**CIDM-6350 Project**

Miguel Sanchez

1059824

CIDM-6350

**Section 1** - Introduction describing your database, including current data management problems in your chosen area, motivation for your DB, potential benefits and potential users of your DB, etc.

**Chosen Area:** Finance / Financial Securities

**Holdings Unified Bank “HBU” Database Creation**

**Current Data Management Issues:** The current data issues that exist within our database architecture is the replication of data points throughout the current landscape of our financial organization. Our organization has grown organically and therefore as the business line identified that certain data points / data sets were needed, new databases were created. The problem here is the lack of data governance that took place. Instead of validating if current data sets exist or not the business took it upon themselves to request IT to create new data locations and data integrations without asking the why behind it all. In addition, IT failed to identify if these data sets are replicated in other tables. The goal here is to have a structured database where data is not replicated but also governed in a way that all data sets have a data dictionary explaining the usage of each individual column, field, and data store. Our organization did not do this and what ended up happening is that the original curators of the data have either left the company or it has been so long since the data was originally created that there is no way to identify the true reasoning/source of the data. As a financial institution this becomes a high-risk problem because when dealing with a publicly traded company all financials must be vetted and can be audited and therefore data inconsistencies would have a negative impact on the organizations bottom-line and stock performance. Data security is also a constant issue in an ever-changing technological world. Our data storage is currently on premises based and data leaks / breaches are a constant fear, especially when dealing with personal identifiable information. Any of the issues stated above would impact customer retention and shareholder satisfaction. The aforementioned issues also pose risks to security, data integrity/quality/redundancy which can cause wasted internal capital, fines from the federal government due to mismanagement of funds (audit) and data breaches. Translating the data has also proven to be an issue in the past due to the lack of data governance, redundancy of data and poor overall quality.

**Motivation:** For the financial sector, monetary assets (currency and real assets) is the “skeleton” of a financial organization. Data is essentially the “lifeblood” that keeps everything moving forward and churning. Data not only drives decisions but also provides potential trends, analysis, and potential avenues for new business to grow by reviewing and translating data. As W. Edwards Deming’s quote states “Without big data, you are blind and deaf and in the middle of the freeway”. This is so true because you could have unlimited sources of income/capital but without data to be able to make impactful decisions you are ‘blind and deaf in the middle of a freeway’. In the highly competitive financial sector, competitors will be flying past you on the ‘freeway’ while we are stuck in neutral and being surpassed by our rivals. With data redundancy it was identified that there are approximately three separate data stores that hold deposit data information for potentially the same group of clients. This replication of data proved to be costly for us because our on-premises costs for data storage increased exponentially and by simply validating non distinct values within the database over 60% of the data stored can be removed and freed-up for other potential databases/entities. In addition, a living and breathing document such as a “Data dictionary” is critical for the creation of any new entities. This data dictionary could potentially be housed within a Confluence page that can be updated periodically as new fields or databases come about. This new database will have source input files from the business, source field names, source descriptions (these must be extensive and well detailed) and the datatypes that will be expected (along with the format of each). By having a well-designed foundation for the data store the quality of said data increases, exponentially.

**Potential Benefits:**  As previously stated in the current data management issues section replication of data is a major issue within our organization. Duplicate data can and will throw off any queries that are produced and without knowing which data is in fact “correct” the data you may review is completely wrong. Each financial organization uses various applications that house their data whether it be on premises or cloud. Examples of data that reside on either cloud or on-prem can potentially include human resources data (i.e. ADP), Trade blotter sales, Credit/Risk ratings (i.e. Moody’s analytics), and procurement contracts, just to name a few. Various business lines would benefit greatly from this solution to be able to identify potential business opportunities (via trend analysis), predict employee churn/turnover (review employee compensation benefits to ensure top talent is kept) and identify potential commercial clients risk ratings for new loan approvals (Moody’s Risk Rating analytics). When using this newly created database the users will have peace of mind that the data is not replicated, the data is well defined and any reports that extract these specific data fields can be manipulated via a semantic layer that will present the information in a readable format for any audience to ingest, translate and share. Reporting will not only be easier and more efficient but also more accurate.

**Potential Users:** The users of this new database would be C-level executives, Project Management Office (PMO’s), Treasury Management, Financial Analysis, Resource Allocation Managers, Human Resources and Accounting.

**Section 2** – Business Rules

Holdings Unified Bank (HUB) requires an extensive organization that must maintain all departments and financial streams within its data architecture. There are a total of nine department/business lines where data must be captured; Human Resources (HR/ADP), Accounting, Project Management Office (PMO), Treasury, Trade Blotter Sales, Securities Analysis Scenario, Credit Risk, CNI Summary and CRE Summary. The goal of this database is to track all the possible revenue streams and identify securities and commercial loan trend analysis, health of internal projects/initiatives and correlating that information into our accounting general ledger to report accurate financial health and metrics to both auditors and shareholders. For ease of understanding, Primary Keys will be shorted with “PK” and foreign keys as “FK” for readability purposes.

