Datawarehouse Design for part sales of Hyundai Motor America (HMA)

Prepared for:

Professor Turel

Prepared by:

Chris Barcinas
Edward Chang
Cyndi Chee
Kedar Kulkarni
Daniel Ulmer
Eryu Wang

Table of Contents

Overview

The end goal of this system proposal is to enhance reporting and analysis of parts sales key performance indicators to help Hyundai Motor America (HMA) and its dealers develop ways to increase parts sales revenue. Parts sales is not only an integral component in helping Hyundai owners maintain and repair their Hyundai vehicles, but also an important profit center for HMA. Thus, increasing parts sales is one of the top objectives for HMA's Service Division every year and improving reporting and analysis of parts sales will help HMA achieve its parts sales goals.

Description of Organization and Need for Business Intelligence

Established in 1986, HMA has been marketing and selling new Hyundai vehicles in the United States through its network of over 835 Hyundai dealerships. HMA also supports and services its franchise dealers and customers by wholesaling and distributing vehicle parts to Hyundai dealers.

There are separate business systems for storing parts sales and service transaction data like SAP and IBM AS/400. These entities make up HMA's source systems. Various industry standard ETL services are needed internally to copy all the sales, vehicle, dealer, and service data from these disparate systems into a combined enterprise data warehouse or relational database management system (RDBMS) like Oracle to support various business intelligence (BI) needs. It will be ideal if this ETL process denormalized all the necessary data since performance is an issue when querying and analyzing large volumes of parts and service data.

Analysts should start using newer BI and data visualization tools like Alteryx, SAS and Tableau to access to the RBDMS' denormalized data marts they need to analyze parts sales by dealer, location, vehicle type, vehicle age, mileage, etc. Management can use data insights from new standard reports and dashboards to develop new promotions or programs aimed at increasing parts sales.

The HMA organization is capable of implementing this proposed solution because the solution enhances existing ETL processes and provides a query optimized data mart for aforementioned BI needs. Table 1 below shows the high level enterprise data warehouse bus matrix for HMA.

	Common Dimensions						
Business Processes	Date	Vehicle	Dealer	Location	Customer	Service Rep	Parts Detail
Parts Sales	x	x	x	x	x	x	x
Parts Inventory	x	x	x	x			x
Repair Orders	x	x	x	x	x	x	x
Appointments	x	x	x	x	x	x	x

Functionality and Features of Desired BI System

The key business process that we will focus on in this BI system is Parts Sales. We choose to focus on this KPI because the profit margin for parts sales is higher than that of new cars sales. And, there are significantly more service customers purchasing Hyundai parts than new car buyers purchasing new Hyundai vehicles. For example, HMA sells around 700,000 new cars each year, whereas customers performing repair or maintenance on their vehicle exceed 9,000,000 service visits per year. During a customer's lifecycle, a customer may potentially service their car at a Hyundai dealership for many years after buying a brand new car.

The types of reports that the BI system produces will be both ad-hoc and auto-generated. Parts sales reports and service transactions reports are examples of auto-generated reports that will be available on a daily, weekly and monthly basis. HMA also would like to improve data visualization in reports because current analytic process creates reports with basic tables and simple charts in Microsoft Excel. It would also be useful for HMA and dealership management to be able to view a dashboard that shows real-time information such as parts sales velocities and parts sales opportunities to allow for better parts order planning and parts sales promotion management.

These reports will mostly be used by analysts and upper management of HMA, but it will also be shared and distributed to dealerships to help parts and service managers forecast future parts sales and service needs, and increase the parts sales performance of service consultants. HMA will also develop a model that analyzes a vehicle's service history and present information to service consultants about cars that customers bring in for service and alert the service consultants to recommend the customers to replace certain parts as needed based on age of vehicle.

Dimensional Modeling

The design technique of dimensional modeling is selected to structure part sales data for four reasons. First, dimensional models are more understandable to business users compared to normalization models. Although the model structure is simple, it represents the complexity of the business. Second, the query performance of dimensional models is more efficient because the predictable framework allows database systems and query tools to make strong assumptions about the data. Third, dimension models can withstand unexpected changes in query patterns because of its symmetrical structure. In addition, dimension models are extensible, which means the existing tables in the model can be easily changed either by adding new rows or executing SQL ALTER TABLE command.

