





Build Your Modern Data Landscape with our Data Lakehouse Solution on Microsoft Fabric

Our MVP Solution Offering

Powered by Microsoft Fabric

Content

- O1 Challenges & Opportunities | Implementing a Modern Data Platform
- 02 | Our Offering
- 03 Use case Implementation
- 04 | Approach & Timelines
- O5 | Scope, Deliverables & Assumptions
- 06 | Commercials & FAQs

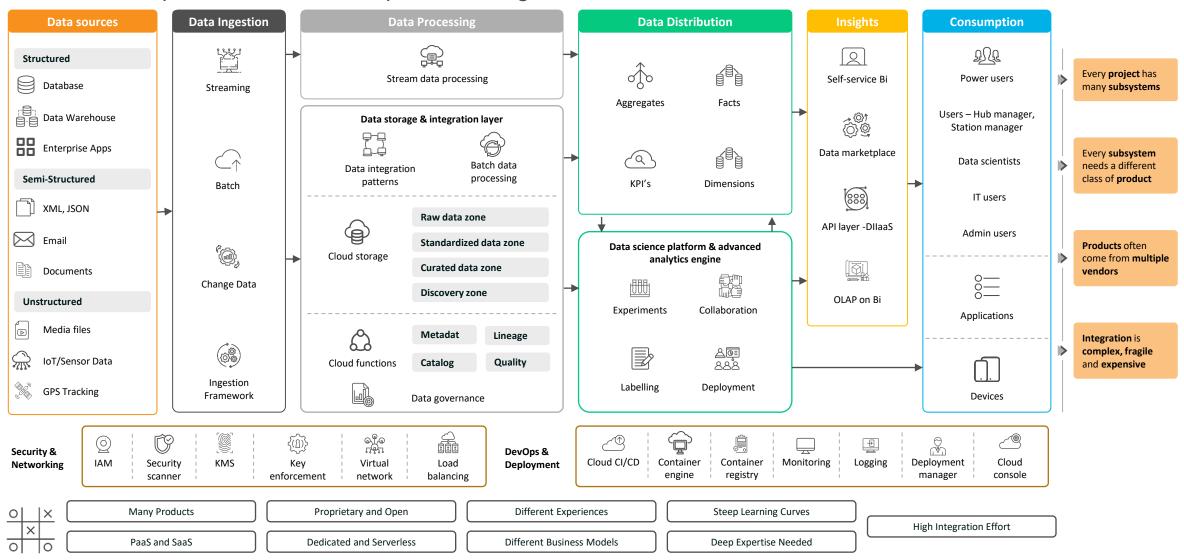


01

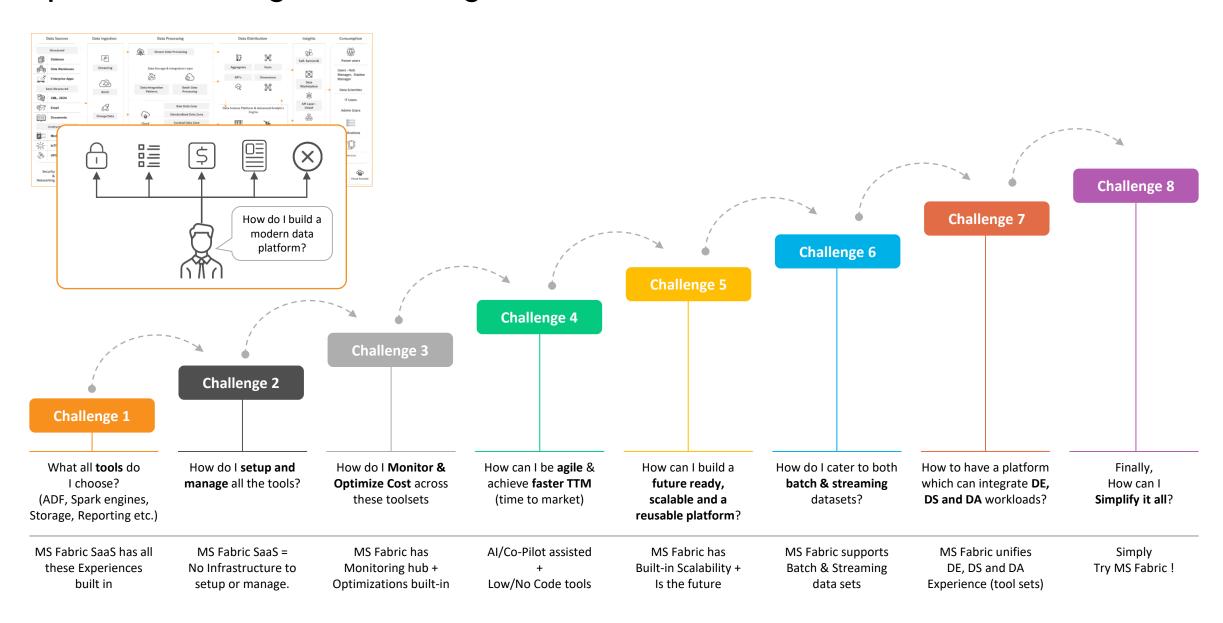
Challenges & Opportunities: Implementing a Modern Data Platform

Modern Data Platform Implementation Challenges

Data and Analytics Platform is Complex and Fragmented



Specific Challenges in Building a Modern Data Platform



Microsoft Fabric | The Data Platform for the Era of Al

A unified **SaaS** platform for your Analytical needs..

