**Capstone Project Manual**

**Data Engineering Academy**

# Table of Contents

[Table of Contents iii](#_Toc98722883)

[Capstone Context Information 1](#_Toc98722884)

[Case Study Overview 1](#_Toc98722885)

[Milestones 2](#_Toc98722886)

[The Instructor's Role 3](#_Toc98722887)

[Case Study Roles 5](#_Toc98722888)

[CASE STUDY PRESENTATION 6](#_Toc98722889)

[Establishing the Business Need: Part I 7](#_Toc98722890)

[Establishing the Business Need: Part II 8](#_Toc98722891)

[Deliverables 8](#_Toc98722892)

[Potential Steps 8](#_Toc98722893)

[Introduction to Data Management: Part I 14](#_Toc98722894)

[Deliverables 15](#_Toc98722895)

[Steps 15](#_Toc98722896)

[Intermediate Data Management: Part I 20](#_Toc98722897)

[Deliverables 20](#_Toc98722898)

[Steps 20](#_Toc98722899)

[Intermediate Data Management: Part II 22](#_Toc98722900)

[Deliverables 22](#_Toc98722901)

[Steps 22](#_Toc98722902)

[Advanced Data Management: Part I 24](#_Toc98722903)

[Deliverables 24](#_Toc98722904)

[Steps 24](#_Toc98722905)

[Advanced Data Management Part: II 26](#_Toc98722906)

[Deliverables 26](#_Toc98722907)

[Steps 26](#_Toc98722908)

[Data Integration 28](#_Toc98722909)

[Deliverables 28](#_Toc98722910)

[Steps 28](#_Toc98722911)

[Data Analytics and Reporting 29](#_Toc98722912)

[Deliverables 29](#_Toc98722913)

[Glossary 30](#_Toc98722914)

[APPENDIX 32](#_Toc98722915)

[Appendix A: Business Requirements Document Template 33](#_Toc98722916)

[Appendix B: Use Case Examples and Graphics 49](#_Toc98722917)

[Appendix C: Calculating a Fund’s Weighted Average Risk Score 51](#_Toc98722918)

[Appendix D: Monte Carlo Simulation Background 52](#_Toc98722919)

[APPENDIX E: Risk Questionnaire 54](#_Toc98722920)

[Appendix F: Additional Resources 56](#_Toc98722921)

# Capstone Context Information

## Case Study Overview

How many times do you hear the word data… daily? Weekly? Monthly? Conversations about data were once limited to researchers and technically inclined individuals. Today, these conversations are commonplace and with good reason. Data feeds every part of an organization from product design to quality control. An organization's ability to transform data into information and information into business intelligence makes a direct impact on its potential growth, ability to avoid risk, and maintain a competitive advantage. No longer can businesses afford to rely solely on intuition or gut feel. If you have not heard the terms analytic culture or data-driven, you will soon, especially as more and more industries recognize the need for robust data management and analytics programs.

How does this affect you? As a new associate, you will serve a variety of customers, both internally and externally. This means you have to be data savvy, analytical, and able to communicate effectively to a variety of constituents the results of your analysis. It is not only about understanding data for data’s sake, but learning how it moves through an organization, its tie to the products, services and operations of the business, and most importantly how data solves business problems.

One of the best ways to build these skills is through a case study or capstone project-based learning. The case study for this track is designed to develop your data expertise, to get you to think critically about the effect of data on business, as well as to refine your problem solving and analysis skills. Although this is a simulation, expect to encounter similar, albeit more complex, problems in your work going forward.

For this case study you will focus on risk, specifically investment risk. Why risk? Risk affects many parts of the organization. It is part of managing workforce policies, asset planning, developing operational standards and procedures, as well as corporate investment strategy just to name a few. At some point or another you are bound to encounter a conversation about risk.

Though there are many facets to risk, you will focus on individual risk and investments in your case study. Understanding investment risk will require you to manipulate and work with data from a variety of sources. As part of your simulation, you will use risk profiles for individual investors and high-level assets classes to create a custom investment fund strategy and to determine the potential pool of customers who might be interested in your investment strategy. Funds can be a mix of different asset types (stocks, bonds, money market, large cap, mid cap etc.). This document either contains the information you need to complete your case study including expected deliverables, steps, and supporting documentation, or the information will be provided to you using the appropriate tools and applications.

Your instructor will guide you through the case study and may provide additional requirements as appropriate during the capstone project, just as a manager or other customer within the organization would do. Expect the instructor(s) to assume a variety of fictitious roles as you go through your case study. These roles are described in the “Case Study Roles” section.

You will work in a team throughout the project simulation. You will remain with the same team for the duration of the project.

## Milestones

At the end of each case study working session, you will be asked to report on the progress of your case study (your daily standup)*.* Each team member will be asked to provide answers to the following questions*:*

* What was accomplished?
* What remains to be done?
* Are there any outstanding issues?

These questions will help you remain on track, allow the instructor to provide you guidance as needed, and give you an opportunity to practice and refine your agile teamwork skills. Each day, you will build upon the progress from the previous session. This track concludes with your team presenting your solution to your instructor on the final day of the academy. The expectation of the final presentations is that your team will present their investment strategy as if performing a startup pitch. Each team member should present on a portion of the presentation. The presentation should divide itself into at least 3 sections…

* The work done to acquire and integrate the data into your chosen platform.
  + This is a technical overview for the company’s data architects.
  + It should include an overview of the work done by the team to achieve its objectives.
* The technology and solution architecture decisions made to facilitate analysis.
  + The case study is broken into sections… the solution implemented by the team for each section should be presented
* The investment fund strategy chosen and a study on the target customers identified from the customer list (including the reason for the choices made and analytical justification for said choices.
  + This is the analytical culmination of the work done to acquire, clean, integrate, model, migrate, and analyze the data provided to create a solution your team can pitch to the audience.

## The Instructor's Role

During the case study sessions, the instructor will play a variety of roles. No matter what role the instructor is playing for the case study, he or she will always serve as a technical resource and mentor. When teams have questions or concerns, the instructor should assume the role of the Senior Project Manager. The Senior Project Manager (SPM) overseas the development of the fund strategy and coordinates work between lines of business. The SPM can guide the team and provide input on designing the data model, constructing the database, analyzing the data, and generating/building reports.

The goal of the case study is for your team to autonomously solve the business problem; however, many of you are new to these types of activities and will require some guidance. Instructors may modify or add additional requirements to a case study section as needed or appropriate. All modifications should simulate the real-life modifications and changes that inevitably take place during a project. Adjustments should not deter nor derail any of the teams from the primary objective. Be prepared to demonstrate an agile response to change.

One critical instructor role will be as a Financial Advisor, In this capacity the instructor may need to explain various realities of the financial world, including but not limited to, why various assets would be important to a customer, the individual’s characteristics and behaviors that affect risk, and the role of a Financial Advisor in managing the customer experience. The Financial Advisor is also the individual most likely to offer a fund strategy to the customer. They need to have knowledge of the customer’s risk tolerance and know how to contact and connect with customers across market segments. In addition to the information provided in the activity description, the instructor and associates can (and should) refer to material in the Appendix. A glossary of key terms is also provided at the end of this document.

During each case study session, the teams should check-in with the their stakeholders. Every team member should work independently and maintain focus. Case study time is limited and efficiency is key. If you feel that your team is off-track and requires additional assistance, notify your instructor immediately. The sooner you acknowledge the need for help, the easier it will be for your instructor to help you get back on track. Questions are encouraged and looked upon favorably. Communication, collaboration, and outreach will be critical to the success of your project and work as an associate.

Your instructors will conduct casual check-ins throughout the case study time periods. He or she will be assessing the following:

* *How well are they progressing?*
* *Are there any persistent issues?*
* *Is the team able to communicate concerns, issues, or ideas effectively to you and to each other?*
* *What activities/topics did teams engage in the most?*
* *Which are the most challenging?*

At any given time, associates can request that breakout rooms be created to facilitate the subdivision of the teams work amongst the team members. The goal is to divide and conquer while also leveraging the shared skills and experiences the team has developed over the course of the academy. Associates can always reach out to the case study’s Senior Project Manager. This role is a fictitious position played by a real person. The Senior Project Manager will provide technical guidance where needed. This individual will also serve as the project sponsor with a title of Wealth Strategy Director.The Senior Project Manager will check-in with team members periodically regarding the progress of the team.