Granularity, Dimension Tables, and Fact Tables

Figure 1 above shows the dimensional model we design for the business process of Hyundai part sales. The line item sales transaction is declared as the grain because it's the lowest level of detail available in the model. Since atomic data is the most expressive, most dimensional, and most flexible, building the fact table at atomic level will always allow the model to be rolled up to higher levels of details. Based on bus matrix and business requirements, the dimensional model consists of the transaction-oriented sales fact table and six dimension tables including date, vehicle, dealership, customer, service representative, and part details. Surrogate key generator can be used to create artificial identifiers for both fact and dimension tables. Special dimensions will be taken into account for the model design as following.

Snow-Flaking Dimensions

The snowflaked dimension is created for location information of dealerships. Dimensions of city, state, service district, and region are linked back to the dealership dimension with artificial keys. Although, the snow-flaking dimension requires more joins on query performance, it enables easier updates and structure changes, and improves the drill-down and drill-up capabilities of the model. This will help Hyundai BI users generate sales report and administer the information at different level of geographic details. It will also help decision makers better adjust operational strategies based on part sales forecasts for different location grain.

Degenerate Dimension

Based on Hyundai part sales data, the part sales invoice may contain one or more line items. The invoice number should not be discarded because sometimes it is useful for analysis. For example, it can be used to analyze average part sales per invoice by day of week to see if any sales pattern exists within a week interval. The invoice number is dimensional information. However, if we create a dimension table to store it, the table will contain only the header number without any remaining attributes. Thus, we put the dimension of part sales invoice number and pay type into the fact table as degenerate dimensions.

Slowly Changing Dimensions

Type 1 SCD- Type 1 slowly changing dimension is used to overwrite the old value with the most current value. It will be applied when the historical data has no business value or values in the source system need to be updated by processing corrections such as special character removal and spelling error correction. For example, if a customer changes his or her phone number, and storing customer's previous phone number create no more business value to the company, Type 1 SCD will be used to overwrite old phone number with the new. However, if the fact data has been aggregated previously based on the historical value, the pre-aggregate or any ad-hoc summarization related should be updated as well to ensure consistency.

Type 2 SCD- In type 2 slowly changing dimensions, a new dimension row is inserted so that both current and historical value can be preserved. Type 2 SCD will be used when old values are meaningful for business analysis. The latest values are captured through ETL process by maintaining a current flag or adding start date and end date attributes. There could be only one flag indicating current values, and surrogate keys will be assigned to each unique dimension row that represents different span of time. For example, the attributes of operation date and termination date will be added to dealership dimension so that the system can store sales related data of dealership within different time span if the dealership collaborate with HMA more than once. A similar technique can be used to store employee's position changes.

Data Analysis

Here is a sample of some data fields available with data source and field mapping:

