

Project:

Predictive Analytics for the Online Retailer Kohl's Corporation

1. Company Background:

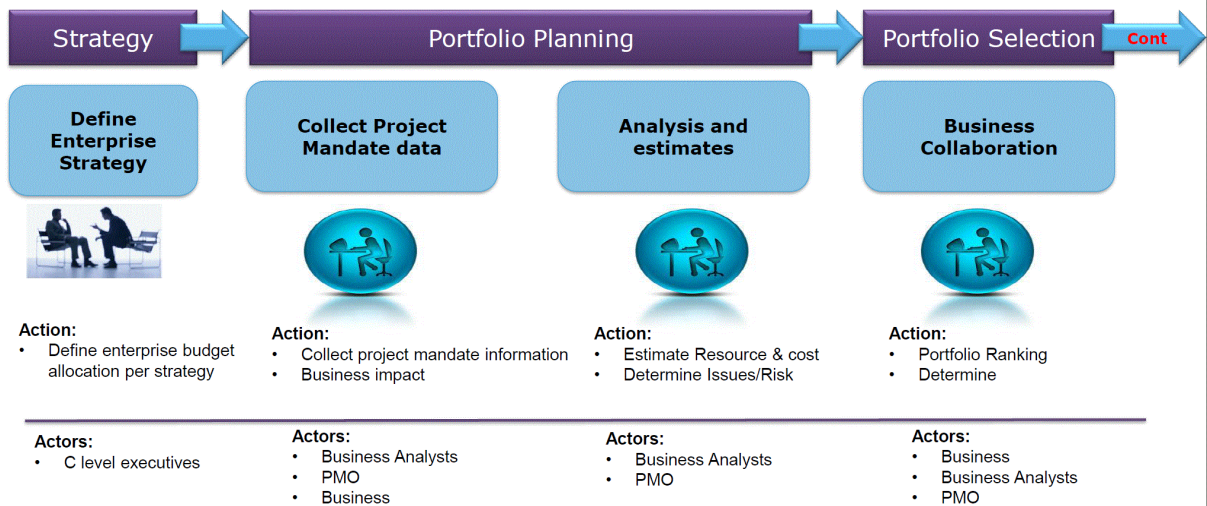
Kohl's Corporation incorporated on March 23, 1993. It is one of the biggest discount retail chains in the United States. The chain targeting middle-income consumers and maintains low retail prices and limited staffing. The Company has 1,160 department stores in 50 states and sells private label, exclusive and national brand apparel, footwear, accessories, beauty and home products. It has an e-commerce website (kholes.com) includes products available in its stores, and products available only online. Kohl's mission is to be the leading family-focused and value-oriented department store.

2. Project Description:

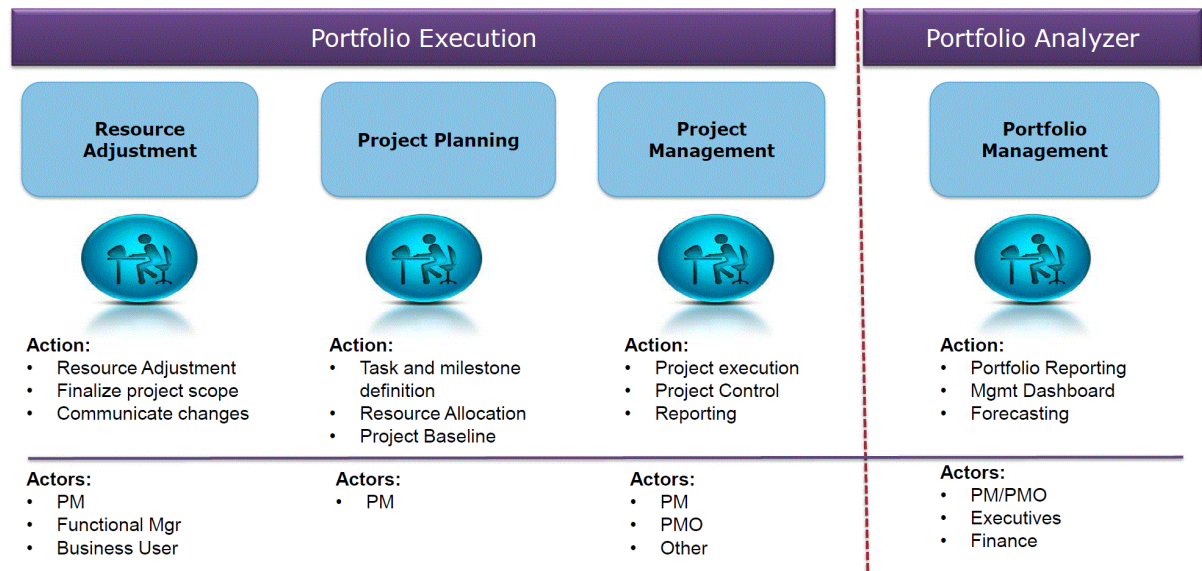
Predictive Analytics models to predict the inventory demand and consumers interests in order to optimize the online shopping experience for Kohl's consumers.

3. Portfolio Management Life Cycle:

vPPM Portfolio Life Cycle



Portfolio Life Cycle (Continue)



4. Kohl's Stakeholder Groups:

I. Investors and Executive team

Investors and CEO's are mainly interested in profits. They want Kohl's to have more profits, which translate to higher dividends or earnings per share. In relation, they are also interested in minimizing the operational costs of the company. Lower costs usually lead to higher profits, which are beneficial for Kohl's investors and executive officers.

