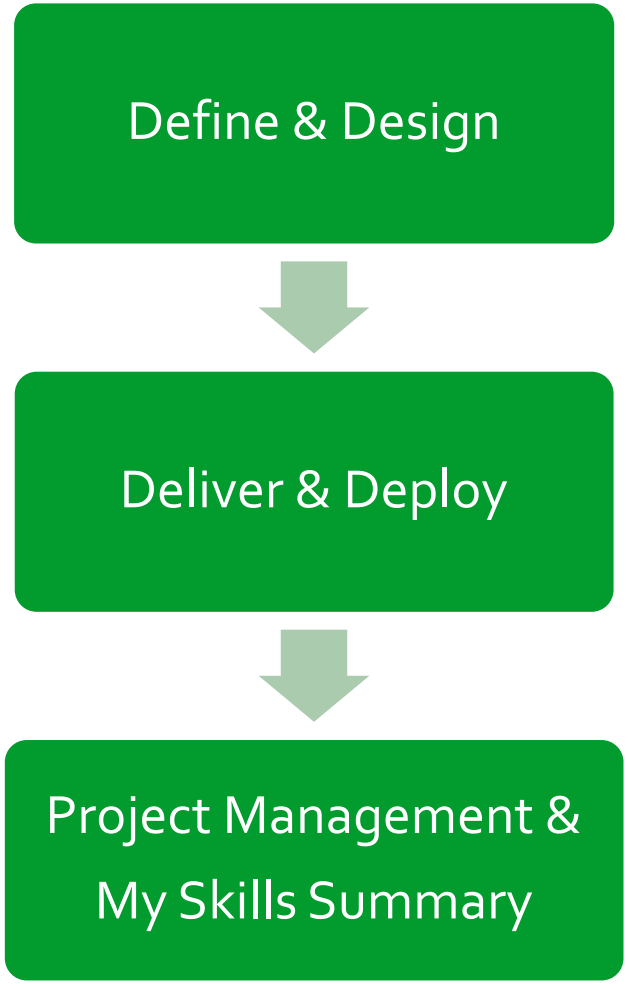
Digital Transformation (DX) of Three Lines of Defense Model

(Business Case : User's Network Access Controls)

- Sateesh Babu

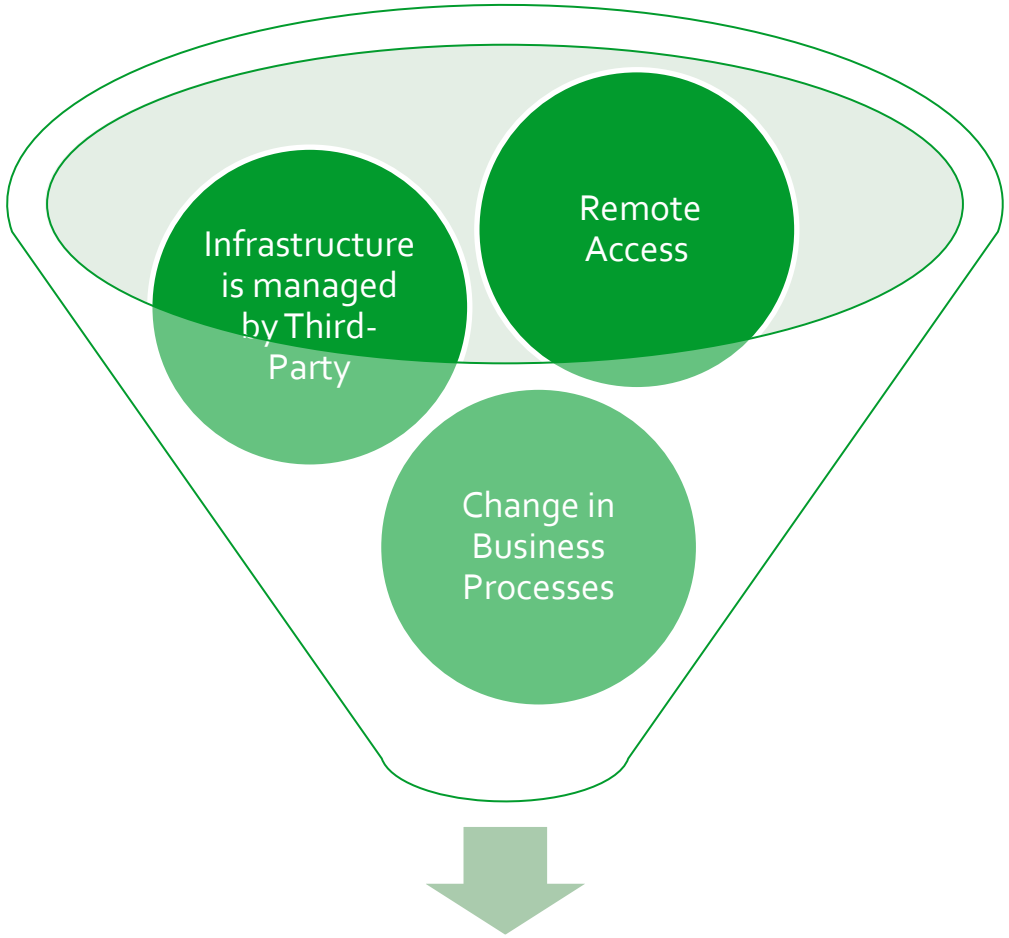
High Level Agenda



Define & Design

- Introduction & Problem Statement
- Current State Analysis and Architecture
- Identify Key Stakeholders
- Future State Analysis and Architecture
- Gap Analysis & Requirements
- Scope Statement & Solutions
- Value Proposition

Intro: Provoking question during/post Pandemic



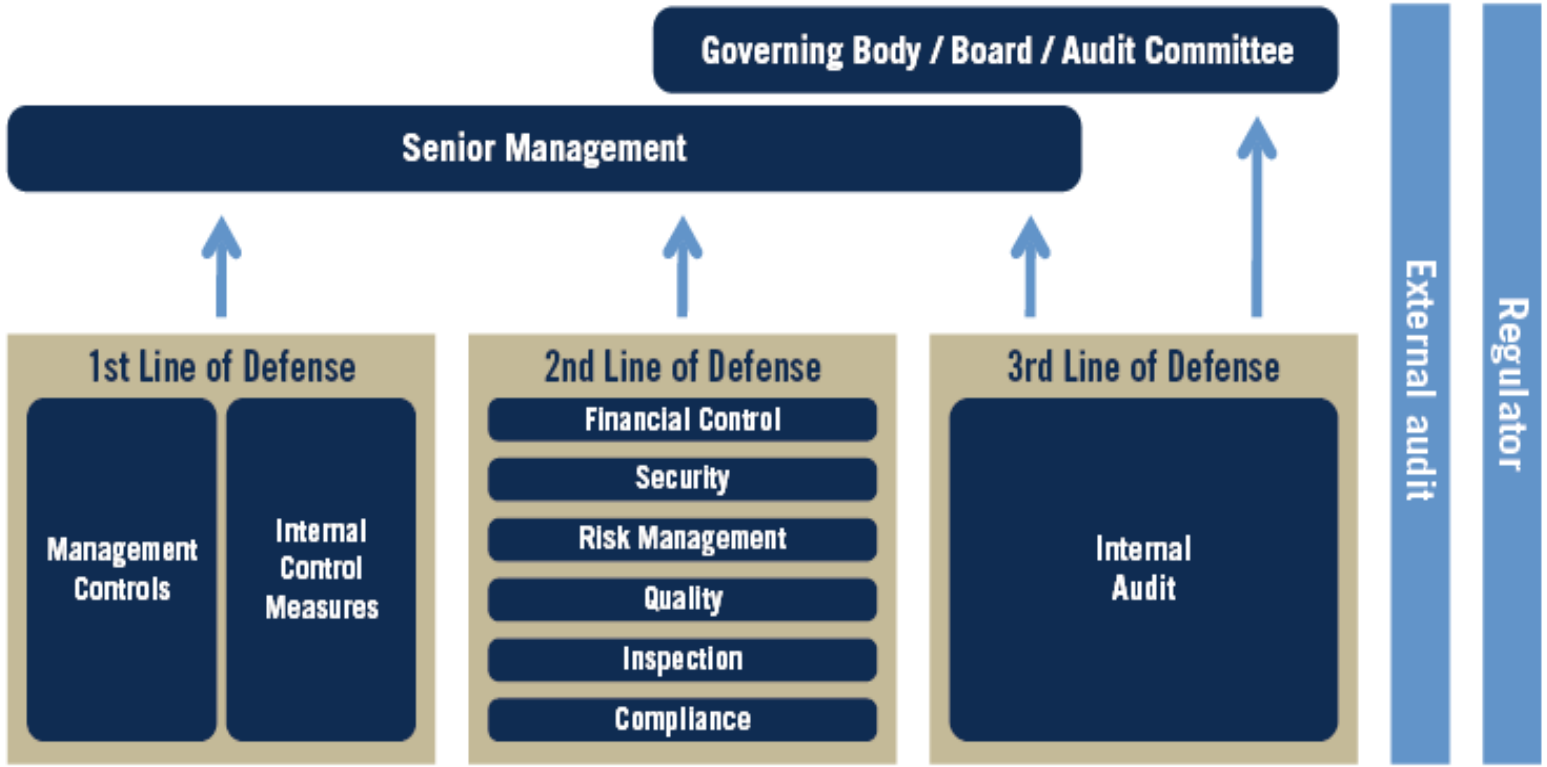
How do we REVIEW

our User Network Access?

The pandemic accelerated existing trends in remote work,e-Commerce,virtual meetings and other automations, organization need to streamline IT operations and enhance remote work experiences while balancing security measures at same time.

Intro: Three Lines of Defense Model

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*

Decision-making at our xyz organization is challenged by data quality, slow access to data, a lack of universal understanding of data best practices, and a general hesitancy to trust in the reliability of data-derived insights

Problem Statement

"As stakeholders of the three lines of defense mechanism of "Users' IT network access review", we don't have an integrated data and analytics platform or capabilities that focused on driving results for risk-based decisions and assurance reporting. Digitalization would help us to set risk control mechanism with strong corporate governance for better decision making and assurance reporting in monitoring "User Network Access" of xyz enterprise."