Data | Storage, Integration, Engineering, Warehousing, Analytics (real time), Science, Visualization all of these in one unified SaaS platform.



OneLake

Unified LakeHouse with an Open Format



Data Integration

Low-Code No-Code support for NoteBooks for Ingestion Jobs

Schedule & Run Jobs without managing Clusters



Data Warehouse

Powered by Delta Parquet Accessible through SQL and Spark Decoupled Storage & Compute Organized around Workspaces



Data Science

Run NoteBooks within seconds without managing Clusters.

Train Models and Experiment between different Models Versions



Real-Time Analytics

Ingest & Query Realtime Data within seconds and Stored Real-Time Data in OneLake



Business Intelligence

Reports &
Dashboard with
Power BI



Unified Security

Across all types of Data Stores & Engines



Shortcut

Enables Poly-Cloud with easy Access to Data Sored in Other Clouds

Features



Simplicity

Less complex. Just simple

Microsoft

Fabric

One Lake accessible in win-explorer

No install.

Just Internet & Web browser needed.



Time & Effort

Reduced manual effort

Instant provisioning

No infra to manage



Scalable & Optimized

Auto-scalable

Auto optimized + Auto integrated

Optimal use of cloud infrastructure



Unified & Open Standards

Centralized security & monitoring

Based on open standards

Delta-Parquet open standard storage



Business Value

Faster time to market

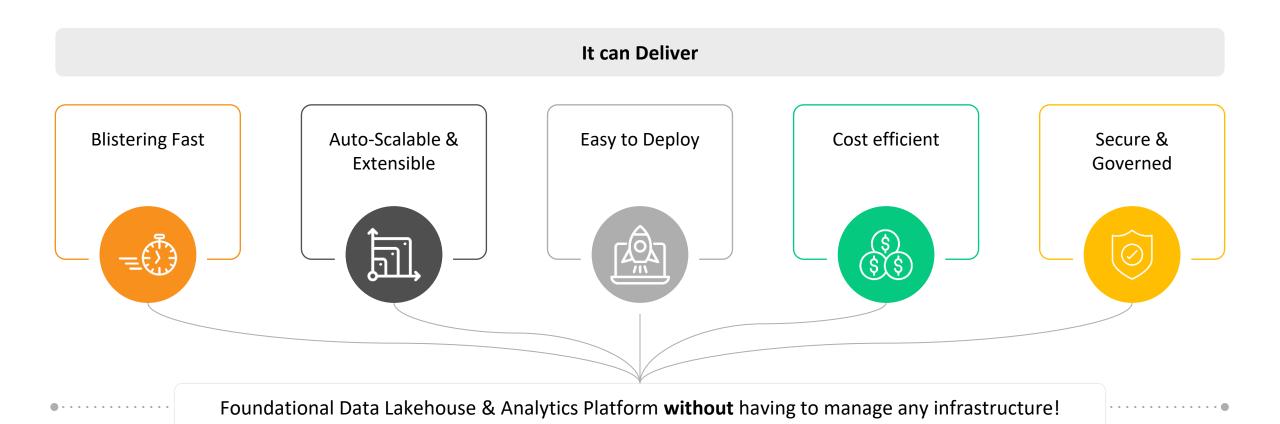
Future-ready & re-usable platform

Integrates with existing Azure components

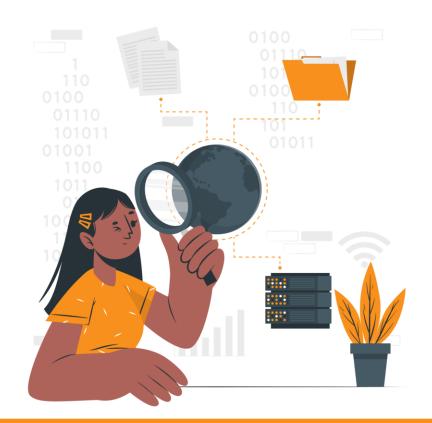
6

Our Offering

Our MVP Solution Offering Powered by Microsoft Fabric



Our Offering



2 Use cases delivered



Deploy a Modern Data Lakehouse & Analytics foundation MVP on Microsoft Fabric in

10 Weeks

Utilizing up to **15** data sets

Overview



- 2 priority use cases identified and mutually agreed upon that deliver high business value but have low complexity. Defined by business and BI-analysts through workshops & assessments
- Produce business outcomes that drive decision making and or drive cost savings
- These use cases cannot be achieved using existing systems and processes or lack better delivery SLA's

15 Data Sets

- Establish data pipelines from up to 15 data sets around the identified use cases
- Setup re-usable, metadata driven framework with ability to scale to future use cases and data sets
- Data available for consumption and to power use cases/data driven decision-making in a Secure manner