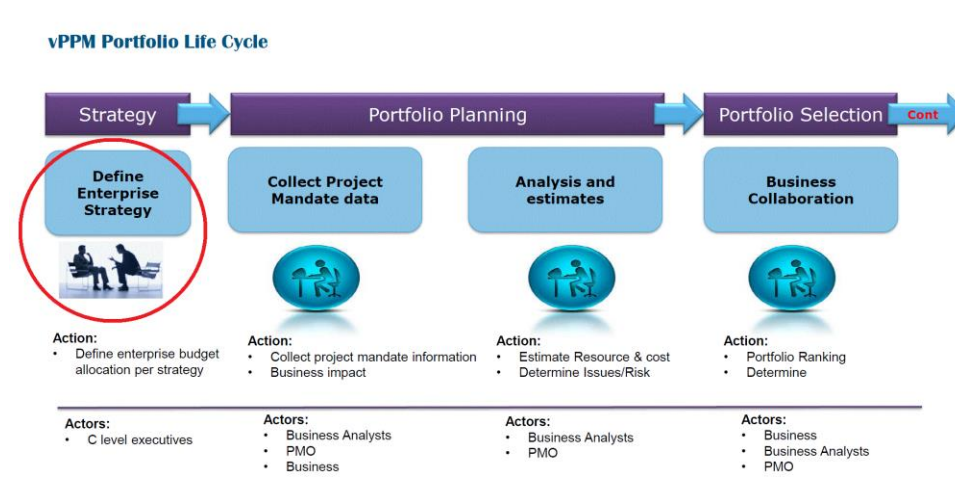
II. Customers

Kohl's addresses the interests of customers as stakeholders. The company maintains its cost leadership generic strategy, which involves offering the lowest possible prices. In fact, the company is popular because of its low prices. In this regard, Kohl's is effective and successful in addressing the interests of customers as a stakeholder group.

III. Suppliers

Suppliers are interested in getting more of their products sold at Kohl's stores in a profitable way. This interest involves not just the selling of the suppliers' products, but also the selling of these products at acceptably profitable prices.

5. Enterprise Strategy:

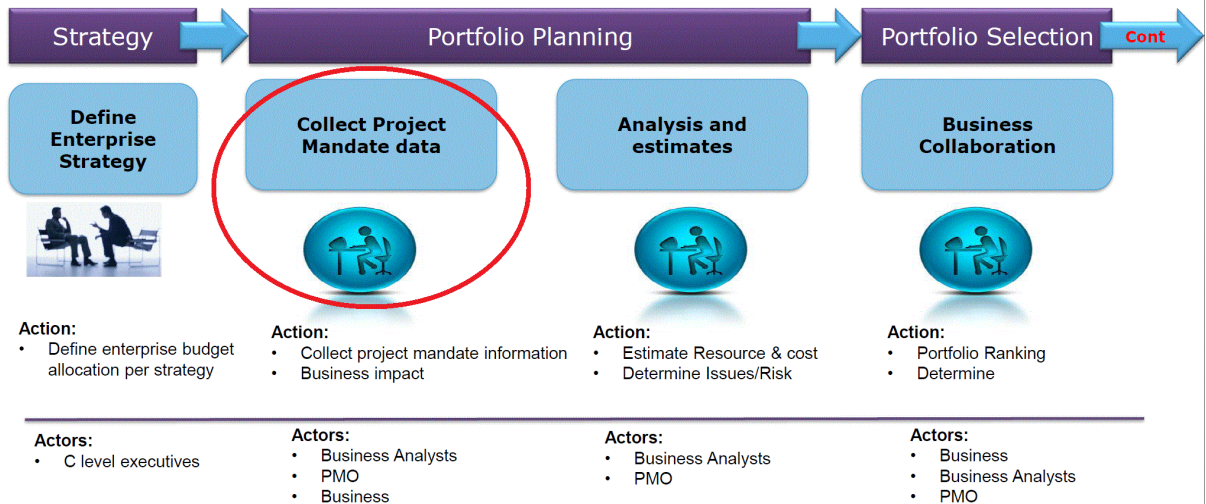


5.1. Project Budget:

Project* Budget			
Hardware	Software	Consulting	Operations
50,000.0	90,000.0	130,000.0	500,000.0
Other	Total Budget		
0.0	770,000.00		USD

6. Project Mandate Data and Business Impact:

vPPM Portfolio Life Cycle



6.1. Mandate:

I. Purpose

Kohl's predictive analytics is a data-driven project that will analyze past and current performance to predict the likelihood of future outcomes based on historical data. The project will play a vital role in improving the decision-making process, customer service strategies, and marketing campaigns and promotions.