Everyone has a virtual environment that provides access to the technologies necessary to work through the various tasks and challenges put forth during this case study. It is recommended that the team develops a methodology for sharing their daily work with their fellow team members. In many cases the work done by one group could be an important step in the work being performed by others. Developing a workflow that allows this work to be shared and executed in each of the virtual machines will ensure that people can work independently and in parallel throughout the case study time. As always the choices made by the team should reflect the learning and technologies to which they were exposed throughout the academy and justification for the choices made should be a part of the presentation at the end.

## Case Study Roles

It is important that you understand the difference between the fictitious roles designed for this case study. The table below illustrates all the individuals you will work with as part of the case study. Your instructor will “act” as each of the fictitious roles described below. When you ask the instructor a question about the narrative of the case study, you must tell him/her what role they should be playing. For example, your team could ask the instructor, “We would like to know whether the wealth strategy director is interested in adding foreign exchange funds to the fund offering.” The instructor would then answer your team’s question as the wealth strategy director. **You are required to document all the questions you ask and “who” you asked them to.**

|  |  |
| --- | --- |
| Fictitious Role | Description |
| Senior Project Manager | Provides technical guidance to the project managers as needed. Acts as liaison to other fictitious departments in the organization. |
| Financial Advisor | Represents the individual most likely to offer a fund to the customer. They have knowledge of customers risk tolerance and know how to contact and connect with customers across segments. |
| Wealth Strategy Director | Project sponsor for the case study. Is responsible for strategic direction around portfolio and investment management for the Department of Investment Management. |
| Senior Database Engineer | Oversees data management practice and performs database development for software applications. |
| Business Analysts for Department of Investment Management | Focus on delivering key Management Information System (MIS) information via reporting solutions such as web services, OLAP, and relational database(s). |

## CASE STUDY PRESENTATION

The track concludes with the case study team creating and delivering a presentation. Team members should be able to discuss the following topics (the order below does not represent the order of the presentation):

* How they selected customers for the fund
* How they came up with the fund
* How risk influences the fund created (to what extent and why)
* The potential advantages of offering a particular fund (e.g., revenue, etc.)
* The potential advantages of the fund to the customer (e.g., better engagement, etc.)
* How they chose to assign/divide tasks to each team member
* What they learned from the project
* Were there any unresolved issues (what was attempted, how could the issues be resolved)
* What they would do differently next time

Team members should be able to communicate the technical components of the project in a way that will be understood by everyone. You are encouraged to include relevant summaries, screenshots, or data visualizations that would help explain your work product. Visual representations such as models, diagrams, and concept maps can also help clarify often complex topics and bridge the technical/non-technical divide.

# Establishing the Business Need: Part I

How do we offer our customers the right investment products and services? What does the bundle look like? Why might it change over time? Why might it change from one person to another? The answers to these questions are not always simple to find, but critical if you are to maintain your customers (retention), attract new ones (acquisition), and prevent customers from leaving (attrition).

In the real-world, there is no single investor profile. Every investor has different goals and objectives. A major part of what drives an individual to choose one set of investments over another is their tolerance of risk. Risk tolerance is, “The degree to which an investor is willing and able to accept the possibility of an uncertain outcome to an economic decision.”1 In other words, how willing the person is to accept swings or changes in the value of his/her investments. Risk tolerance is a key component of financial planning. Some factors that affect risk tolerance are age, lifestyle, overall financial health, net wealth, and the type of investment in question (savings, bonds, stocks, exchange traded funds etc.).

Rationally driven individuals understand that there is also risk in the act of taking no chances or avoiding risk completely. When designing investment products, the key is to differentiate between how someone feels and how they will react to risk factors. While emotionally charged and highly individualistic, it is still an exercise in mathematics, as you will see in later activities. Risk is not just about the individual's characteristics or risk factors. Levels of inherent risk within a product will also change over time (e.g., asset class, sector of investment, market volatility, fund goals, macroeconomic conditions etc.).

For this project, you will focus on creating an investment fund strategy. An investment fund strategy is a way of investing money along with other investors. In other words, you buy into collectively purchased securities. There are many types of funds including, mutual funds, exchange trade funds, money market funds, and hedge funds. Each fund has a different mix of assets such as bonds, stocks, and currency (all of which can be subdivided into smaller categories).

Your goal is to create a fund strategy that appeals to a segment of the customer base that you select. To select the customer base, you will need to understand your customers (demographics), the nature of their relationship to our company (preferences and engagement), and willingness to assume risk (available assets, personal objectives/goals, and risk tolerance). To complete this project, you will construct the data architecture required for the analyses and construct any necessary reports and dashboards that support your pitch of the fund strategy and target customer selection.

The potential power of the solution you produce is not trivial. A variety of wealth managers will utilize your fund offering to drive new and improve business with the customers they represent. Your offering may also serve as the foundation for broader marketing and outreach efforts for prospective customers. As with the dynamic nature of our investors, the stakeholders within the company are also unique and may require interactive reporting to maximize the benefits of your solution. Be prepared to amend your project to meet the needs of the business.

Source: <http://faculty.mccombs.utexas.edu/keith.brown/Research/FAJ-12.90.pdf>

## Establishing the Business Need: Part II

At this point, you will need to gather the requirements for the case. When working through the business requirements you will need to think about who might be involved with the project, determine what it will take to solve the problem, and distinguish between business and technical inputs and outputs. The requirements define and lay out the activities that must take place to achieve your objective. At this point you need to ensure that the team understands the core objectives presented in the preceding pages. It may behoove you to also analyze the strengths of your team. To this end you can choose to do a SWOT analysis and fill out the requirements document template below. These are not required, but they may serve to help you focus your efforts and lead you to questions that you need to ask of the various stakeholders (roles) provided. The instructor will leave the team with some time to read and digest the early parts of the document and then will host a project kickoff where the team can ask the available stakeholders clarifying questions before the team launches into an internal discussion of how they will divide the initial tasks set for them. Recall you must address your question to one of the fictitious roles outlined in the “Case Study Roles” section above. Assign someone to document all questions you asked and who you asked it to.

Your project starts with the project kickoff and completes on the last day of the capstone week. While you are being supervised by your Senior Project Manager (also known as your instructor), the overall project is being sponsored by the Wealth Strategy Director for the Department of Investment Management. For the purposes of this case study, there are no other initiatives like this taking place. This project is a pilot program for the organization driven by a need to refine product offerings, to increase customer retention, and to improve revenues through the selection of more relevant and profitable fund strategies. To complete this project you will work with members of IT including the Senior Database Engineer and business analysts from the Department of Investment Management.

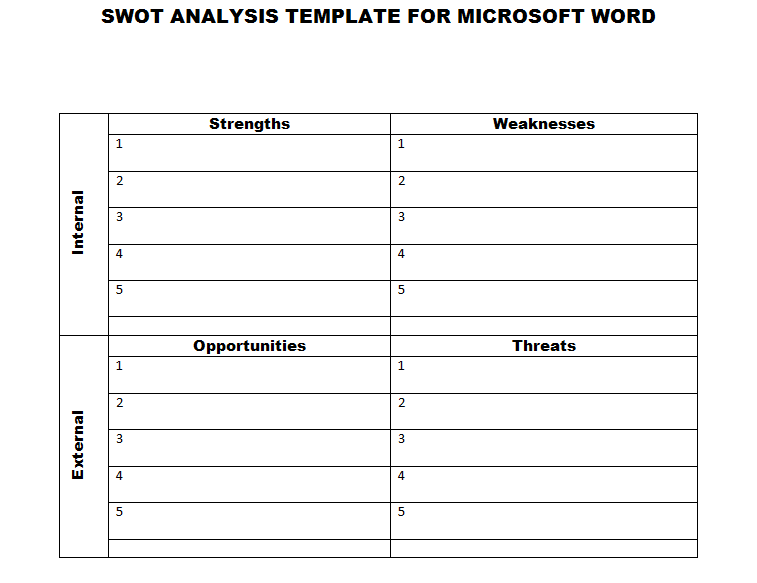
## Deliverables

* Complete your first team meeting
* Evaluate the components of the challenge and discuss them with your team

## Potential Steps

* Use the provided template to conduct a SWOT analysis
* Discuss with your team members the objectives and goals of the case study (also referred to as a project)
* Use the provided template below to fill in information about the project

Note: The following sections are part of a larger business requirements document template. The complete template can be found in Appendix A. Business requirement documents vary from one organization to another but this can help focus your efforts.