Variable	Description	Sample Data	Data Source (MDM: Master Data, DMS: Dealer Management System)	Dimensional Model Mapping
Dealer_Number	Dealer's ID number	CO027	DMS	Dealership dimension and Parts sales fact table
Dealer_Region	Region where dealer is located (CE = Central, EA = Eastern, SC = South Central, SO = Southern, WE = Western	WE	Dealer MDM	Dealership dimension
Dealer_State	State where dealer is located	CO027	Dealer MDM	Dealership dimension
Dealer_Service_District	District where the dealer services	WE3	Dealer MDM	Dealership dimension
Dealer_Name	Dealer's name	AUTONATION HYUNDAI 104	Dealer MDM	Dealership dimension
Customer_City	City where customer is locate	DENVER	DMS	Customer dimension
Customer_State	State where customer is located	CO	DMS	Customer dimension
Customer_Zip	Customer's zip code	80229	DMS	Customer dimension
Vehicle_ID_Number	Vehicle's ID number (VIN)	5NPDH4AE8GH744972	DMS	Vehicle dimension and Parts sales fact table
Vehicle_Model_Year	Vehicle's model year	2016	Vehicle MDM	Vehicle dimension
Vehicle_Model_Name	Vehicle's model name	ELANTRA	Vehicle MDM	Vehicle dimension
Vehicle_Retail_Date	Vehicle's retail sale date	20160416	Vehicle MDM	Vehicle dimension
Part_Sales_Invoice_Number	Invoice number for part sales	348031	DMS	Part sales fact table (degenerate dimension)
Part_Number	Part number that was sold	2630035503	DMS	Parts detail dimension and Parts sales fact table
Part_Number_Description	Part number's description	FILTER ASSY-ENGINE OIL	Parts MDM	Parts detail dimension
Part_Product_Line_Description	Part product line's description	OIL FILTER-ENGINE	Parts MDM	Parts detail dimension
Part_Sales_Date	Date on parts sales invoice	20161001	DMS	Date dimension and Parts sales fact table
Part_Sales_Year	Year of part sales date	2016	Date MDM	Date dimension
Part_Sales_Month	Month of part sales date	10	Date MDM	Date dimension
Part_Sales_Day	Day of part sales date	1	Date MDM	Date dimension
Part_Sales_Week	Day of week of part sales date	SAT	Date MDM	Date dimension
Part_Pay_Type	Part's pay type (C = Customer Paid, W = Warranty Paid, I = Internal/Dealer Paid)	C	DMS	Part sales fact table (degenerate dimension)
Part_Price	Part's selling price	8.5	DMS	Part sales fact table
Part_Quantity	Part quantity sold	1	DMS	Part sales fact table
Part_Sales	Total part sales revenue (price x quantity)	8.5	DMS	Part sales fact table

The grain is transaction, which in this case is the parts sales invoice line item. HMA's parts sales transaction data is extracted from primary sources made up of individual Hyundai dealership management systems (DMS) by a third party data extraction vendor. The DMS keeps track of repairs and parts sales to customer. This vendor performs preliminary ETL before the normalized data is transmitted and loaded into HMA's operational data stores, data warehouse, master data (MDM) and data mart. Some filtering and cleansing is needed due to large volumes of historical parts sales records and non-standardized dealer level data that may not be 100% error free. For example, the system or an analyst will most likely need to apply additional filtering and cleansing rules like selecting 12 months of parts sales invoices dating from October 1, 2016 to September 30, 2017, only valid VINs from HMA's Vehicle MDM and only valid part numbers from HMA's Parts MDM, only parts sales line items where both price and quantity are greater than zero and without missing customer information (city, state and zip) to analyze a more manageable and complete subset of data.

Extraction, Transformation, and Loading

A lot of attention will be focused on the extraction, transformation, and loading (ETL) process, as it will deliver the data from the source systems to the front room BI applications where the added business value of the data warehouse is created. When the process is done correctly, the analysts and management teams will be able to use the BI applications to derive useful facts and information to improve operations and increase revenues for a higher ROI from the data warehouse implementation and maintenance costs. The ETL process is extensive, and includes 34 subsystems that are involved. The main parts of these subsystems are extracting, transforming, delivering, and managing.

The source systems for the HMA parts sales data architecture include internal sales and services Operational Data Stores, Master Data and external Dealership Management Systems running on third-party software. Data originates at these source systems, specifically SAP and IBM AS/400, where HMA receives sales and service transactions from dealers. It is critical for accurate analytics to have clean, complete data records. The source operational data management systems will be built with data validation checks to make sure records have all the fields entered appropriately without making data entry too cumbersome. If the source system data is clean, then the extraction process will have a greater chance of running smoothly. Data validation checks alone are not enough, however, as a data conscious and analytics culture must be enforced from the top down of the entire organization. For example, if a clerk responsible for data entry does not understand how important data is to the rest of the company then they may figure out ways to cut corners by supplying false data to get past the validation checks (ex. entering a false phone number). On the other hand, if they understand the importance of data and how it affects the company's operations then they will be more likely to make sure of its accuracy. The data from the external source which is handled by a third-party vendor will go through a preliminary ETL process and data cleansing before it is loaded with the other source

systems. The extraction process occurs prior to the data transformation, where the data is then merged and conformed into conformed dimensions and metrics.