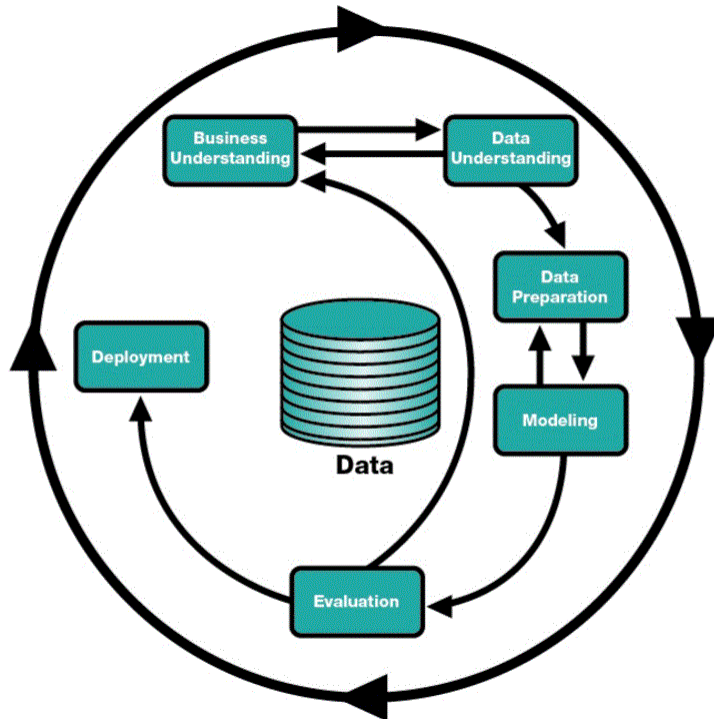
II. Authority Responsible

- **Chief Data Officer (CDO):** a senior executive who is responsible for Kohl's enterprise wide data and information governance, control, and protection.
- **Finance Department:** For authorizing cost.
- **Supply Chain and Logistics Department:** For managing the inventory.

III. Background

The predictive analytics project will build statistical and predictive models to optimize the online shopping experience for Kohl's consumers.

The project will follow the recommended **CRISP-DM** (Cross Industry Standard Process for Data Mining) methodology. It has six major phases:



1. **Business Understanding:** To understand the project objectives.
2. **Data Understanding:** To identify data quality problems.
3. **Data Preparation:** To clean the data and prepare it for modeling.
4. **Modeling:** To select from various modeling techniques.
5. **Evaluation:** To validate the model on testing data.
6. **Deployment:** To use the models in Kohl's website.

The project will complement other implemented BI solutions such as reporting, OLAP and data visualization.

IV. Project Objectives

- **Marketing initiatives:** The results of predictive analytics will lead to improving the decision making process and coming up with marketing campaigns and promotions to target the customers need and increase return on investment (ROI).

- **Customer Sentiment Analysis:** The project will use the data collected from communication channels like social media and product review forums to analyze customers' preferences and behaviors.
- **Personalized product recommendations:** Kohl's can increase their sales by recommending the best product for the consumer after predicting his/her behavior from the purchase history.
- **Predictive prices:** Kohl's can determine the right price at the right time after analyzing the pricing trends in the market, sales, and historical data.
- **Predicting inventory demand:** Predictive analytics can identify consumer demands, hence, manage the inventory
- **Detecting fraud:** by analyzing consumer's behaviors, predictive analytics can eliminate potential fraud before the consumer complete the transaction.

V. Scope

The major deliverable of the project are different predictive models to address all the project objectives mentioned above. During the modeling phase, for example, Kohl's depends on historical data to build its predictive models. After the project deployment phase, whenever Kohl's decides for sales during holiday seasons, the predictive model should be updated with historical data related to previous festive seasons.

VI. Constraints

Managing the constraints is always a critical part of any project. However, it will be challengeable to find professional resources with experience in analytics applications and knowledge of data science programming languages like R and Python to build predictive models.

VII. Interfaces

The predictive analytics project will interface with other BI implemented solutions, such as querying and reporting, OLAP and data visualization.

VIII. Quality Expectations

Today's consumers are looking for personalized experiences, seasonal discounts, high quality customer services, and notifications for new products that match their interests. However, our predictive models will be designed to satisfy their expectations and target their needs.

IX. Outline Business Case

There are many reasons stress the need for predictive analytics project. First and foremost, the predictive analytics will allow us to learn from the past in order to predict the future. Second, it will optimize the business by predicting inventory demand and consumers interests. Third, it will narrow and detect fraud cases. Fourth, the analytics will be practical in terms of meeting today's consumer expectations and providing personalized recommendations. Finally, the project will increase the revenue and retain consumers.

X. Associated Documents

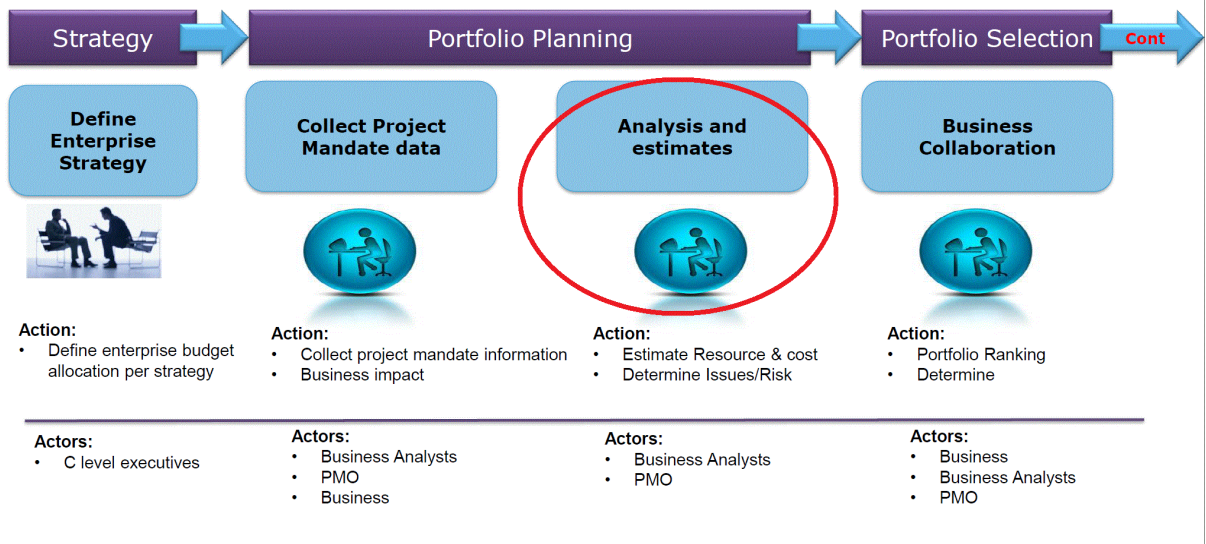
Walmart, for example, is one of the largest online retailers. However, Walmart Labs collects 2.5 petabytes of data from 1 million customers every hour. One of the risk associated with Walmart predictive analytics project is making predictions with limited historical data.