**Project Details**

|  |  |
| --- | --- |
| **Project Name** | Enter Project Name |
| **Project Type** | ***(e.g. New Initiative or Phase II)*** |
| **Project Start Date** |  |
| **Project End Date** |  |
| **Project Sponsor** |  |
| **Primary Driver** | ***(e.g. Mandatory or Efficiency)*** |
| **Secondary Driver** |  |
| **Division** |  |
| **Project Manager/Dept** |  |

**Overview**

This document defines the high level requirements **[insert project name].** It will be used as the basis for the following activities:

<Examples>

* Creating solution designs
* Developing test plans, test scripts, and test cases
* Determining project completion
* Assessing project success

**Document Resources**

|  |  |  |
| --- | --- | --- |
| **Name** | **Business Unit** | **Role** |
| <Identify all stakeholders and resources involved in gathering requirements> |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Glossary of Terms**

|  |  |
| --- | --- |
| **Term/Acronym** | **Definition** |
| <Identify any terms and acronyms used within this document or terminology that is specific to this project> |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Project Overview**

**4.1 Project Overview and Background**

<**This information can be developed from case study description**. This is a brief description of the project. It includes the current situation, the problem, and the objectives. This section serves as the vision statement for the requirements. Each requirement should bring the project closer to the vision.>

**4.2 Project Dependencies**

<List any related known projects that relate in whole or in part, or has a dependency on this project.>

**4.3 Stakeholders**

The following comprises the internal and external stakeholders whose requirements are represented by this document:

|  |  |
| --- | --- |
|  | **Stakeholders** |
| 1. |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Key Assumptions, Constraints and Open Issues**

**5.1 Key Assumptions and Constraints**

|  |  |
| --- | --- |
| **#** | **Assumptions** |
|  | List any assumptions the requirements are based on including technical assumptions |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
| **#** | **Constraints** |
|  | List any constraints the requirements are based on |
|  |  |
|  |  |
|  |  |
|  |  |
| **#** | **Unresolved Issues or Questions\*** |
|  |  |
|  |  |
|  |  |

# Introduction to Data Management: Part I

You have been tasked with creating a risk profile and creating fund strategies from a pool of available asset classes. To do this, you will need data to support your analysis and help you devise the right asset mix for your strategy, assign risk scores, and calculate the potential revenue. The Wealth Strategy Director is also looking to improve how our company engages with its individual investors. He would like to learn more about current engagement channels, the customers preferred contact method, and how often they prefer to be contacted. Some data will be provided to you to help you in this process including the following:

* Customer
* Risk Questionnaire
* Asset class
* Customer assets
* Asset allocation (personal objectives and goals

Begin thinking about how these data sources relate to each other at a high-level. The data we provide you is a starting point. You are not limited to these sources and can add sources you deem appropriate. As you contemplate the objectives of the case study and possible uses of the data consider the following:

* What data will we need?
* Where will the data come from?
* What data do we have in our current sources?
* What is the connection between the data and the objectives?
* What is or might be missing?
* How will individuals interact with the data?
* What will they need or want to see?

There should always be a clear business application for the data you have. Use Cases lay out the interaction of the user with systems and data. For instance, consider how a fund manager views asset classes and calculates future revenue or simulates future pricing. Other examples might include how a financial advisor obtains reports on customer segments while away from the office or how a business analyst uses data to track changes in financial indices.

Considering the case study. *What is one issue you are trying to solve? How will the data help solve the issue you choose? What systems and applications are involved? What does the user’s interaction with the data look like?*  Examples of “big data Use Cases” and use case graphics can be found in Appendix B.

## 

## Deliverables

* Create a data source inventory
* Design a use case using the provided template

## Steps

* Review all the data sources provided
  + - What does the data look like?
    - How much data is there?
    - What is the data describing?
    - Are the field names understandable?
  + Create a data inventory using the template on the next page
  + Note any supporting data you might need to complete your case study based on the requirements you outlined and the information provided in the first paragraph of this section
* Map the data sources, the overall project objectives
  + - List the key concepts that relate to the topic/process/objective/goal
    - Elaborate on each concept (what is essential to understanding the topic or achieving the goal?)
    - Identify links between topics/process/objective/goal
  + Perform any necessary data profiling to understand the data you are working with
* Design a Use Case for the project using the template below. For the purposes of this case study, your customer is internal not external. Your fund will ultimately be marketed to external customers but it is being developed internally first.

**TEMPLATE: DATA INVENTORY**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source Name** | **Format** | **Description of Fields (types)** | **Data types present** | **Method of Capture** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Supporting or additional data requirements:**

**USE CASE TEMPLATE**

**Use Cases**

< The primary purpose of the Use Case is to capture the required system behavior from the perspective of the internal customer achieving one or more desired goals. A Use Case contains a description of the flow of events describing the interaction between actors and the system. Typically these are used for software application design. For our purposes, focus on how the data must connect and relate. How does it tie to what the analyst will see? What about the wealth manager? What is the relationship of the data and systems to the internal constituents? >.

**Use Case Narrative**

<Each Use Case should be documented using this template. Refer to the Appendix A for Use Case Narrative instructions and example>

|  |  |  |  |
| --- | --- | --- | --- |
| Use Case ID: |  | | |
| Use Case Name: |  | | |
| Created By: |  | Last Updated By: |  |
| Date Created: |  | Date Last Updated: |  |

|  |  |
| --- | --- |
| Actors: |  |
| Description: |  |
| Preconditions: |  |
| Postconditions: |  |
| Normal Course: |  |
| Alternative Courses: |  |
| Exceptions: |  |
| Includes: |  |
| Priority: |  |
| Frequency of Use: |  |
| Business Rules |  |
| Special Requirements: |  |
| Assumptions: |  |
| Notes and Issues: |  |

**Use Case Graphic**

< See pages 44-45 and Appendix B for examples>

**TEMPLATE: DATA INVENTORY**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Source Name** | **Origin** | **Format** | **Description of Fields** | **Data types present** | **Method of Capture** | **Data refresh publication cycles** | **Data Quality Measures** |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

# Intermediate Data Management: Part I

You must also be sensitive to how customers want to engage and interact with you. In other words, some customers will be very “high touch” and want continuous contact as it relates to their investments. Other customers will want minimal contact. We also have to consider the types of contact methods (channels) whether in-person, by phone, by mail, by email or online. These preferences may influence the products and services you offer them. The only way to know if there is a meaningful relationship between engagement types, frequency and preferences, is to look for patterns in the existing data. For example, do individuals who prefer in-person, weekly, contact tend to invest in high risk funds? Or are we designing funds that will require high touch or more frequent interaction with the customer? If so, to which customers do we offer the fund?

After profiling your data you will need to develop a data model that accurately represents the available data and will assist you in constructing a database to store and manage the data provided.

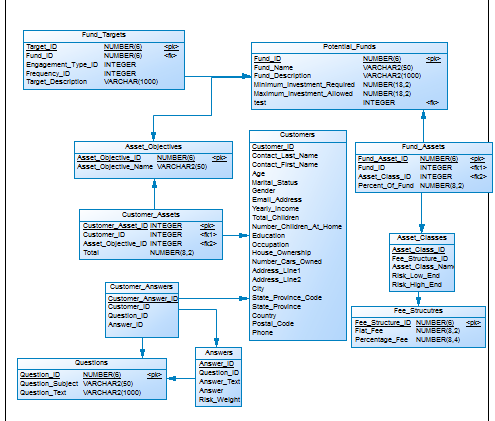
## Deliverables

* A data model for the provided data.