Once the data is collected from the source systems and ready to be integrated, it will go through a process of conforming to the needed conformed dimensions used in the presentation server DBMS. The conformed dimensions we are using include date, vehicle, dealer, location, service rep, parts detail, and customer. A surrogate key generator will be used to create the keys for the dimensional tables of the data being loaded into the presentation server. The fact tables will be loaded according to whether they are transaction grain, periodic snapshot, or accumulating snapshot fact tables. To avoid referential integrity issues, surrogate keys will replace the operational source system keys. The keys in the fact table and dimension tables must match to be meaningful. One issue that may occur is late arriving data, for example a dealership has connectivity issues and cannot enter data until next business day or week. This can potentially cause problems with the aggregates and partitions, so we will create a subsystem that deals with late arriving data with minimal disruption.

ETLTransformations

Data transformations are an integral part of the process and need to take place before loading the data into the presentation server. Even with data validation checks and a company culture of analytics there will still be some data cleaning required. Checking for null values, duplicate values (de-duplicating), appropriate data types, erroneous entries such as 999-999-9999 for phone number, missing customer information, values less than zero, checking for valid VIN numbers, and other data with these issues will need to be cleaned. Once the data is extracted into the back room and cleaned, it will need to be de-normalized since it originated from a relational system. The presentation server uses dimensional modeling architecture with denormalized fact and dimension tables for enhanced query performance. With the number of records kept in the data management system being in the millions, slow performance would be catastrophic for analysts who need to run ad hoc reports as quickly as possible. De-normalizing the data would increase query performance and understandability with less complicated table joins. Another complication of the loading process is dealing with slowly-changing dimensions (SDC), such as customer demographics information. The three most common methods for dealing with SDC are overwriting, creating new rows, or adding a column for a soft change. The most common approach is creating a new row, which the transformation process will need to facilitate when a change in demographic data is detected. Process metadata will be created to explain how this works to analysts and anyone else who works on the system.