6.2. Project Impact:

- Enhancement to the existing Bus/Ops **Yes** ✓
- Current Process Improvement? **Yes** ✓
- Service Improvement? **Yes** ✓
- Cost Saving? **Yes** ✓

7. Analysis and Estimates:

vPPM Portfolio Life Cycle



7.1. Issues:

I. Lack of skilled resources

It is challengeable to find professional resources with experience in analytics applications and knowledge of data science programming languages like R and Python to build predictive models.

Maintain Issue I						
Issue Date 02/01/17	Designation Issue ▼	Probability[0-100] 100 (%)	Impact[1-10/100] 5	Action Consolidate ▼	Date Expected 05/31/2017	
Exposure Recipient				Print Include ▼		
Date Assigned 02/04/2017	Assigned to NA	Department Employee Alsalamah,Hanan [A6000] ▼				
Description* Lack of skilled resources						

II. Limited quantities of historical data

The biggest challenge/barrier for retailers like Kohl's is to make predictions with limited historical data.

Maintain Issue						
Issue Date	Designation	Probability[0-100]	Impact[1-10/100]	Action	Date Expected	
02/01/17	Issue ▼	100 (%)	5	Consolidate ▼	05/31/2017	
Exposure Recipient				Print		
				Include ▼		
Date Assigned	Assigned to		Department Employee			
02/03/2017	NA		Alsalamah,Hanan [A6000] ▼			
Description*						
Limited quantities of historical data						

III. Consumers Privacy

The project poses substantial risks to consumers as their sensitive data will be collected from social media and browsing history without obtaining their consent.

Maintain Issue						
Issue Date	Designation	Probability[0-100]	Impact[1-10/100]	Action	Date Expected	
02/01/17	Issue ▼	80 (%)	2	Consolidate ▼	05/31/2017	
Exposure Recipient				Print		
				Include ▼		
Date Assigned	Assigned to		Department Employee			
02/02/2017	NA		Alsalamah,Hanan [A6000] ▼			
Description*						
Consumers Privacy						

7.2. Risks:

I. Data Quality

The velocity and volume of structured and unstructured data may result in data quality issues such as lack of consistency, accuracy, and integrity.

Maintain Issue						
Issue Date	Designation	Probability[0-100]	Impact[1-10/100]	Action	Date Expected	
03/20/17	Risk ▼	90 (%)	9	Identified as Risk ▼	06/01/2017	
Exposure Recipient				Print		
Kohl's Corp				Include ▼		
Date Assigned	Assigned to	Department Employee				
03/23/2017	NA	Alsalamah,Hanan [A6000] ▼				
Description*						
Data Quality						

II. Inaccurate statistics

Predictions will be less accurate with poor statistical modeling

Maintain Issue						
Issue Date	Designation	Probability[0-100]	Impact[1-10/100]	Action	Date Expected	
06/21/17	Risk ▼	50 (%)	8	Identified as Risk ▼	08/21/2017	
Exposure Recipient				Print		
Kohl's Corp				Include ▼		
Date Assigned	Assigned to	Department Employee				
06/23/2017	NA	Alsalamah,Hanan [A6000] ▼				
Description*						
Inaccurate statistics						

III. Big assumptions

Big assumptions should be avoided as it's not necessary that the future will continue like the past and past performance does not guarantee future results

Maintain Issue						
Issue Date 06/15/17	Designation Risk ▼	Probability[0-100] 40 (%)	Impact[1-10/100] 7	Action Identified as Risk ▼	Date Expected 08/30/2017	
Exposure Recipient Kohl's Corp and Consumers				Print Include ▼		
Date Assigned 06/18/2017	Assigned to NA		Department Employee Alsalamah,Hanan [A6000] ▼			
Description* Big assumptions						

8. Resources:

Resource Management for: Predictive Analytics for the Online Retailer Kohl's Corp (Start Planning year: 2017)													
Application Architect	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Application Engineer	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
Data Analyst	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00
Data Engineer	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00
Database Admin	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Project Manager	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Systems Architect	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

9. Project Life Cycle Dates:

Client* Kohl's Corp Program* NEU - Big Data Projects Add a new Program*

Predictive Analytics for the Online Retailer Kohl's Corp 123 01/20/2017 43200 770000.00 1