Who: Stakeholders of Three Lines of Defense

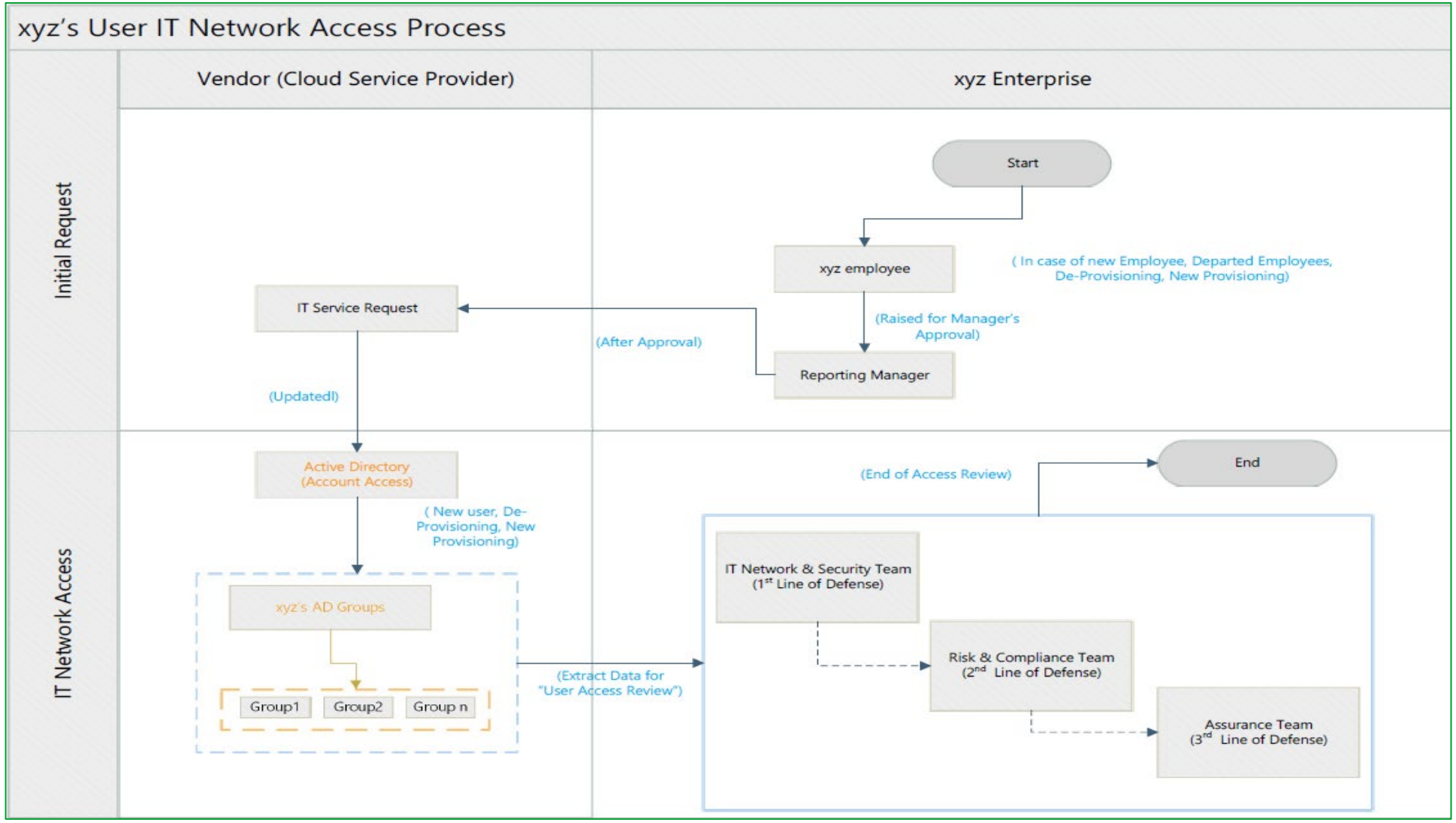
What: No integrated digitalized IT User Access review process. Inaccurate reporting led to wrong management decisions and conflicting assurance opinions.

Where: No Centralized integrated Data and Analytics Platform

Why: To make better decision making and assurance reporting in monitoring "User IT Network Access".

When: Daily, Periodic and Adhoc DIGITAL reporting.

High Level Current State Business Process

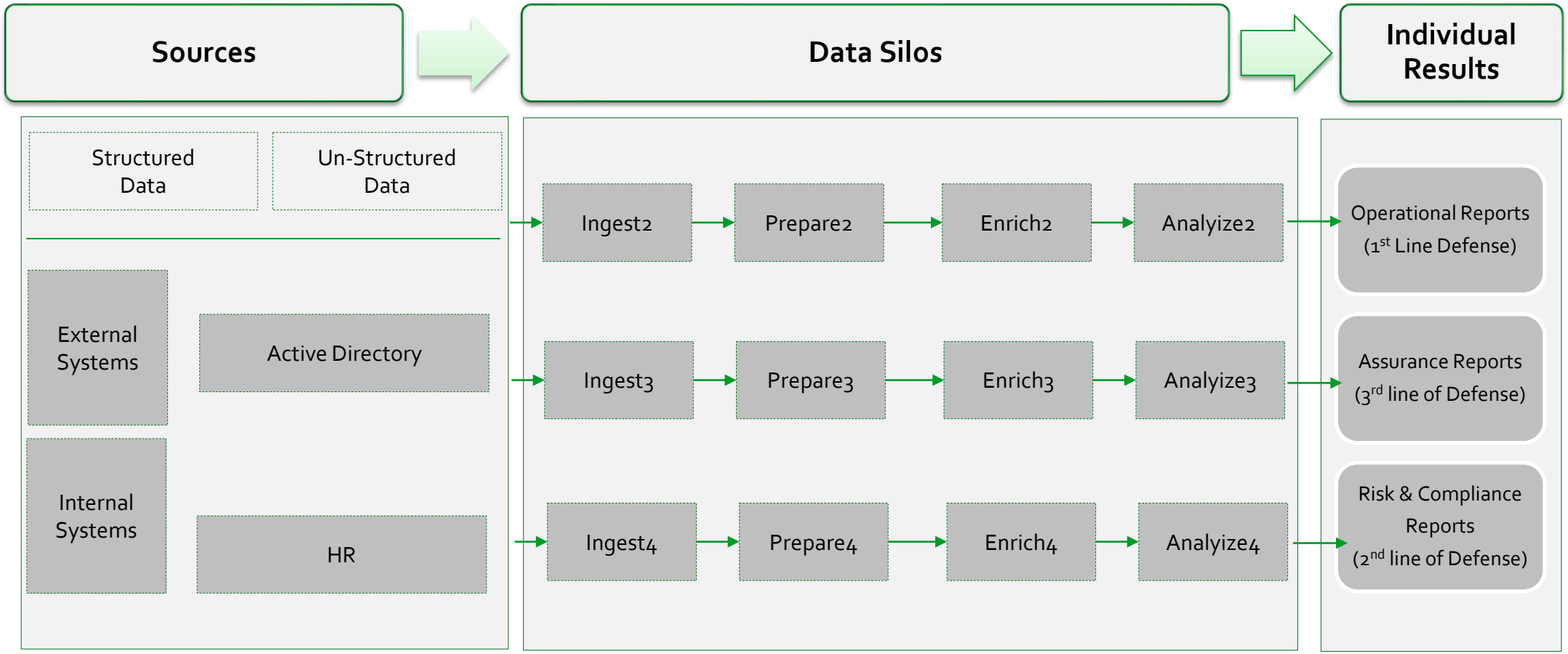


Key Business Stakeholders

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

Current State Analysis and Architecture



Current State Analysis and Architecture

As a part of a current state assessment of the organization, the following "**User's Network Access Controls**" data challenges were identified across business units:

- **Unable to Access the Data :** Due to data silos, the data collected by one team (1st line of defense) is not fully accessible to the other teams (Risk Committee and Internal Audit). It has created barriers to information sharing and collaboration across the key stakeholders.

- **Data Silos**: Each Business unit implement their own processes and transformations in collecting active directory data from the vendor. There is no standardized management practice or data governance. Usage of outdated HR data.

- **Lack of Data and Analytics Visibility:**Business Units often take on data and analytics initiatives with limited collaboration due to a low visibility into cross-BU data and analytics learnings and best practices. No single source of truth (SSOT).

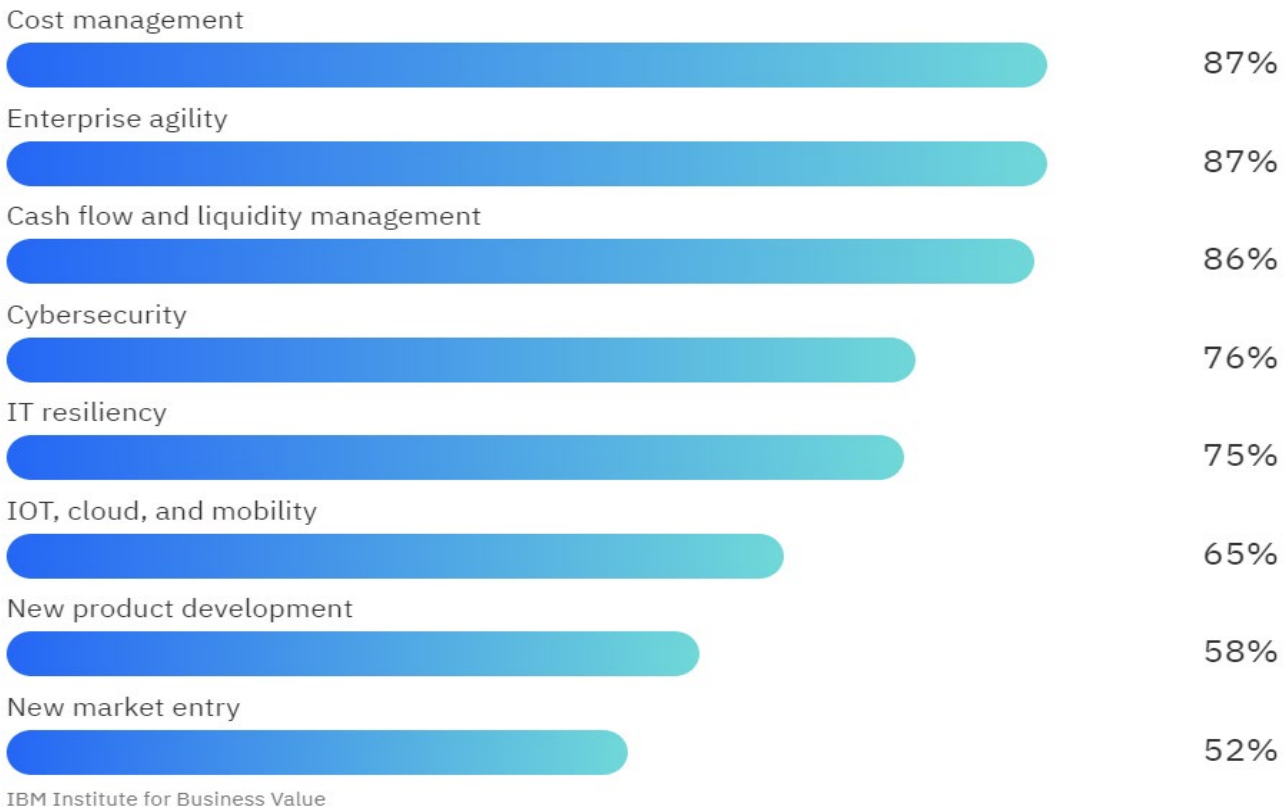
- **Wastage of Resources :** There are difficulties in collecting data and using it for reporting / analytics due to manual processes. This leads to high amounts of effort spent on data cleansing and wrangling, limiting the ability to extend into more advanced data and analytics use cases.

- **Data Quality Issues:** Lack of defined processes to measure, monitor, and report on data quality levels

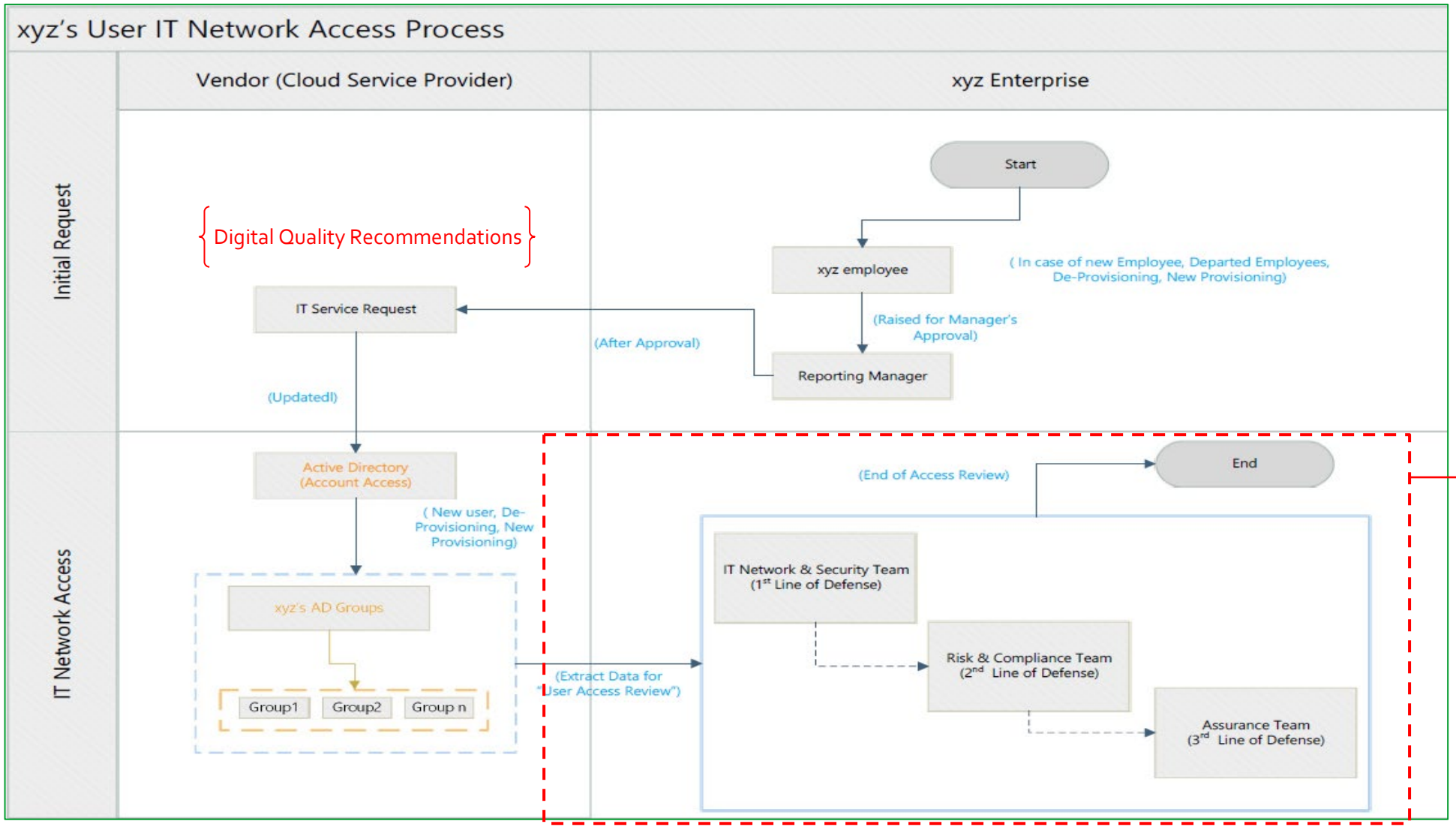
Intro: Digital Transformation & Key drivers