50 Working

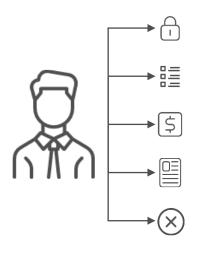


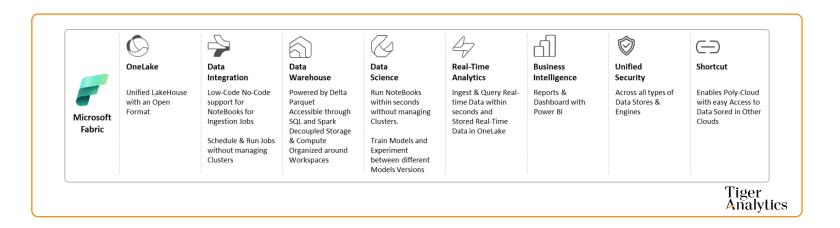
- Required business insights delivered for 2 use cases
- Foundational medallion Data Lakehouse – ready to use and future-ready for implementing use cases at scale
- 2 weeks of Discovery, 6 weeks for Implementation and 2 weeks for UAT (in the same MVP environment)

Member

- Proj. Manager 1
- BA/Functional QA 1
- Senior Arch 1
- Senior DE 1
- DE 3
- Senior Power BI Dev. 3

Value Addition | Our Offering





Challenge 1

What all **tools** do I choose? (ADF, Spark engines, Storage, Reporting etc.)

Challenge 2

How do I setup and manage all the tools?

Challenge 3

How do I Monitor

& Optimize Cost

across these

toolsets

Challenge 4

How can I be agile & achieve faster TTM (time to market)

Challenge 5

How can I build a future ready, scalable and a reusable platform?

Challenge 6

How do I cater to both batch & streaming datasets?

Challenge 7

How do have a platform which can integrate **DE**, **DS and DA** workloads?

Challenge 8

Finally, How can I Simplify it all?

Tiger's MS Fabric CoE team brings expertise

Know-how to Implement best fit MS Fabric options as per need Bring MS Fabric best practices for building a modern data lakehouse Have re-usable Ingestion and Transformation Frameworks

Will continue to evolve the Modern Data Platform post MVP. Tiger team brings experience on MS Fabric Batch & Streaming options

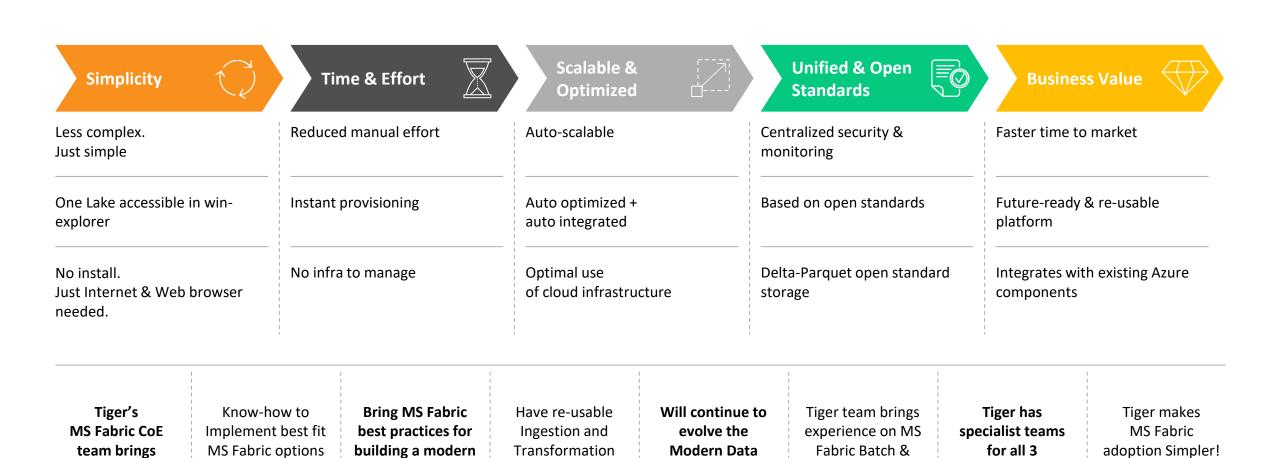
Tiger has specialist teams for all 3 (DE, DS & DA) Tiger makes MS Fabric adoption Simpler!

Key Benefits | Our Offering

as per need

data lakehouse

expertise



Platform post

MVP

Streaming options

Frameworks

(DE, DS & DA)

Where can this be Used



SMC (Greenfield Implementation)

- Data Foundation
- Organization with leaner Prod support
- Self Service enablement

- Reusable platform scale to future needs
- Data Democratization
- Simple and Cost-effective solution



Enterprise (Existing Azure Platform)

- Simplify complexities in existing platform
- Reduce Time to market
- No infra to manage
- Existing team can be re-purposed

- Improved agility
- Optimal Resource utilization
- Self Service enablement



On-prem (Migration) – upcoming

- Move on-prem data using Data Gateway
- Migration toolkit is work in progress.

