

**Ellie Mae Data Enrichment**

**Technical Design Document**



Contents

[**Introduction** 3](#_Toc6923)

[**Summary** 3](#_Toc6924)

[**Web Page Layout** 4](#_Toc6925)

[**Data Enrichment Azure Infrasctructure** 4](#_Toc6926)

[High level Data Flow: 6](#_Toc6930)

[**Database principals and design** 11](#_Toc6934)

[Inventory Details: 11](#_Toc6930)

[Data Management, Disaster recovery and compliance 14](#_Toc6931)

# Introduction

The EllieMae Data scrubbing and billing process was originally designed as an execution of 40+ SQL Server stored procedures with business logic to find billing transactions and calculate total charge per partner/product on monthly base. Over the period of time, business logic had has evolved into more complex form with multiple exclusions and scrubbing rules.

The purpose of Data Enrichment project is to:

1. Review and list all existing rules per partner/product.
2. Transform existing logic (code) into defined, parametrized execution blocks.
3. Establish UI for EM management to modify/create data enrichment rules.
4. Visualize Data enrichment process into monthly transaction summary page
5. Move Data Enrichment process into the Cloud (Azure).

# 

# Summary

The new EllieMae Data Enrichment process will consist of following execution steps:

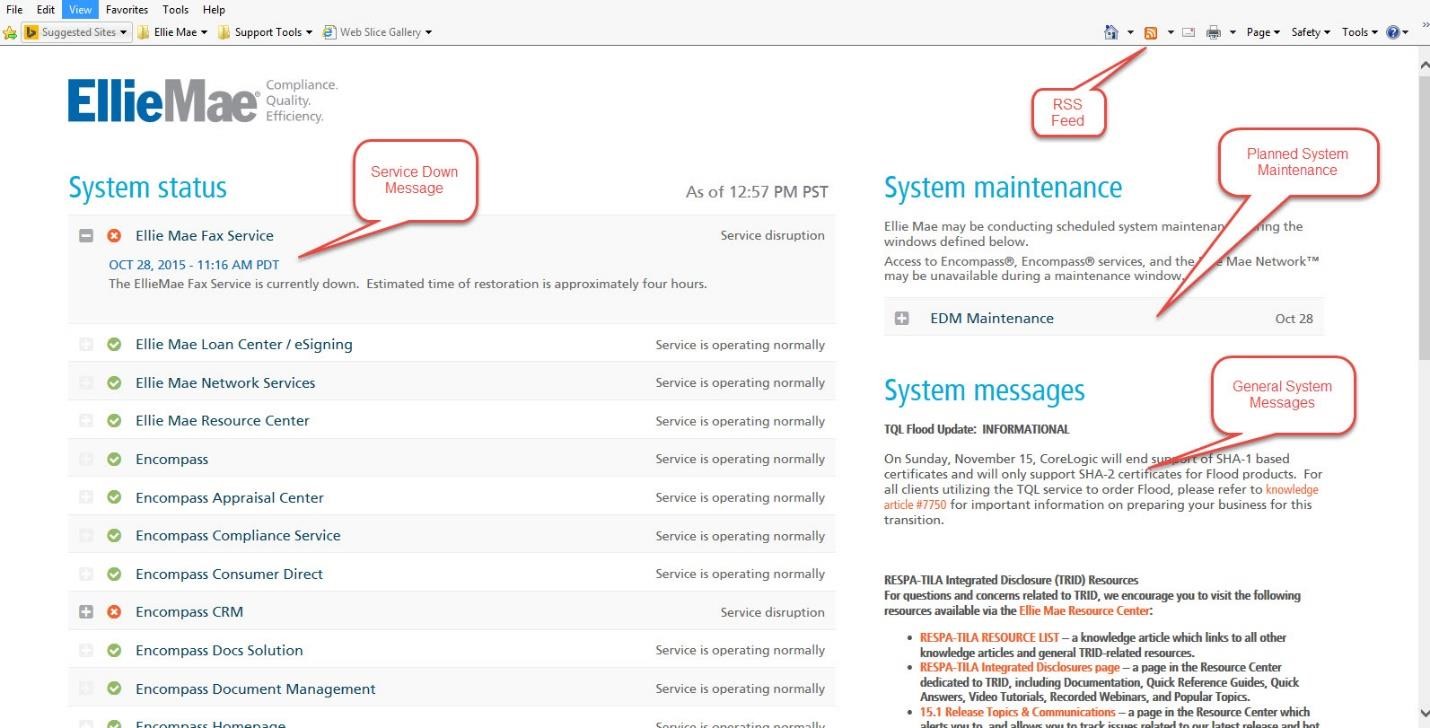
1. Move transactional records from on premises storage to Azure database.
2. Extract latest metadata from Salesforce into Azure database.
3. Execute Data Enrichment logic blocks (stored procedures) to refine data and determine billing records only.
4. Generate aggregated numbers representing enrichment stats.
5. Isolate billing records for Workday transfer.