Digital transformation is not just change in technology, its a rethinking of how an organization uses technology, information, people, and processes to improve efficiency, value or innovation.

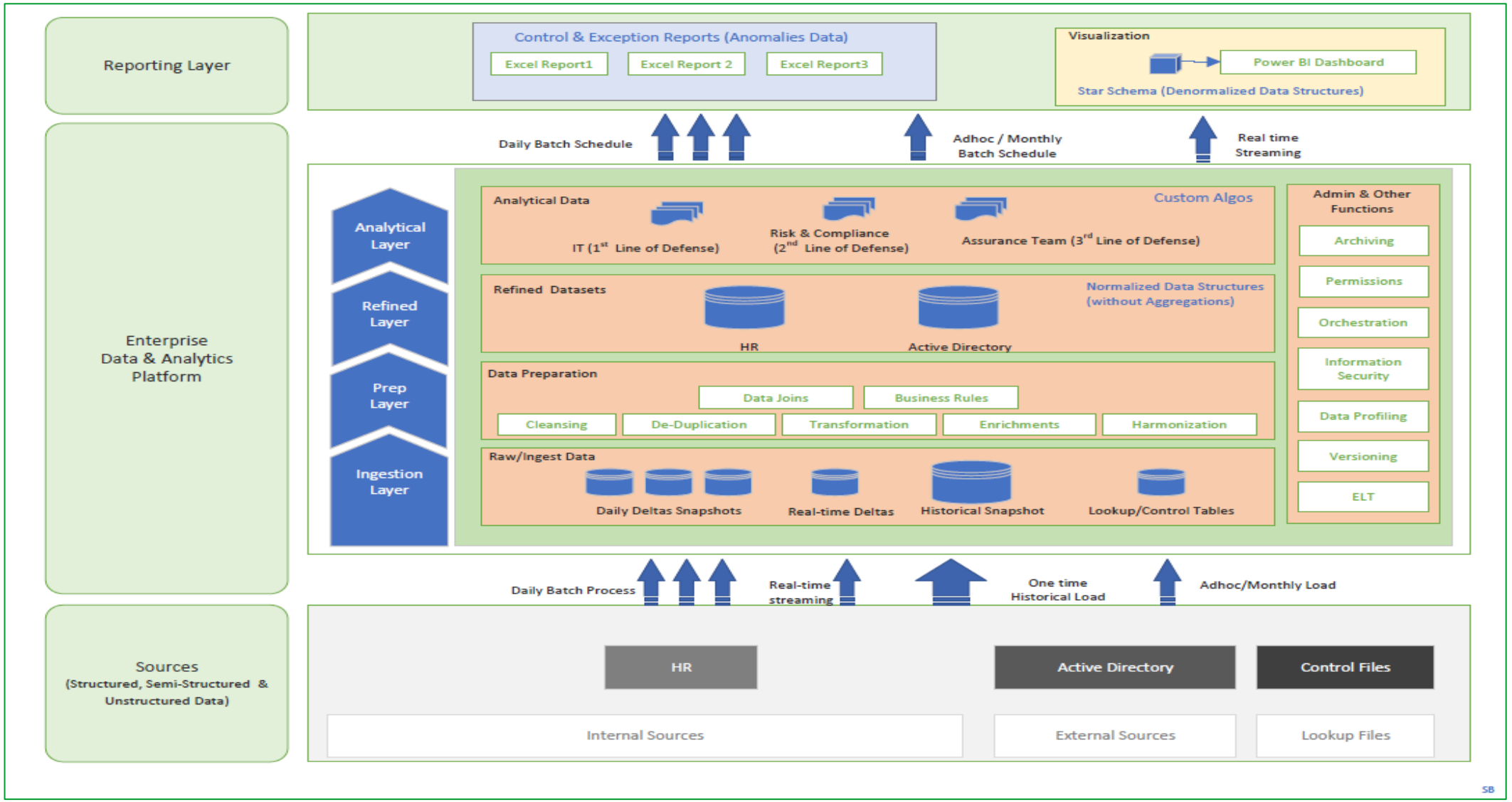
As per IBM survey, the COVID-19 pandemic has accelerated digital transformation initiatives and the prioritization of the key drivers as follows



High Level Future State Business Process



Future State Analysis and Architecture



Future State Analysis and Architecture

The following are the key points that should be addressed to achieve the “Target State” of the organization.

- **Bridge Gaps in Enterprise Data Communication:** Business units should collaborate to resolve data related issues and should create a reusable data analytical assets. Facilitate issue resolution and change management as a neutral body.

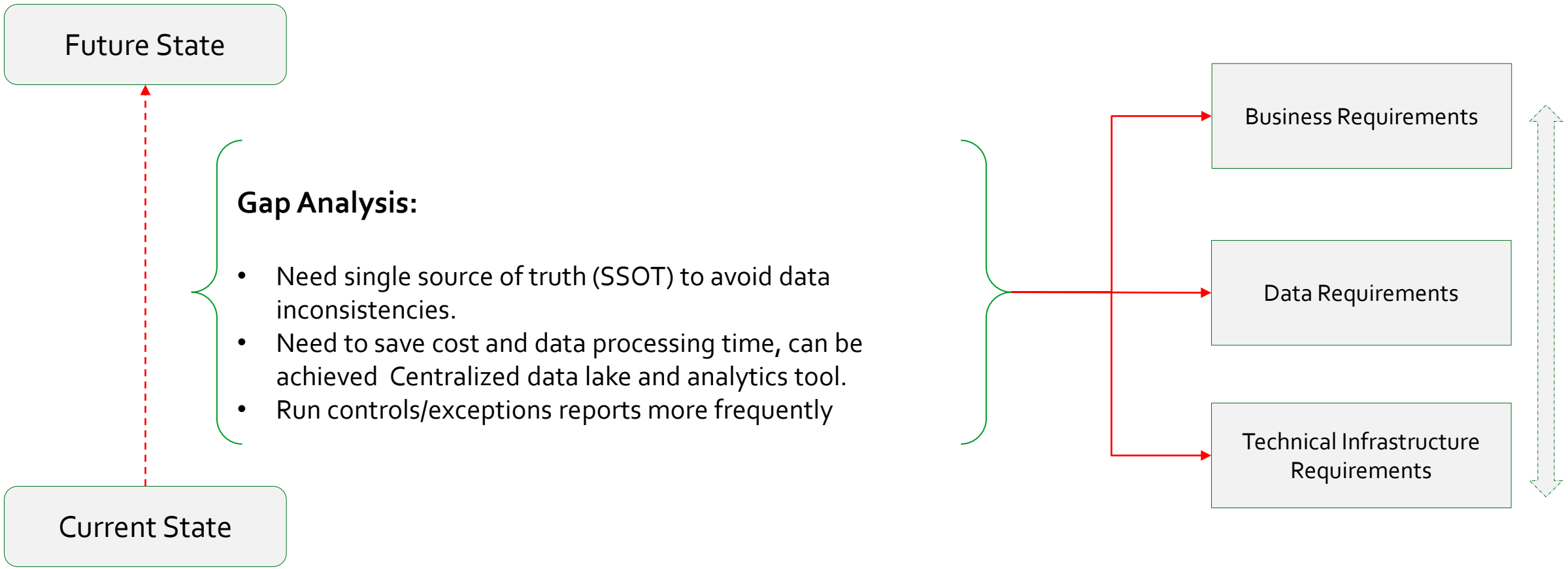
- **Break Data Silos with Centralized Data Management Services**: The best way to bust silos is to pool all corporate data into a cloud-based data lake— a central data repository optimized for efficient analysis. Data from disparate sources will be homogenized and consolidated, and access can be easily granted to individuals or groups to balance business need against privacy and security.

- **Enhance Baseline Analytics Capabilities and Streamline “User Access” Reporting:** Develop internal analytics and reporting capabilities in the context of the CoE and specific business needs. Provide necessary tools and support.

- **Data Quality Recommendations** : 1.Clean revoke user access and assignments if account is not active more than 90 days. 2. Daily cross validation with HR data will be helpful in removing the access of unauthorized users and terminated employees. 3. Frequent reviews on AD master data.

- **Build Data Governance Gateways:** Create robust data access policies and define strong data governance framework.

Gap Analysis



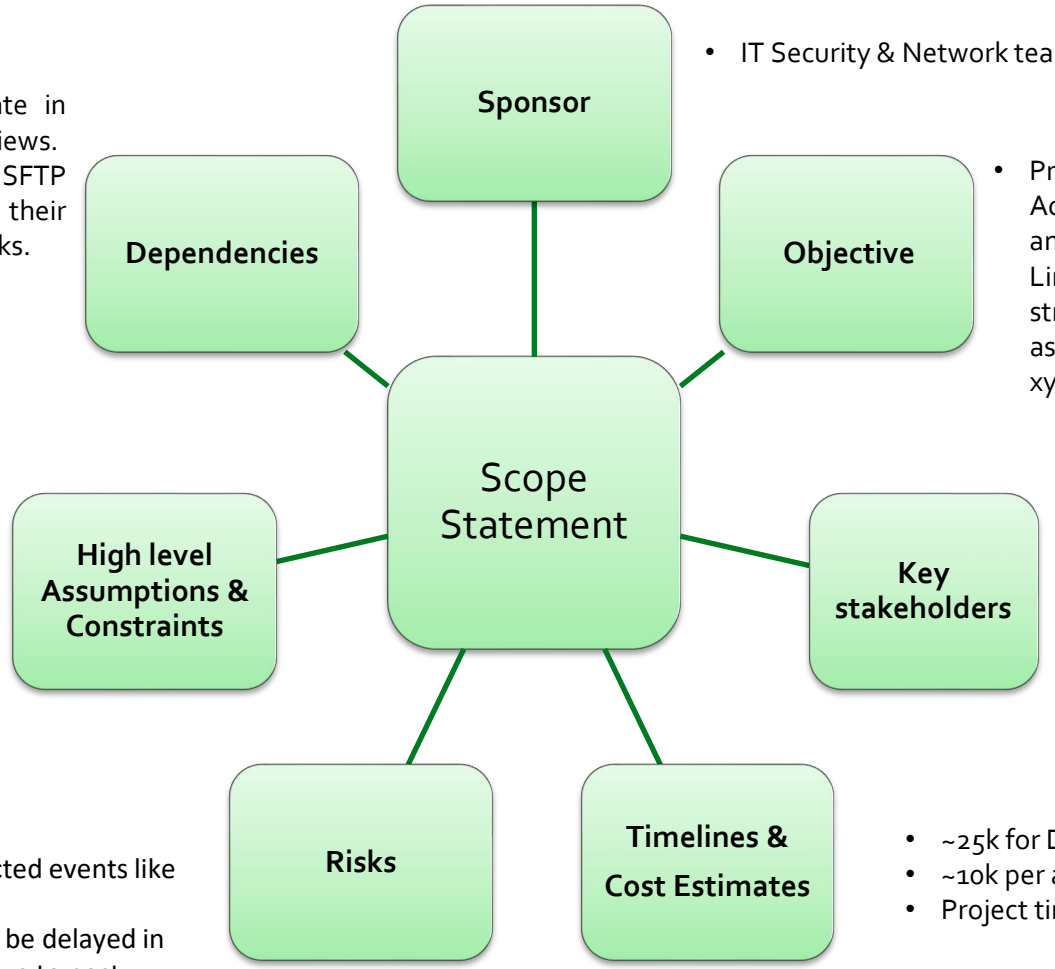
Business Requirements

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	Daily 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	Monthly 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	Adhoc 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)

BRD

Scope Statement



- 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.