03 Use Case Implementation

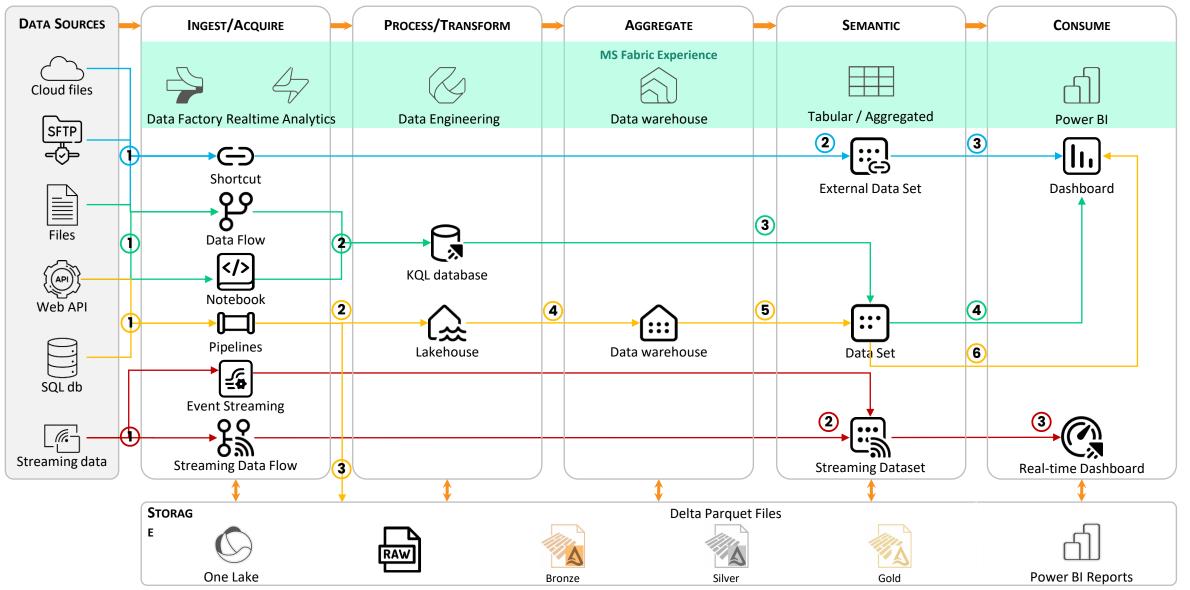
Reference Architecture

Data Analytics via Shortcut

Data Engineering & analytics using KQL

Data Engineering & analytics using Lakehouse

Streaming analytics



Reusable Components Made Available in Our Offering

Pipelines & Data Ingestion f/w

Framework for ingesting data from multiple sources.

Data Quality f/w

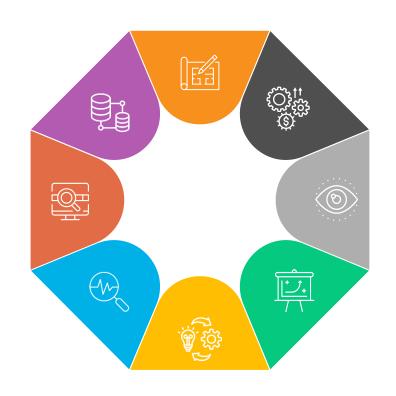
Reusable light and metadata driven framework to implement Data Quality checks.

Audit Log Monitoring

Functionality that ingests audit log information into metadata tables to augment auditing and troubleshooting.

Architecture Blueprints

Architecture blueprints available for different use cases



Our knowledge & experience

Our knowledge and experience on having worked on several Data Engineering engagements can be leveraged.

Frameworks for Transformations

Reusable transformation f/w & blueprints using different Microsoft Fabric Experience tool sets.

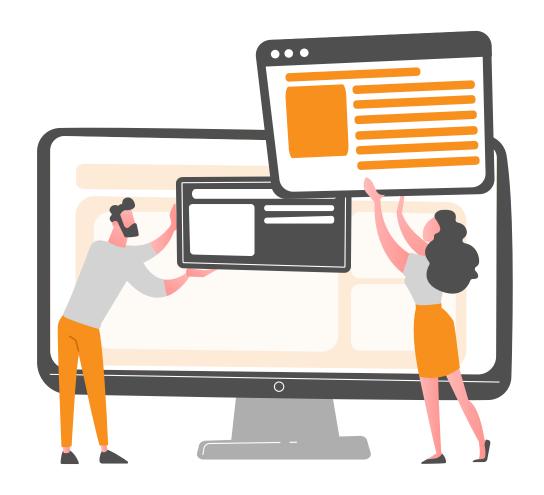
Our PoV

Our Point of view on the available features and functionalities on Microsoft Fabric with workarounds where required.

Demo-able Use Case

A demoable use case solution from a Banking domain including pipelines, data quality, transformations and aggregations with Power BI dashboards.

