ETL, Datawarehouse & Analytics

Ed Kiiru

Table of Contents

[Revision History 1](#_Toc436286629)

[Business Background 3](#_Toc436286630)

[Project Information 3](#_Toc436286631)

[Project Diagram 4](#_Toc436286632)

[Project Methodology: 5](#_Toc436286633)

[Agile Methodology stories: 6](#_Toc436286634)

[R1:SP1: Story 1: Gather 1003 Mortgage application input data 6](#_Toc436286635)

[R1:SP2: Story 2: Read input data 7](#_Toc436286636)

[R1:SP2: Story 3: Insert data into staging tables 7](#_Toc436286637)

[R1:SP3: Story 4: Validate & Error Handling staging data 8](#_Toc436286638)

[R1:SP3: Story 5: Insert data into ODS database 8](#_Toc436286639)

[R1:SP4: Story 6: Insert data into OLAP / Data warehouse / Dimensional Model 9](#_Toc436286640)

[R2:SP5: Story 7: Generate Reports – Loans to Date 10](#_Toc436286641)

[R2:SP5: Story 8: Generate Reports – Loans to Date / Demographics 11](#_Toc436286642)

[R2:SP6: Story 9: Generate Reports – Loan Overview 12](#_Toc436286643)

[R2:SP6: Story 10: Report Functionality 13](#_Toc436286644)

[Appendix 1: Form Fields Information 15](#_Toc436286645)

[Appendix 2: Questions to explore as you do the project 16](#_Toc436286646)

[R1:SP1: Story 1: 16](#_Toc436286647)

[R1:SP2: Story 2: 16](#_Toc436286648)

[R1:SP2: Story 3: 16](#_Toc436286649)

[R1:SP3: Story 4: 17](#_Toc436286650)

[R1:SP3: Story 5: 17](#_Toc436286651)

[R1:SP4: Story 6: 17](#_Toc436286652)

[R2:SP5: Story 7, 8 & R2:SP6: 9,10: 18](#_Toc436286653)

[Additional Questions: 19](#_Toc436286654)

# Business Background

**Client:** Revolution Mortgage Group (CAPTIDE MORGAGE)

**About CAPTIDE MORGAGE:** CAPTIDE MORGAGE has built a strong reputation as an outstanding mortgage lender, serving the lending needs of individual homebuyers, real estate professionals and builders throughout their lending network.  With offices in New York, Boston, Raleigh, Dallas, Houston, Austin, Atlanta, and Tampa, CAPTIDE MORGAGE is able to service customers in New York, Massachusetts, North Carolina, Texas and Florida.

CAPTIDE MORGAGE is a full-service mortgage lender with an experienced mortgage loan offers offering expertise in every area of mortgage lending, from purchase to refinance to construction lending. CAPTIDE MORGAGE have access to a full range of mortgage sources and all CAPTIDE MORGAGE lending specialists are dedicated to finding the right loan-with the best rates, terms and costs-to meet their customer’s unique needs.

As a subsidiary of North Beach Bank, Revolution Mortgage Group has the resources and support of a community bank with over 250 years of experience. CAPTIDE MORGAGE customers not only have access to the best loans available in the marketplace, but also can review loan alternatives, and even apply for loan, at their convenience, online - 24 hours a day.

# Project Information

**Project Description:** CAPTIDE MORGAGEreceives 1003 mortgage applications from customers. CAPTIDE MORGAGE receives application from multiple sources such as online, email/fax and from Zillow. An ETL solution should be developed to consolidate the loan applications that are received into one Loan Application Database. ETL process should validate the loan applications for all the required fields and for field format. If there are any problems with the applications, loan officers are notified about the errors. Loan officers will review the application errors, fix the errors, and resubmits the application for further validation. Loan officers will review the valid applications daily and either approves the loan or rejects the loan. Both approved and rejected loans are then loaded into data warehouse for reporting purposes.