## Steps

* Develop a logical data model
* The engagement preference model represents how you would need to understand frequency of communication and type of communication
* Frequency of communication: Daily, Weekly, Monthly, Quarterly, Annually
* Type of communication: In-person, Email, Online, Phone, Letter, Mail
* Link the engagement entities you create to the Fund Targets and Customers entities
* Each Fund Targets record will have a distinct engagement type and frequency associated with it
* Each customer should also have a preferred engagement type and frequency
* Refer to the data model provided on the next page for guidance

**DATA MODEL FOR RISK PROFILE AND FUND OFFERING**



## Intermediate Data Management: Part II

Take the data model turn it into a reports requirement assessment. This assessment will be used throughout the remaining case study activities.

The Wealth Strategy Director loves data, but doesn’t always understand where it comes from or what is required to turn raw data into meaningful reports. You will be responsible for surfacing the information the Director requires. In particular, he is interested in learning more about the following scenarios:

* Scenario 1: Viewing customer segments based on demographic information
  + What is the age and gender distribution of investment customers?
  + What is the regional distribution of investment customers?
  + Are customers married, divorced, or single?
  + Is there a relationship between available assets and the customer’s total dependents?
    - The Director wants you to exclude individuals who have more than 2 dependents and available assets of less than $100,000
  + What is the range of assets available across various segments? What are the average available assets? What are the customer’s maximum investable assets available for each segment?
* Scenario 2: Viewing customer segments based on risk tolerance category
  + What is the relationship between risk category and age?
  + What is the relationship between the customer’s available assets and risk category?
  + Are there regional variations in the distribution of risk category?
* Scenario 3: Determining potential revenue for the mix of asset classes for the created funds
  + Once a mix of asset classes (the fund) is chosen, what is the risk category it would be assigned? (hint: what is the weighted average of the risk scores of your chosen fund mix based on the percentages of each asset class selected)
  + For created funds calculate the approximate minimum and maximum revenue per customer (hint: multiply fee structure by minimum investment required and maximum allowed by an individual)
  + What is the potential revenue by customer segment? (hint: multiply the number of customers in a particular segment by minimum and maximum revenue)

The Senior Project Manager has been working with the Senior Database Engineer (SDE), who is responsible for managing all the data systems for the Department of Investment Management, to gather the architectural information you will need to build your investment solutions. The SDE has provided a database model diagram to help you define your report requirement assessment.

The exercise of creating a report requirement assessment will serve as a core part of developing queries, building functions/procedures, and building reports/dashboards in later case study activities.

## Deliverables

* Report requirement assessment for each scenario (3)

## Steps

* Review the data model
* Review the report scenarios above (3)
* Apply the five step report requirement process to each of the scenarios
  + What is the final output of the requested report for each scenario?
  + Where am I going to get the data?
  + What relationships will I need to define to get that data?
  + What are the constraints on the relationships?
  + Additional analysis rules (e.g., how we are going to filter, aggregate etc.)

# Advanced Data Management: Part I

In your last activity, you were tasked with developing the report requirements for three scenarios. You will now use those assessments to create views (or at least saved sql statements) and to support basic inquires by business users. Each team has access to data files and must implement the creation of the database and populate it with the provided data.

The Wealth Strategy Director also wants to limit the data used in reports based on some additional insight he has obtained regarding some of the customer and fund assets. In particular, he is only interested in viewing segments that have at least 1000 customers (e.g., individuals under the age of 30). He is only interested in funds that have risk tolerance categories of 2 (Below Average), 3 (Average), and 4 (Above Average).

## Deliverables

* Develop a SQL solution/code-set of report-ready sql statements based on the 3 scenarios from the database concepts section and the additional requirements provided above

## Steps

* Recall the three scenarios laid out by the Director in the Database Concepts section
* Scenario 1: Viewing customer segments based on demographic information
  + What is the age and gender distribution of investment customers?
  + What is the regional distribution of investment customers?
  + Are customers married, divorced, or single?
  + Is there relationship between customer’s available assets and number of dependents?
    - * Do not include individuals who have dependents but available assets of less than $100,000
  + What is the range of assets available across various segments? What are the average available assets? What are the customer’s maximum investable assets available in each segment?
* Scenario 2: Viewing customer segments based on risk tolerance category
  + What is the relationship between risk category and age?
  + What is the relationship between available assets and risk category?
  + Are there regional variations in the distribution of risk category?
* Scenario 3: Determining potential revenue for the mix of asset classes for the created funds
  + - * Once a mix of asset classes (the fund) is chosen, what is the risk category it would be assigned? (The percentage of the fund attributed each asset class multiplied by its risk known risk category number. Round the output to a whole number. Call this the “weighted average risk score” See Appendix C.)
  + For created funds calculate the approximate minimum and maximum revenue per customer (hint: multiply the fee structure by minimum investment required and maximum allowed by an individual)
  + What is the potential revenue by customer segment? (hint: multiply number of customers in segment by minimum and maximum revenue)
* Build sql statements to pull data from your data model without altering data in the underlying tables

When thinking of what asset classes to include in your fund strategy, use the following data as a point of reference. The data is the result of a simulation model and is only meant to be used for explanatory purposes. The total fund mix should add up to 100%. Appendix D provides background information on the simulation model used to generate the numbers in the table.

|  |  |  |
| --- | --- | --- |
| Name | Average Return - 10 Year | Std. Dev 10 Year |
| Bonds | 4.44% | 3.29% |
| Large Cap | 7.85% | 14.32% |
| US Mid Cap | 9.55% | 17.68% |
| US Small Cap | 9.22% | 19.55% |
| Large Foreign | 2.26% | 18.21% |
| Emerging | 5.57% | 23.60% |
| Commodities | -2.62% | 18.11% |
| S&P | 7.89% | 14.74% |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Risk Tolerance | Large Cap | US Mid Cap | US Small Cap | Large Foreign | Emerging | | Commodities | Bonds |
| High Risk |  |  |  |  |  |  | |  |
| Above Average |  |  |  |  |  |  | |  |
| Average |  |  |  |  |  |  | |  |
| Below Average |  |  |  |  |  |  | |  |
| Low Risk |  |  |  |  |  |  | |  |

# Advanced Data Management Part: II

In your last activity, you were tasked with developing the views for your three report scenarios. You will now use those views to establish a more dynamic toolset based on functions and procedures. The Wealth Strategy Director wants to analyze the data by changing the combination of filters in a variety of ways. As a result you will build functions and procedures that are more general than the already established views and have parameters that enable surfacing different types of data as needed.

## Deliverables

* Produce a series of functions and procedures to support dynamic reporting in later activities

## Steps

* Create functions for each of the calculations used in the views in the previous activity. At minimum include functions for the following:
  + Assigning customers to a risk tolerance category by using a rounded average of the customer’s answers to the 8 questions in the risk questionnaire data (the risk questionnaire questions can be found in Appendix E)
  + Calculating the minimum and maximum potential revenue per customer for the potential fund view(s)
  + Calculating the weighted average risk score of your fund samples
* Create procedure toolsets for the views you created in the previous activity (Introduction to SQL)
  + The parameters for the procedures should be based on how the views need to be dynamically filtered
  + As time permits put in the following triggers and check constraints to manage or improve data integrity
    - Ensure the percentage distribution within the fund assets table cannot exceed 100%
    - Create a formatted mailing address which has first and last name on the first line, street address on line two, and City, State Zip on line three
    - Create a duplicate detection and remediation routine

Use the following sample data to think through the potential mix of assets of funds (total mix 100%)

|  |  |  |
| --- | --- | --- |
| Name | Average Return - 10 Year | Std. Dev 10 Year |
| Bonds | 4.44% | 3.29% |
| Large Cap | 7.85% | 14.32% |
| US Mid Cap | 9.55% | 17.68% |
| US Small Cap | 9.22% | 19.55% |
| Large Foreign | 2.26% | 18.21% |
| Emerging | 5.57% | 23.60% |
| Commodities | -2.62% | 18.11% |
| S&P | 7.89% | 14.74% |

# Data Integration

Your models look great! Your project manager wants you to move to the next phase. Now that you have profiled and interacted with your available data, your project manager wants you to finish populating your data tables. Any data not yet incorporated into your database should now be integrated in. Be sure to consider any cleansing or transformation tasks that would facilitate your ongoing efforts.