Core Information Categories Controls Custom Planning

Project* Delivery Dates

Project* Life Cycle Dates

Development*	QA*	Production*	Parallel Test*	Custom Controls Dates	
Est Comp. Date 07/20/2017	Start Date 05/01/2017	Est. Date 07/20/2017	Start Date 05/01/2017	Date(1)*	Date(2)*
Start Date 02/24/2017	End Date 05/19/2017	Revised Date	End Date 05/19/2017	Date(3)*	Date(4)*
End Date 07/17/2017		Actual Date		Date(5)*	Date(6)*

10.Activities:

- Admin

Activity Management Management

Define Activities for Project*

Use Activity Category: Admin

--- Select one or more Activity ---

- Asset Mgt Administration
- Associate Performance Mgt
- Billing
- Change Control
- Daily Run Sheets
- Employee Recruiting
- General Secretarial
- Hardware Asset Mgt
- HR Issues
- Merit/Bonus/Promo Planning
- People Management

Activities to be added

Valid Activities for Project*

--- Select one or more Activity ---

- Contract Negotiations & Upkeep
- Meetings (dept, division, company)
- Upgrades/Maintenance
- Vendor Relationship Management

- Analysis & Design

Activity Management Management

Project* Valid Activities for Predictive Analytics for the Online Retailer Kohl's Corp

Define Activities for Project*

Use Activity Category: Analysis & Design

--- Select one or more Activity ---

- Research
- Visual Design

Activities to be added

Valid Activities for Project*

--- Select one or more Activity ---

- Analysis
- Business Process/Workflow
- Functional Specification
- Requirements Analysis & Design

- **Architecture**

Activity Management Management

Project* Valid Activities for [Predictive Analytics for the Online Retailer Kohl's Corp]

Define Activities for Project*

Use Activity Category: Architecture ▼

--- Select one or more Activity ---

- Architecture & Research
- Architecture Review
- Security Architecture
- Technology Evaluation
- Technology Research

--- Activities to be added ---

Valid Activities for Project*

--- Select one or more Activity ---

- Architecture Design
- Information/Data Architecture

- **Data Management**

Activity Management Management

Project* Valid Activities for [Predictive Analytics for the Online Retailer Kohl's Corp]

Define Activities for Project*

Use Activity Category: Data Management ▼

--- Select one or more Activity ---

- Data Movement
- Data Rejects
- Decision Support

--- Activities to be added ---

Valid Activities for Project*

--- Select one or more Activity ---

- Data Analysis
- Data Correction & Updates
- Data Mapping
- Data Modeling
- Database Administration

- **Development**

Activity Management Management

Project* Valid Activities for [Predictive Analytics for the Online Retailer Kohl's Corp]

Define Activities for Project*

Use Activity Category: Development ▼

--- Select one or more Activity ---

- Coding/Configuring
- Design

--- Activities to be added ---

Valid Activities for Project*

--- Select one or more Activity ---

- Code Development
- Code Review

- **Documentation**

Activity Management Management		
Project* Valid Activities for [Predictive Analytics for the Online Retailer Kohl's Corp]		
Define Activities for Project*		Valid Activities for Project*
Use Activity Category: Documentation ▼		
---- Select one or more Activity ----	---- Activities to be added ----	---- Select one or more Activity ----
Content Updates Documentation		Process Definition & Documentation

- **Management & Planning**

Activity Management Management		
Project* Valid Activities for [Predictive Analytics for the Online Retailer Kohl's Corp]		
Define Activities for Project*		Valid Activities for Project*
Use Activity Category: Management & Planning ▼		
---- Select one or more Activity ----	---- Activities to be added ----	---- Select one or more Activity ----
Annual Planning Business/Strategic Planning Client Meeting Project Management Status Meeting		Budget Planning & Review Project Planning Resource Mgmt Status Reporting

- **Quality Assurance**

Activity Management Management		
Project* Valid Activities for [Predictive Analytics for the Online Retailer Kohl's Corp]		
Define Activities for Project*		Valid Activities for Project*
Use Activity Category: Quality Assurance ▼		
---- Select one or more Activity ----	---- Activities to be added ----	---- Select one or more Activity ----
System Test Test Script Development Test Script Maintenance Testing - Automation Testing - Manual Unit Test		Performance Testing & Tuning Test Planning & Preparation User Acceptance Test

- **Support**

The screenshot displays the 'Activity Management Management' interface. At the top, a blue header bar contains the text 'Activity Management Management'. Below this, a section titled 'Project* Valid Activities for [Predictive Analytics for the Online Retailer Kohl's Corp]' is visible. The main area is divided into two primary panels. The left panel, titled 'Define Activities for Project*', contains a dropdown menu for 'Use Activity Category:' set to 'Support'. Below this is a list of activities: 'End User Support', 'Environment Setup', 'General Support', 'Problem Identify and Resolve-After Hr', 'Problem Management', 'Processing Jobs', 'Production Support', 'Release Management', and 'Security Assessment'. A '--- Select one or more Activity ---' button is positioned above this list. To the right of this list is an empty box labeled '--- Activities to be added ---'. The right panel, titled 'Valid Activities for Project*', contains a '--- Select one or more Activity ---' button above a list of activities: 'Application Administration', 'Incident Handling', 'Post Installation Support', and 'Security Control Validation'.