Aggregations and Partitioning

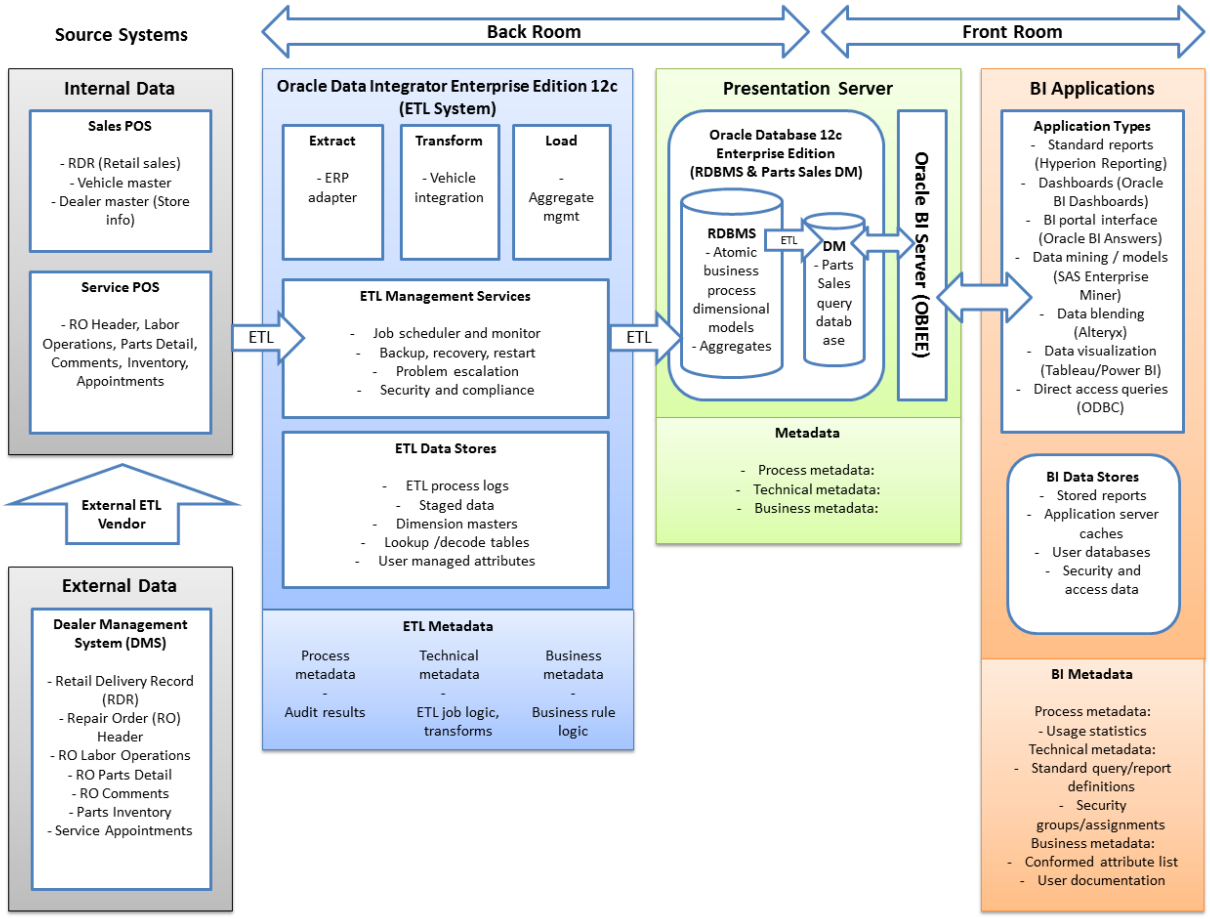
One of the subsystems in the ETL process deals with building aggregates. It is important to aggregate data for query performance in the same way that creating indexes does. For data that is frequently queried in the aggregate form, such as monthly sales volumes per part type, it will help performance to have the data pre-aggregated in aggregate fact table rows. In the beginning of the data warehouse implementation our team will create certain aggregates such as sales volumes by month per part type and total sales per dealership, and then going forward monitoring of performance will be needed to determine if more aggregates are necessary for data that is frequently pulled together. The size of the data that make up some aggregates are too large for the load window, so we will use incremental loads and append only new data to the aggregate fact tables.

Another tool we are using to help with performance is partitioning. Some of the part's sales tables will be very large, so partitioning these into separate tables will help to decrease the number of records that need to be looked through to find the appropriate ones based on the chosen partition. Date is a good partition to use in this case because it is easily understandable. Using date as a partition is also good in helping maintenance, system back-up, and loading activities. Scripts will be created for incremental loads to automatically create partitions under certain conditions such as a new day or when a table reaches a certain number (an artificial maximum) of rows.

Job Scheduling

The best time for the ETL process to occur will be during non-operating hours. The data warehouse will be used by employees from across the country, so typically we would want the process to be completed no later than 6 am ET. This will allow for the teams working late on the West Coast to be able to run ad hoc queries until around 10pm PT which should be sufficient, and not interfere with teams starting early on the East Coast. The entire process should take no more than three hours. The data will be pulled and loaded on a daily basis. The ETL scheduling will sufficiently support the needs of the analysts and management teams who are looking for trends daily. There will be a metadata repository manager containing process, business, and technical metadata to document and define all processes and logic involved. Like any system there is potential for errors that could bring the process to a halt, so an error-processing and event tracking problem escalation system will be put in place to arrange work-arounds to the errors and log them to assess how to fix the issue going forward. A backup system and restart and recovery system will also be implemented in the ETL process to deal with failed jobs due to any number of issues including network failures, disk failures, and memory failures.