**Project Input data:** Applications are received in the following formats:

1. Online: Customers can visit CAPTIDE MORGAGE website and fill the form at: <https://apply.revolutionmortgage.com/apply/name>, The data captured using online application is stored in a OLTP database.
2. Email or Fax: Customers can fill Fannie Mae application and email or fax it directly to CAPTIDE MORGAGE loan service representative. Sample Fannie Mae application is located at <https://singlefamily.fanniemae.com/media/15141/display>. Loan officers will manually fill this data into a spread sheet daily and sends it to FTP location daily.  This CSV file will have all the applications that are received during the day.
3. Zillow: Customers can go to [Zillow.com](http://Zillow.com), search for mortgage rates and apply for mortgage with CAPTIDE MORGAGE. Zillow will send the 1003 application in an XML format real time. This xml file is placed in an FTP location. This xml file always has one loan application.

**Data validations:** All the required information on the received mortgage application data should be in a format that can be loaded into the OLAP and the Data warehouse. For data validation requirement, refer appendix 1 of the 1003 Application Requirements Document.

**Data warehouse:** Data warehouse should be created to meet future growth of the group across USA.

**Reports:** Reports developed should provide intelligence to business on an hourly, daily, weekly, monthly, quarterly, and yearly basis.

# Project Diagram



# Project Methodology:

CAPTIDE MORGAGE is very particular about getting high quality project deliverables. Agile methodology will be used to implement this project. Agile methodology will divide project into Releases, Sprints and User Stories.

* A Project has multiple releases, with each release delivering a business value to its customers.
* A Release contain one or more sprints.
* A Sprint, with a time boxed scope, implement one or more stories in that sprint.
* A Story, an end user driven business function may contain one or more use cases. Often in the business, a story is interchangeably used with a use case (two terms from different generations). A story tends to be driven more by the business user perspective whereas use cases used to represent the system perspective.

This mortgage project will use the following Agile methodology structure:

|  |  |  |
| --- | --- | --- |
| Release | Sprint | Story |
| Release 1 (R1) | Sprint 1 (SP1) | Story 1 |
| Sprint 2 (SP2) | Story 2  Story 3 |
| Sprint 3 (SP3) | Story 4  Story 5 |
| Sprint 4 (SP4) | Story 6 |
| Release 2 (R2) | Sprint 5 (SP5) | Story 7  Story 8 |
| Sprint 6 (SP6) | Story 9  Story 10 |

*Note: You are encouraged to allocate a max of one week of time for each sprint and complete the project within six weeks or less time.*

# Agile Methodology stories:

## R1:SP1: Story 1: Gather 1003 Mortgage application input data

**Narrative:**

As a Developer

I want to gather multiple types of 1003 mortgage application input data from various sources

So that I can process the data using ETL

**Scenario 1 - Excel Data:** Get 1003 mortgage application data in excel format from Loan officers

Loan officers receive 1003 mortgage applications from borrowers in an email or fax or via phone. They will capture each application data into an excel file and upload that excel file to an ftp folder at the end of the day.

Developers should create this input excel file with dummy data for practice purpose. 1003 PDF application can be found at <https://singlefamily.fanniemae.com/media/15141/display>. You can also find an online html application at <https://apply.revolutionmortgage.com/apply/name>. This application will give you information about required fields and other data validations.

Tools / Technologies: Microsoft Excel, ftp

**Scenario 2 – OLTP Data:** Online users can fill and submit the 1003 Mortgage application located at application at <https://apply.revolutionmortgage.com/apply/name>. When the users submit the application, it is stored in an OLTP database.

Developers should create the OLTP database, populate it with dummy data and get this data periodically by directly connecting to the database or by exporting this data into a flat file.

Tools / Technologies: SQL Server, MS Access, Microsoft Excel, SQL Server Export, ftp

**Scenario 3 – Zillow Data:** CAPTIDE MORGAGEreceives 1003 Mortgage applications from Zillow.com marketplace. Zillow sends data in XMl format to an ftp location.

