Digital Transformation of
"Three Lines of Defense Model"

(Business Case: User's IT Network Access Controls)

Functional Specification Document

Sateesh Babu

Dec 30,2021

Version 1.0

1 Document Purpose

1.1 Project Information

Project ID	xyz2021
Project Manager	Sateesh
Project Sponsor	xyz IT team
Business Lead	Sateesh

1.2 Revision History

Revision	Date	Status	Contribution By	Summary of Changes
1	2021/12/30	Draft	Sateesh	
2	2021/12/30	Final	Sateesh	

1.3 Document Approval List

Revision	Date	Status	Reviewed By	Summary of Changes
1	2021/12/30	Approved	Sateesh	

1.4 Document Distribution List

Revision	Date	Status	Reviewed By	Summary of Changes
1	2021/12/30	Final	Sateesh	

Contents

2   Introduction

2.1   Executive Summary

Digital transformation is the integration of digital technology into all areas of a business, fundamentally changing how you operate and deliver value to customers. It's also a cultural change that requires organizations to continually challenge the status quo, experiment, and get comfortable with failure.

In the post COVID era, as most of workforce started connecting to the IT network remotely and IT Infrastructure managed by cloud service vendors, our enterprise has implemented "Three lines of Defense Model" control measures in reviewing user network access on timely manner.

Digitalizing these control measures will bring the following benefits:

1. It will improve efficiency, transparency, and coordination among the stakeholders of "Three lines of Defense Model"
2. It will break the data silos and saves cost as well as processing time.
3. It will help in quicker problem identification (including unauthorized access) and enables a secured environment.
4. Finally, it provides better experience and motivation to the workforce while connecting to the IT network remotely.

2.2   Three lines of Defense Model and Project stakeholders

The three lines of defense framework is a fundamental pillar of corporate governance structures and enterprise risk management. Its objective is to provide a right information at right time to the governing body and the senior management for their risk-based decisions. Also, provides assurance to regulators and external auditors.
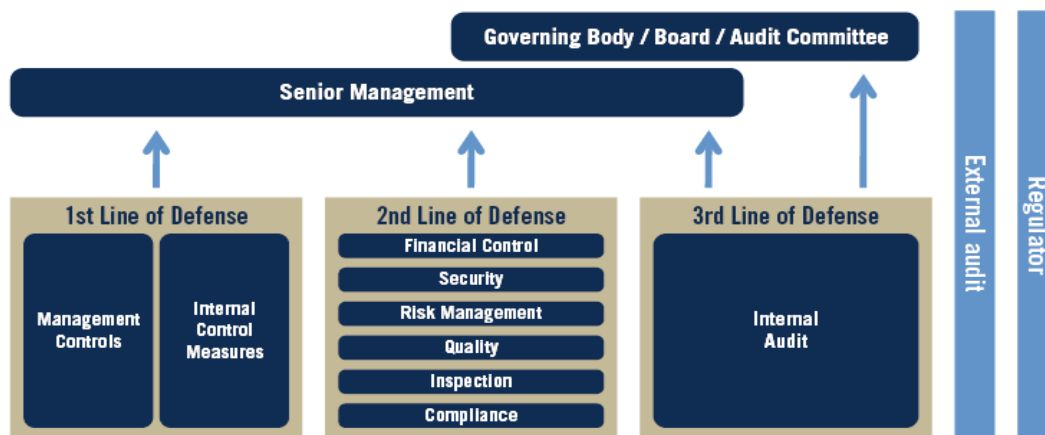
Adapted from ECIIA/FERMA *Guidance on the 8th EU Company Law Directive, article 41*

2.2.1   Primary Stakeholders

Business stakeholders in the "Three lines of defense model" play a distinct role within the three lines of defense. But they use the same organizational data for different perspectives/or reporting. Inaccurate reporting will lead to wrong management decisions and conflicting assurance opinions.

Line of Defense	Responsibility/Description in	Business Stakeholders/Users	Primary Contact Email id	RACI Chart
1st Line: IT Operational Management	Responsibility to own and manage risks associated with day-to-day operational activities of the "IT network access and security". Other accountabilities assumed by the first line include design, operation, and implementation of controls.	IT Security and networking team	abc@xyz.com	Accountable, Consulted, Informed
2nd Line: Risk Control and Compliance	Specialize in the oversight of risk management and compliance. It does this by providing compliance and oversight in the form of frameworks, policies, tools, and techniques to support risk and compliance management.	Risk & Compliance committee; Cybersecurity team	qwer@xyz.com	Accountable, Informed
3rd Line: Risk Assurance	Ensure whether the first- and second-line functions are operating effectively. It is charged with the duty of reporting to the board and audit committee, in addition to providing assurance to regulators and external auditors that the control culture across the organization is effective in its design and operation.	Assurance team/auditors	yoyoyyoy@xyz.com	Accountable, Informed

Note: IT Security and networking team are the project sponsor and act as a point of contact for all network systems processes.