The Recommended Architecture:



The goal of this recommended application architecture model (see image above) is enabling Hyundai Motor America (HMA) to analyze and report parts sales and helping Hyundai Dealerships to develop tactics to increase parts sales revenue. Currently, all of HMA's sales and service operational data are extracted from individual Hyundai dealership management systems (DMS) by a third party data extraction vendor. This vendor performs initial Extract Transform Load (ETL) before the normalized data is transmitted and loaded into HMA's Operational Data Stores (ODS). These ODS entities make up HMA's source systems. In HMA's existing back room, there are separate ODS' for storing vehicle sales and service transaction data like SAP and IBM AS/400. Various industry standard ETL services migrate all the sales, vehicle, dealer, and service data from these disparate ODS' into a normalized Oracle data warehouse.

The recommended architecture will create a Parts Sales decision support system by processing parts sales related data from Oracle data warehouse and loading consumable data into a query optimized (i.e., denormalized) Parts Sales data mart accessible to the organization's analytical community and stakeholders. A secondary ETL process will denormalize, transform and load a subset of data from Oracle Database into a Parts Sales query database since performance is an issue when querying and analyzing large volumes of sales and service data. Oracle is a proposed end-to-end solution provider from secondary ETL services that maintain the HMA Parts Sales Data Mart housed in Oracle Database in the back room to BI Presentation Services that provide

the framework and interface for presentation of business intelligence data to BI applications and end users in the front room.

In the back room, multiple Oracle products including Oracle Data Integrator Enterprise Edition 12c, Oracle Database 12c and Oracle Business Intelligence Enterprise Edition (OBIEE) 12c will work with each other to process and present data seamlessly to front room BI applications and end users.

1. Oracle Data Integrator Enterprise Edition 12c is the latest version of a recommended ETL solution that integrates various parts sales data collected from dealers, evaluates the quality of data, removes anomalies in data and loads cleansed data into the query database. Essentially, Oracle Data Integrator provides data integration and data quality service for cross-platform, multiple sourced data.
2. Oracle Database 12c is the latest version of a popular and best-in-class relational database management system in the software industry. The salient features of the database are scalability, aggregation, flexibility, and efficiency in handling large volume data. This user-friendly database gives robust, simple and fast results on data storage and retrieval. The unique quality of the Oracle database is the consistent data retrieval capability which is the most critical factor in the recommended HMA parts sales architecture. Extended capacity of data tables, schema, and processing performance has given an enormous advantage to the enterprise BI processes. This product represents the foundation for data and analytical needs of HMA parts sales data architecture.
3. OBIEE 12c's latest suite of products including BI Server, Oracle BI Dashboards and Oracle BI Answers complement the Oracle Database and enhance the BI presentation capability of HMA's parts sales data architecture. These products are used for making parts sales data in the data mart available to front room BI applications and end users, creating dashboards/scorecards and standardizing web-based reports to HMA management and dealerships. OBIEE is a multi-user platform that acts as the central hub for users to access data and modify reports, analyses, and dashboards based on their needs. It is also one of the best tools to get insights from the HMA part sales data mart.

In the future front room, our recommended solution will empower more analysts by providing more advanced self-service BI applications like Alteryx, SAS Enterprise Miner and Tableau and granting more analytical users with access to the denormalized data marts they need to more easily perform data mining and analyze parts sales by dealer, location, vehicle type, vehicle age, mileage, etc.

1. Alteryx is a vital BI application in HMA parts sales front room architecture for data analysis on the multi-sourced data. Alteryx's graphic user interface enables users to quickly access, prepare and clean HMA Parts Sales data for data exploration and analyses. Moreover, this software efficiently reduces data redundancy. It has extensive capabilities to support creativity, design and custom solutions. The benefit of the product is low cost which makes it ideal for empowering many analysts in the organization. Data prepared using Alteryx is easily imported into other BI applications like SAS Enterprise Miner and Tableau for advanced analytics and data visualization.