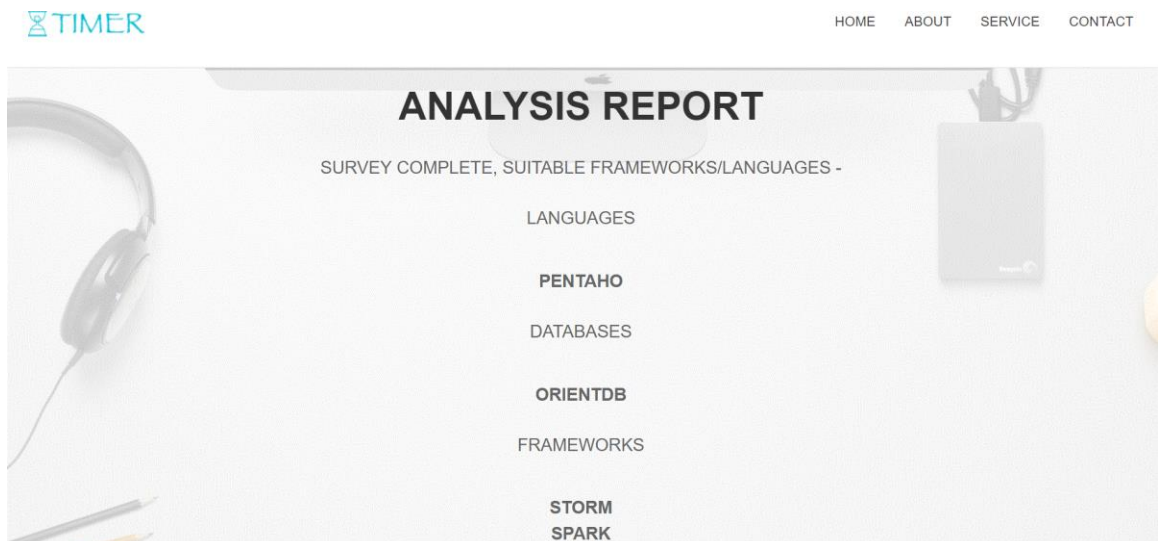
11. Functional Requirements:

- I.** The results of predictive analytics shall improve the decision making process and come up with marketing campaigns and promotions to target the customers need and increase return on investment (ROI).
- II.** The system should be able to use the data collected from communication channels like social media and product review forums to analyze customers' preferences and behaviors.
- III.** The system shall be able to recommend the best product for the consumer after predicting his/her behavior from the purchase history.
- IV.** The results of predictive analytics shall determine the right price at the right time after analyzing the pricing trends in the market, sales, and historical data.
- V.** The results of predictive analytics shall be able to identify consumer demands, hence, manage the inventory.
- VI.** By analyzing consumer's behaviors, predictive analytics shall be able to eliminate potential fraud before the consumer complete the transaction.

12. Big Data Architecture Decision Making Process:



- **Assessment Results:**



12.1 Language and Orchestration:

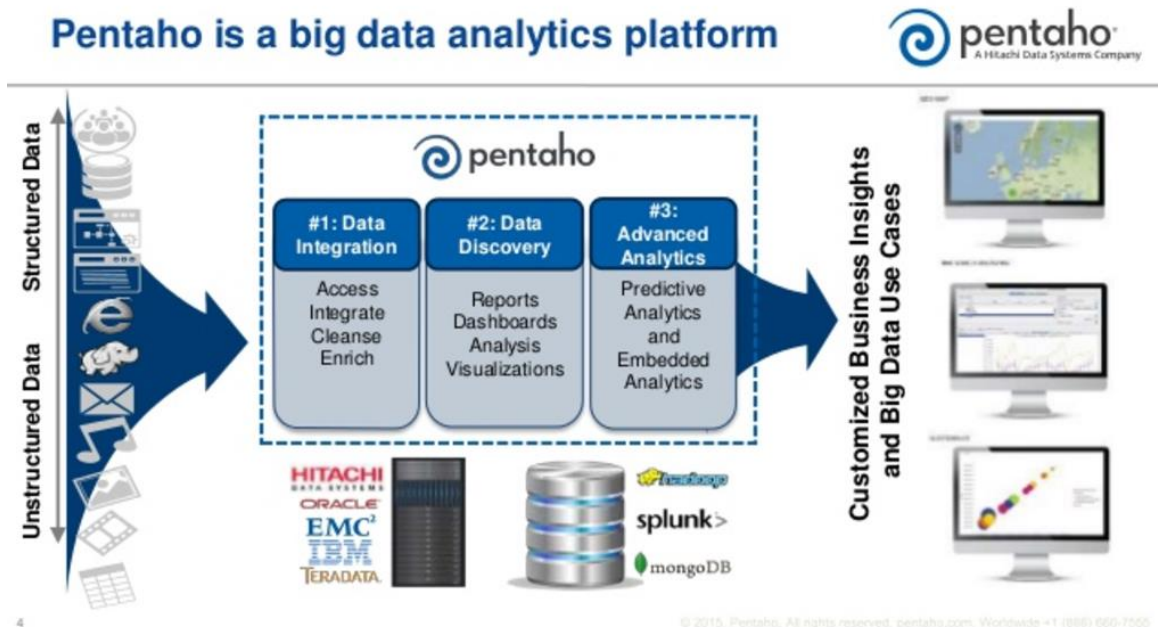


Pentaho:

A comprehensive data integration and business analytics platform that brings together IT and business users for easy access, integration, visualization and exploration of any data.