- IT Security & Network team (1st line of Defense Model)

- 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.

Assumption: Assume that data owners/stewards/custodians will provide complete and validated daily datasets in tandem to the business processes

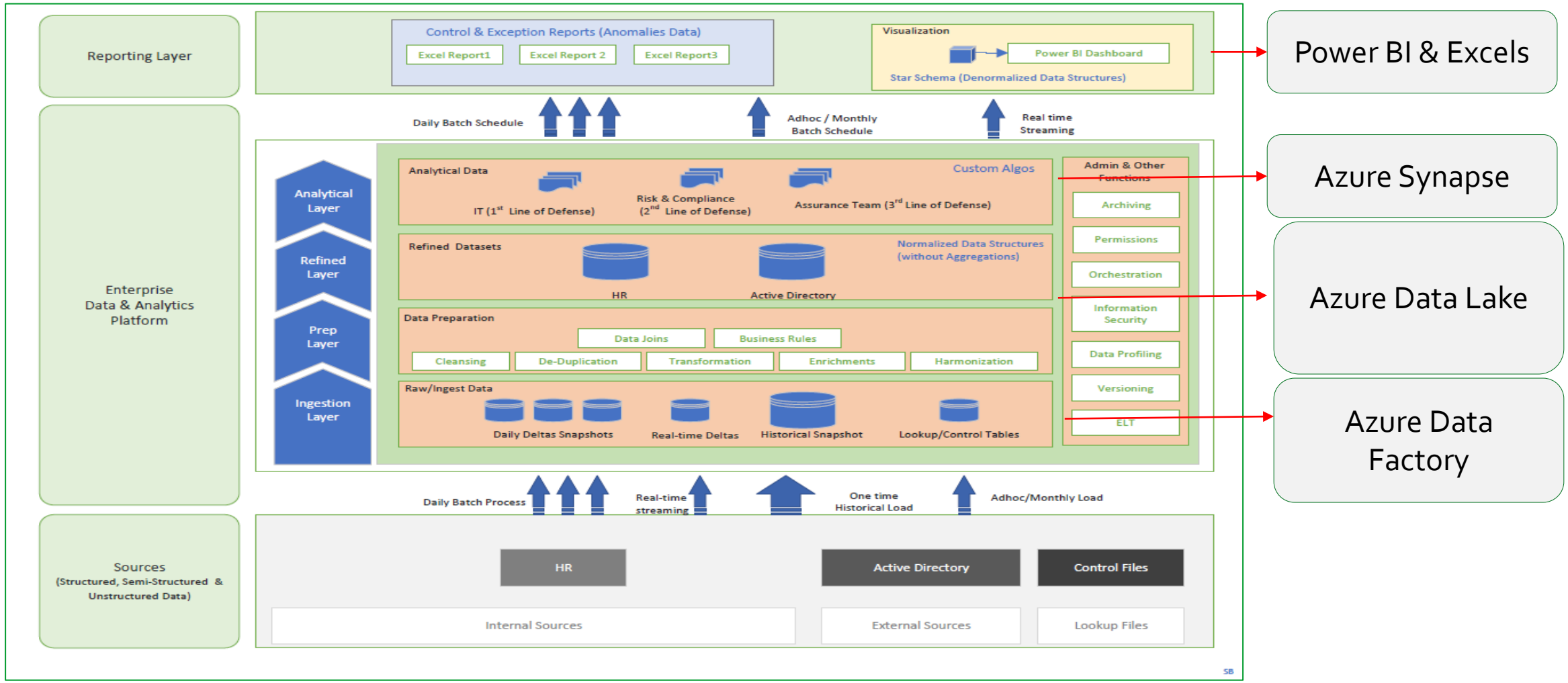
Constraint: Data quality issues (like email mismatches, user id mismatches, no standard master data etc.) in the vendor AD, might be a constraint in designing and maintaining the dashboard visuals.

- **Business Stakeholders:** IT Security & Network team (owner); Risk –Compliance team ; Internal Audit.
- **Other Stakeholders**: IT ; Data Governance Sponsor ; Enterprise Architect team ; Enterprise Data Governance Committee; Vendor; HR team

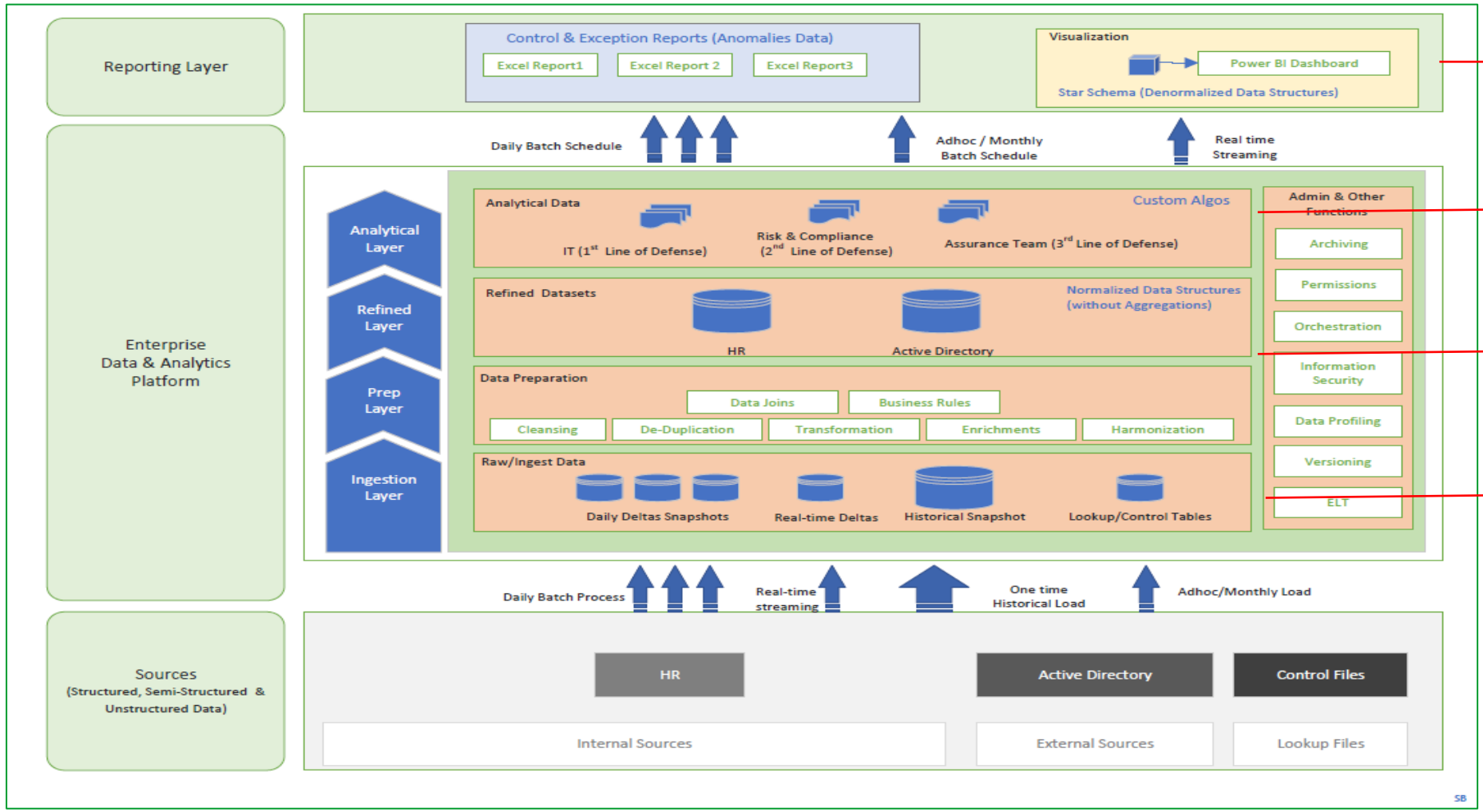
- No control over staff priorities in unexpected events like COVID-19 pandemic.
- Vendor's statement of work (SOW) might be delayed in getting approved by procurement team due to post COVID-19 changes in the business processes.

- ~25k for Design ,Build, Test and Deploy (One time)
- ~10k per annum (for Maintenance support)
- Project timelines are approx. 6-8 months

Strategic Solution - Future State Architecture for Continuous Analytics



Tactical Solution – Standalone Analytics



Value Propositions

At Traction on Demand, we use best technology and innovative ideas to empower your business and IT Teams.

- We will guide you in implementing **"future fit"** strategy in addressing the emerging business/IT needs (like "User Network Access") of your company.
- We **empower your team** to maintain a healthy org that's aligned to platform best practices, letting you focus on investing in forward-looking functionality and scalable innovation, and reducing your technical debt long term.
- Our data experts provide clients with a proven approach for **managing data with tailored business rules,** data definitions, and processes for validating.
- Our innovative data analytics solutions :
  - Will **reduce the costs** though reusability and modularity
  - Will help in **breaking silos** with cross-functional collaboration
  - Has the **ability to evaluate massive data** quickly to provide comprehensive audits.
  - **Will save time** and reduces manual burden of in enhancing business processes.
  - **Identifies additional risks** and flags anomalies by better understanding the existing risks
  - Will provide **predictable insights** for strategic risk-based decisions .

Deliver & Design

- Requirements & Data Analysis
- In-Scope & Out-of-Scope
- Assumptions & Constraints
- Non-Functional Requirements (NFRs)
- Logical Data Modeling & Data Lineage
- Construct Data Pipelines
- Power BI Visualization

Business Requirements to 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

Data Profiling Analysis

Sample 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.

Functional_Spec

In-Scope & Out-of-Scope

In-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.

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.

Assumptions & Constraints

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.

