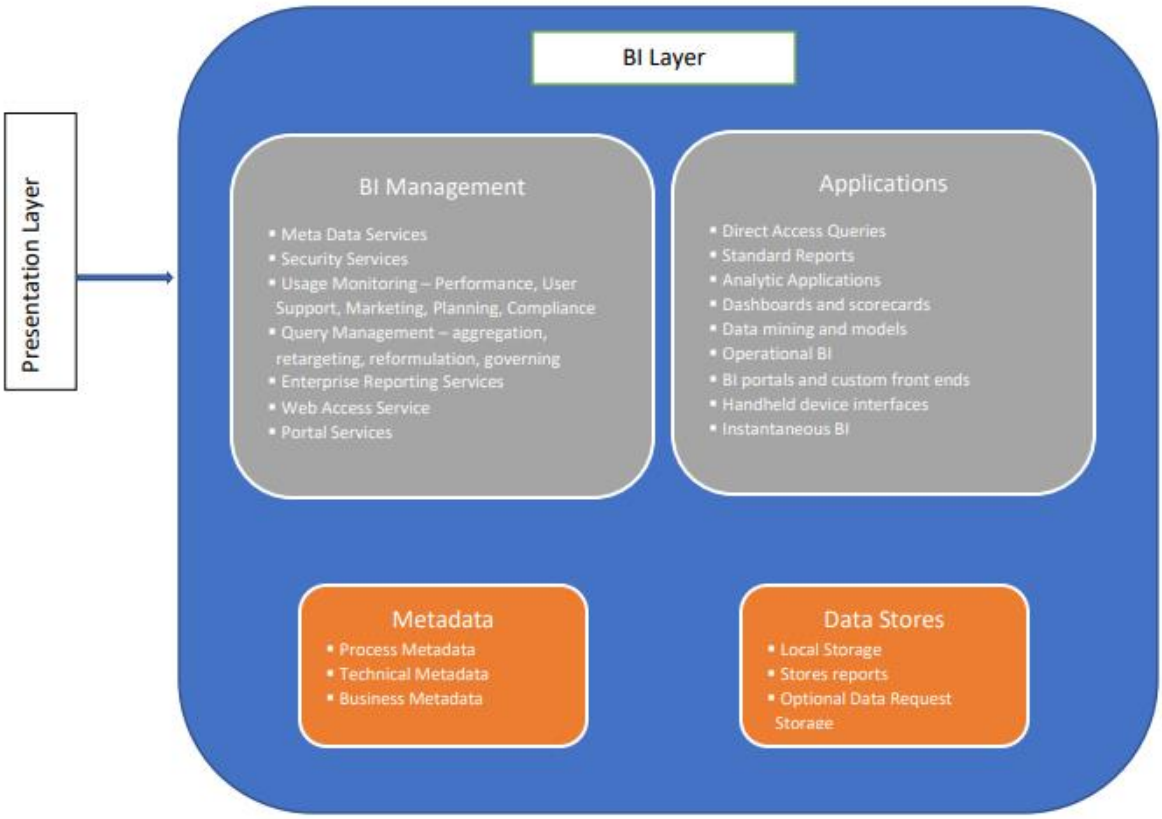
b) Source Systems

Source	External/Internal	Business Owner	IS Owner	Platform	Location	Description
Billing	Internal	Elon Musk	Marion Christensen	Windows	HQ - LA, California	Customer billings
Comptetitor Sales Data	External	Regina Cole	Adrian James	Windows	Chicago	External spreadsheet of competitor sales
Customer Service	Internal	Elon Musk	Johanna Craig	Windows	HQ - LA, California	Customer ratings and complaints
Giga Berlin	Internal	Elon Musk	Tricia Rice	Unix	Factory - Berlin	Made in Europe M3/Y manufacturing and shipment database
Giga Nevada	Internal	Elon Musk	Darren Jensen	Unix	Factory - Nevada	Battery Pack manufacturing - supply chain - inventory
Giga New York	Internal	Elon Musk	Marion Moore	Unix	Factory - New York	Solar panel and energy product manufacturing database
Giga Shanghai	Internal	Elon Musk	Brent Dawson	Unix	Factory - Shanghai	Made in China M3/Y manufacturing database
Giga Texas	Internal	Elon Musk	Deborah Williamson	Unix	Factory - Texas	Model Y and Cybertruck manufacturing database
Market Analysis	External	Dwayne Franklin	None	Windows	Denver	Externally sourced EV market analysis spreadsheet
Orders and Sales	Internal	Elon Musk	Henry Walton	Windows	HQ - LA, California	Customer order configurations and sales figures/ forecasts
Panasonic	External	Kazuhiro Tsuga	Stacey Wallace	Windows	Plant - Nevada	Battery shipments from plant
Tesla Fremont	Internal	Elon Musk	Lynne Park	Unix	Factory - California	Model 3 manufacturing database