Demo



Risk Management Solution for Banking Segment



Objective

Implement a robust risk management solution tailored for banking sector, to provide 360° coverage of banking fraud, credit default and liquidity risks.



Key Challenges addressed

- Analyzing the trend & pattern of fraudulent transaction w.r.t originating country.
- Analyzing liquidity risks in relation to operational expense and credit lines.
- Understanding spread of credit risk across key dimensions: business segments, customer demographics etc.



Dataset Considered

- · Fraud analytics Data
- Customer Details
- General Ledger Data
- Other Master data
- Firewall network logs
- Account statement data



Business Value



Credit Risk Mgmt.

Better credit risk management w.r.t Credit to Repayment ratio.



Risk Mitigation

Advance risk Identification & mitigation to minimize impact on business operations and stability



Forecast Liquidity

Discovery of data trends & to make timely forecasts to cover liquidity risks



Compliance

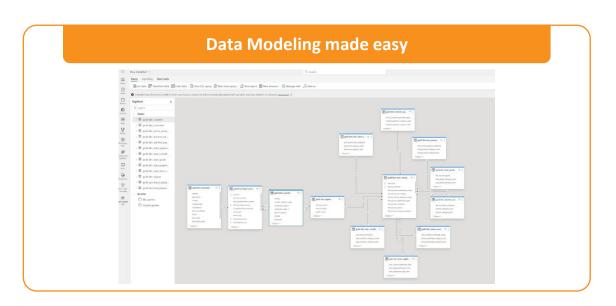
Compliance with industry regulations and standards.

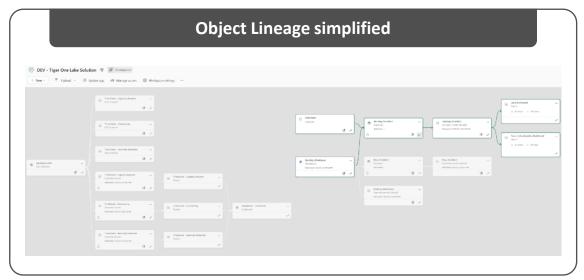


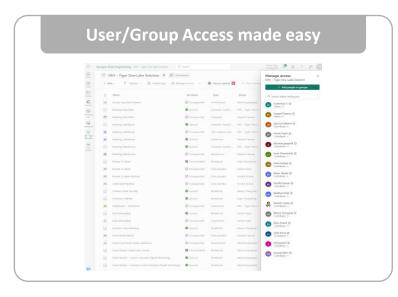
Access Governance

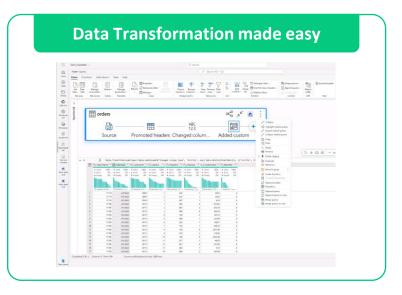
Enabling data democratization & a culture of 'responsibility & accountability' by governing access to risk related information

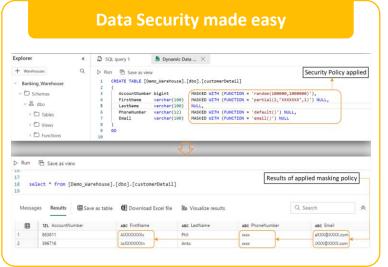
Microsoft Fabric Features used in Our Offering (sample screen shots)











Microsoft Fabric Features Used in Our Offering



Monitoring Monitoring

- Monitoring of pipelines, dataflows and notebooks through central **Monitoring Hub**
- Data action alerts generated through **Data Activator**
- Pipeline email alerts through Outlook activity in Data Factory









Real time integration

- Real time web traffic data ingestion with **Event Stream**
- **Spark streaming** used for time window aggregation



Time Travel

- Time Travel in Delta lake feature is helpful for performing data rollback for any load failure scenarios
- Lakehouse tables: Can be restored as of timestamp
- Warehouse tables: Can be restored using **clone** tables as of a particular version







- Usage of Lakehouse **Shortcuts** to make remote data accessible
- Medallion Arch implemented through lavers in Fabric Lakehouse & Warehouse



- Using Smart Al heuristics **Dataflow Gen2** gives capability to generate extraction code with provided data patterns
- **Unstructured** text data in MT940 messages extracted in table format using Dataflow Gen2













Microsoft Fabric Features Used in Our Offering (Cont'd)



Time series analysis

- KQL Databases provide the features to create time series on GL data for seasonality analysis
- In-built models help in forecasting and anomaly detection



Data Privacy

- Sensitive Customer data has been protected using Rowlevel and Column-level security in Warehouse
- Dynamic data masking applied to obfuscate sensitive data



Data Discovery & Lineage

- Data assets such as Lakehouse, PowerBI dataset generated have been endorsed and made searchable
- Lineage helps provide visibility on dependencies between pipeline, notebooks and Lakehouse
- Granular data lineage and data catalog can be further realized with Purview integration