Assumptions & Constraints

Constraints

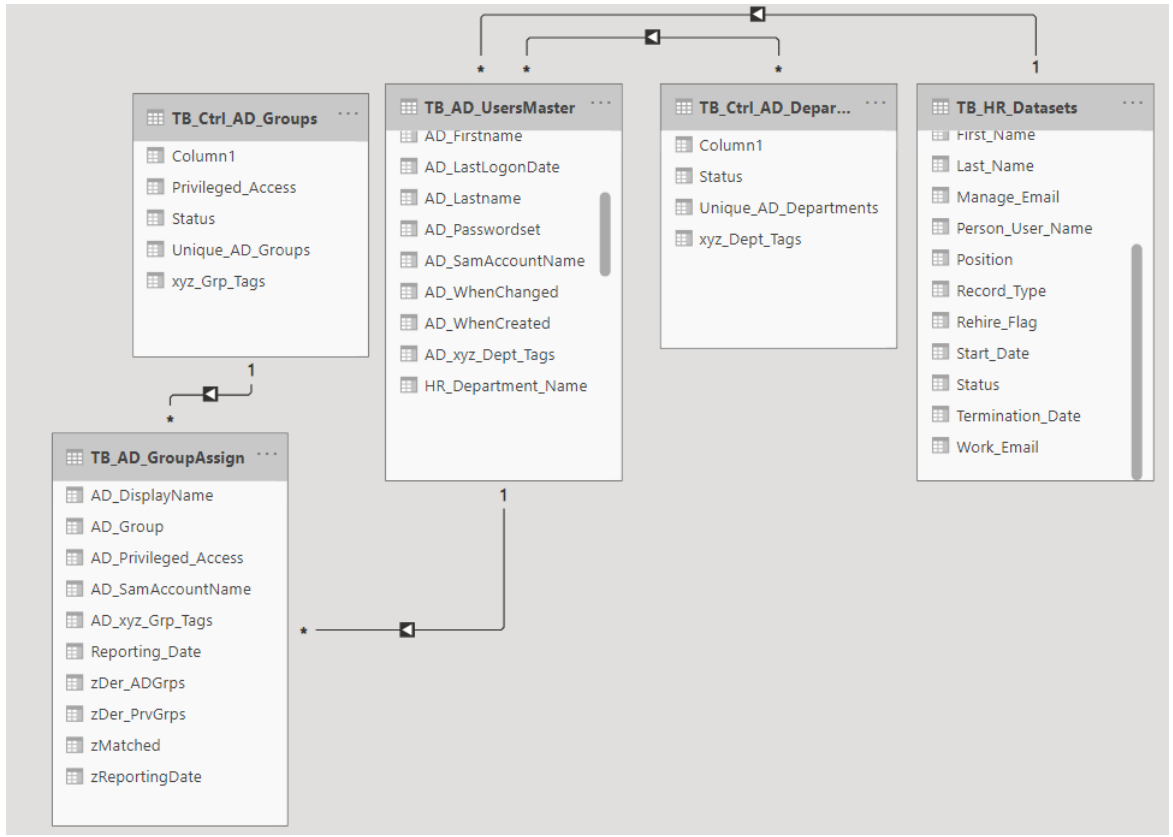
- As AD don't have metadata information on the version history in its dataset, it's not feasible to provide version changes of each user account.
- No primary key for AD user/ group assignments, there will be some challenges in maintaining the delta extractions.
- Vendor's Active Directory does not delineate 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.

Non-Functional Requirements (NFRs)

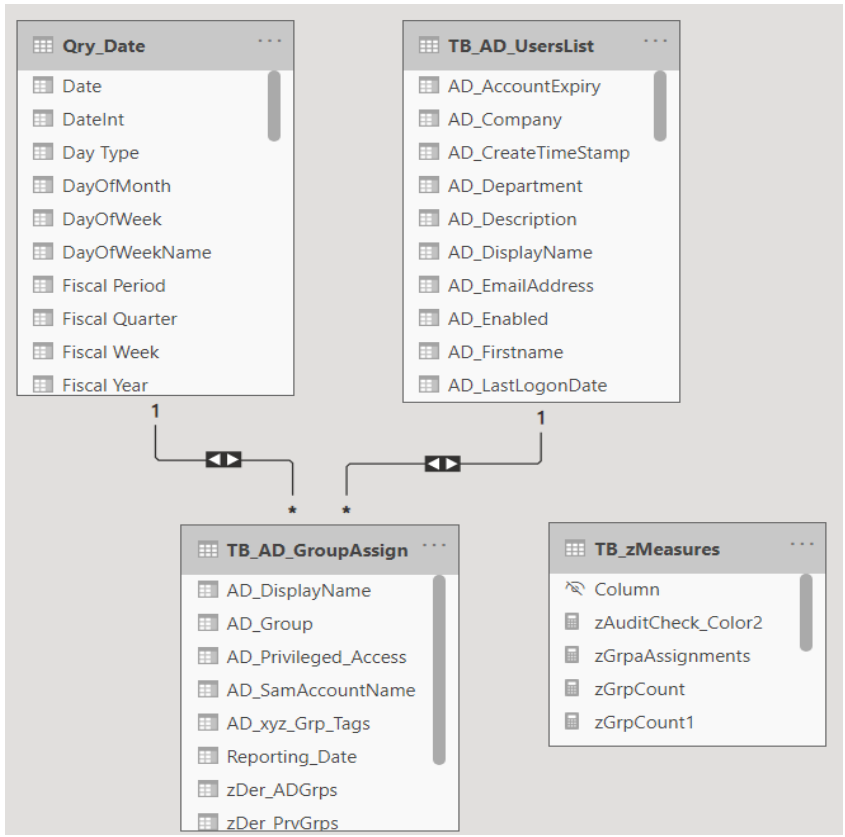
#NFR	NF Requirement	Description
1	Frequency	Schedule the data pipelines in sequence at EOD between 10pm to 12am. Please ensure that all dependencies jobs in source systems are completed.
2	Performance	While processing the data pipelines or access the visual dashboard, ensure there is no performance issues 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.

Logical Data Modeling

"ER model" for the refined layer in the Data Lake



"Star Schema" for the reporting in POWER BI

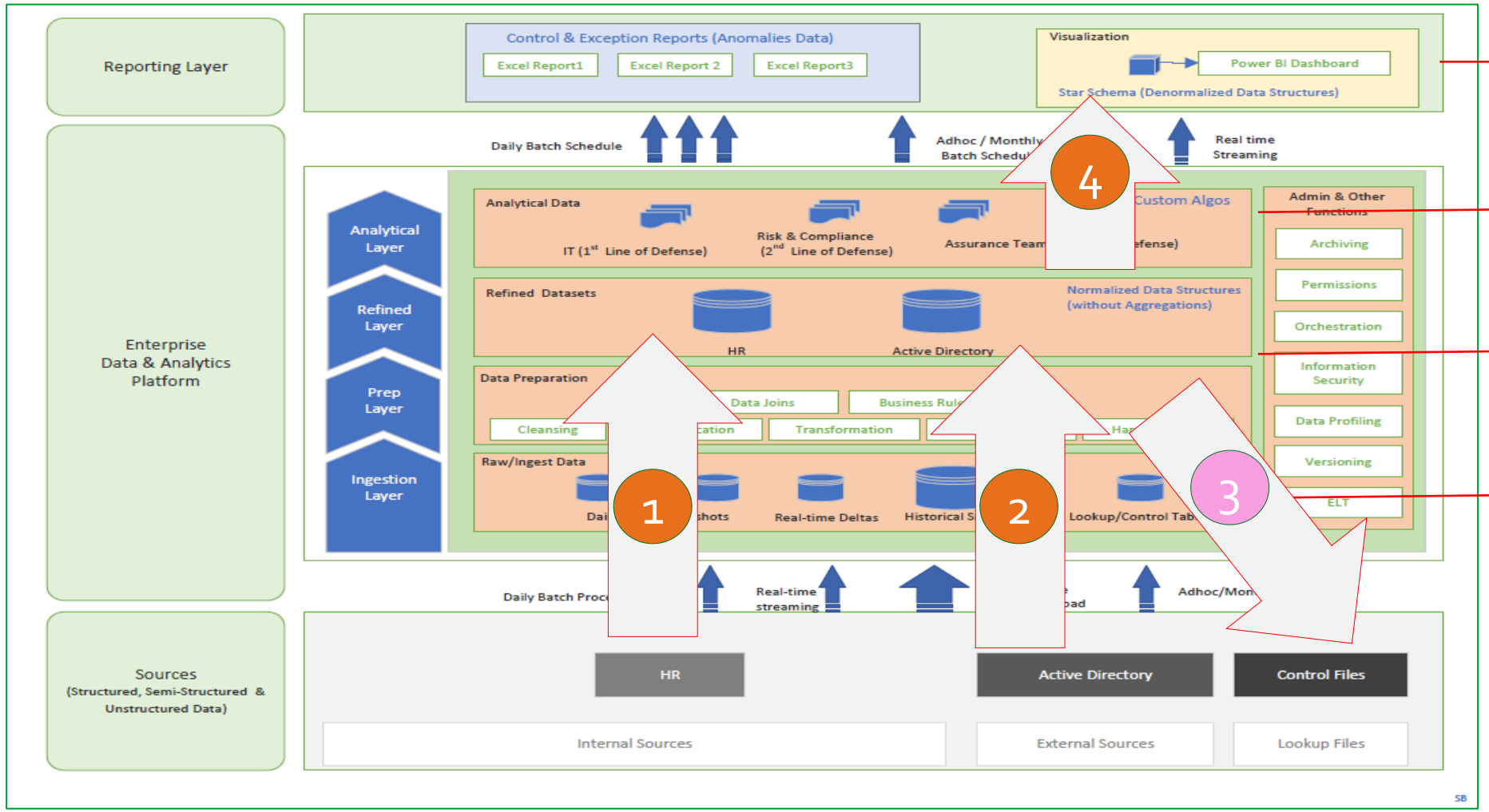


Data Lineage (Source-2-Target Mapping)

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.

Source-to-Target Mapping									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 (Analytic	Description	Direct/Derived	DataType	Mandat	Power BI Dashboard	Monthly User Access Review	Terminated User Review Report
HR_Status_Active_yyyymmddHR_	First Name	df_final_hr_dataset_yyyymmdd	First_Name	HR_First_Name	FTE first name	Direct	Varchar	O	X	X	
HR_Status_Active_yyyymmddHR_	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_yyyymmddHR_	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_yyyymmddHR_	Division	df_final_hr_dataset_yyyymmdd	Division	HR_Division	HR Division	Direct	Varchar	M	X	X	X
HR_Status_Active_yyyymmddHR_	Department Name	df_final_hr_dataset_yyyymmdd	Department_Name	HR_Department_Name	HR Department Name	Direct	Varchar	M	X	X	X
HR_Status_Active_yyyymmddHR_	Start Date	df_final_hr_dataset_yyyymmdd	Start_Date			Direct	Datetime	M			
HR_Status_Active_yyyymmddHR_	Person User Name	df_final_hr_dataset_yyyymmdd	Person_User_Name			Direct	Varchar	M			
HR_Status_Active_yyyymmddHR_	Email Address	df_final_hr_dataset_yyyymmdd	Work_Email	HR_Work_Email	FTE work email	Direct	Varchar	M	X	X	X
HR_Status_Active_yyyymmddHR_	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_yyyymmddHR_	Record Type	df_final_hr_dataset_yyyymmdd	Record_Type	HR_Record_Type		Direct	Varchar	O	X		
HR_Status_Active_yyyymmddHR_	Employee #	df_final_hr_dataset_yyyymmdd	Employee_No	HR_Employee_No	FTE employee number	Direct	Varchar	M	X	X	X
HR_Status_Departure_yyyymmd	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			