The entity HumanResources (HR/ADP) has the attributes Employee\_ID (PK), Name, address, city/state/zip, birthdate, contactnumber, salary, tenure, role\_title and costcenter. One Employee can manage many employees and all employees will have one supervisor which is tracked by supervisor\_ID.

Accounting keeps track of every workstream/business line within the organization and when completing a general ledger each accounting line item is given a GL\_REF#. Each GLREF# is unique to the individual business line to track profitability and overall cost. Accounting has the attributes GLREF# (PK), Costcenter (FK), DEAL\_ID (FK), CUSIP# (FK), ProjectID (FK), TreasuryRefID (FK), DailyAvgCost, FMNRunRef#, NPV, CurrentCost, BudgetVariance, Gross. Treasury deals will be constantly flowing into accounting and various GL\_REF#’s will be associated to those deals but every treasury deal will only have one GLREF#. Project Management Office (PMO) can potentially have multiple GLREF# associated to the department but each project will only have one GLREF# to track project health and budget variances.

Project Management Office (PMO) has the attributes Project\_id (PK), EmployeeID (FK), ResourceAllocation, ProjectName, ProjectCost, Contract Cost, Budget Amount. The PMO office tracks all potential vendor Statement of Works, vendor contracts and initiates business line projects for enhancements and the introduction of new technologies or processes into HUB.

Treasury consists of the attributes; TreasuryRefID (primary key), CUSIP# (FK), RiskRatingRun\_ID (FK), DailyAvgCost, FMV\_RunRef#, NPV and PortfolioBookValue. TreasuryRefID tracks each individual “deal” whether it be a security, CNI loan or CRE loan (with a credit risk associated to it(. The TreasuryRefID is the way HBU tracks which line of business is creating the specific GLREF# as previously stated within the accounting entity.

Trade Blotter Sales is a trade record which provides details of all possible trade transactions performed over a period. The attributes for this entity are CUSIP# (PK), Analysis\_Run\_Ref# (FK), Gain\_Loss\_On\_Sale, CreditStress Category, Asset\_Class. There can be one to many CUSIP#’s associated to a Treasury RefID because it will be tracked by the most recent trade date since trades happen multiple times throughout the day, week and even month the overall price of the blotter sale will fluctuate quite a bit. This is why there is a YieldRateDesignation which which will track overtime the Asset\_Class and the Market\_Yield that was associated to each individual CUSIP# throughout the blotter sale record time period.

Securities Analysis Scenario has the attributes; Analysis\_Run\_Ref# (PK), CUSIP (FK), AnalysiS\_Run\_Date, MarketYield, SECReportingGroup. Not all CUSIP’s will have a specific run date because it may have not been processed yet and therefore there could be CUSIP’s missing a run date until all proper MarketYields and SECReportingGroup data is provided. All completed SecuritiesAnalysisScenario runs will always only have one individual run date by CUSIP because at the time of the scenario run there are various factors that could potentially change its status.

The Credit Risk entity consists of the attributes; RiskRatingRun\_ID (PK), Deal\_ID (FK), RiskRating\_Value, RiskRating\_Desc, RiskCal\_Assessment, Facility\_Type and Portfolio\_Type. The Deal\_ID serves as either a CNI\_DEAL\_ID or CRE\_DEAL\_ID, it cannot have both CNI and CRE deal id’s since they are different financial instruments. Commercial Real Estate (CRE) loans comprise of the acquisition, development, and construction lending for income-producing real estate. Commercial and Industrial (CNI) loans provide working capital or finance capital expenditures (like hardware/machines) to companies. Essentially CRE loans are commercial mortgages while CNI are loans provided to a business corporation. It is important to differentiate the two in separate entities to identify the more revenue driving asset and there are different factors to the risk rating valuations for each.

The CNI Summary entity consists of CNI\_Deal\_ID (PK), RiskRatingRun\_ID (FK), Loan SummaryDesc, Rating\_Date, LoadDate, Bankrupt and Stress Indication.

The CRE Summary entity consists of CRE\_Deal\_ID (PK), RiskRatingRun\_ID (FK), Loan SummaryDesc, Rating\_Date, LoadDate, Bankrupt and Stress Indication.

**Section 3** – Entity Relationship Diagram



**Section 4** – 3rd Normal Form Relations Diagram



**Section 5 – SQL Code Script of Table Creations**

CRE\_Summary TableA screenshot of a computer

Description automatically generated

Treasury Table

A screenshot of a computer

Description automatically generated

CNI\_Summary Table

A screenshot of a computer

Description automatically generated

Credit\_Risk Table

A screen shot of a computer

Description automatically generated

Accounting TableA screenshot of a computer

Description automatically generated

Securities\_Analysis\_Scenario Table

A screenshot of a computer program

Description automatically generated

Project\_Manangement\_Office\_PMO Table

A screenshot of a computer

Description automatically generated

Trade\_Blotter\_Sales Table

A screenshot of a computer

Description automatically generated

Human\_Resources TableA screenshot of a computer

Description automatically generated

Section 7 – Custom SQL Queries

1 - This query joins the CNI\_Summary, CRE\_Summary and Accounting tables to identify the ProjectID’s, GL\_Line and cost of each CNI and CRE deal. Not every CRE and CNI deal have a ProjectID because there may be other revenue streams we want to confirm that are not associated to a project.