c) ETL Components

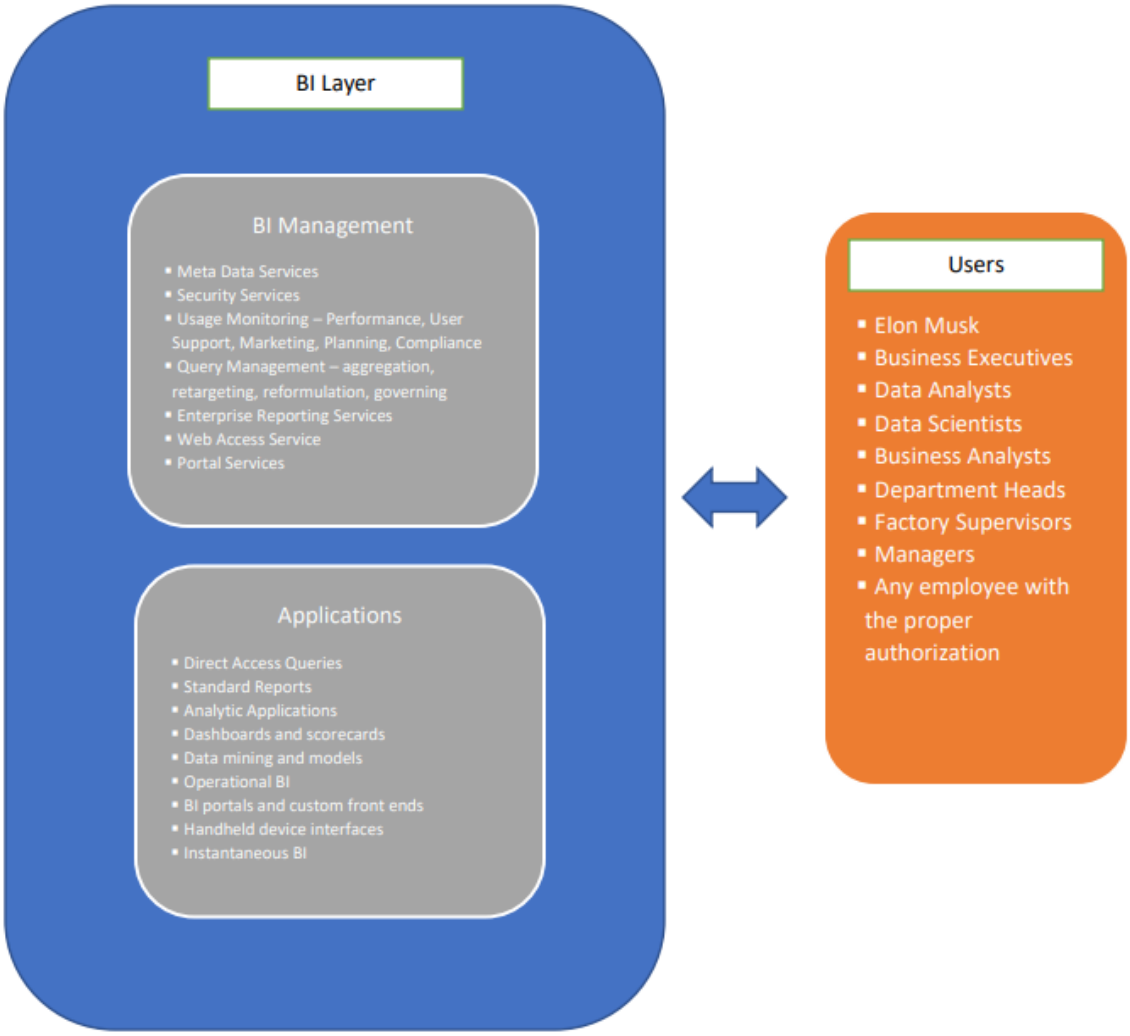


d) BI Tools

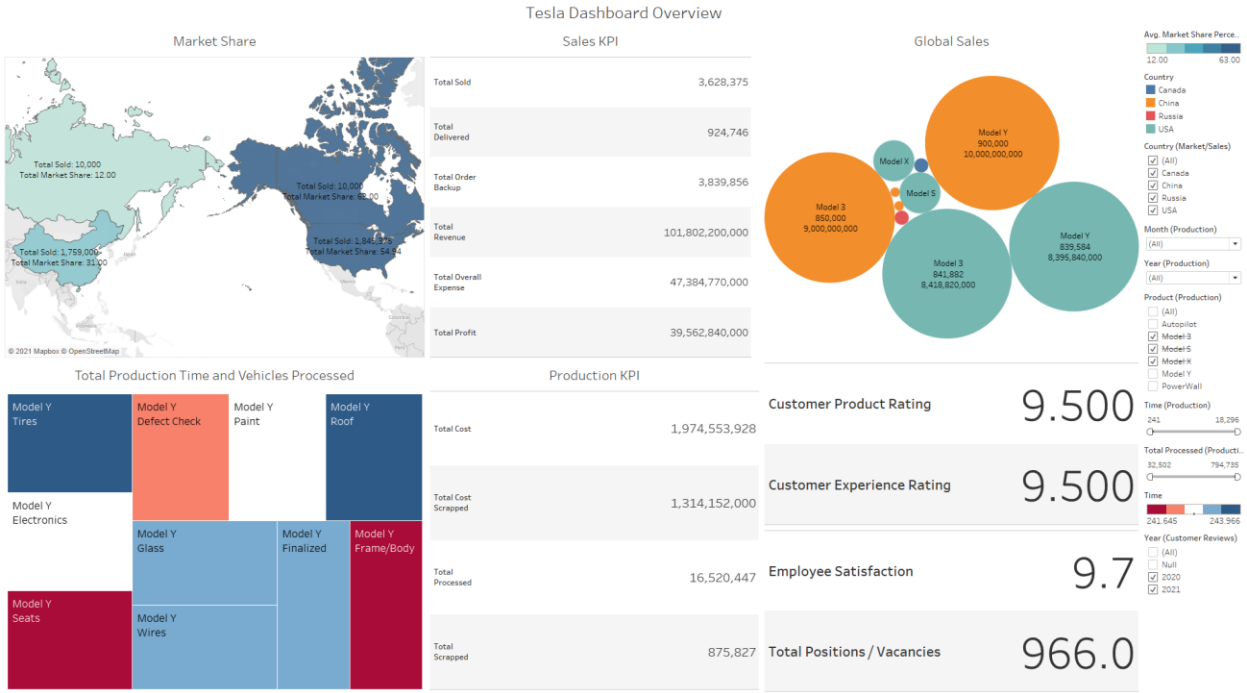


The BI Tool used for this project is Tableau. However, other BI tools such as Power BI or custom-made BI Tools can easily connect with the Data Warehouse.