2.2.2 Secondary Stakeholders

Data Owners/Stewards/Custodians, Domain Experts, Project team and Enterprise Data Governance are the secondary stakeholders, who play an important role in the delivering the project.

Description/Context	Secondary Stakeholders	Primary Contact Email id	RACI
HR datasets	HR reporting team	hrrep@xyz.com	Responsible, Consulted
Active Directory (network) datasets	Vendor (Cloud Service Provider)	vendor@vendy.com	Responsible, Consulted
Work order/ RFI approvals	Procurement team	proc@xyz.com	Informed
Solution & Data architecture review	Enterprise Architect team	ea@xyz.com	Consulted
Project delivery team	Third party IT vendor	vend@ibxy.com	Responsible
Enterprise Data Governance	Data Governance Committee	edg@xyz.com	Informed

2.3   Business Needs

Summarized the business needs as below:

- Need access to accurate and organized data analysis of "Users' IT network access". This gives the ability to spot risk trends/alerts and make effective decisions in monitoring organization's IT resources.
- Centralized integrated data platform (digital asset) should have capabilities to collect, store, process, analyze, and visualize high volumes of a wide variety of data, drive value in many ways by considering standard enterprise data taxonomies, data integrity checks and data security measures.
- The integrated data and analytical platform should be flexible, scalable, cost economical and the freedom to design and deploy standard as well as custom controls that fit business context at any given time.

3   Project and Scope

3.1   Project Objective

Project objective is to digitalize the "User's IT Network Access Controls" by leveraging the enterprise data lake and analytics. Thereby, it would help key stakeholders of "Three Lines of Defense Model" to set risk control mechanism with strong corporate governance for better decision making and assurance reporting in monitoring "User Network Access" of xyz enterprise.

3.2   Project Scope

- All users in vendor's active directory who are tagged to xyz departments and xyz groups. Both active and inactive users.
- Digitalize the "User's IT Network Access Controls" by leveraging the enterprise data lake and analytics
- Train stakeholders of "Three Lines of Defense Model" on the analytics outcomes and exceptions handling.
- Design operational visual dashboard in Power BI.
- Reviewing IT network access of rehired FTEs (employee) is part of the scope

3.3   Project Out-Of-Scope

- Entities/users/user groups within the active directory of cloud service provider (vendor), which don't belong to XYZ, are not in scope.
- Does not cover security at the database, operating system, network, client stations level, etc.
- Active directory's segregation of duties is not in scope. Unable to identify conflict responsibilities.
- The source datasets don't cover processing of Manager's requests to the vendor.
- Current IT infrastructure as well as business processes do not support real-time analytics. To achieve that all IT system should be well integrated on real-time basis and should provide their real-time/near-time feeds to enterprise data lake.
- AD's daily EOD snapshot doesn't have historical versions and their metadata changes, like who approved, who created etc. Hence, the analytical reports don't include any of these metadata attributes.

3.4   Project Dependencies

- Availability of key stakeholders to participate in necessary business processes and technical reviews.

- Establishing the data ingestion connection or SFTP connection to vendor system after adhering to their security checks as well as internal security checks.

3.5 Project Assumptions

- Ensure all necessary and appropriate stakeholders are available to participate in necessary business processes and technical reviews.
- Assume that the work email in active directory could be possible match with HR Email id of each FTEs
- Unmatched emails or users with no emails in active directory are treated as contractor or mismatch
- Within the unmatched email list, if the user's email id has suffix as xyz.com than they will be considered as "FTEs mismatch"
- All datasets are available on every day between 10pm to 12pm after completion of the dependent jobs in the respective source systems.
- Vendor and xyz's security teams will implement a clean revoke in scenarios like - Termination, Transfers or any other events that are linked to role change.
- Assume that data owners/stewards/custodians will provide complete and validated daily datasets in tandem to the business processes.
- Project virtual servers or environments are configured as expected.
- The project scope will not change once the stakeholders sign off on the BRD.
- The system of the project is compatible, functions properly and stable for the project to take place smoothly.
- Project will be completed in expected timelines and within the budget allocated.