# Web Page Layout

The Status Center is split into three main areas: Services, Maintenance, and System Messages. The web page is built to be responsive, and will adapt to different screen widths. Both the Service and

Maintenance sections will have messages that are collapsible, allowing more items on the page.

If a user has a RSS enabled web browser or a RSS reader installed, the user can subscribe to a RSS feed of the 25 services as well.



# Data Enrichment Azure Infrastructure

Currently, Data Enrichment project team have established following principals to build Azure infrastructure:

Security principals:

1. Building Services to Protect Data

1.1 Firewall - 2 levels security:

a. Cloud infrastructure firewall rules - Microsoft Azure Active Directory (SAML2.0, WS-Federation)

b. Azure SQL Database server level firewall rules ( To help protect your data, firewalls prevent all

Access to EllieMae database server until you specify which computers have permission.

The firewall grants access to databases based on the originating IP address of each request.)

1.2 Logically segment subnets - will create a single private IP address space-based network on

Which EllieMae can place all Azure Virtual Machines

1.3 Deploy DMZs for security zoning - will help enable DDoS prevention, Intrusion Detection/Intrusion

Prevention systems (IDS/IPS),

Firewall rules and policies, web filtering, network antimalware

2. Protecting Data in Service Operations:

2.1 Security groups

2.2 Access Control List - data access controls (replaced with AD B2C)

2.3 SSL data transfer

3. Protecting Data at rest:

3.1 availability, redundant and disaster recovery: Enable data replication in multiple regions (Regional Pair - US West, US West 2)

3.2 data encryption (Azure Key Vault)

Audit principals

1 Audit activities: enable audit logs - control operations

EllieMae resources (includes logs such as creation of VMs, starting websites, dropping database,

Success and failure of deployments and etc.)

2. SQL database auditing - tracks database events (define categories of database actions to be audited)

And writes them to an audit log in EllieMae Azure storage account.

Performance and Scalability principals:

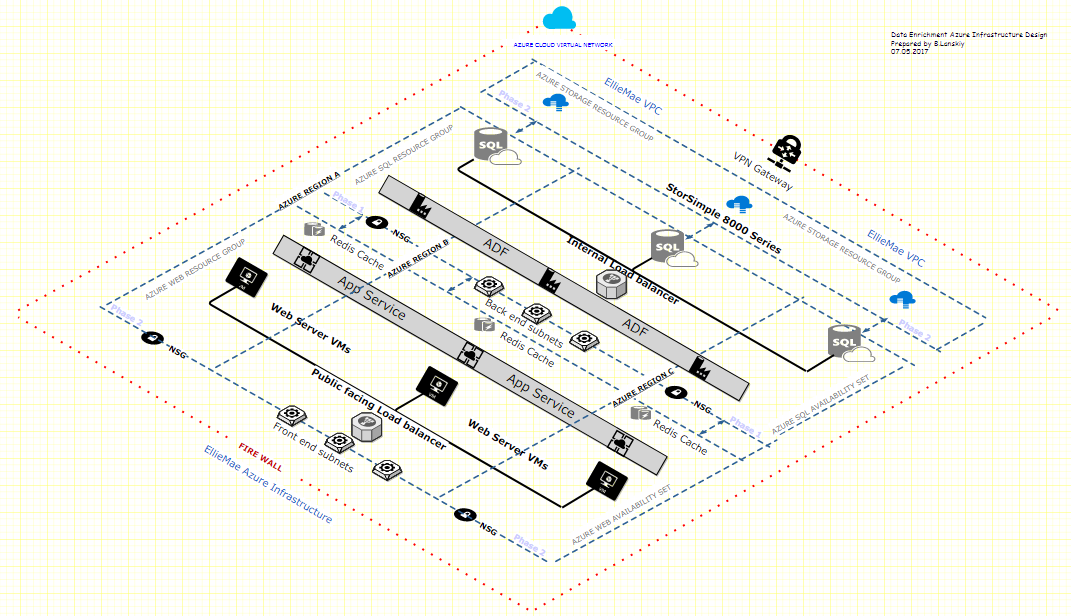
1. Use built-in auto scaling features - scale out on a schedule. For example, scale out during business hours

2. Avoid instance stickiness - Stickiness, or session affinity, is when requests from the same client are always routed to the same server.