I. Why did we Choose Pentaho?

- Huge community support.
- Unlimited visualization and data sources
- Big data analytics platform supports the whole process of predictive analytics.
- Easily scalable.
- Easily integrated on top of existing infrastructure including any big data ecosystem.
- Complete Suite of analytics tools.
- Pentaho includes data discovery, data integration and predictive analytics.



Data Discovery, Analysis and Visualization:

- Interactive visual analysis allows decision makers to drill into data for greater insight.
- A rich library of interactive visualizations to find patterns and anomalies, including geo-mapping, heat grids and scatter/bubble charts.
- Drill-down functionality into supporting reports and dashboards provides deeper analysis.
- Extreme scale in-memory data caching for speed-of-thought analysis with large data volumes

Data Integration:

Pentaho Data Integration delivers powerful data preparation capabilities including extract, transform and load (ETL). An intuitive and rich graphical design environment minimizes complexity and time invested in specialized scripts and coding to prepare data.

Predictive Analytics:

- Ability to support the whole process of predictive analytics
 - Preparation of input data
 - Statistical evaluation of learning schemes
 - Visualization of input data and the result of learning
- Includes powerful algorithms such as classification, regression, clustering and association
- Allows import of Ability to import third-party models using Predictive Modeling Markup Language (PMML)
- Allows storing and versioning of models using the Pentaho repository
- Incorporates algorithms into Pentaho's visual interface

II. Pentaho Non-Functional Requirements:

-Server (enterprise level download):

- RAM: at least 4GB
- Hard drive space: at least 2GB for the software, and more for solution and content files.
- Processor: dual-core AMD64 or Intel EM64T

-Workstation (client side):

- RAM: at least 2GB
- Hard drive space: at least 1GB for the software, and more for solution and content files.
- Processor: dual-core AMD64 or Intel EM64T

- Supported Platforms: Windows, Linux, Mac

- Manageability:

Pentaho Data Integration provides powerful data preparation capabilities including extract, transform and load (ETL). It includes support for multiple data sources including over 25 open source and proprietary database platforms, flat files, Excel documents, and more. The architecture is extensible with a plug-in mechanism.

- Security:

Supports and integrates with the following third party authentication systems:

- Active Directory
- Central Authentication Service (CAS)
- LDAP (Lightweight Directory Access Protocol)

- Scalability: It provides a highly scalable solution for enterprises.

12.2 Database:



OrientDB:

Is an open source NoSQL database management system that supports graph, document and object models but just like graph databases relationships are managed through direct connections with the records.

I. Why did we Choose OrientDB?

- In general, the Graph DB is ideal for connected data require recommendation, reputation, and predictive analysis.
- OrientDB is a 2nd generation of graph database and its 10X faster than 1st generation.
- Using OrientDB as a graph database: The vertices and edges can be read from/written to Spark GraphX graphs.
- OrientDB has Apache 2 license Free for any purpose, even commercial.

II. OrientDB Non-Functional Requirements:

- Java Version 1.6 or higher needs to be installed before running this database.
- **Supported Platforms:** All Linux distributions, Mac OS X, Microsoft Windows from 95/NT or later, Solaris, HP-UX and IBM AIX.
- **Implementation Language:** Java.
- **Query Languages:** SQL, Tinkerpop Gremlin, API Calls, REST
- **Scalability:** OrientDB is highly scalable database.
- **Security:** OrientDB uses password salt and data-at-rest encryption that makes it one of the most secure DBMS at present.

- **Manageability:**

OrientDB comes with a multi-model outlook that contains a pure graph database. This database is compliant with the Apache Tinkerpop standard that makes this database highly flexible and therefore much easy to manage.

12.3 Framework:



Spark:

Is a fast, scalable, and flexible open source distributed computing framework.

III. Why did we Choose Spark?

- Real-time data processing capability.
- Spark includes its own machine learning libraries, called MLlib (for large scale data processing and advanced streaming analytics).
- Spark is well known for its performance, but it's also somewhat well known for its ease of use.

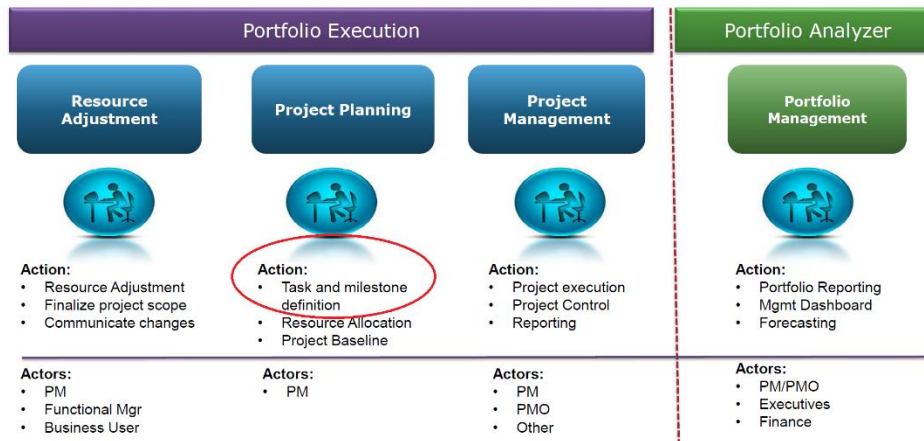
IV. Spark Non-Functional Requirements:

- Java needs to be installed before running this framework.
- **Supported Platforms:** Windows, UNIX, Mac.
- **Performance:** Spark's in-memory processing delivers near real-time analytics for data from marketing campaigns, machine learning, Internet of Things sensors and log monitoring.