e) BI users

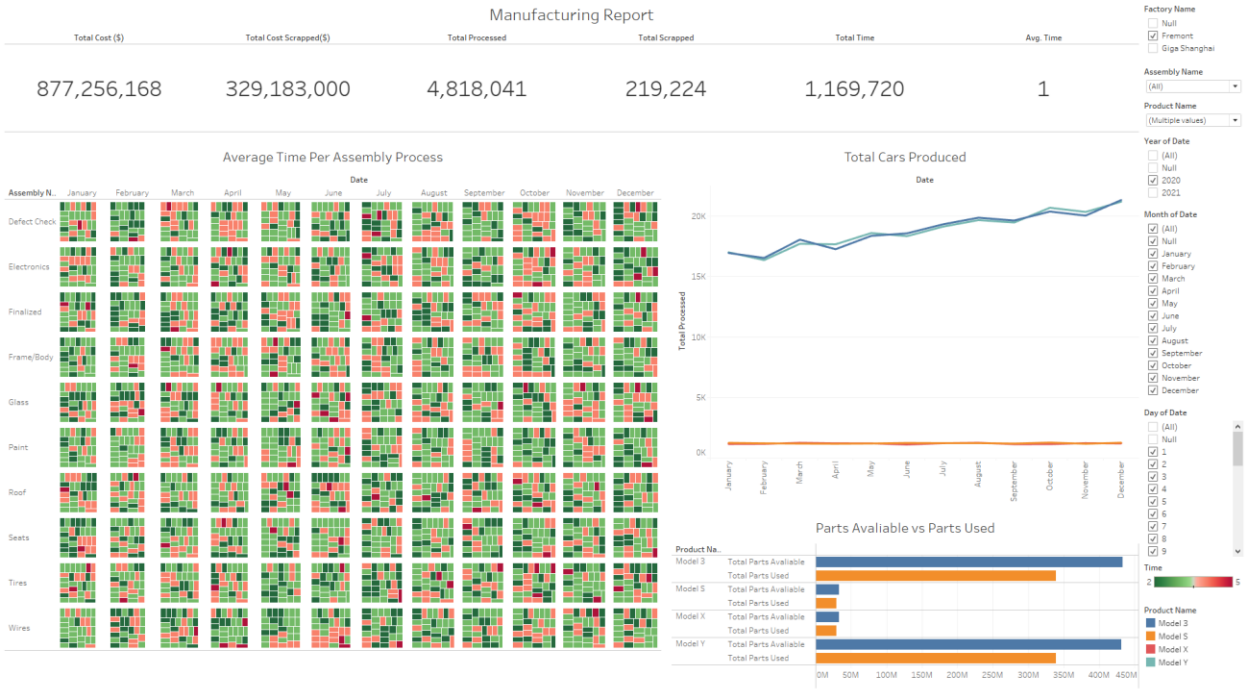

BI Layer
BI Management
Meta Data Services
Security Services
Usage Monitoring – Performance, User Support, Marketing, Planning, Compliance
Query Management – aggregation, retargeting, reformulation, governing
Enterprise Reporting Services
Web Access Service
Portal Services
Applications
Direct Access Queries
Standard Reports
Analytic Applications
Dashboards and scorecards
Data mining and models
Operational BI
BI portals and custom front ends
Handheld device interfaces
Instantaneous BI
Users
Elon Musk
Business Executives
Data Analysts
Data Scientists
Business Analysts
Department Heads
Factory Supervisors
Managers
Any employee with the proper authorization