It is your job to analyze the data sources and develop an Extract Transform Load (ETL) solution using Pentaho that will populate your table structure from the sources provided by the Senior Database Engineer. A few things you should consider:

* During ETL, what data cleansing steps can you and should you implement to improve the quality of your data?
* What will it take to bring new data into your existing customer and engagement model when your source data doesn’t have proper normalization or well defined keys?

Your project manager may assign you additional activities as time permits.

## Deliverables

* Create an ETL Project using Pentaho to handle data flows and transformations

## Steps

* Analyze the source data (hint: similar to data profiling)
  + What form does the data take?
  + Are there data quality and integrity issues?
  + Is the data normalized?
  + Are there well defined keys?
* Using Pentaho DI create a source extraction and transformation project
  + Extract
    - How will you connect to your source data within Pentaho?
    - Are there any pre-processing steps required before the data can be consumed?
  + Transform
    - Do you need data cleansing tasks?
    - Does the data need to be manipulated before it can be fitted to your data model?
  + Load
    - How will you connect your flow to the destination?
    - Are there are any post load processing steps required?
* Create a workflow to make your ETL routine repeatable

# Data Analytics and Reporting

The principle way the data will be consumed by users is through dynamic reporting. You now need to develop and deploy interactive reports and dashboards that meet the requirements of the users. You also know based on your requirements gathering that a variety of constituents will use the reports you develop.

For this activity, you will focus on developing reports and dashboards for three roles, (1) the Business Analysts in the Department of Investment Management, (2) the Financial Advisor, and (3) the Wealth Strategy Director. Keep in mind that analyst reports are often the testing ground for what metrics should be passed through to a dashboard. In other words, what pieces of information should be continuously monitored and tracked over time. For example, organizations commonly have Key Performance Indictors for revenue, sales, and productivity included in their dashboards that were originally identified in analyst reports.

Your dashboards should be comprised primarily of data visualizations (charts and graphs). Data visualizations help users understand and interpret what is often voluminous and complex data. The more you can move away from tabular data the easier it will be for the users to consume the information you present to them. Data visualization best practices and resources are provided to you in Appendix F.

## Deliverables

* Use Power BI to develop interactive reports that will help key stakeholders visualize the reasons you chose a particular asset mix, particular subset of customers to target, key risk information, and key engagement information… along with any other visualizations you deem valuable to understanding your solution and your choices.
* Determining which contents from the reports will be included in the dashboard for the Wealth Strategy Director and the Financial Advisor
  + What graphs/charts/tables do you keep from the analysts’ reports?
  + What still needs to be added?
  + What are the metrics that the Director will want monitor and track over time in the dashboard?
* While the quantity and nature if the reports you build are at the discretion of the team… You should ensure that (at a minimum) you create reports that are…
  + focused on customer analysis
    - Should be able to segment customers on different parameters (e.g., risk tolerance, age and gender, geographic location, available assets, etc.)
    - Include engagement preferences in your dashboard where appropriate
  + focused on fund strategy analysis
* Should be able to view funds based on the what the customer’s assets are earmarked for and by engagement preferences
* Be able to filter on available assets, risk tolerance category, etc. (hint: think of how the customer pool for a fund changes based on the filters you choose)

# Glossary

|  |  |
| --- | --- |
| Cloud | Access and availability of online, on-demand computing resources. Commonly divided into software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS). Cloud environments can be public (e.g., Amazon Web Service), private (hosted on internal servers), or hybrid (public and private mix). |
| Data Governance | The organizing framework for establishing strategy, policy and objectives for managing your data as a corporate asset.  Source: <https://www.youtube.com/watch?v=iV_bT8eVP3s> |
| Dashboard | A graphical or visual summary of important measures, variables, or metrics that is used commonly in business settings |
| Data Lifecycle Management | The process of managing business information throughout its lifecycle, from requirements through retirement across applications, databases, and storage media.  Source: <http://www-01.ibm.com/software/data/lifecycle-management/> |
| Data Management | The development and execution of architecture, processes, practices and procedures to effectively manage information as an asset to the organization. |
| Data Stewardship | The formalized oversight and accountability of enterprise data assets.  Source: <http://blogs.sas.com/content/datamanagement/2014/01/07/demystifying-data-stewardship/> |
| Drill-through | An operation in which an end user selects a single cell from a cube and retrieves a result set from the source data for that cell in order to get more detailed information. In other words, passing from summary to a more detailed view such as viewing detailed transactions for a specific time frame.  Source: <https://msdn.microsoft.com/en-us/library/ms145964.aspx> |
| Drill-down | Accessing data that is in a lower level of a hierarchically structured database (e.g., passing from Country to State to City levels of detail).  Source: Google Definitions |
| Data Visualization | Data visualization is the representation of data in a pictorial or graphical format. The purpose of data visualization is to simplify data values, promote the understanding of them, and communicate important concepts and ideas. |
| Extract Transform Load (ETL) | Three functions of data processing combined into a single programming tool. Extraction, reads the source data and extracts the specified subset. Transform, works with acquired data to convert it to a desired state. Load, writes the resulting data to a target database, which may exist already or be newly created.  Source: <http://searchdatamanagement.techtarget.com/definition/extract-transform-load> |
| Line of Business (LOB) | A corporate subdivision focused on a single product or family of product**s.**  Source: <http://www.gartner.com/it-glossary/lob-line-of-business> |
| Segmentation | Divisions or subdivisions of homogenous market groups having similar characteristics (age, needs, wants, demand, expectations, behaviors, outcomes etc.). |
| Stakeholder | A person with vested interest or concern in a particular area, product, service, initiative etc. |

# APPENDIX

Appendix A: Business Requirements Document Template

Appendix B: Use Case Examples and Graphics

Appendix C: Calculating a Fund’s Weighted Average Risk Score

Appendix D: Monte Carlo Simulation

Appendix E: Risk Questionnaire

Appendix F: Additional Resources

## Appendix A: Business Requirements Document Template

**Source:** New York University Program Service Offices-Toolkit <https://www.nyu.edu/content/dam/nyu/prgmServices/documents/27_PSO_Business_Requirements.doc>

Template begins on the next page of this document.

**Appendix A: Business Requirements Document Template**

**[Insert Project Name]**

Business Requirements Document (BRD)

Version 46

**Appendix A: Business Requirements Document Template**

**Version and Approvals**

|  |  |
| --- | --- |
| **Version History** | |
| **Version #** | **Date** | | **Revised By** | **Reason for change** |
|  |  | |  |  |
|  |  | |  |  |
|  |  | |  |  |
|  |  | |  |  |

This document has been approved as the official Business Requirements Document for <project name>, and accurately reflects the current understanding of business requirements. Following approval of this document, requirement changes will be governed by the project’s change management process, including impact analysis, appropriate reviews, and approvals.

|  |  |
| --- | --- |
| **Document Approvals** | |
| **Approver Name** | **Project Role** | | **Signature/Electronic Approval** | **Date** |
|  |  | |  |  |
|  |  | |  |  |
|  |  | |  |  |
|  |  | |  |  |

**Appendix A: Business Requirements Document Template**

**[Table of Contents Goes Here**]

**Appendix A: Business Requirements Document Template**

**Project Details**

|  |  |
| --- | --- |
| **Project Name** | Enter Project Name |
| **Project Type** | ***(e.g. New Initiative or Phase II)*** |
| **Project Start Date** |  |
| **Project End Date** |  |
| **Project Sponsor** |  |
| **Primary Driver** | ***(e.g. Mandatory or Efficiency)*** |
| **Secondary Driver** |  |
| **Division** |  |
| **Project Manager/Dept** |  |

**Overview**

This document defines the high level requirements [insert project name]. It will be used as the basis for the following activities:

* Creating solution designs
* Developing test plans, test scripts, and test cases
* Determining project completion
* Assessing project success

**Appendix A: Business Requirements Document Template**

**Document Resources**

| **Name** | **Business Unit** | **Role** |
| --- | --- | --- |
| <Identify all stakeholders and resources involved in gathering requirements> |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Glossary of Terms**

| **Term/Acronym** | **Definition** |
| --- | --- |
| <Identify any terms and acronyms used within this document> |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Project Overview**