3.6 Project Constraints

- As AD don't have metadata information on the version history in its dataset, it's not feasible to provide version changes of each user account.
- No primary key for AD user/ group assignments, there will be some challenges in maintaining the delta extractions.
- Vendor's Active Directory does not delineate between xyz FTEs and contractors.
- HR datasets don't have information on the contractors. HR system is only for the FTEs
- Data quality issues (like email mismatches, user id mismatches, no standard master data etc) in the vendor AD, might be a constraint in designing and maintaining the dashboard visuals.

3.7 Project Risks

- No control over staff priorities in unexpected events like COVID-19 pandemic.
- Vendor's statement of work (SOW) might be delayed in getting approved by procurement team due to post COVID-19 changes in the business processes.

4 Requirements

4.1 Business Requirements

The following are the business requirements of the business stakeholders to digitalize three lines of defense checks:
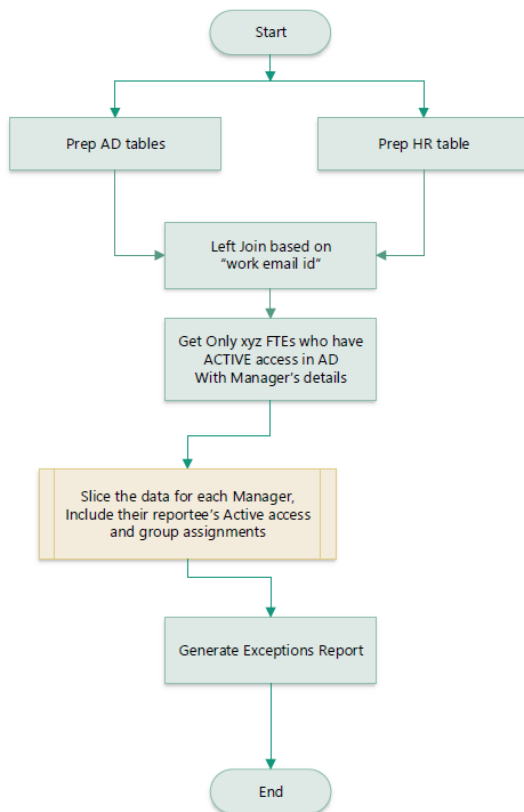
#BR	Business Requirement	Description	Business Stakeholder/Owner
1	<u>Daily</u> Operational Review of network access of XYZ's users.	To provide operational visual dashboard so that IT security team can review the user's network access and can make decisions in controlling user's network access in their day-to-day activities. And also, should able to manage the master date, especially for attributes – department and groups	IT Security operations team (1st line of defense)
2	<u>Monthly</u> validation of active Network access of XYZ's users by their managers	To generate excel files at end of each month for each manager in XYZ so that they can review and validate their reportee's network access.	Risk & Compliance team (2nd line of defense)
4	<u>Adhoc</u> review of active Network access of XYZ's terminated users.	To generate excel file with list of terminated users having network access to XYZ entities/groups, for assurance/audit team on adhoc basis.	Assurance team (3rd line of defense)

4.1.1   User Stories

#US	User Story	#BR mapping
#US1	As IT security team lead, I need a POWER BI visual dashboard on details of active directory (AD)'s users who have access to XYZ groups along with their HR details (like - his/her work email, position, department, manager's email etc) for daily operational review. Would be helpful if this dashboard gets updated daily at end of the day (EOD).	#BR 1
#US2	As IT security team lead, I want an adhoc excel report with all master data changes (new, amend and deleted) in active directory (AD), especially for the attributes – department names, groups, user ids. This report will be helpful in defining the daily operational review controls.	#BR 1
#US3	As risk and compliance team lead, I need a excel report for each manager with the details of their reportees and their active network access assignments in active directory. Would be helpful if these excel reports are available for managers' validation at end of each month.	#BR 2
#US4	As Assurance and audit team lead, I need a excel report with list of terminated/departed users having network access to XYZ entities/groups, for validation checks on adhoc basis.	#BR 3

4.2 Flow charts

BR#2

BR#3A

BR#3B

Rules:
- User is not part of xyz HR list
- User should be assigned to one of xyz AD Department or xyz AD group
- Terminated Contractor who has active AD Account
- Contractor who has account expiry date < today
- Contractor who hasn't logged into AD Account > 90 days
- New Contractor who hasn't logged into AD Account > 90 days since creation
- Contractor who hasn't reset their AD Account pswd > 90 days
- Contractor who's AD account has last update > 90 days

4.3   Data Requirements

The following are the data requirements of this project:

4.3.1   Source systems

#DR	Data Requirement	Description
1	Need daily ingestion from the sources – active directory and HR.	To create a daily ingestion pipeline in the enterprise data lake for each of the source system – Active directory and HR.