Dashboard

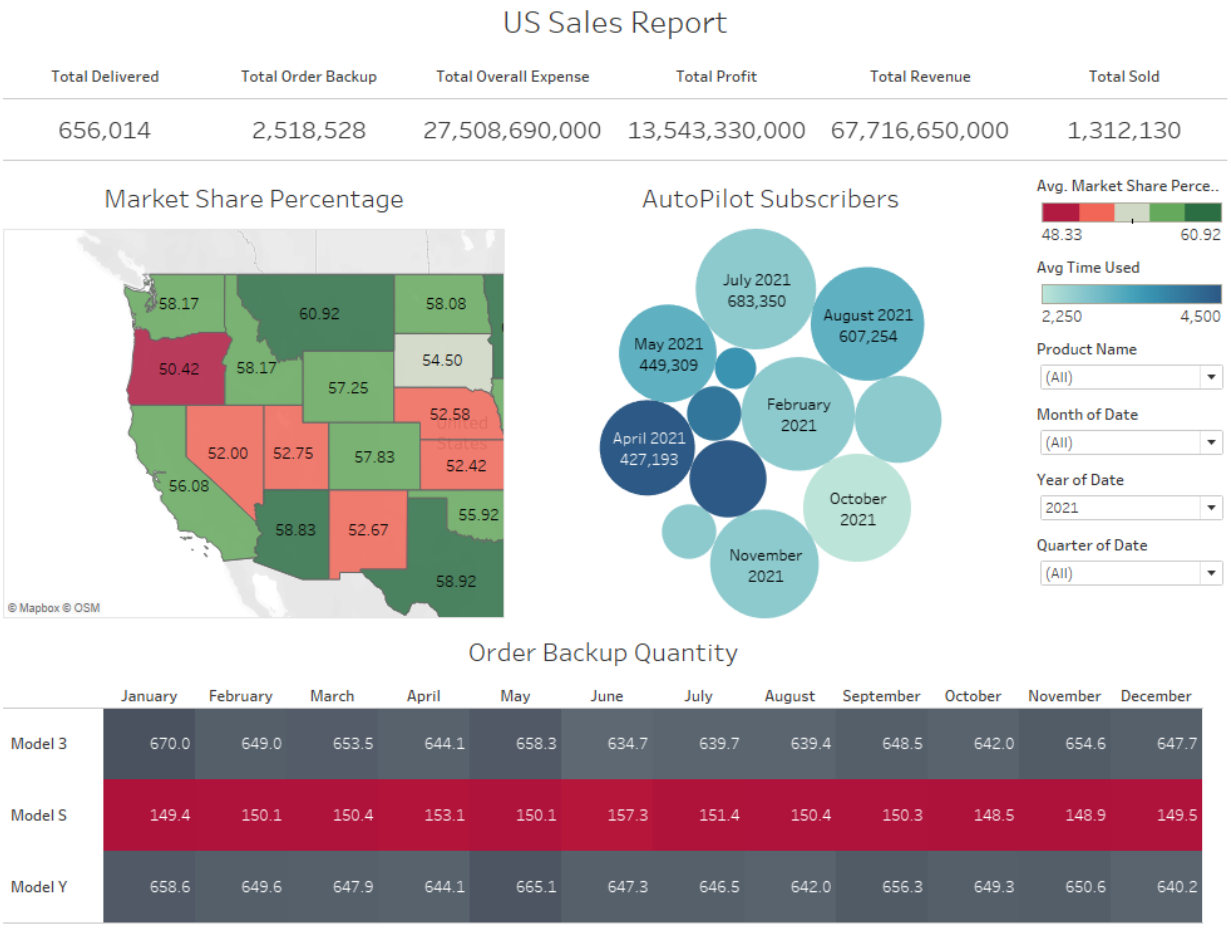


This dashboard was made in Tableau and showcases several key performance indicators from a broad holistic perspective of the company. There is Map showing international market share on a per country basis. The map also showcases total sales in that region. There are tables showing sales KPIs such as Total Overall Cost, Total Orders Pending (Backed), and Total Profit. There is a bubble chart that is showcasing global sales per Model of vehicle, per country. There is a tree diagram that breaks down production processes by model and by assembly process. This is color coded based on the time taken in each assembly process. The size is determined by the number of cars that pass through that process during the given time. There is likewise a Production KPI table showcasing Total Production Cost, Scrapped Costs, and Production Quantities. Finally other KPI are Customer Ratings, Employee Ratings, and the total number of vacancies that urgently need to be filled.