Construct Data Pipelines



Power BI & Excels

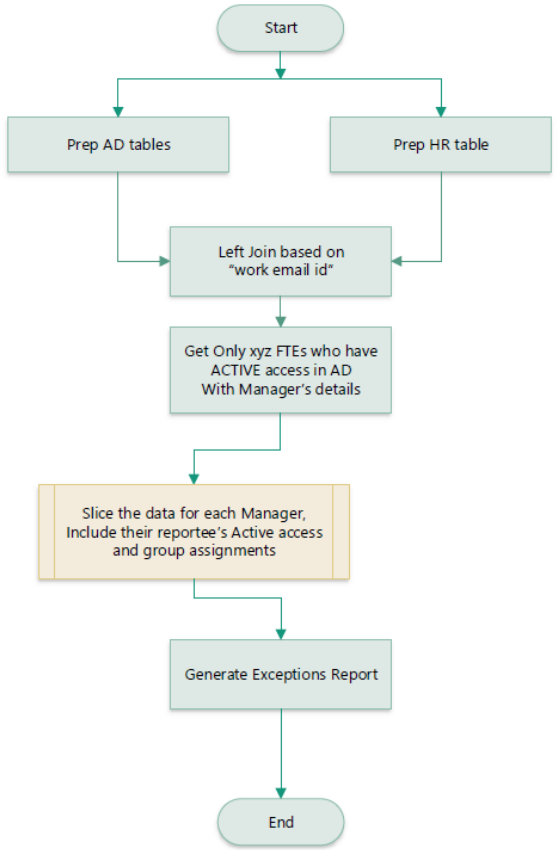
Jupyter - Analytical WB

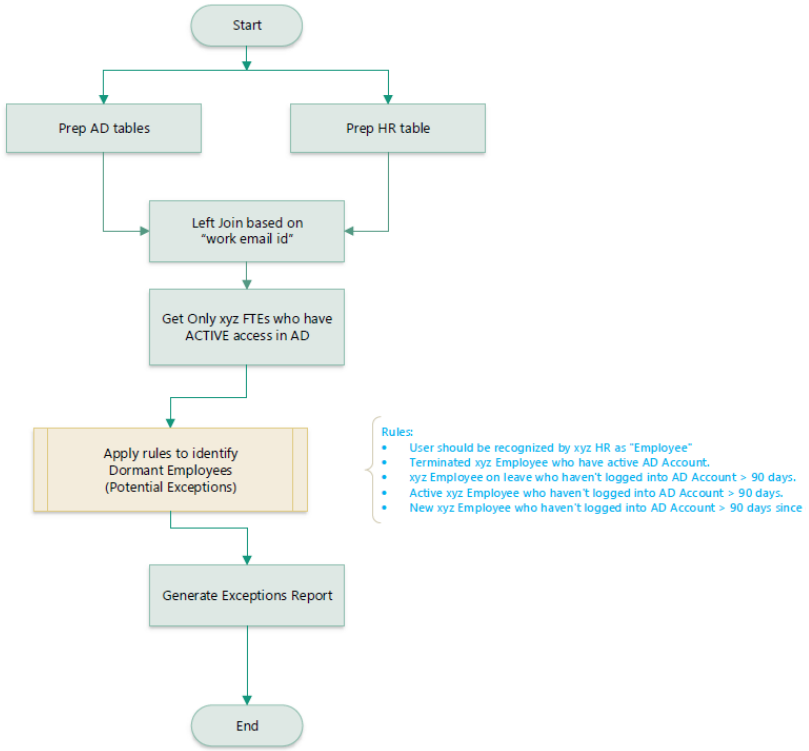
Local Network Drive

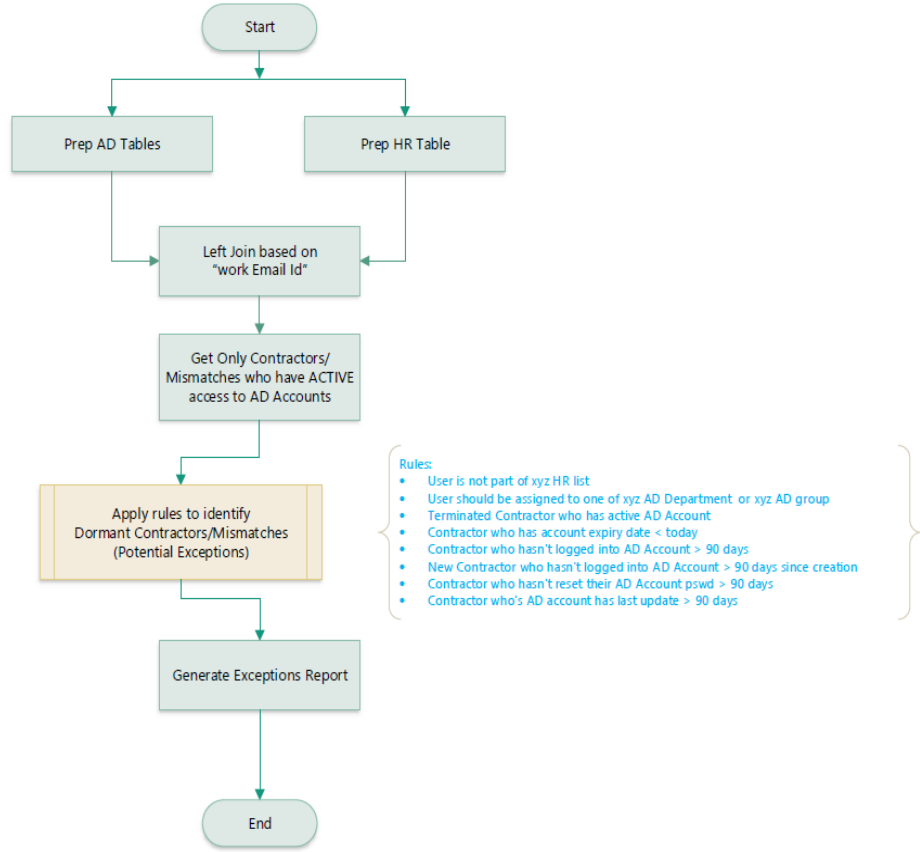
Jupyter - Analytical WB

1. HR data pipeline
2. AD data pipeline
3. AD master data changes
4. Analytical pipeline

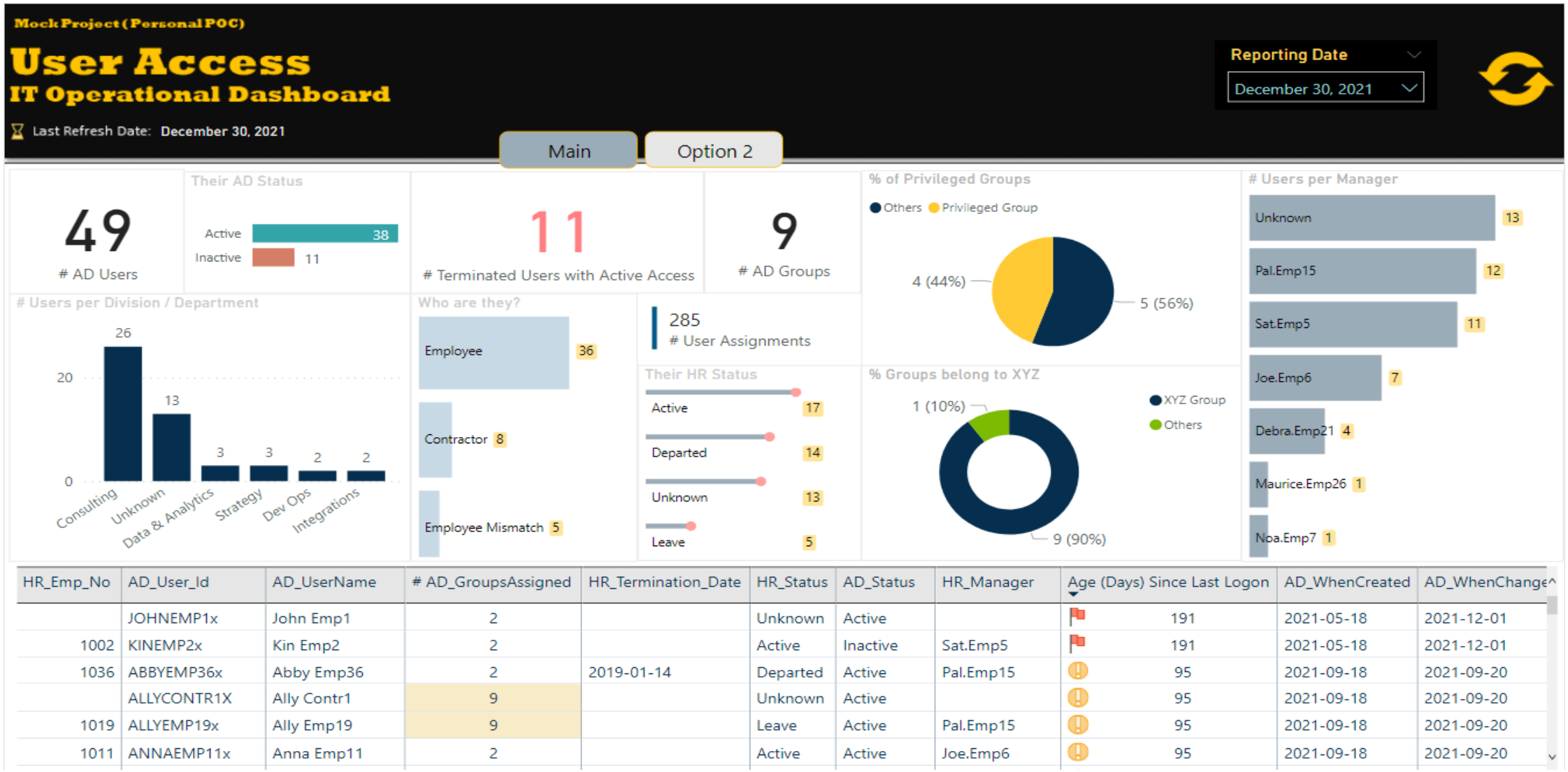
Analytical Reports – Flow Charts





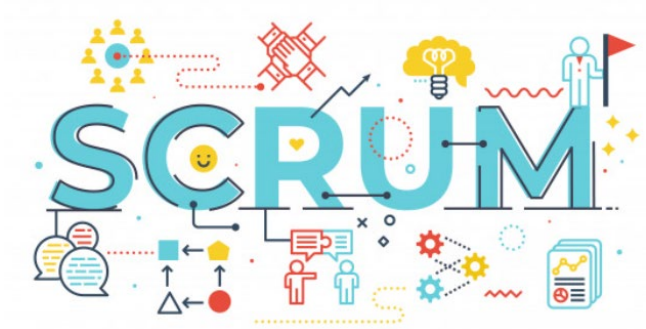


Power BI Dashboard



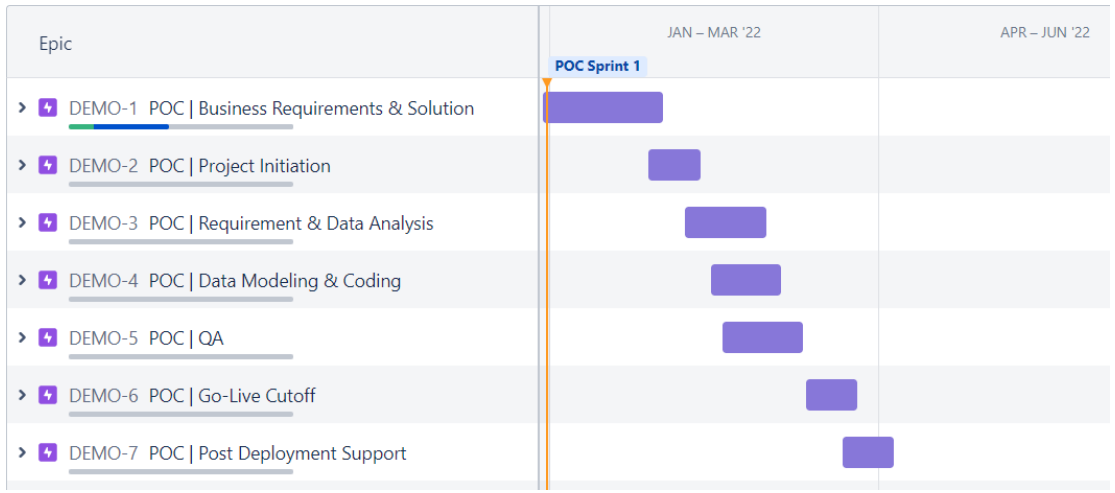
HR_Emp_No	AD_User_Id	AD_UserName	# AD_GroupsAssigned	HR_Termination_Date	HR_Status	AD_Status	HR_Manager	Age (Days) Since Last Logon	AD_WhenCreated	AD_WhenChange
	JOHNEMP1x	John Emp1	2		Unknown	Active		191	2021-05-18	2021-12-01
1002	KINEMP2x	Kin Emp2	2		Active	Inactive	Sat.Emp5	191	2021-05-18	2021-12-01
1036	ABBYEMP36x	Abby Emp36	2	2019-01-14	Departed	Active	Pal.Emp15	95	2021-09-18	2021-09-20
	ALLYCONTR1X	Ally Contr1	9		Unknown	Active		95	2021-09-18	2021-09-20
1019	ALLYEMP19x	Ally Emp19	9		Leave	Active	Pal.Emp15	95	2021-09-18	2021-09-20
1011	ANNAEMP11x	Anna Emp11	2		Active	Active	Joe.Emp6	95	2021-09-18	2021-09-20