4.3.2   Data Profiling Analysis

The source datasets are analyzed based on dimensions –Number of datasets, Completeness, Uniqueness, Timeliness (point of time), Validity, Accuracy, Consistency, Relationship etc

Dimension	Active Directory Dataset	HR Dataset	Comment

# of Datasets	1 Excel file: 1. IT_AD_XYZ_Users_yyyymmdd.xlsx	2 Excel files: 1. HR_Status_Active_yyyymmdd.xlsx 2. HR_Status_Departure_yyyymmdd.xlsx	
Point of time	It's an EOD dataset which is generated and shared between 10pm-12am	It's an EOD dataset which is generated and shared between 10pm-12am	Both datasets are generated on the same point of time at EOD after all important data changes.
Completeness	User and their group assignments are provided as a full data snapshot on every day. But, metadata attributes – who created, who approved, version history – are missing. Master data attributes (Department, work id) have some missing values.	FTEs and their key details are provided as a full data snapshot on every day. But, metadata attributes – who created, who approved, version history – are missing.	Extracts' record counts are matching with the source systems
Uniqueness	There is no primary key defined. But, "SamAccountName" is a unique attribute which can used to define the incremental changes (records addition and deletions) in the dataset.	There is no primary key defined. But, "Employee Number" is a unique attribute which can used to define the incremental changes (records addition and deletions) in the dataset.	
Validity and Accuracy	Some users work email ids are not matching with their corresponding HR work email ids.	Rehired employees are not defined or flagged in the HR dataset.	Both datasets are validated by the data owners/stewards.

Consistency	Did a consistency check on the last 5 days snapshots. There is no change in the data structures. User's Group data value should be parsed to a table format.	Did a consistency check on the last 5 days snapshots. There is no change in the data structures of both HR datasets. Noticed duplicate records when merging the two excels (active and departed). Also, the data structures of the both excel files are not matching. We need to merge, normalize the data structures, de-duplicate the records and identify rehired employees.	This duplication issue won't exist if established direct data ingestion connection between source systems and enterprise data lake.
Relationship	EmailAddress is used as join key	Work_email is used as join key	HR data is used as reference data to enrich the AD user info.

4.3.3 Data transformations / enrichments

| #DR	Data Requirement	Description
2	Normalize the data structures	Merge the datasets of each, if received multiple files. Choose the latest updated record in case of duplication records.
3	Enrich AD data with corresponding HR details of xyz employee.	Use work email to join the datasets. If matched, then mark as record as "HR_Matched", else it should mark as "MisMatch".
4	Rehire flag of the XYZ employees	If XYZ employee's Start Date is greater than their Termination date, then flag them as "rehired employee".
5	Parse AD user assignments into a separate table	Parse the values in the "Group" column into each row for each user.
6	Add prefixes to the attributes	Add the following prefix to the source and derived attributes, 1. "AD_" in case of Active Directory dataset 2. "HR_" in case of HR dataset 3. "z" in case of derived/calculated attributed
7	Transform AD status of each user	If AD_Enabled = "True", then transform as "Active" user. Else, transform as "Inactive" user.

8	Flag the terminated users with active network access.	If the XYZ employee's HR Status is "Departed" and his/her AD status is "Active", then flag as "AuditCheck".
9	Aging in days	Calculate the days difference between run date and AD source attributes - 'AD_LastLogonDate', 'AD_Passwordset', 'AD_WhenChanged', 'AD_WhenCreated'.
10	Reporting Date	Is the date when both the source system datasets are ingested into the enterprise data lake.

4.3.4   Reporting Attributes

#DR	Data Requirement	Description
8	User Group Assignments table	['AD_SamAccountName','AD_DisplayName', 'AD_Group', 'AD_xyz_Grp_Tags', 'AD_Privileged_Access',"zMatched","Reporting_Date"]

9	User Dimension table	['AD_SamAccountName','AD_DisplayName','AD_AccountExpiry', 'AD_Company', 'AD_CreateTimeStamp', 'AD_Department', 'AD_Description', 'AD_EmailAddress',
		
		    'AD_Enabled', 'AD_Firstname', 'AD_LastLogonDate', 'AD_Lastname', 'AD_Passwordset', 'AD_xyz_Dept_Tags', 'AD_WhenChanged', 'AD_WhenCreated', 'HR_Department_Name', 'HR_Division', 'HR_Employee_No', 'HR_First_Name', 'HR_Last_Name', 'HR_Manage_Email', 'HR_Record_Type',
		