2. SAS Enterprise Miner excels at data mining and will build accurate predictive and descriptive models using large volumes of data from HMA parts sales data mart. It provides the ability to analyze the data and makes it easy to identify and remove outliers in the data. It compares different models and scores model performance using training and validation dataset. The various data mining models such as regression, cluster analysis, decision trees help to find variables that are important, the trends in data, and insights that help in decision making for executives, senior management, and stakeholders. This data mining tool can predict customer needs, market demand, parts sales and revenue for HMA.
3. Tableau helps HMA to visualize and understand the data. It enables data exploration, identify market trends, visualizes geographic sales and caters to the needs of the organization. Tableau can import data from different sources like Alteryx and even connect directly to the backend Parts Sales data mart. Also, it can join data from many various sources.

Preliminary Budget Plan

Line Item	**Unit Price**	**Units**	**Total**
Oracle Data Integrator Enterprise Edition User License	$ 900	50	$ 45,000
Oracle Data Integrator Enterprise Edition User Support	$ 198	50	$ 9,900
Oracle Data Integrator Enterprise Edition Processor License	$ 30,000	2	$ 60,000
Oracle Data Integrator Enterprise Edition Processor Support	$ 6,600	1	$ 6,600
Oracle Database 12c Enterprise Edition User License	$ 950	50	$ 47,500
Oracle Database 12c Enterprise Edition User Support	$ 209	50	$ 10,450
Oracle Database 12c Enterprise Edition Processor License	$ 47,500	2	$ 95,000
Oracle Database 12c Enterprise Edition Processor Support	$ 10,450	1	$ 10,450
Oracle BI Enterprise Edition Foundation Processor License	$ 300,000	2	$ 600,000
Oracle BI Enterprise Edition Foundation Processor Support	$ 66,000	1	$ 66,000
HMA Parts Sales Reporting Website and Maintenance	$ 200,000	1	$ 200,000
Microsoft Power BI & Power Pivot	$ -	600	$ -
Alteryx Designer & Server User License	$ 6,802	25	$ 170,050
SAS Enterprise Miner with Maintenance fee	$ 202,575	1	$ 202,575
SAS Enterprise Miner 25 User License with Maintenance	$ 10,375	1	$ 10,375
Tableau Professional & Server License	$ 1,260	25	$ 31,500
Project Manager	$ 110,000	1	$ 110,000
Data Architect	$ 110,000	1	$ 110,000
Website Developer	$ 100,000	1	$ 100,000
Business Systems Analyst	$ 90,000	1	$ 90,000
Data Scientist	$ 120,000	1	$ 120,000
Systems Consultant	$ 35,000	1	$ 35,000
Total Annual Cost			**$ 2,130,400**

Using the products outlined in the architecture, the pricing structures were researched. Tableau and Alteryx are at a fixed price on a per-user basis. The number of licenses for Alteryx and Tableau were based on 25 power users out of a group of 600 total employees.

Oracle Data Integrator and Database 12c have a minimum of 25 named users per processor license; 50 total using a standard server with two processors.

The Oracle Business Intelligence Enterprise Edition suite of BI software is the most expensive line item and carries a high support and update fee as well. In fact, all Oracle Line Items have additional fees for software updates and support in the first year.

Aside from the initial costs of the technology itself, hiring the right roles to help implement the data architecture was considered. Annual salaries for each position was researched along with an estimate of the minimum number of employees needed to enhance the data architecture.

It is also very likely that for the successful implementation of the project, an outside consultant is needed. The typical consultants cost in the $150-$300 per hour range and a roughly 4 week engagement was estimated. This cost could fluctuate greatly, especially if permanent employees are able to ramp-up quickly.

Many of these costs, such as the number of processors required and per-user licenses will grow as the scale of the project increases overtime, but a budget of $2.1 million is needed for development, implementation and software licensing during the first year and about $1.6 million for following year assuming no changes.

Summary, Conclusion & Recommendation For Action

Hyundai Motors America is a prime candidate for a data warehouse, parts sales data mart and parts sales BI system implementation that will help them improve their analytic capabilities surrounding part sales KPI. Part sales make-up a large portion of the profits seen by HMA, however they are not currently utilizing the latest BI solutions with more advanced analytics to identify missing sales opportunities and increase parts sales even more. Our data architecture proposal will allow HMA and its dealers to report, monitor and analyze part-sales data on a daily-basis throughout the nation. The scale of the proposal uses the Kimball approach, by starting out with parts sales KPI, and is designed in a way that will allow additional data added later while avoiding the silo data mart issue. A dimensional model was developed with fact and dimension tables to be used in the presentation server that will allow fast query access of denormalized information in parts sales data mart with minimal data table joins.