**4.1 Project Overview and Background**

<**This information can be taken from the Project Charter**. It is a brief description of the project including the current situation, the problem, and the objectives. This section serves as the vision statement for the requirements. Each requirement should bring the project closer to the vision.>

**4.2 Project Dependencies**

<List any related known projects that relate in whole or in part, or has a dependency on this project.>

**Appendix A: Business Requirements Document Template**

**4.3 Stakeholders**

The following comprises the internal and external stakeholders whose requirements are represented by this document:

|  | **Stakeholders** |
| --- | --- |
| 1. |  |
| 2. |  |
| 3. |  |

***Example of a completed use case:***

|  |  |  |  |
| --- | --- | --- | --- |
| Use Case ID: | 1 | | |
| Use Case Name: | View Interactive Campus Map | | |
| Created By: | Dan Sward | Last Updated By: |  |
| Date Created: | 4/19/09 | Date Last Updated: |  |

|  |  |
| --- | --- |
| Actors: | User |
| Description: | This use case describes the main way this interactive campus map will be used – as a web browser accessed application. The user accesses the appropriate URL and interacts with the functionality made available. |
| Preconditions: | Web browser opened, and interactive campus map URL accessed. |
| Postconditions: | User navigates from interactive campus map web site. |
| Normal Course: | 1. Open browser 2. Navigate to campus map URL 3. Interact with the campus map using available functionality |
| Alternative Courses: | None |
| Exceptions: | None |
| Includes: |  |
| Priority: | High |
| Frequency of Use: | Once per visit. |
| Business Rules | TBD… |
| Special Requirements: | * 24/7 access * Response times comparable to common web mapping solutions (e.g. Google Maps) * U of M accessibility requirements * U of M eCommunications requirements |
| Assumptions: |  |
| Notes and Issues: |  |
| Use Case Graphic | UseCase1 |

**Appendix A: Business Requirements Document Template**

**Business Requirements**

The following sections document the various business requirements of this project.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Requirement Type | ID – Prefix ?? | ID – Number | Function – Feature - Requirement | Use Case Reference | Required | **??** | **??** | **??** | Comments |
|  | Business User Requirements | | | | | | | | |
|  | f | 0001 |  |  |  |  |  |  |  |
|  | f | 0002 |  |  |  |  |  |  |  |
|  | f | 0003 |  |  |  |  |  |  |  |
|  | f | 0004 |  |  |  |  |  |  |  |
|  | F | 0005 |  |  |  |  |  |  |  |
|  | F | 0007 |  |  |  |  |  |  |  |
|  | f | 0007 |  |  |  |  |  |  |  |
|  | f | 0008 |  |  |  |  |  |  |  |
|  | Reporting Requirements | | | | | | | | |
|  | f | 0001 |  |  |  |  |  |  |  |
|  | f | 0002 |  |  |  |  |  |  |  |
|  | f | 0003 |  |  |  |  |  |  |  |
|  | f | 0004 |  |  |  |  |  |  |  |
|  | F | 0005 |  |  |  |  |  |  |  |
|  | F | 0007 |  |  |  |  |  |  |  |
|  | f | 0007 |  |  |  |  |  |  |  |
|  | f | 0008 |  |  |  |  |  |  |  |
|  | User Access/Security Requirements | | | | | | | | |
|  | f | 0001 |  |  |  |  |  |  |  |
|  | f | 0002 |  |  |  |  |  |  |  |
|  | f | 0003 |  |  |  |  |  |  |  |
|  | f | 0004 |  |  |  |  |  |  |  |
|  | F | 0005 |  |  |  |  |  |  |  |
|  | F | 0007 |  |  |  |  |  |  |  |
|  | f | 0007 |  |  |  |  |  |  |  |
|  | f | 0008 |  |  |  |  |  |  |  |
|  | Service Level/Performance Requirements | | | | | | | | |
|  | f | 0001 |  |  |  |  |  |  |  |
|  | f | 0002 |  |  |  |  |  |  |  |
|  | f | 0003 |  |  |  |  |  |  |  |
|  | f | 0004 |  |  |  |  |  |  |  |
|  | F | 0005 |  |  |  |  |  |  |  |
|  | F | 0007 |  |  |  |  |  |  |  |
|  | f | 0007 |  |  |  |  |  |  |  |
|  | f | 0008 |  |  |  |  |  |  |  |
|  | Scalability Requirements | | | | | | | | |
|  | f | 0001 |  |  |  |  |  |  |  |
|  | f | 0002 |  |  |  |  |  |  |  |
|  | f | 0003 |  |  |  |  |  |  |  |
|  | f | 0004 |  |  |  |  |  |  |  |
|  | F | 0005 |  |  |  |  |  |  |  |
|  | F | 0007 |  |  |  |  |  |  |  |
|  | f | 0007 |  |  |  |  |  |  |  |
|  | f | 0008 |  |  |  |  |  |  |  |
|  | Support and Maintenance Requirements | | | | | | | | |
|  | f | 0001 |  |  |  |  |  |  |  |
|  | f | 0002 |  |  |  |  |  |  |  |
|  | f | 0003 |  |  |  |  |  |  |  |
|  | f | 0004 |  |  |  |  |  |  |  |
|  | F | 0005 |  |  |  |  |  |  |  |
|  | F | 0007 |  |  |  |  |  |  |  |
|  | f | 0007 |  |  |  |  |  |  |  |
|  | **f** | **0008** |  |  |  |  |  |  |  |

**Template Appendixes**

*Appendix A – Business Process Flows*

<Describe the current existing process workflow using flow diagrams (i.e.,Visio Flowcharts or other process flow) and/or a detailed narrative.>



**Appendix B – Business Rules Catalog**

<Instructions: Use the following template for each business rule. >

|  |  |
| --- | --- |
| Business Rule Name: | <The name should give you a good idea about the topic of the business rule.> |
| Identifier | <Defines unique identifier.> *EXAMPLE: BR1* |
| Description | <Defines the rule in detail.> *EXAMPLE: “All employee labor is tracked, reported and billed in 15 minute increments.”* |
| Example | <(Optional) An example of the rule> |
| Source | <Source of the rule. E.g. stakeholder> |
| Related Rules | <List of related rules, to support traceability> |

**Use Case Narrative Instructions**

<Instructions for completing the Use Case Narrative are included here. Remove these instructions from the completed Business Requirements Document>.