Stickiness limits the application's ability to scale out

3. Parallel and asynchronous processing loading/refining data

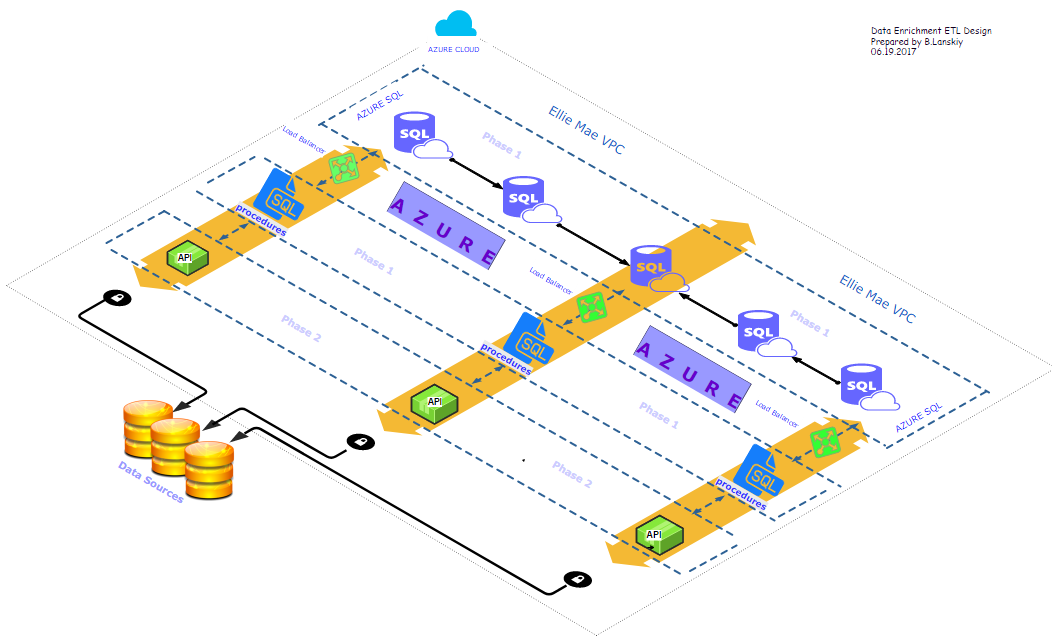
4. Sharding - Divide a data store into a set of horizontal partitions or shards



# High Level Data Flow

There are multiple Data Flow sources for Data Enrichment load:

1. EllieMae On Premises Data
2. Salesforce Data



Data Enrichment Access level

The following groups has been identified for Data Enrichment Access Level

1. Data Enrichment Team
2. Product Management Team
3. Finance Team
4. IT Team

# 

# Database principals and design:

# Enrichment Tables Inventory Details

The inventory groups\ tables:

1. Global for scrubbing window: table TBL\_Product\_types
2. Exclusion:

DBO.lu\_Test\_By\_ClientID -- TBL\_lu\_test\_integrated\_Exclusions,

DBO.lu\_test\_by\_borname --- TBL\_lu\_test\_integrated\_Exclusions

DBO.lu\_test\_by\_company --- TBL\_lu\_test\_integrated\_Exclusions

DBO.lu\_Test\_By\_Access\_Code -- TBL\_lu\_test\_integrated\_Exclusions

DBO.lu\_mapfiles ---vendorid

DBO.TBL\_Scrub\_Exclusion\_Records

1. Parsing:

dbo.lu\_Vendor\_Scrub\_Control

dbo.TBL\_Parse\_Misc\_LOOK\_UP

1. Enriched: DBO.Temp\_AllVendorData
2. Reporting: TBL\_Transaction\_Usage\_Data
3. Supporting UI:

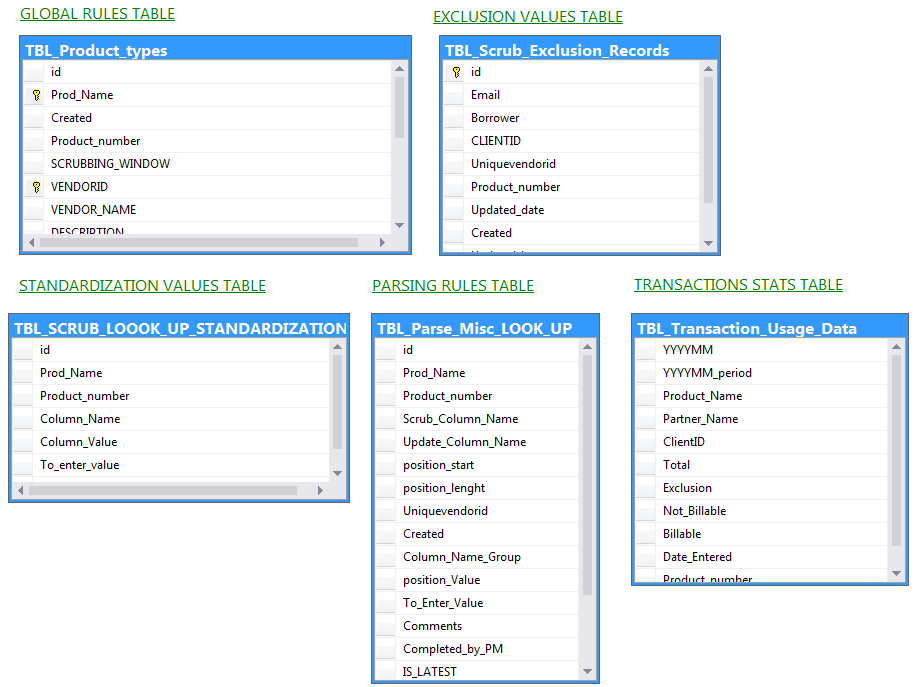
dbo.TBL\_Product\_types

dbo.TBL\_Parse\_Misc\_LOOK\_UP

dbo.TBL\_SF\_Partner\_Client\_Product\_Details

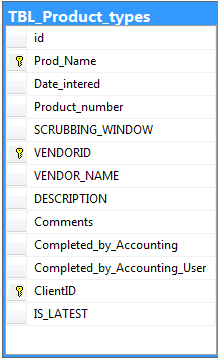
1. Salesforce Integration: TBL\_SF\_Partner\_Client\_Product\_Details
2. Auditing:

# Table Groups

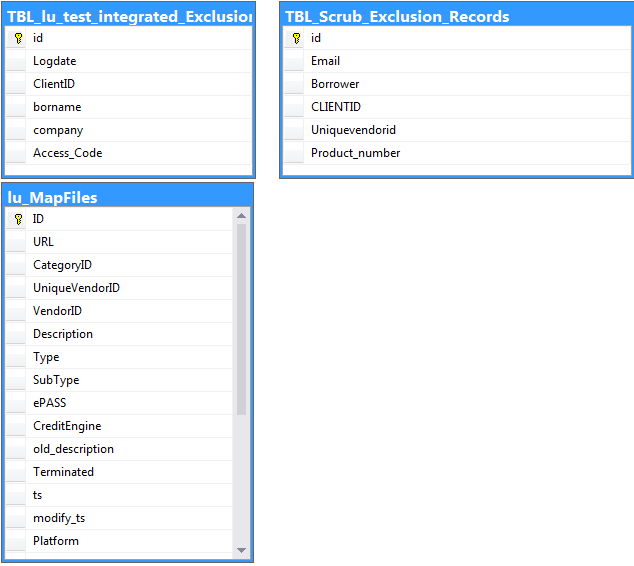


# Table Details

1. Global for scrubbing window: table TBL\_Product\_types



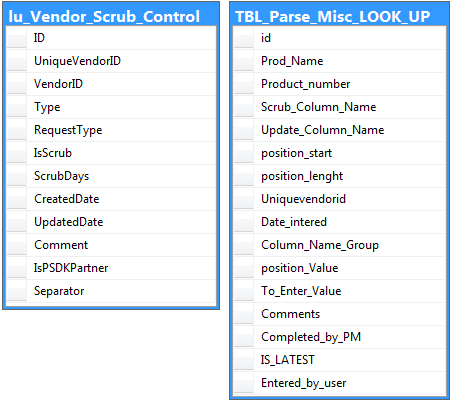
1. Exclusion:



1. Parsing:

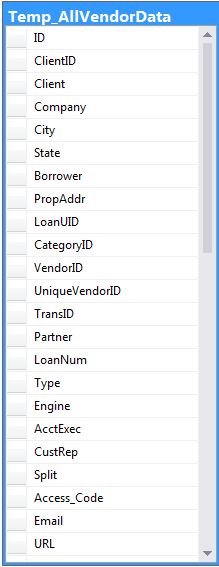
dbo.lu\_Vendor\_Scrub\_Control

dbo.TBL\_Parse\_Misc\_LOOK\_UP

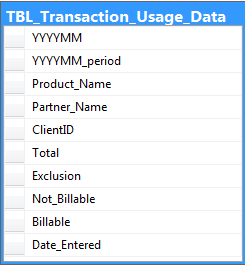


1. Enriched:

DBO.Temp\_AllVendorData



1. Reporting: TBL\_Transaction\_Usage\_Data



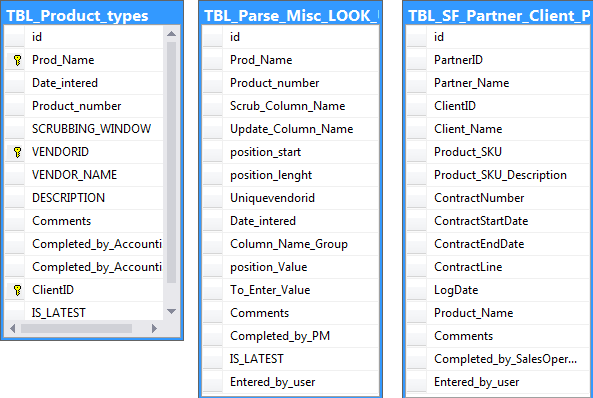
1. Supporting UI:

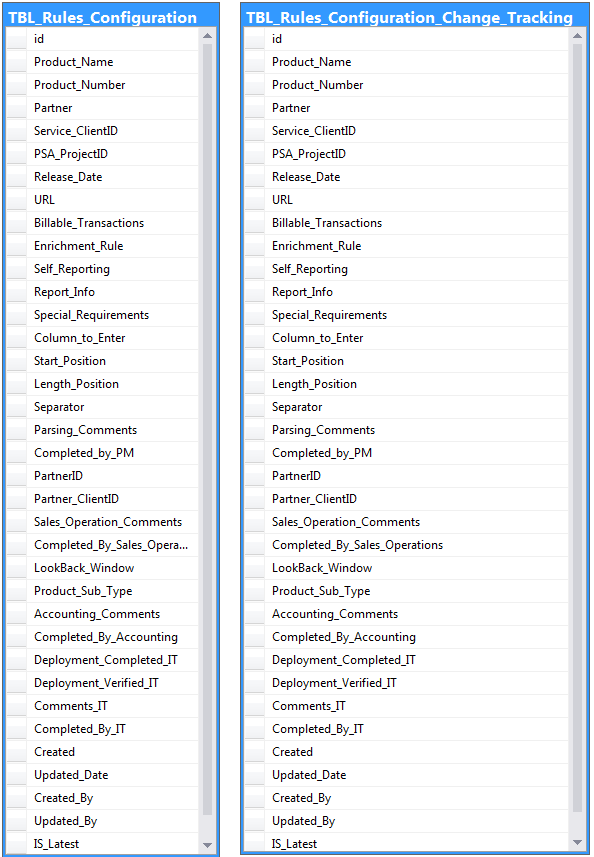
[dbo].[TBL\_Product\_types],

dbo.TBL\_Parse\_Misc\_LOOK\_UP

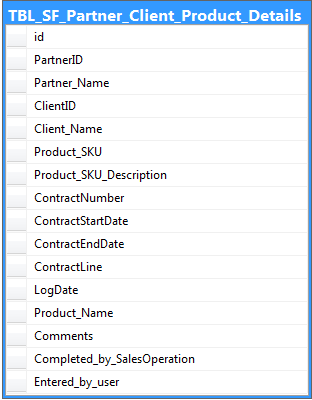
dbo.TBL\_SF\_Partner\_Client\_Product\_Details

dbo.TBL\_Rules\_Configuration





1. Salesforce Integration: TBL\_SF\_Partner\_Client\_Product\_Details



1. Auditing: Will use Azure SQL database auditing features to tracks database events and writes them to an audit log in EllieMae Azure storage account.

# Data Management, Disaster Recovery and Compliance

1. To manage data growth more effectively, Data Enrichment team will use StorSimple:
   1. Automatically archive inactive primary data.
   2. Remove the need for separate back up infrastructures.
   3. Eliminate the need for remote replication.
2. To accelerate disaster recovery, improve compliance:
   1. StorSimple gives immediate data availability during disaster recovery.
   2. Ensures compliance with EllieMae corporate policies for data retention and disaster recovery.