		    'HR_Status', 'HR_Termination_Date', 'HR_Work_Email', 'zAD_User_Status', 'zTerminated_User_ActiveAD', 'zDays_SinceLastLogonDate', 'zDays_SincePswdReset', 'zDays_SinceWhenChanged', 'zDays_SinceWhenCreated',"zMatched","Reporting_Date"]

4.3.5   ER Diagram

Inmon approach is followed in designing the normalized data structures in the refined layer of the data lake. This layer serves as the single source of truth for the enterprise. As the business requirements change or source data changes, it is easy to update the data warehouse as one thing is in only one place. The following is the entity - relationship diagram of the refined tables.
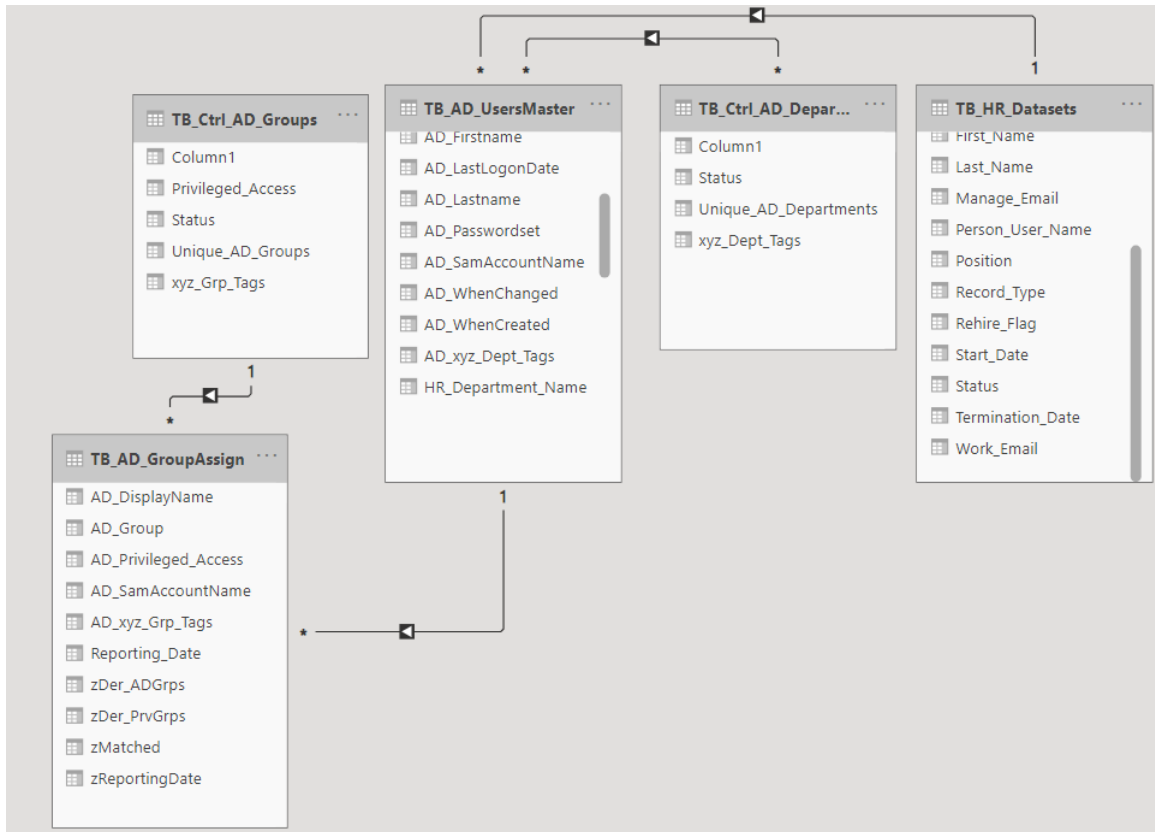
4.3.6   Reporting Model