Cost **Observability**

- Microsoft Fabric **Capacity Metrics App** built as a PowerBI app can be installed to track capacity utilization
- Capacity Unit (CU) % utilization can be tracked across different Fabric SKU capacities, workspaces, items and users
- OneLake Storage costs can be similarly tracked too







Semantic Link

- With help of **Semantic Link** feature, we can do dependency analysis between columns in a dataset
- This has been utilized while identifying DQ rules and their applicability as well as sensitive data identification

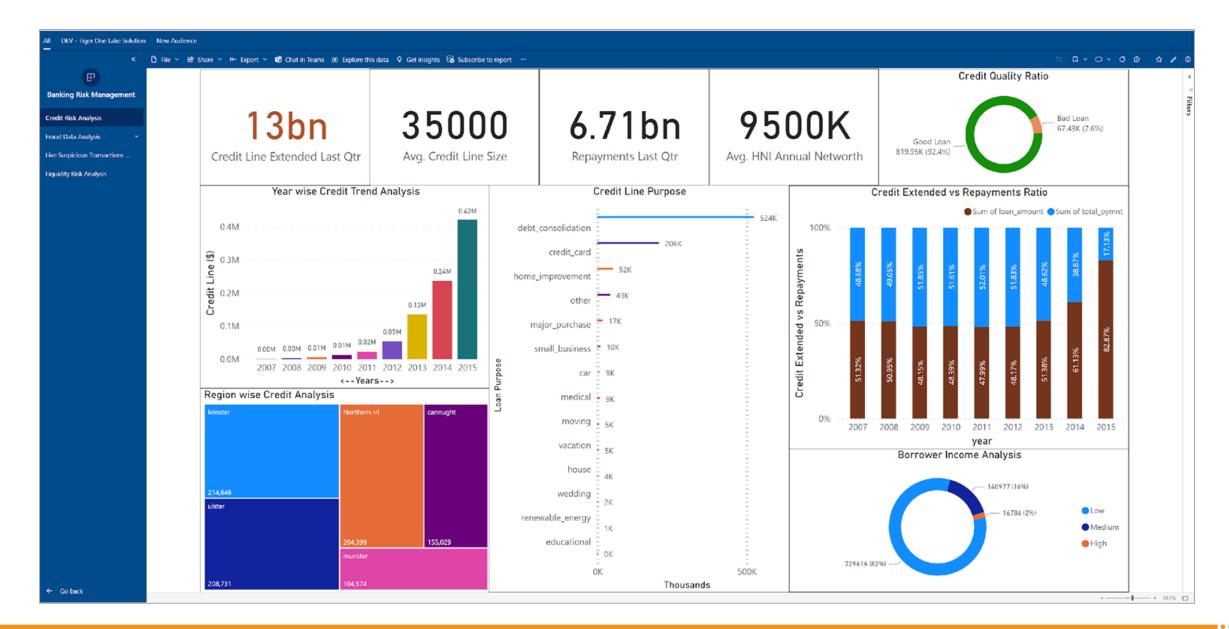




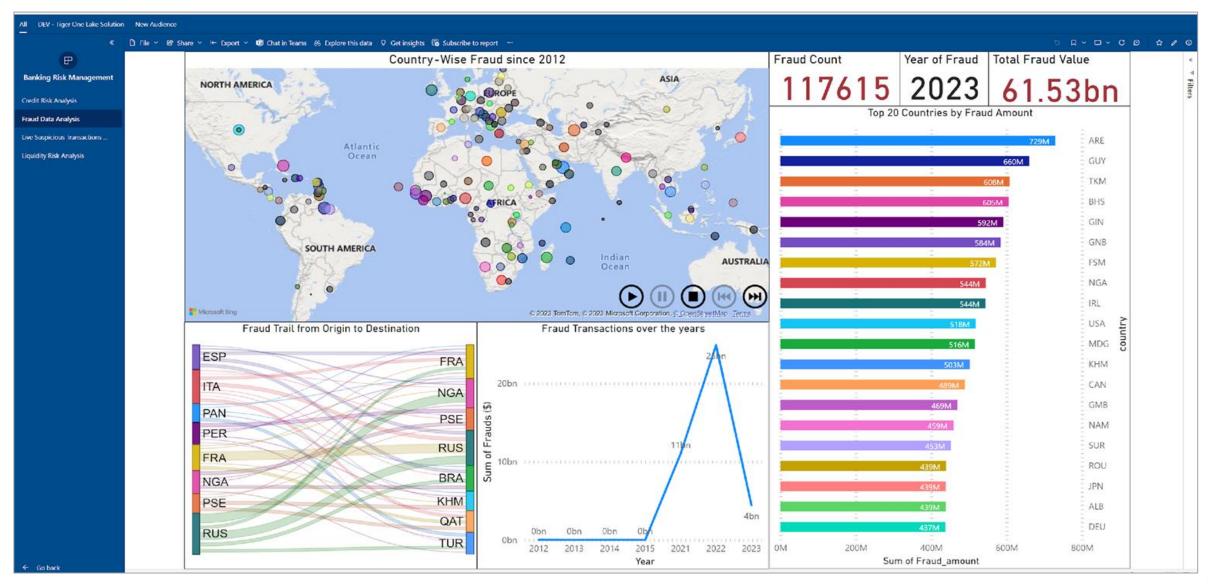




Demo Use case | Sample Reports

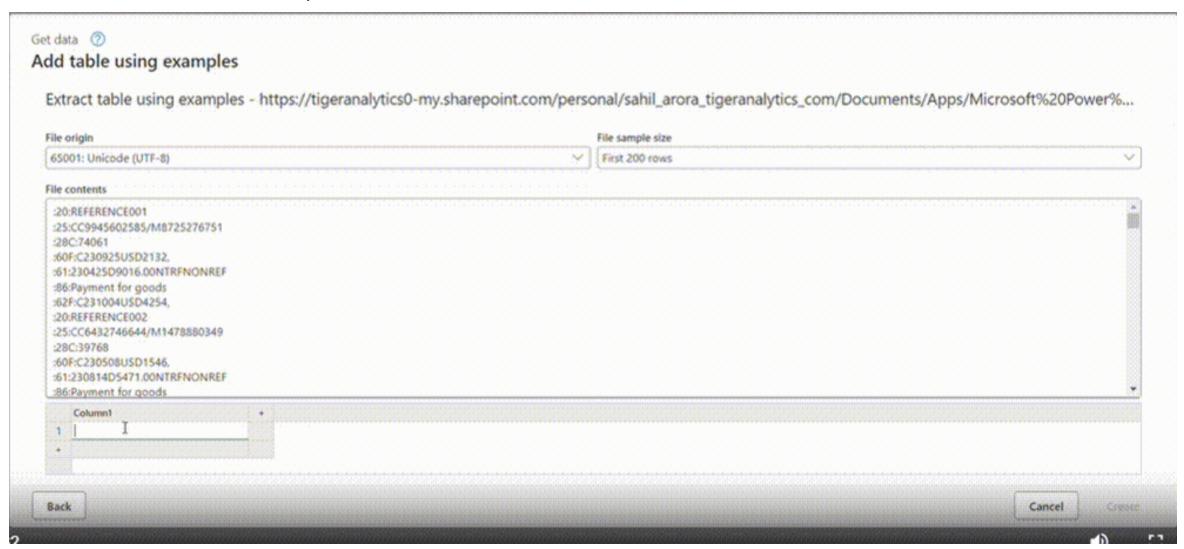


Demo Use case | Sample Reports



Microsoft Fabric | AI based Code Generation

Code Generation made easy



04 Approach and Timelines

Solution Approach

Data Sources

On-prem databases

Web-API

Streaming data

Enterprise data

3rd party data sources

Structured/ semi-structured

Build Solution

Framework development

- Metadata & config. driven parameterized pipelines for automated onboarding of data sources
- Create folders in data lake zones [Landing, Bronze, Silver (valid, invalid)] & Gold (in warehouse)

STTM

- Populate SQLmetadata with details of source metadata.
- Create source-totarget map.
- Create aggregated data model for consumption (gold)

Usecase Development

- Utilize common framework to develop usecase based ingestion & transformation pipelines.
- Create pipelines to apply data quality validations.
- Start building BI dashboards in parallel.

Observability & Monitoring

- Establish a framework to monitor & audit data pipelines
- Enable applications to capture logs to improve operational visibility.
- Implement alerts and notifications.

Testing + go-Live

- Integrate and test the overall solution.
- Onboard foundational data ready across various use cases for downstream consumption
- Productionize the solution post bugfixing.