The database structure of dimensional modeling will be built at transactional grain level in HMA data warehouse. The model would consist of a part sales fact and six dimensions including date, part detail, vehicle, customer, service representatives, and dealership. A snow-flaking structure will be used for the location information of dealerships. Part sales invoice number and pay type data will be stored as degenerate dimensions in fact table. Type 1 slowly changing dimension

will be applied to capture no business value changes, and type 2 slowly changing dimension will be used to update meaningful business changes.

The primary data used in this architecture is extracted from parts sales transactions that happens in all HMA dealerships and sales offices across the country. HMA loads data for vehicle's information, dealership's information, customer's information, part's information and parts sales information into operational data stores and master data. These records will need to be managed well and ensure that new data are being consistently entered into the system to reduce cleansing efforts and losing important data information.

The implementation of the ETL system takes careful planning and a lot of time dedicated to getting it right. Our design will start by taking the data from the internal and external source systems and lastly deliver it to the BI applications for the analysts and management teams to find business value. The entire process involves 34 subsystems, each dedicated to a specific aspect related to either extracting, transforming, loading, or managing the data. Validation checks, cleaning, de-normalizing, conforming dimensions, aggregating, and partitioning are key points in the progression. Proper job scheduling and error-handling will ensure that business processes are not interrupted. Process, business, and technical metadata are used in both back and front room environments for audit results, ETL job logic, business rule logic, usage statistics, and more making it easier for users and technicians to understand and develop in the future. In the end, business users will find standard reports, BI dashboards, and ad hoc queries that deliver clean, reliable data from the daily ETL process.

The recommended architecture will create a Parts Sales decision support system by processing parts sales related data from Oracle data warehouse and loading consumable data into a query optimized (i.e., denormalized) Parts Sales data mart accessible to the organization's analytical community and stakeholders. Oracle is a proposed and best in class end-to-end solution provider from Oracle Data Integrator's ETL services that maintain the HMA Parts Sales Data Mart housed in an Oracle Database in the back room to OBIEE's BI presentation services that provide the framework and interface for presentation of business intelligence data to BI portals and dashboards like a HMA Parts Sales reporting website and analytical applications like Alteryx, SAS Enterprise Miner and Tableau in the front room.

The development, implementation, and software required for this project has significant variable costs. The costs of the software from Oracle, Tableau, Alteryx, and SAS can be reasonably estimated for what will be required, assuming no negotiation of the pricing. The variable costs of the human resources required are somewhat volatile. The salaries and headcount for the required roles is easily estimated given the market demand for developers, DBAs, and data scientists. However, the engagement of an external consultant can be estimated, but the length and scope of that engagement has the potential to balloon. That said, given the initial scale of the project, a first year estimate of $2.1 million is reasonable, with an ongoing annual cost of $1.6 million.

By implementing this Parts Sales BI solution for HMA, the corporation could benefit from various tangible and intangible aspects. Having a standardized data warehouse architecture that could be used across dealerships and service centers will save time and increase efficiency in

terms of locating data, downloading data, and analysing data. The time saved could be used to service the customers better and increase customer's satisfaction. The architecture will also allow data to be analyzed in different ways that would reveal new information which could direct HMA to allocate its resources to the right direction and increase the company's part's sales. This proposed architecture could also improved decision making process, allowing management to view information in a more analytical way and able to identify and act on any issues more effectively. Given these benefits, the next step for HMA is to lay out the tangible benefits and calculate the net present value, payback period, and return on investment on building this data warehouse. Then, a better picture of the benefits of implementing this BI solution could be presented to gain project approval and budget allocation. With HMA expected to exceed 1 billion dollars in parts sales revenue this year, even a modest 1% increase in parts sales as a result of new sales opportunities realized from this system will recoup this system's 2.1 million dollar initial investment.