The below star schema is a denormalized data structure is used for better performance of the Power BI dashboard. Dimensional tables – Enriched AD Users Master and Date Dummy table, are linked to the AD user assignments fact table. Table "zMeasures" is placeholder for all the calculated measures that were used in designing the visual dashboard.
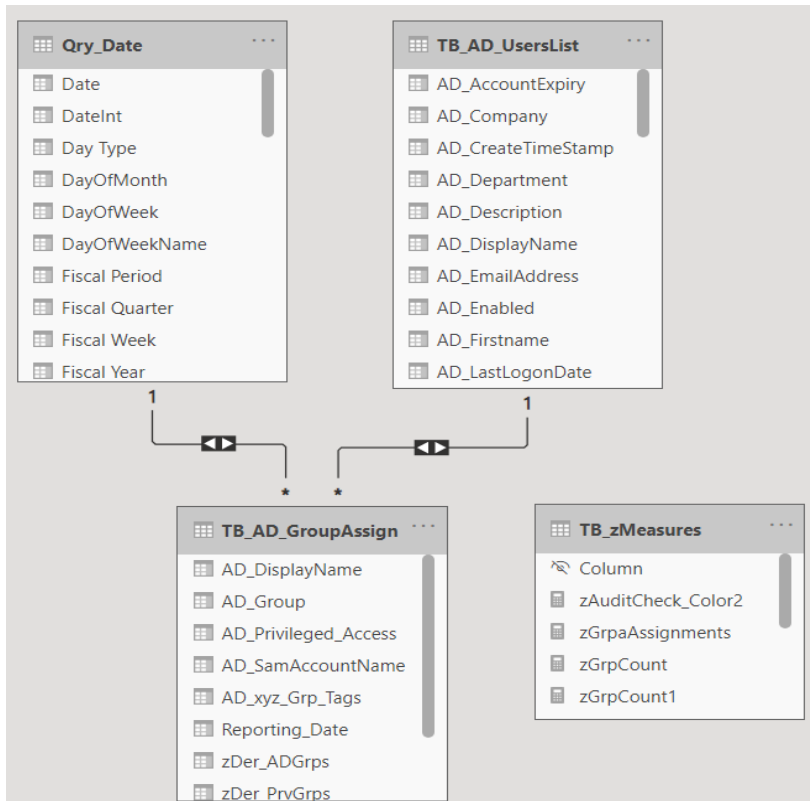
Relationships and Cardinalities are stated as below:

AD_Group	AD_xyz_Grp_Tags	AD_Privileged_Access	zMatched	Reporting_Date	zReportingDate	zD
grp1	X	null	HR_Match	20211230	December 30, 2021	Ot
grp3	X	null	HR_Match	20211230	December 30, 2021	Ot
grp1	X	null	HR_Match	20211230	December 30, 2021	Ot

TB_AD_GroupAssign

Qry_Date

Date	Year	QuarterOfYear	MonthOfYear	DayOfMonth	DateInt	Month Name
Wednesday, January 01, 2020	2020	1	1	1	20200101	January
Thursday, January 02, 2020	2020	1	1	2	20200102	January
Friday, January 03, 2020	2020	1	1	3	20200103	January

Cardinality

Many to one (*:1)

Cross filter direction

Both

TB_AD_GroupAssign

AD_SamAccountName	AD_DisplayName	AD_Group	AD_xyz_Grp_Tags	AD_Privileged_Access	zMatched
KINEMP2x	Kin Emp2	grp1	X	null	HR_Match
KINEMP2x	Kin Emp2	grp3	X	null	HR_Match
JUNEMP3x	Jun Emp3	grp1	X	null	HR_Match

TB_AD_UsersList

AD_SamAccountName	AD_DisplayName	AD_AccountExpiry	AD_Company	AD_CreateTimeStamp	AD_Dep
KINEMP2x	Kin Emp2	null	null	5/18/2021 3:20:58 AM	xyz
JUNEMP3x	Jun Emp3	null	null	9/18/2021 12:31:40 PM	XYZ_Dep
GALEMP4x	Gal Emp4	null	null	9/18/2021 12:31:40 PM	XY-Dept

Cardinality
Many to one (*:1)

Cross filter direction
Both

4.3.7 Source-to-Target Mapping (Data Lineage)

Source-to-Target Mapping (Data mapping) is where a map of the source data and how it is transformed and derived to reach its target destination in analytical reports. This Data lineage process gives visibility while greatly simplifying the ability to trace errors back to the root cause in a data analytics process.