Developers should create this input XML file with dummy data for practice purpose.

Tools / Technologies: XML, ftp, Textpad editor from <http://www.textpad.com/>

## R1:SP2: Story 2: Read input data

**Narrative:**

As a Developer

I want to write ETL components that will read various types of 1003 mortgage application input data from various sources

So that I can insert that data into staging tables

**Scenario 1 - Excel Data:** Read excel data from ftp location. Use script task to validate excel to make sure all the required columns are there.

Tools / Technologies: Excel source, Script Component, Microsoft Excel, ftp

**Scenario 2 – OLTP Data:** Read OLTP data from SQL Server or Access database.

Tools / Technologies: OLEDB, SQL Server, MS Access, Excel, SQL Server Export, ftp

**Scenario 3 – Zillow Data:** Read XML data from ftp location.

Tools / Technologies: XML Source, Script Component, XML, ftp

## R1:SP2: Story 3: Insert data into staging tables

**Narrative:**

As a Developer

I want to write ETL components that will insert 1003 mortgage application input data from various sources into staging database

So that I can validate that data for any errors

**Scenario 1 – Staging Database:** Design and create staging tables

Tools / Technologies: SQL Server

**Scenario 2 - Excel Data:** Write excel data into staging database. You may have to transform excel data into format that can be inserted into staging. For example, you may have to split some data or combine some data or change the format of some data so that the data can be inserted into staging database.

Tools / Technologies: Script Task, Excel Source, Conversion Components, OLEDB, Views

**Scenario 3 – OLTP Data:** Write OLTP data into staging database. You may have to transform OLTP data into format that can be inserted into staging. For example, you may have to split some data or combine some data or change the format of some data so that the data can be inserted into staging database.

Tools / Technologies: Script Task, Conversion Components, OLEDB Source, Views

**Scenario 4 – Zillow Data:** Write XML data into staging database. You may have to transform XML data into format that can be inserted into staging. For example, you may have to split some data or combine some data or change the format of some data so that the data can be inserted into staging database.

Tools / Technologies: Script Task, Xml Source, Conversion Components, OLEDB Source, Views

## R1:SP3: Story 4: Validate & Error Handling staging data

**Narrative:**

As a Developer

I want to validate the 1003 mortgage application data that is received to make sure that we have all required information in the format that is acceptable

So that I can load the data into OLAP database and data warehouse for reporting.

**Scenario 1 – Read Staging Data and Validate:** Read staging data and validate it for required fields, data types and other types of validations. If there are errors raise errors.

Tools / Technologies: Script Component

Note: Refer appendix for validation requirements.

**Scenario 2 – Do Error Handling:** If there are errors in any application data, please notify loan officer about the error so that they can fix errors.

Tools / Technologies: Script task, Send mail task / SQL command

## R1:SP3: Story 5: Insert data into ODS database

**Narrative:**

As a Developer

I want to write validated 1003 mortgage application input data into ODS database

So that I can create business intelligence reports

**Scenario 1 – ODS Database:** Design and create ODS database.

Tools / Technologies: SQL server

**Scenario 2 – Load ODS Data:** Insert validated data into ODS database.

Tools / Technologies: SP, OLEDB, SQL Server

## R1:SP4: Story 6: Insert data into OLAP / Data warehouse / Dimensional Model

**Narrative:**

As a Developer

I want to write validated 1003 mortgage application input data into data warehouse

So that I can create business intelligence and analytics reports

**Scenario 1 – Dimensional Model:** Design and create SQL Server OLAP / Data warehouse / Dimensional Model (Facts / Dimension).

Dimensions:

1. Dim\_Borrower
2. Dim\_Property
3. Dim\_Loan

Fact:

1. Fact\_Financials

Tools / Technologies: SQL server

**Scenario 2 – Load Dimensions:** Load the following dimensions:

* Fixed dimension and
* Slowly changing dimension.