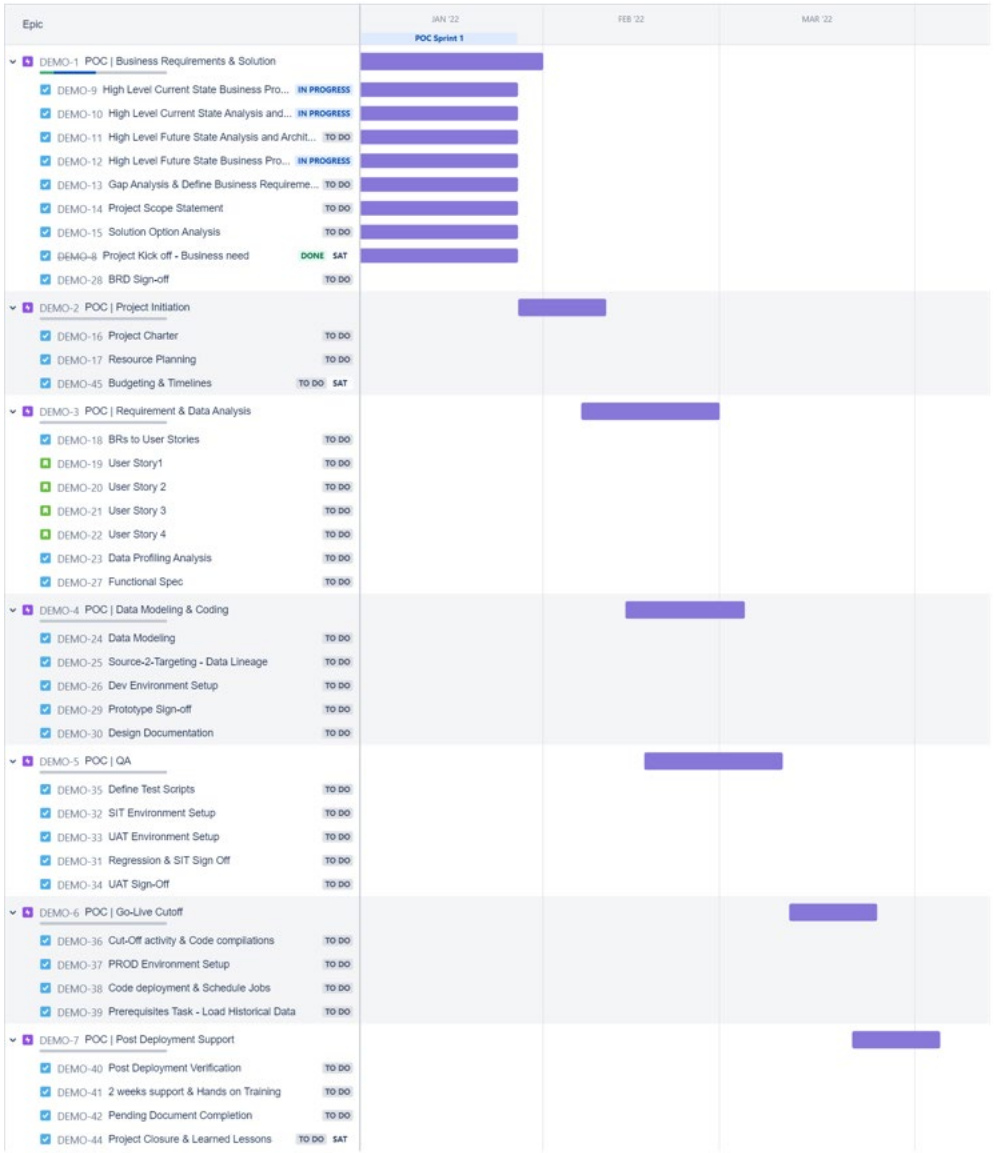
Project Management & My Skills Summary



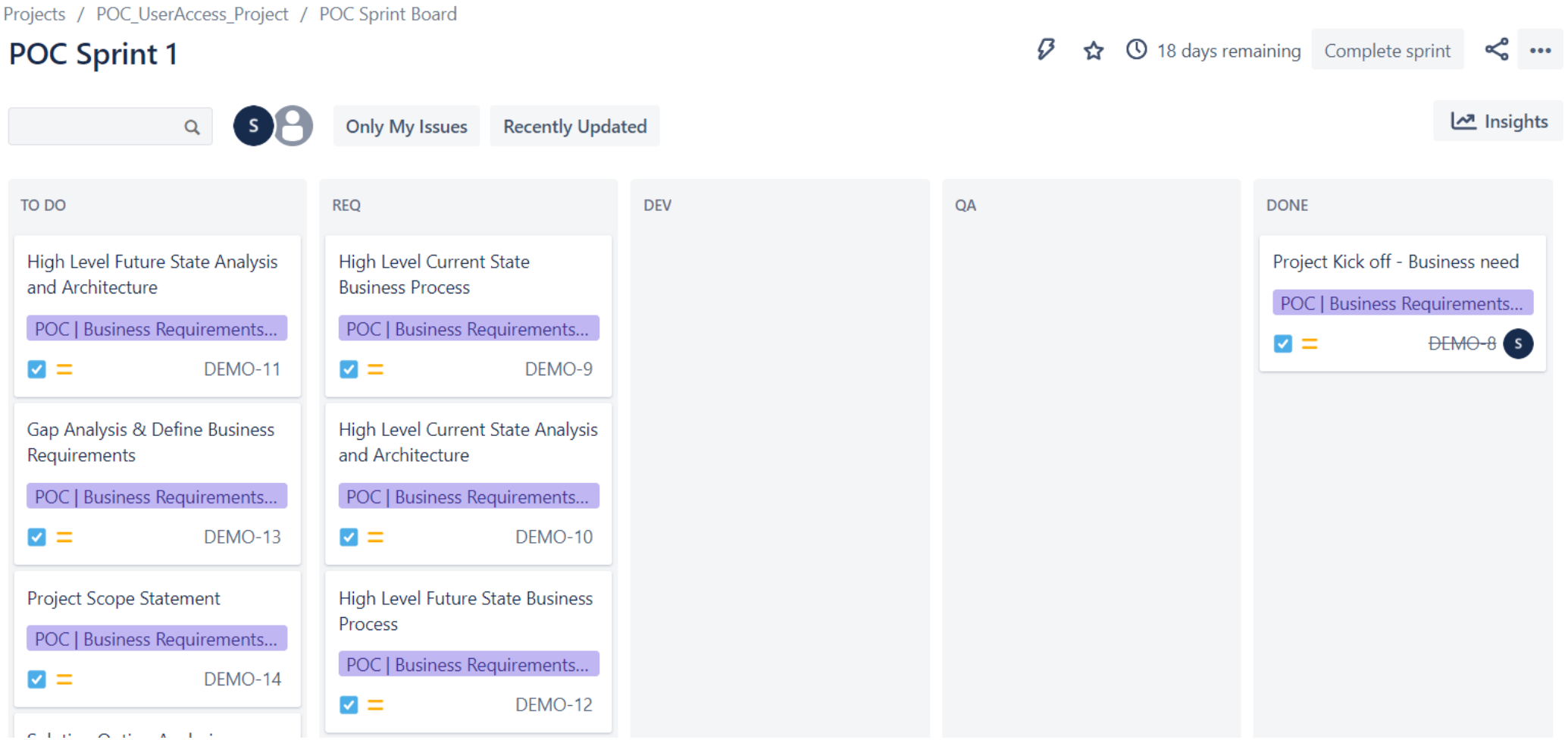
- Jira EPICs & Work Break-down Structure
- Agile SCRUM Board
- Project Backlog
- My Skills Summary

Project Management : Jira Epics & Work-breakdown Structure (WBS)

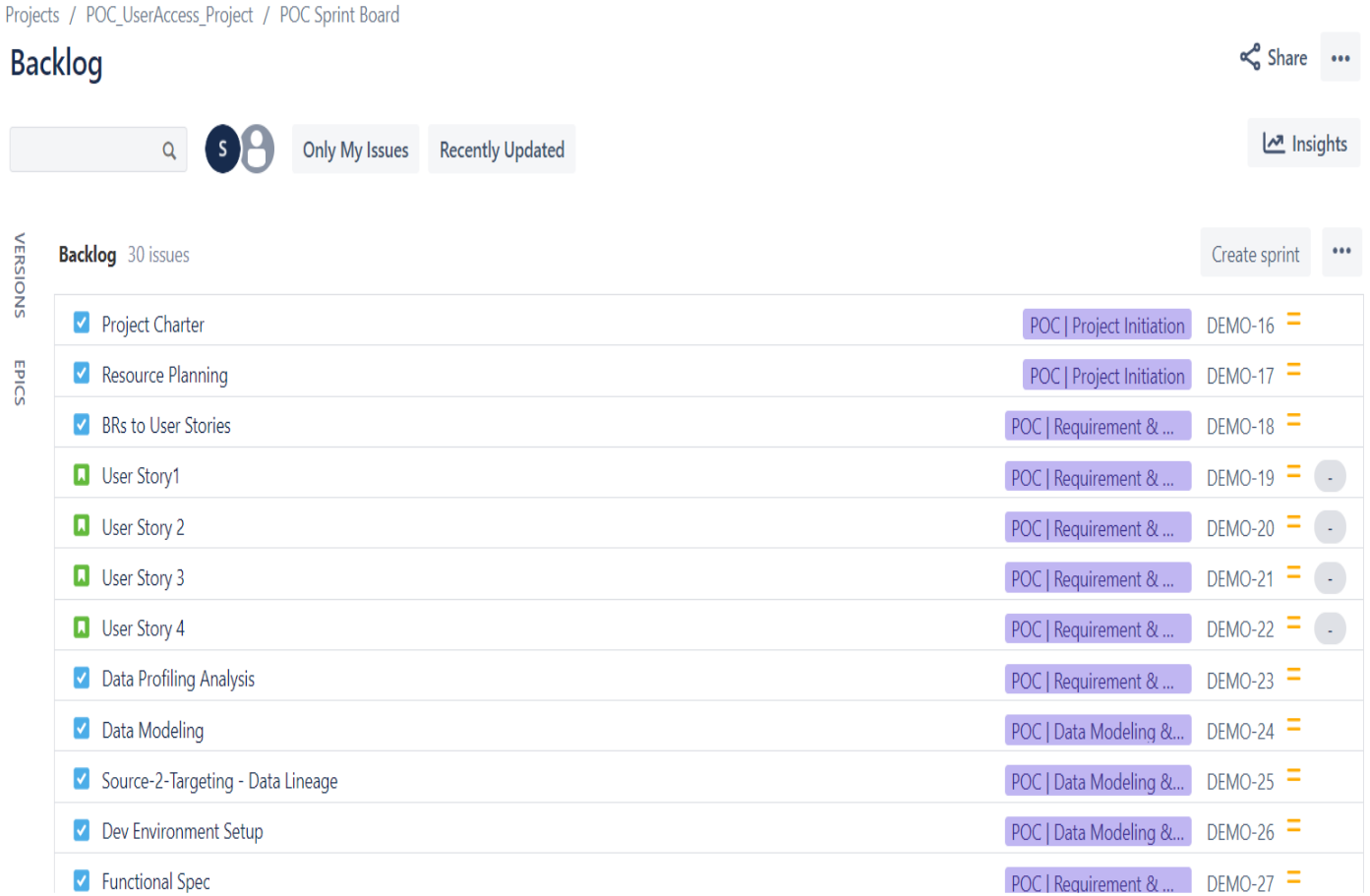




Project Management : SCRUM Board



Project Management : Backlog & Resource



Project Cost Estimates, Team and Timelines

Strategic Solution	• $25k for Design, Build, Test and Deploy (One time) • $10k per annum (for Maintenance support) • Project timelines are approx. 6-8 months • Automation of continuous analytics and process digitalization • **Project team:** Integration/ETL specialist, Data Engineer, Data Architect, Business Analyst, Power BI Developer, QA and Project Manager.
Tactical Solution	• $15k for Design, Build, Test and Deploy (One time) • $5k per month (for Maintenance support) • Project timelines are approx. 2-3 months • Resources are required to run data pipelines manually. Prone to errors. • **Project team:** Data Engineer, Data Architect, Business Analyst, Power BI Developer.