| **Use Case Field Name** | **Definition** |
| --- | --- |
| Use Case ID | Give each use case a unique numeric identifier, in hierarchical form: X.Y. Related use cases can be grouped in the hierarchy. Functional requirements can be traced back to a labeled Use Case. |
| Use Case Name | State a concise, results-oriented name for the Use Case. These names reflect the tasks the user needs to be able to accomplish using the system. Include an action verb and a noun. Some examples:   * View part number information. * Manually mark hypertext source and establish link to target. * Place an order for a CD with the updated software version |
| Created By | Include the name of the person who initially documented this Use Case. |
| Date Created | Enter the date on which the Use Case was initially documented |
| Date Last Updated | Enter the date on which the Use Case was most recently updated |
| Last Updated By | Include the name of the person who performed the most recent update to the Use Case description. |
| Actor | Enter the person or other entity external to the software system being specified who interacts with the system and performs Use Cases to accomplish tasks. Different actors often correspond to different user classes, or roles, identified from the customer community that will use the product. Name the actor(s) that will be performing this Use Case. |
| Description | Provide a brief description of the reason for and outcome of this Use Case, or a high-level description of the sequence of actions and the outcome of executing the Use Case. |
| Preconditions | List any activities that must take place, or any conditions that must be true, before the Use Case can be started. Number each precondition. Examples:   * User’s identity has been authenticated * User’s computer has sufficient free memory available to launch task |
| Post conditions | Describe the state of the system at the conclusion of the use case execution. Number each post condition. Examples:   * Document contains only valid SGML tags * Price of item in database has been updated with new value |
| Normal Course | Provide a detailed description of the user actions and system responses that will take place during execution of the Use Case under normal, expected conditions. This dialog sequence will ultimately lead to accomplishing the goal stated in the Use Case name and description. This description may be written as an answer to the hypothetical question, “How do I <accomplish the task stated in the Use Case name>?” This explanation is best represented as a numbered list of actions performed by the actor, alternating with responses provided by the system. |
| Alternative Courses | Document other, legitimate usage scenarios that can take place within this Use Case separately in this section. State the alternative course and describe any differences in the sequence of steps that occur. Number each alternative course using the Use Case ID as a prefix, followed by “AC” to indicate “Alternative Course”. Example: X.Y.AC.1 |
| Exceptions | Describe any anticipated error conditions that could occur during execution of the Use Case and define how the system is to respond to those conditions. Also, describe how the system will respond if the Use Case execution fails for some unanticipated reason. Number each exception using the Use Case ID as a prefix, followed by “EX” to indicate “Exception”. Example: X.Y.EX.1 |
| Includes | List any other use cases that are included (“called”) by this Use Case. Common functionality that appears in multiple use cases can be split into a separate Use Case that includes by the ones that need that common functionality. |
| Priority | Indicate the relative priority of implementing the functionality required to allow this Use Case to be executed. The priority scheme utilized must be the same as that used in the software requirements specification. |
| Frequency of Use | Estimate the number of times this Use Case will be performed by the actors per some appropriate unit of time. |
| Business Rules | List any business rules that influence this Use Case. |
| Special Requirements | Identify any additional requirements, such as nonfunctional requirements, for the Use Case that may need to be addressed during design or implementation. These situations may include performance requirements or other quality attributes. |
| Assumptions | List any assumptions that were made in the analysis that led to accepting this Use Case into the product description and writing the Use Case description. |
| Notes and Issues | List any additional comments about this Use Case or any remaining open issues or TBDs (To Be Determined) that must be resolved. Identify who will resolve each issue, the due date, and the ultimate resolution. |

## Appendix B: Use Case Examples and Graphics

**USING DATA TO ADDRESS BUSINESS PROBLEMS**

**Retail** The retail industry along with others wants to have a more holistic view of its customers. They are seeking ways to track and measure customer interactions across channels (web, mobile, social media). For example: Where do people shop the most? Do they make recommendations? Are there behavioral triggers that can tell us something about purchasing behaviors (frequency of abandoning a cart)? Most of this data is unstructured, clickstream data. To create this holistic view requires integrating clickstream, web analytic, and other data with existing customer data marts and warehouses.

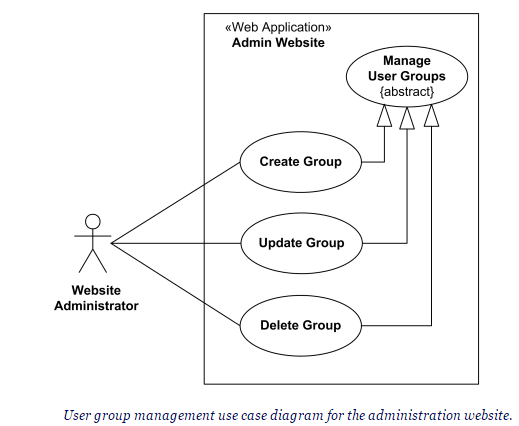
**Transportation** GE Aviation uses streaming data from sensors connected to engines to flag existing maintenance issues and predict future maintenance needs. The algorithms for processing data and flagging non-standard services have even been used to update GE’s customer portal to find information on service updates.

**Public Sector** Repeat offenders and parole violators remain a major problem for law enforcement. Often the databases of these systems are disparate or not integrated. If a criminal commits an act in one county and then a different crime in the neighboring county, the different police departments may not be able to see both offenses. This disconnect occurs since many systems are not integrated and lack advanced analytic capabilities that could flag the offender regardless of where the crime is committed. States across the United States are working with technology firms and federal agencies to combat this problem.

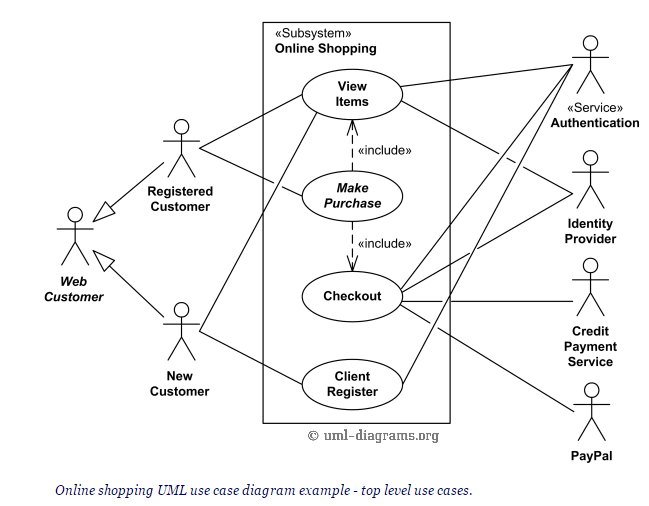
**Healthcare** The Ebola epidemic is a prime example of how data is used to track, measure and understand disease patterns. Clinics, hospitals, and nurses throughout the country log cases as they happen. “Flash Messages to Freetown” is a newly established facility that cares for children who may be infected with Ebola. Workers review notes and respond to questions about the children to the Ministry of Social Welfare, Gender, and Children’s Affairs via SMS. <http://www.unicef.org/infobycountry/sierraleone_81144.html>

**Insurance** More and more insurance companies are mining the language of historical policies to understand how their services and product offerings evolve over time. This information helps them determine which aspects of policies are still relevant and which need to be updated to reflect current issues or standards. Health insurance companies are also using text analytics to mine electronic health records to better understand why patients come back to the ER or have repeat visits to primary care physicians. They look for triggers and mechanisms that cause a person to be compliant or fail to obey physician recommendations. Then, they use this information to determine if they need to provide ancillary services, such as sending nurses to visit high risk diabetes patients at home. If they can lower the person’s likelihood to need emergency services or additional medications, it saves the patient and insurer money.

**Graphics**



Source: <http://www.uml-diagrams.org/examples/website-admin-use-case-diagrams-example.html>



Source: <http://www.uml-diagrams.org/examples/online-shopping-use-case-diagram-example.html>

## Appendix C: Calculating a Fund’s Weighted Average Risk Score

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Risk Tolerance | Large Cap | US Mid Cap | US Small Cap | Large Foreign | Emerging | Commodities | Bonds |
| High Risk |  |  |  |  |  |  |  |
| Above Average |  |  |  |  |  |  |  |
| Average |  |  |  |  |  |  |  |
| Below Average |  |  |  |  |  |  |  |
| Low Risk |  |  |  |  |  |  |  |

You are being asked to attribute a percentage fund to each class. Your decision for what percentage to include should be based on the average returns for that class and any other indicators you feel are relevant. For example, you could have a high risk fund with 50% large cap, 50% Emerging, and nothing else.

|  |  |  |
| --- | --- | --- |
| Name | Average Return - 10 Year | Std. Dev 10 Year |
| Bonds | 4.44% | 3.29% |
| Large Cap | 7.85% | 14.32% |
| US Mid Cap | 9.55% | 17.68% |
| US Small Cap | 9.22% | 19.55% |
| Large Foreign | 2.26% | 18.21% |
| Emerging | 5.57% | 23.60% |
| Commodities | -2.62% | 18.11% |
| S&P | 7.89% | 14.74% |

The weighted average is the percentage of the fund for each asset class multiplied by the risk category value. The risk category value can be found in your schema. You will need to round to the closest whole number and your fund should have at least 4 assets. Be able to justify your assignment of the risk category value and percentage weight.

From the example above 50% Large Cap (risk category value) + 50% Bonds (risk category value) .5(5)+.5(1)= 3

## Appendix D: Monte Carlo Simulation Background

The Monte Carlo theory holds that when a simulation is performed enough times, given a random probability of success, a fairly predictable outcome can be established. Although no tool can predict the future with complete accuracy, the Monte Carlo is often used in financial planning to perform market and portfolio simulations.