									1st Line of Defense	2nd Line of Defense	3rd Line of Defense
Source Tables(Ingestion Layer)	Source Attributes (Ingestion)	Target Tables (Refined Layer)	Target Attributes (Refined La	Reporting Attribute (Analyti	Description	Direct/Derived	DataType	Mandat	Power BI Dashboard	Monthly User Access Review	Terminated User Review Report
HR_Status_Active_yyyymmdd	First Name	df_final_hr_dataset_yyyymmdd	First_Name	HR_First_Name	FTE first name	Direct	Varchar	O	X	X	
HR_Status_Active_yyyymmdd	Last Name	df_final_hr_dataset_yyyymmdd	Last_Name	HR_Last_Name	FTE last name	Direct	Varchar	O	X	X	
HR_Status_Active_yyyymmdd	Status	df_final_hr_dataset_yyyymmdd	Status	HR_Status	HR Status	Direct	Varchar	M	X	X	X
HR_Status_Active_yyyymmdd	Position Name	df_final_hr_dataset_yyyymmdd	Position			Direct	Varchar	O			
HR_Status_Active_yyyymmdd	Employee Assignment Type	df_final_hr_dataset_yyyymmdd	Employee Assignment Type			Direct	Varchar	O			
HR_Status_Active_yyyymmdd	Division	df_final_hr_dataset_yyyymmdd	Division	HR_Division	HR Division	Direct	Varchar	M	X	X	X
HR_Status_Active_yyyymmdd	Department Name	df_final_hr_dataset_yyyymmdd	Department_Name	HR_Department_Name	HR Department Name	Direct	Varchar	M	X	X	X
HR_Status_Active_yyyymmdd	Start Date	df_final_hr_dataset_yyyymmdd	Start_Date			Direct	Datetime	M			
HR_Status_Active_yyyymmdd	Person User Name	df_final_hr_dataset_yyyymmdd	Person_User_Name			Direct	Varchar	M			
HR_Status_Active_yyyymmdd	Email Address	df_final_hr_dataset_yyyymmdd	Work_Email	HR_Work_Email	FTE work email	Direct	Varchar	M	X	X	X
HR_Status_Active_yyyymmdd	Manager E-Mail Address	df_final_hr_dataset_yyyymmdd	Manage_Email	HR_Manage_Email	FTE manager's email	Direct	Varchar	M	X	X	X
HR_Status_Active_yyyymmdd	Record Type	df_final_hr_dataset_yyyymmdd	Record_Type	HR_Record_Type		Direct	Varchar	O	X		
HR_Status_Active_yyyymmdd	Employee #	df_final_hr_dataset_yyyymmdd	Employee_No	HR_Employee_No	FTE employee number	Direct	Varchar	M	X	X	X
HR_Status_Departure_yyyymmdd	Termination Date	df_final_hr_dataset_yyyymmdd	Termination_Date	HR_Termination_Date	FTE termination date	Direct	Datetime	M	X	X	X
IT_AD_XYZ_Users_yyyymmdd	Rehire_Flag	df_final_hr_dataset_yyyymmdd	Rehire_Flag			Derived	Varchar	M			
IT_AD_XYZ_Users_yyyymmdd	Department	df_ad_users_master_yyyymmdd	Department	AD_Department	Department Name	Direct	Varchar	M	X	X	X
IT_AD_XYZ_Users_yyyymmdd	Passwordset	df_ad_users_master_yyyymmdd	Passwordset	AD_Passwordset	Password reset date	Direct	Datetime	M	X	X	
IT_AD_XYZ_Users_yyyymmdd	DisplayName	df_ad_users_master_yyyymmdd	DisplayName	AD_DisplayName	User display name	Direct	Varchar	M		X	X
IT_AD_XYZ_Users_yyyymmdd	AdminCount	df_ad_users_master_yyyymmdd	AdminCount			Direct	Varchar	O			
IT_AD_XYZ_Users_yyyymmdd	Firstname	df_ad_users_master_yyyymmdd	Firstname	AD_Firstname		Direct	Varchar	O	X		