Tools / Technologies: Stored Procedures, SSIS, OLEDB, SQL Server, Agents

R2:SP5: Story 7: Generate Reports – Loans to Date

**Narrative:**

As a Loan Department Executive

I would like to have business intelligence analytical reports on a week-to-date, quarter-to-date and year-to-date basis so that I can manage the loan processing business better. I would like to know Loans to date for total loans, purpose of the loan, loan amount breakdown and property usage.

**Scenario 1 – Loans processed to date:** This is the 1st part of the dashboard. Loans processed to Date will consist of 4 graphs

1. Loans to Date
2. Loans to Date by Purpose of Loan
3. Loans to Date by Loan Amount
4. Loans To Date by Property Usage

## R2:SP5: Story 8: Generate Reports – Loans to Date / Demographics

**Narrative:**

As a Loan Department Executive

I would like to have business intelligence analytical reports on a week-to-date, quarter-to-date and year-to-date basis so that I can manage the loan processing business better. I would like to know Loans-to-date for the following demographics: Marital / Status / Race / Sex

**Scenario 2 – Loans processed to date / Demographics:** This is the 2nd part of the dashboard. Loans processed to Date will consist of 4 graphs

1. Loans to Date by Marital Status
2. Loans to Date by Age
3. Loans to Date by Race
4. Loans To Date by Sex

## R2:SP6: Story 9: Generate Reports – Loan Overview

**Narrative:**

As a Loan Department Executive

I would like to have business intelligence analytical reports on the loan overview process so that I can manage the loan processing business better. I would like to know visually appealing dashboard with gauges and different charts.

**Scenario 2 – Loan Overview:** This is the 3rd part of the dashboard. Loans processed to Date will consist of 3 gauges and 3 graphs.

1. Gauges – All gauges will be displayed in %’s. The scale will be from 0 to 100%. The scale will have ranges: Red: 0 to 50% Yellow: 51 to 75% Green 76%+
   1. MTD vs Prior 6 Month Avg
   2. Last Mnt vs Prior 6 Mng Avg
   3. QTD vs Last Full Quarter
2. Avg Loan - % Monthly Income By Month
3. Loan Count By Month
4. Loan Count By Quarter

## R2:SP6: Story 10: Report Functionality

**Narrative:**

As a Loan Department Executive

I would like to have business intelligence analytical reports on the loan overview process so that I can manage the loan processing business better. Preferably, I’d like one interactive dashboard so Story 7/8/9 should be linked together with tabs. I would like the following filters on every report in the dashboard.

*Filters:*

1. **Report Date** – Every loan has a loan date. I want to see a list of all dates and I want the report to default to the most current loan date.
2. **Loan Amount** – Categorize the Loans into 3 groups
   1. Less Than $100k
   2. $100k to $200k
   3. More than $200k
3. **Loan Purpose**
4. **Property Usage**
5. **Demographics Group** – Age / Marital Status / Race / Sex
6. **Demographics Details** – This should display all demographic options. If I select a demographics Group, I would like this list to automatically filter by the attributes in on the groups I selected. For example, if I select Marital Status for the Demographics Group filter, this filter should only show the different Marital Status’.
   1. Age – Categorize the Age into the following groups
      1. <= 25
      2. 26-35
      3. 36-45
      4. 46+
   2. Marital Status
   3. Race
   4. Sex

When I select a tab, I want to be able to navigate to the associated report, and pass my same filters that were already selected. Active tables should be white with black font (bold). Inactive tabs should be LightGrey with white font.

When I click on any of the graphs for the *ToDate Report* and the *Demographics Report*, I would like it to drill through to a matrix report so I can see the details of the group I selected. All previously selected filters should also be filtered in the Matrix Report. For example, if I click on the Divorced Group in the Marital Status graph in the Demographics report, the matrix report should show all loans from Divorcees along with any other parameter that was included at the time I clicked on the Divorced Group. The matrix will resemble the following picture:

***Once completed, the Dashboard should be loaded to a local Report Server where they will be ready for report users and the creation of Automated Subscriptions.***