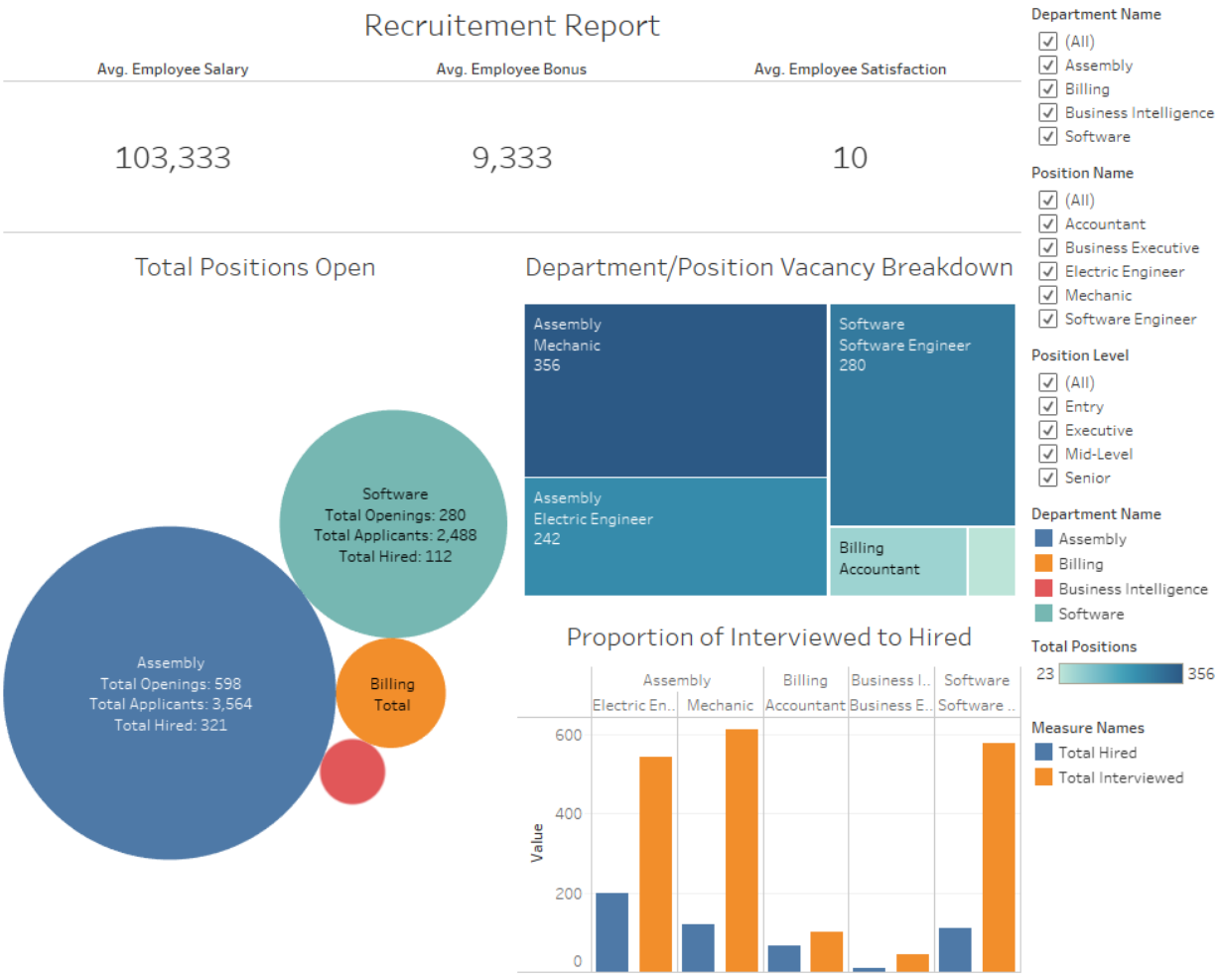
Reports



This report is made in Tableau and represents the Manufacturing side of the company. The KPIs displayed here are the Total Cars Produced, the production time at each of the assembly phases, the parts available to used ratio, the total cost of production, and the total cost and quantity of all scrapped product. I used a heatmap to display average time for each assembly process. This plot shows every day, every assembly process, and uses color to represent time. It can easily be used to show trends or hiccups in the production process. You can easily see days with production problems highlighted in red.