POC_Source-2-Target
_Mapping.xlsx

4.4 Operational Dashboard Prototype

The below operational dashboard prototype is designed in Power BI
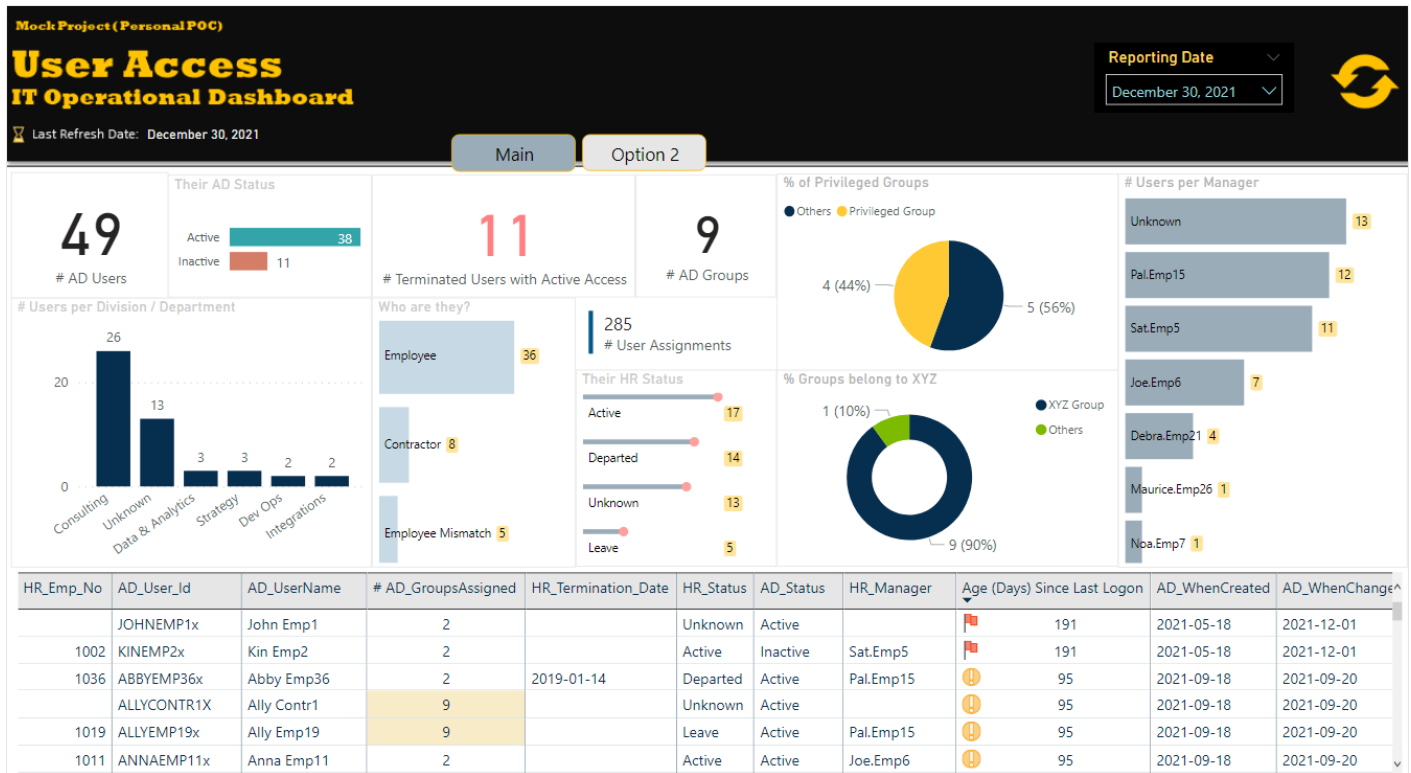
4.5   Non-Functional Requirements

The following are the non-functional requirements of this project:

#NFR	NF Requirement	Description
1	Frequency	Schedule the data pipelines in sequence at EOD between 10pm to 12am. Please ensure that all dependencies jobs in source systems are completed.
2	Performance	While processing the data pipelines or access the visual dashboard, ensure there is no performance issues occurred.
3	Scalability	Optimal resources are available at given time and should able to accommodate change requests/or upgradations in the near future.
4	Security and SLA	All standard security and compliance measures are implemented while running the data pipelines and accessing the visual dashboard. Also, ensure RTO (Recovery Time Objective) and RPO (Recovery Point Objective) measures are implemented as per the SLA agreement.

5 Appendices

5.1 List of Acronyms

Acronyms	Meaning
ETL	Extract Transform and Load
DW	Data warehouse
DB	Database
API	Application Programming Interface
AD	Active Directory

5.2 Glossary of Terms

Data wrangling (ELT), sometimes referred to as data munging, is the process of transforming and mapping data from one "raw" data form into another format with the intent of making it more appropriate and valuable for a variety of downstream purposes such as analytics. A data wrangler is a person who performs these transformation operations.

Extract-Transform-Load (ETL) is different from data wrangling. Here, we extract source data and apply the required transformations and then, load the refined data to target system.

5.3 Related Documents

Link(s)
Three lines of Defense Model: https://na.theiia.org/standards-guidance/Public%20Documents/PP%20The%20Three%20Lines%20of%20Defense%20in%20Effective%20Risk%20Management%20and%20Control.pdf
Three lines of Defense Model to Cloud Operations Segment: https://guidehouse.com/insights/financial-services/2021/public-sector/garp-three-lines-of-defense
Digital Transformation: https://www.i-scoop.eu/digital-transformation/
COVID-19 impact on business strategy :https://www.ibm.com/thought-leadership/institute-business-value/report/covid-19-future-business