The coin flip is an easy way to begin thinking through the Monte Carlo simulation. On average the probability of heads is 50% tails 50%. If you only flip the coin a few times, you may get all heads or all tails. As you flip the coin more and more the probability of any one side migrates to 50/50. Monte Carlo can be used to simulate thousands of coin flips.

Now use this theory to run a simulation based on the market’s overall potential return. For this exercise we are viewing the market as the aggregate of all listed securities. This idea is not meant to diminish the role of other important indicators and influences such as commodities, bond markets, foreign markets, private investment, and non-standard investments etc., but intended to keep the example simple.

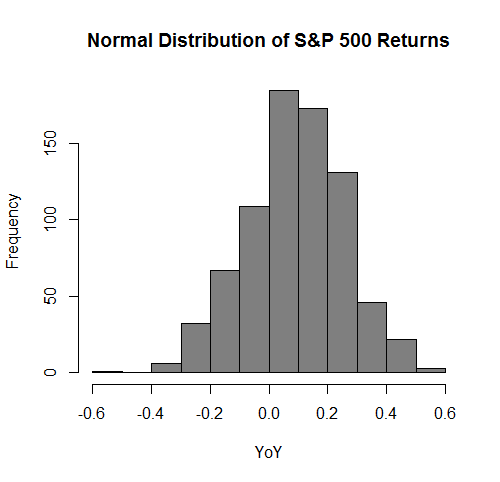
Using a history of an index such as the S&P 500 we can set our basis for the predicted market return in a given month. Analyzing the returns since 1951 we get the following summary statistics for Year-over-Year returns:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| MIN | 1ST QU. | MEDIAN | MEAN | 3RD QU. | MAX | STD DEV |
| (44.8%) | (1.5%) | 9.8% | 8.8% | 19.0% | 53.0% | 16.1% |

***Information courtesy of Jeff Burns.***

The results show a wide range of YoY. In over 63 years, the market has swung in a range of almost 100 percentage points. The standard deviations show us the volatility. A standard deviation of 16.1% indicates we would expect to find 68% of the values centered on the mean (in a normal distribution).

A normalized distribution of the S&P 500 returns looks like this:



Use this normalized model to skew the randomness towards what is probable. Running the simulations many times will result in a smoothed and predictive data set.

As an example, using PL/SQL, R, or some other statistical tool, you can create a random number. Using the normal distribution as my potential return set, run a single random number generation. Imagine your return was 1.2%.

If an investor had $10,000 in investable assets 1 year ago, they would now have $10,120 after Running the random number sequence 10 more times, we get the following:

|  |  |
| --- | --- |
| YoY Return | Ending Balance |
| -10% | $ 9,091 |
| -23% | $ 7,817 |
| -12% | $ 8,859 |
| 27% | $ 12,859 |
| -17% | $ 8,370 |
| 31% | $ 13,261 |
| 14% | $ 11,538 |
| 8% | $ 10,900 |
| 3% | $ 10,426 |
| -26% | $ 7,462 |

The YoY Return is the randomly generated return in the total market based on our normal distribution. Never conduct a Monte Carlo with less than 1,000 cases. The example above is purely illustrative, and provides an average return of 0.58% ( in this limited sample). This is calculated by taking the average of the Ending Balance column as a percentage growth over the original $10,000.0.58% (average of the Ending Balance column as a percentage growth over the original $10,000). If the same simulation is run 1000 times, we get an annual return of 9.05%.

## APPENDIX E: Risk Questionnaire

How many years from now do you expect to begin taking income from your investments?

1. 1-5 Years
2. 5-10 Years
3. 10-15 Years
4. 15-20 Years
5. 20+ Years

Choose the **BEST** match to your feelings on the following situation: Your account value after rising 15% in the previous quarter is down 20% this quarter.

1. Sell it all, cash is the place for me
2. Start selling but not all of it
3. I’m reaching for the phone to call my advisor
4. I’m very nervous
5. I feel a little nervous about it
6. I’m neutral
7. Doesn’t really matter, I’m in it for the long haul
8. I’m ready to look into buying
9. Start buying
10. Buy, buy, buy!

What if your portfolio’s value declined another 10% in the subsequent quarter?

1. Sell it all and move to safety
2. Start selling heavily
3. Start divesting the declining positions
4. Wait it out/Do nothing
5. Consider investing in declining positions
6. Start investing in declining positions
7. Start investing heavily
8. Invest in the positions that have declined the most

What role does your retirement account play in your overall savings strategy? Choose the best answer for you:

1. Who has the funds to save for retirement?
2. It’s all I have
3. I have an emergency fund as well
4. I’m thinking about family planning not retirement
5. I’m thinking about home ownership
6. It’s a piece of the picture, I also have real estate
7. I have retirement in hand and I’m looking to diversify
8. I have a fully diversified portfolio

What do you expect your total income to do over the next 10-15 years?

1. Decline substantially
2. Decrease moderately
3. Decrease slightly
4. Increase minutely
5. Fluctuate with a downward trend
6. Stay the same
7. Fluctuate with an upward trend
8. Improve slightly
9. Improve moderately
10. Increase significantly

If given the opportunity to improve your returns by selecting investments whose value may fluctuate significantly over time, you would:

1. Not likely to take on more risk
2. Take a little risk with a small portion of the portfolio
3. Take a moderate risk with a small portion of the portfolio
4. Take a moderate risk with half the portfolio
5. Take a moderate risk with most of the portfolio
6. Take significant risk with half of the portfolio
7. Take a significant risk with most of the portfolio
8. Take a lot of risk with the whole portfolio

When it comes to keeping track of your investments, which statement below BEST describes you?

1. I review my statements regularly and have my portfolio reviewed regularly
2. I review my statements regularly and have my portfolio reviewed annually
3. I review my statements regularly and have my portfolio reviewed every once in awhile
4. I review my statements regularly
5. I review my statements annually
6. I take a look at the statement every once in a while
7. I rarely monitor my investment accounts and I don’t have them reviewed
8. I never monitor my investment accounts and I don’t have them reviewed

How would you describe your investment experience?

1. No experience
2. Little experience
3. Some experience, but only in mutual funds
4. Some experience in mutual funds, stocks, and bonds
5. Experienced in mutual funds
6. Experienced in mutual funds and some individual stocks and bonds
7. Vast knowledge and experience mutual funds and experienced with individual stocks and bonds
8. Vast knowledge of experience with many types of investments

## Appendix F: Additional Resources

**Templates and Documentation**

* + Program Service Offices provides templates to guide users through project planning and implementation <http://www.nyu.edu/about/leadership-university-administration/office-of-the-president/office-of-the-executivevicepresident/finance-and-budget/program-services/redirect/toolkit.html>
* UNC The Odum Institute-Data Management Services <http://www.irss.unc.edu/odum/contentSubpage.jsp?nodeid=661>

**Data and Development Concepts**

* + - O’Reilly Media <http://www.oreilly.com/>
    - W3schools <http://www.w3schools.com/>
    - Diagraming Use Cases- <http://www.tutorialspoint.com/uml/index.htm>
    - Data Visualization-http://www.sas.com/en\_us/insights/articles/analytics/why-your-brain-needs-data-visualization.html

**Analytics and Reporting**

* + Monte Carlo Simulation
    - <https://www.youtube.com/watch?v=3gcLRU24-w0>
    - <https://www.youtube.com/watch?v=8uBxLVvafx0>
* All Analytics Academy <http://www.allanalytics.com/lecture-calendar.asp>
* The International Institute of Analytics <http://www.iianalytics.com/>
* BI Scorecard <http://www.biscorecard.com/webinar-master/>
* Gap Minder <http://www.gapminder.org/>

**Cloud and Big Data**

* Apache™ Hadoop® <https://hadoop.apache.org/>
* Amazon Web Services <http://aws.amazon.com/what-is-cloud-computing/>
* Amazon Web Services Public Data Sets <http://aws.amazon.com/public-data-sets/>
* Cloudera <http://www.cloudera.com/content/cloudera/en/home.html>
* Google Public Data Explorer <http://www.google.com/publicdata/directory>
* Udemy <https://www.udemy.com/>
* 8 Free Hadoop Online Training Sources <http://www.tomsitpro.com/articles/free-hadoop-training-online,1-2074.html>