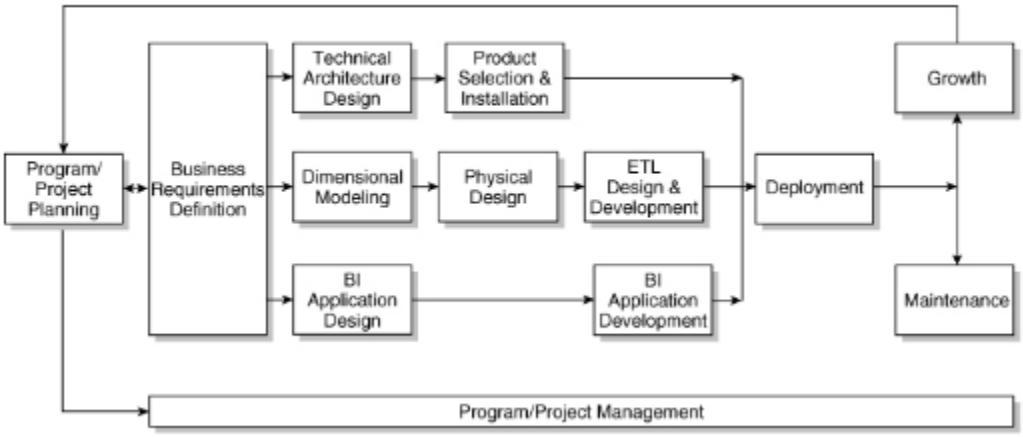


This report was made in Tableau and shows sales in the United States. The KPI represented here are market share, total cars sold, total cars delivered, total revenue, total expenses, total profit, the number of orders that are backed up and unfulfilled, and the number of subscribers. This is fully interactable using the filters on the right and can be sorted based on product name, or the date.



This report was made in Tableau and shows employee and recruitment information. At the top we see overall statistics about the company such as the current average salary, employee bonus, and employee satisfaction scores. This is information recruiters could use as they are key performance indicators in the company that represent how well Tesla treats their employees. Other KPIs showcase the number of openings or vacancies that need to be filled, the number of applicants, and the number already hired. This can be filtered based on the position, department, or position level.

Kimball's Roadmap



Program/Project Planning

A program encompasses multiple projects while a program would refer to just a single iteration of the Kimball Lifecycle. Building a data warehouse is a long and ongoing process that requires adaptability and typically multiple iterations as it did in my case. It may start as a project, but it requires program planning and coordination. Program/Project planning relies heavily on having knowledge of the business's requirements as well as good resource allocation and task identification / delegation. Building my data warehouse was a solo endeavor but these concepts do still pertain. The business requirements were always front, and center as was the thought of who would be using the Data Warehouse.

Program/Project Management

Project and program management is an ongoing process during the entire duration of a data warehouse. This involves monitoring project status, tracking issues, and overall leading the cause. It also requires good communication with the BI people that will be using the data warehouse, and the IT team.

Business Requirements Definition

A Data Warehouse initiative requires a good understanding of the requirements of the business as well as the business users. This typically means conducting a lot of research on the company as well as conducting several interviews with several people working in different departments that will be using the Data Warehouse. It is important to identify the key business processes along with their activities and key performance indicators (KPI). These become the foundation of the Data Warehouse being the central pieces of information and data being stored. These metrics are what business intelligence people and executives can use to evaluate the company's performance and make key business decisions. I did a lot of research on Tesla and similar companies to see what the business processes and KPIs could be as detailed in the process maps above.