My Data Science Skill Summary

The following highlighted skills are used in this POC:

Data Science: Skills

- **Methodologies/Framework:** Kimball, Inmon, Zachman Framework, TOGAF-V9.2, FEA, Agile(SCRUM)
- **Data Modeling:** Relational, Dimensional, Data Vault Modeling, MDM Hub, XML schema, OLAP and OLTP
- **Life Cycle Expertise** Requirements Analysis, BPR, Design, Coding, Testing, Go-Live cut-off and Post Go-Live activities.
- **Management:** Lead Enterprise Data Architect/Data Modeler, Solution Manager and Project Lead
- **Visualizations:** Power BI, Qlik Sense, Tableau.
- **Analytics:** SAP Bank Analyzer, Notebooks (Jupiter, Databricks Google Colab), DataMeer
- **Machine Learning (ML) Lib:** Scikit-learn, SciPy, XGBoost, CatBoost, Numpy, Pandas, NLTK, Beautiful Soup, LXML
- **Languages:** SQL, Python, PySpark, ACL,ABAP
- **Database/DW:** Oracle, MySQL, SAP-BW, MongoDB, Hadoop (Hive,HDFS)
- **Azure Cloud Services:** IaaS PaaS, SQL Azure, Active Directory, Data Lake, Blob storages.
- **Collaboration/Other tools:** JIRA, Confluence, SharePoint, MS Teams, GitBash, GitHub, Erwin, Archimate, PyCharm

QUESTIONS

APPENDICES

Enterprise Architecture Methodologies Considered: TOGAF's ADM and ITIL

Designing and deploying a centralized enterprise data repository (digital asset) is a long-term commitment. It should be built step by step over period by targeting small deliverables (business values) in multiple iterations. This requires a continuous improvement at a strategic level as well as at the service level to align with the business needs and processes. Hence, the following methodologies are considered.

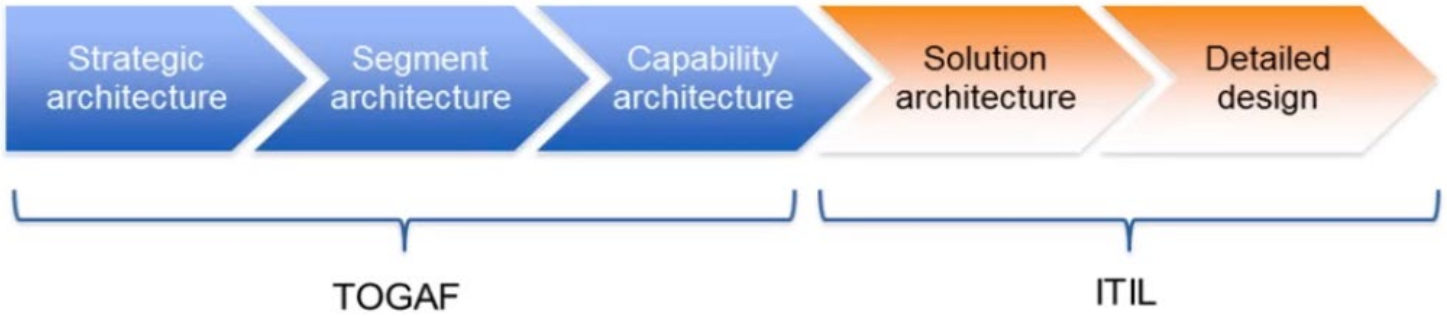
1. TOGAF's ADM

The Architecture Development Method (ADM) is applied to develop an enterprise data lake architecture which will meet the business and information technology needs of the organization. It describes a method for developing and managing the lifecycle of an enterprise architecture and forms the core of TOGAF. It may be tailored to the organization's needs and is then employed to manage the execution of architecture planning activities.

TOGAF's ADM offers the flexibility – to iterate around all eight phases; or select only the relevant phases to iterate; or iterate internally within a single phase, so that target architecture can be aligned to the change in business strategy.

2. ITIL Service Processes

ITIL processes around five service lifecycle stages (:Service Strategy, Service Design, Service Transition, Service Operation, and Continual Service Improvement) can be leveraged in designing the enterprise data lake's IT solution architecture and service model to serve the business needs in more efficient manner. Each of the five stages is focused on a specific phase of the service lifecycle in building and managing the enterprise data lake. Thereby, it helps in providing excellent service support and post Go-Live operations for an enterprise data lake.



Enterprise Architecture Principles

Designing and deploying a centralized enterprise data repository/lake (digital asset) is a long-term commitment. It should be built step by step over period by targeting small deliverables (business values) in multiple iterations. The following enterprise architecture (EA) principles will help to streamline and reduce complexity of IT investment decisions. They are inter-related and must be considered as a set. These EA principles will bridge the gap between high-level strategic intent and concrete designs.

Principles Category	Enterprise Architecture (EA) Principles: Description
Plan	GP1| Reuse before buy, before build: Business applications, system components and data will be reused wherever possible, purchased as commodity solutions if necessary and only built if there is a unique requirement that cannot otherwise be fulfilled.
	GP1| Seek architecture approval: All projects and programmes will be subject to architectural approval at key stages throughout the delivery lifecycle.
Information	IP1| Manage data as asset: Data will be managed to ensure its accuracy and quality to support informed business decisions. All projects that handle a significant amount of data must have a data management plan and information asset register to describe the data used and the security classification of the data.
	IP2| Ensured data is mastered and shared: Standard data models, data elements, and other metadata that defines the data needs to be developed and maintained.
	IP3| Make data is accessible for all decision-making processes: Data must be easy to find and retrieve and present a single version of the truth.

Enterprise Architecture Principles

Application	AP1| Use existing services: Make use of existing Application Program Interfaces (APIs). Don't build redundant services from scratch.
	AP2| Build services not applications: Applications will be built as a collection of services that expose an Application Program Interface (API) enabling them to be combined to deliver users what they need.
	AP3| User interface should be browser based: Ensures that applications are independent of underlying platforms and are easy to use. Deliver them as web-based HTTP applications using using HTML5, CSS, Python and JavaScript.
	AP4| Design apps to meet user needs: Understanding user needs and what functions and features are regularly used by them in existing services helps to prioritize and plan the delivery of new services and make changes to existing ones.
	AP5| Pilot new apps and services: Build a prototype, test it with users and learn from it.
Technology	TP1| Use Open Standard and Open Source after security assessment: Open standards must be used in all solution designs to enable interoperability. Open source software must be compared and considered alongside commercial software when selecting technology solutions. It should be considered after all security assessments.
	TP2| Exploit Cloud Technologies: Solutions will be assessed to determine the benefits of migration to Cloud-based service models (IaaS, PaaS, SaaS) for infrastructure, platform and software.
	TP3| Manage technical debt and obsolescence: Any tactical decisions that introduce technical debt (quick but messy solutions) will only be endorsed if there is a recognized actionable plan to address both of them technically and financially.

Enterprise Architecture Recommended Solution:

Designing and developing an enterprise digital data asset (like enterprise data lake) is a long-term commitment with small deliverables in multiple iterations. Along with TOGAF's ADM methodology, ITIL services should be considered in designing, deploying and maintaining the "Enterprise Data Lake Analytics". Agile's SCRUM methodology is widely used for developing and delivering technology-based projects in an iterative manner.

Solution Phases (in sequence)	Phase description	Tools
Strategy & Motivation	Customize the TOGAF's ADM methodology in defining the scope of the architecture development initiative, identifying the key stakeholders, creating the Architecture Vision, and obtaining approval to proceed with the architecture development of "Enterprise Data Lake". Outcome: Architectural Vision, Stakeholder Matrix	Familiarize in using tools like agile project management tools (JIRA), collaboration documentation tools (SharePoint, Confluence, Teams) and ArchiMate
Business Requirements	Using Architectural vision, document a detailed baseline and target business architecture and perform a full analysis of the gaps between them. Use techniques like Business capabilities catalog to identify, categorize, and decompose the business capabilities. Value Stream/Capability matrix (displays the capabilities required to support each stage of a value stream). A knowledge of the Business Architecture is a prerequisite for architecture work in any other domain (Data, Application, Technology) while creating enterprise digital asset like "Data Lake". Agile's SCRUM enables business modularity, strategic agility through customized and reusable business modules. Factors like – Stakeholder Experience, risk, agility, architecture consistency and business achievability will impact the quality of business architecture. At same time, consider some inputs from ITIL service design processes include service level agreement, service catalog management, availability management and IT service continuity management. Outcome: Business Architecture meta-models, BRD, Business Model Canvas	ArchiMate (Architecture models and viewpoints) Confluence(to document BRD, FRD, process diagrams etc). SharePoint(for Artifacts)

Enterprise Architecture Recommended Solution:

Design Application/Data & Technical Layer	Before designing the blueprint of the technical architecture, consider the following points by using ADM: 1. Develop baseline data-architecture description 2. Review and validate principles, reference models, viewpoints, and tools 3. Create architecture models, including logical data models, data-management process models, and relationship models that map business functions to CRUD (Create, Read, Update, Delete) data operations 4. Select data-architecture building blocks 5. Conduct formal checkpoint reviews of the architecture model and building blocks with stakeholders 6. Review qualitative criteria (for example, performance, reliability, security, integrity) 7. Complete data architecture 8. Conduct checkpoint/impact analysis 9. Perform gap analysis Analyze cost benefits and risk; develop prioritized list of projects to form basis of implementation and migration plan. Finally, remember the importance of unity. Evaluate how to unify your data architecture. Siloed data warehouses and data lakes increase complexity, including complexity in your data integration strategy and tactics. Outcome: Data and Technical Architectures, Application Architecture	ArchiMate (Architecture models and viewpoints) ITSM (for ITIL service processes) Erwin (Data Modeling) Azure Data Lake
Build & Deploy	Using ADM, work break-down structure can be executed in designing and deploying the architectural building blocks (ABB) of Enterprise Data Lake. Ensure that the business value and cost of work packages and Transition Architectures is understood by key stakeholders. Using ITIL's Service Transition, plan and manage the capacity and resources required to package, build, test, and deploy a release in to production and establish the service specified in the business stakeholder requirements. Deliverables should be deployed using In a Continuous Integration / Continuous Delivery (CI/CD) approach. Evolve higher level architectures based on experience of each sprint (capability increment). Monitor changing requirements, but maintain alignment to strategy and linkage to the "big picture". Outcome: Business Architecture, BRD	Azure Data Lake ArchiMate (Architecture models and viewpoints) ITSM (for ITIL service processes) Jira (Agile Scrum user stories & tasks)
Post GoLive Operations	ADM defines the life cycle of the enterprise architecture. Changes to enterprise architecture after project deployment are addressed in the next phase. Using ITIL's Service Operation, ensures that Data management services are being provided efficiently and effectively as per SLAs. It includes monitoring services, resolving incidents, fulfilling data requests and carrying out operational tasks. Actual delivery and consumption of services happen during this stage. Outcome: BAU documentation	Azure Data Lake ITSM (for ITIL service processes) Jira (Defects)
Change Management	Using ADM, modify the architectural change-management process with any new artifacts created in this last iteration and with new information that becomes available. Define metrics and performance reviews using ITIL's Continual Service Improvement. Outcome: Change Requests, Statement of Architecture Work.	ArchiMate (Architecture models and viewpoints) Jira (Change Requests)

References:

Topic	Link
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/
TOGAF	https://pubs.opengroup.org/architecture/togaf92-doc/arch/
COVID-19 impact on business strategy	https://www.ibm.com/thought-leadership/institute-business-value/report/covid-19-future-business
EA integration pattern	https://www.safe.com/blog/2020/10/enterprise-integration-patterns-relevant-organization/

Thank You