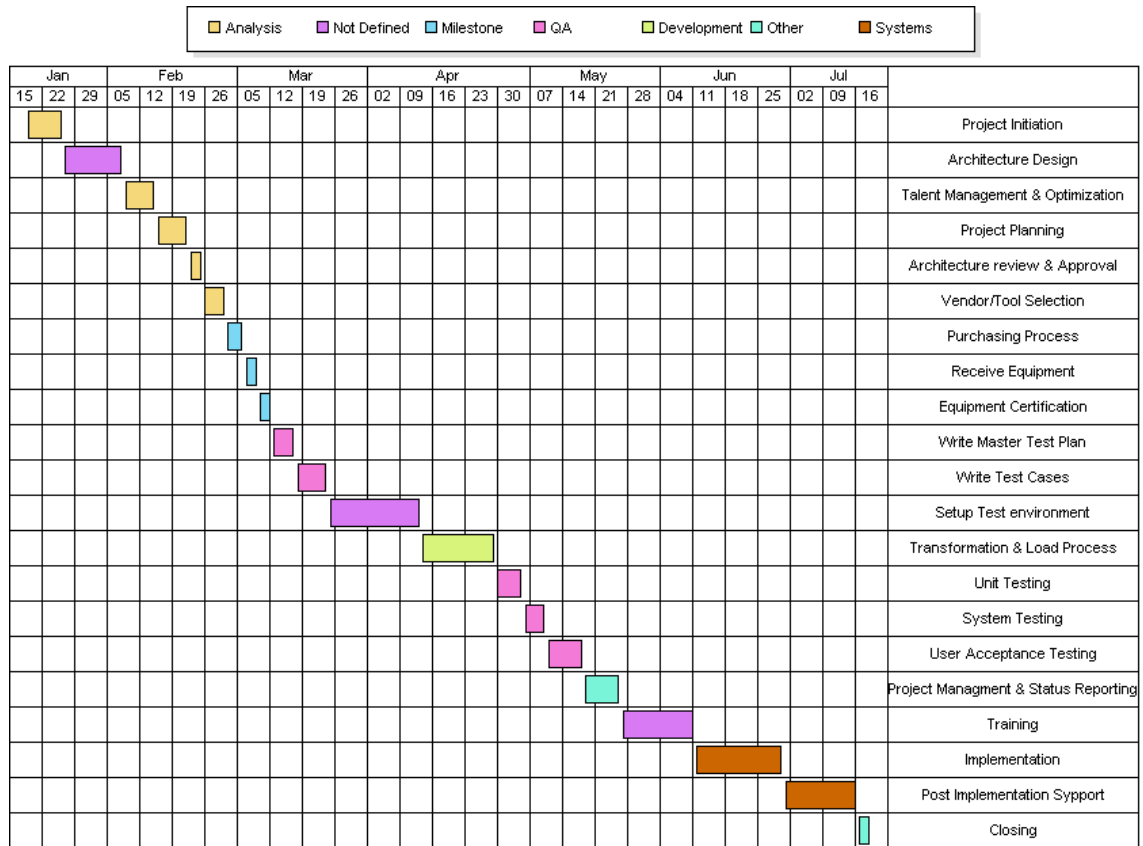
- **Data Processing:** Spark also includes its own graph computation library, GraphX. GraphX allows users to view the same data as graphs and as collections.
- **Fault Tolerance:** Spark uses Resilient Distributed Datasets (RDDs), which are fault-tolerant collections of elements that can be operated on in parallel.
- **Security:** Authentication can be configured to be on via the spark.authenticate configuration parameter.
- **Scalability:** Highly scalable.
- **Manageability:** Spark runs using its standalone cluster mode, on EC2, on Hadoop YARN, or on Apache Mesos.
- **Integration:** MapR (file system and database), Google Cloud, Amazon S3, Apache Cassandra, Apache Hadoop(HDFS), Apache HBase, Apache Hive
- **Supported Languages:** Scala, Java, Python, Clojure, R.

13. Project Milestone:

Portfolio Life Cycle (Continue)



- **Gantt Chart:**



References:

- [1] Anonymous Author. Kohl's Corporation - Company Profile, Information, Business Description, History, Background Information on Kohl's Corporation. Retrieved from <http://www.referenceforbusiness.com/history2/15/Kohl-s-Corporation.html>
- [2] Anonymous Author. Kohl's Corp (KSS). Retrieved from <http://www.reuters.com/finance/stocks/companyProfile?symbol=KSS>
- [3] Mehra. G. (2013). 6 Benefits of Predictive Analytics for Online Retailers, Retrieved from <http://www.practicalecommerce.com/articles/4122-6-Benefits-of-Predictive-Analytics-for-Online-Retailers>
- [4] Anonymous Author. Pentaho for Retail. Retrieved from <http://www.pentaho.com/solutions/retail>
- [5] Bohn, R. (2013). CRM Review – Pentaho Business Analytics. Retrieved from <https://sellmorenow.com/crm-review/pentaho/>