Technical Architecture Design

A Data Warehouse requires the integration of several different technologies. The technical architecture design plan illustrates this showcase both the back end and front end. The Back End consists of data ingestion from the source systems and ETL. Once the data is cleaned and transformed, it is loaded into the dimensional model, and the front end, made up of the presentation layer and the BI layer, has access to the data and its users can analyze it. This is the technical architecture that Kimball recommends a Data Warehouse should follow and it's the one that I have used.

Production Selection & Installation

This simply refers to picking the right tools for the job. This can include the database management systems, the ETL tools, the BI tools, or any of the hardware. For this data warehouse initiative, I utilized SSIS Import Export for ETL, Microsoft SQL Server Management Studio for my database management system and environment, and finally I used Tableau as my BI Tool.

Dimensional Modeling

The Dimensional Model consists of fact (numeric measures / KPI) and the dimensions (context to those facts). My dimensional model follows the star schema recommended for Data Warehousing and doesn't have any snowflaking. It is also easily expandable should the necessity for new fact and dimensions tables arise as it did during the making of this data warehouse. The key part was defining the grain of the facts. I opted to process daily data in most of the fact tables although it was tempting to go as low as to show each individual car produced. However, this granularity would have been inefficient, not offered much more important information, and would have resembled more of a transactional database. Logical design is a part of this as the relationships among the entities is defined.

Physical Design

Physical Design refers to turning a logical model into a physical one as in building the tables, dimensions, and implementing any integrity constraints. This also meant setting up the initial database environment in Microsoft SQL Server Management Studio. The model itself however ends up being virtually identical to the one design during dimensional modeling. Another key aspect at this point is standardization. I came up with proper naming conventions and standards that apply to each table and are important to make sure there is no variation. Nulls are strictly forbidden as we want complete data. There are only a few places in which there isn't a required value such as in the state column for the location dimension. Other countries don't have states so the data must have a string that says "N/A".

ETL Design & Development

My data warehouse utilizes SSIS for ETL. The ETL process takes data from the source systems by extracting it, transforms and conforms it, then loads it into the dimensional model. Data

arrives on flat files to simulate data coming from different source systems. This data is checked for errors specifically relating to data type. Each data row must abide by the constraints of the table it is being inserted into. By default, txt files are all made up of strings, so each data point going to a column that isn't a string has to be transformed during the ETL process. Given that made up data is used, there isn't too much ETL involved here as there would be in a real-world application. Even the simplest table may require extensive data cleanup or at the very least a surrogate key. A lot of my data was generated using python scripts to simulate the data that would arrive from Tesla's factories, this process handles some of the ETL such as assigning surrogate keys.

BI Application Design

This doesn't apply to my project as I am using an existing BI Tool. All I had to do was find a tool that had all the BI capabilities that I required which Tableau did. Of course, my data warehouse can easily connect to other BI tools as well such as Power BI, if other tools were to be required. BI application design consists of figuring out the user's needs and capabilities for the BI tools.

BI Application Development

BI development is building or constructing the BI tools and all required infrastructure or metadata based on those discovered needs and capabilities during the previous step. This did not apply to this initiative as I was using an existing BI tool.

Deployment

Deployment is when everything starts coming together. It generally means that the ETL, BI, and all architecture is ready and just needs to be testing. Documentation is made of the entire system, extensive testing is conducted, and the business users are trained on how to use it. This was the portion of my project before dashboards and reports were created. Data was imported and tested albeit not anywhere near as expensively as in a real-world scenario but I ensured everything was working properly.

Maintenance

Maintaining the data warehouse is key after it is deployed. There need to be support systems in place in case something goes wrong. A data warehouse would be pointless and lose credibility if it went down especially at a key moment. There isn't many steps to keeping it up and running but they must be followed.

Growth

Growth is a sign of success. The great thing about the Kimball approach is that it is very scalable. In my scenario, Tesla recently announced that they would begin offering subscriptions for their autopilot software. I decided to add in additional tables with KPIs relating to software as well as adding autopilot to the existing list of products. This was easy and a good example of how growth needs to occur in a data warehouse environment.