A screenshot of a computer

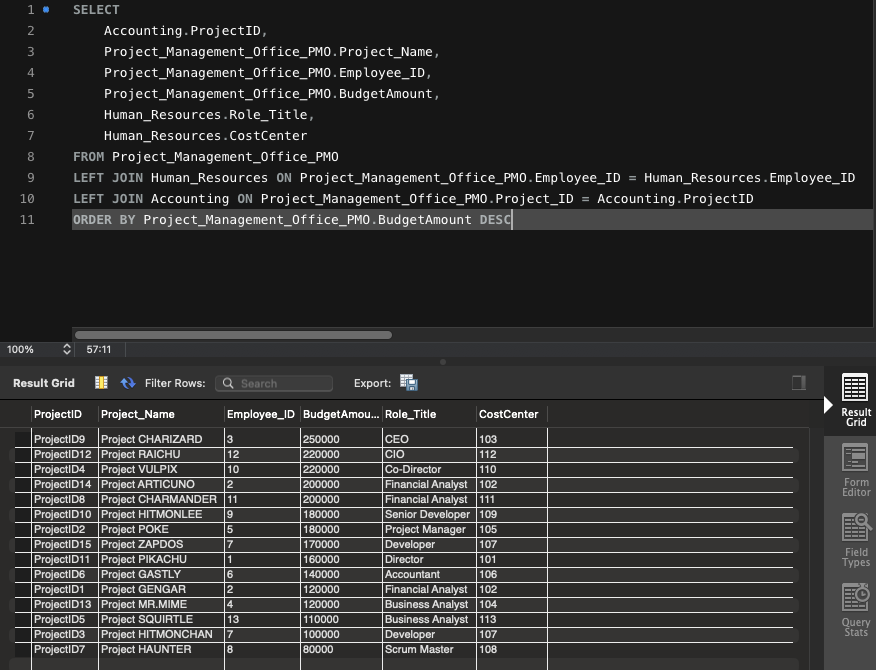
Description automatically generated

2 - This query joins gives the total count of positive versus negative RiskCal\_Assessments for the last week (table data is reviewed weekly). For the last week there were a total of 8 Positive assessments on new loans and 7 total negative loans (these were most likely denied).

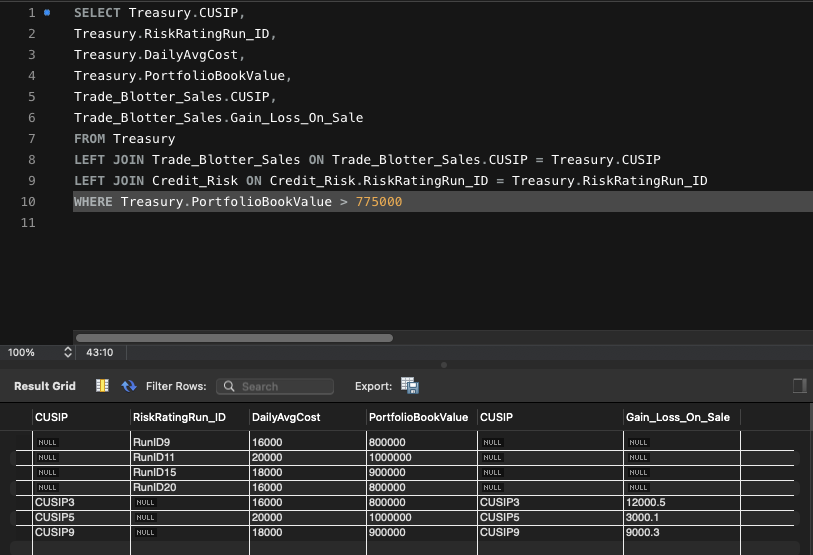
A screenshot of a computer

Description automatically generated

3 - This query joins the Accounting, Project Management and Human Resources Table to identify the banks project details along with specifics of each project such as the Employee ID, Budget Amounts, Role of the assigned Employee ID and the CostCenter associated to each Project.



4 - This query joins the Treasury, Trade Blotter Sales and Credit Risk tables. Our goal with this query was to identify accounts where the accumulated PortfolioBookValue of specific accounts was greater than $775,000. This query actually allows the bank to identify between CUSIP accounts and RiskRatingRun\_ID’s (CRE vs CNI loans), which population is the most profitable. Per this query there is one additional Risk RatingRun\_ID account which could potentially mean we are having more CRE/CNI business.



5 - This query joins provides insight on employees who are assigned to specific projects within the bank. The query also sheds light on the fact that Emily and Miguel both have multiple projects assigned to them as well as their Role, Salary and Tenure. This allows the bank to identify which projects are impacting which cost centers and we can reassign projects to other cost-centers depending on budget availability.

A screenshot of a computer

Description automatically generated