Results

Consume BI dashboards for Analytics

Optimal use of cloud infrastructure

Faster time to market

Reduced manual effort

Future-ready platform

Project Lifecycle

Use case selection & Architecture alignment

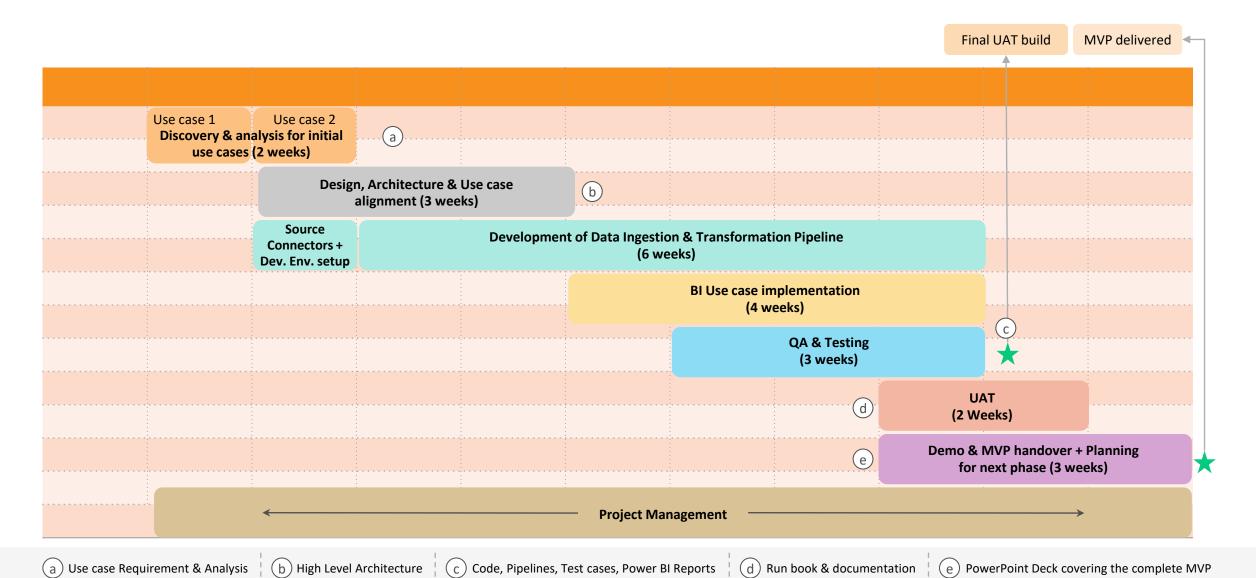
Platform access + Dev. Env. setup

Development (DE & BI)

UAT + Production go-live

35

50 days (10 Weeks) | Timelines & Key Phases



05

Scope, Deliverables and Assumptions

Scope

Use cases

- The use cases would be discussed, finalized and detailed out during the initial 2 weeks
- Details of the use case would include source and entities required for the use cases, relevant business transformation requirements and wireframes of the reports with mapping of the entities required to generate the report
- The use cases being considered would be high impact in terms of business value but would be relatively mid-low in complexity
- ML use cases would not be covered in these

Foundational Platform

- A foundational data platform based on medallion data Lakehouse architecture would be implemented
- The Landing Zone, Bronze Zone and Silver Zone shall be created, and these shall utilize the Lakehouse feature of MS Fabric
- The Gold layer shall be created using the Synapse Data Warehouse feature of MS Fabric
- Other MS Fabric experience tool sets such as Synapse Data Engineering (Spark notebook), Data Factory (orchestration) and PowerBI (BI & Analytics) shall also be utilized in building the foundational data platform

Data & Storage

- Up to 15 (tables from sources) would be considered as part of the scope
- Only structured data shall be considered.
 Any unstructured data or Streaming data would not be in scope for building the foundational data platform
- The data shall be ingested in batch manner. Real-time or near real-time/ streaming data sets shall not be in scope for the foundational data platform
- All data shall be stored in One Lake.
 Medallion Data LakeHouse would store data in Delta Parquet format
- Landing Zone shall be used for acquiring data from source in native format, Bronze Zone for cleansing and standardizing the data and Silver Zone for business data transformations
- Other operational zones or folders such as In-Valid, Valid, Archive etc. as required shall also be created within One Lake

Transformations

- Tiger team shall develop spark transformations for data load, and these shall utilize Synapse Spark clusters within MS Fabric
- Jupyter notebooks shall be utilized to run the code and the same shall be committed to Git associated to the MS Fabric workspace
- Python, Spark-SQL or T-SQL code would be used for implementing the Ingestion & Transformations code
- The data pipelines shall be based on the concept of ELT (Extract-Load Transform) and for loading data into the Bronze Zone, two types of merge strategies would be supported – Truncate & Reload and Append-Insert only









Scope

Monitoring & Ops

- All monitoring and operations of all the components of MS Fabric shall be through the Monitoring Hub station available for MS Fabric
- As of now, Monitoring Hub is available for Power BI, Data Factory, Data Engineering and Data Science

SLA for latency

- There are no hard SLAs for any ingestion load to run within a fixed duration, or User query response time
- The overall performance of the various operations shall depend on the MS Fabric capacity selected
- Tiger shall develop the pipelines by following the available best practices from Microsoft for MS Fabric

Data Ingestion

- Tiger shall develop batch data ingestion pipelines using the Data Factory feature set available in MS Fabric
- These pipelines shall copy data from source to Landing Zone and further into the medallion data lake zones (stored in One Lake)
- Once the ingestion pipelines are completed, data shall be available either for further processing or shall be available for consumption

Metadata framework

- Tiger team shall develop frameworks for ingestion, transformations and data quality in a metadata driven manner
- This shall ensure that future data sources can be ingested into the platform in an easy and reusable manner thereby enabling efficiency
- The metadata for the same shall be stored in a separate Azure SQL database provisioned separately
- Python as programming language would be considered for building any Data engineering workload
- Great Expectations (open-source python) library would be considered for Data Quality & Testing workloads









Scope

Governance

- Implementing custom governance capabilities such as auto data catalog, catalog classification rules etc. is out of scope
- Only the out of box data lineage feature provided by MS Fabric shall be available. This is a built-in feature
- Certain features such as data governance, compliance, sensitive data protection are supported through external service integration viz. Purview, Azure Information Protection etc. These are considered out of scope



Security

- Custom data security such as row or column level access policies are considered out of scope
- Only authenticated users shall be able to access the data in One Lake, data lake, data warehouse or Power BI dashboards
- Security model setup viz.
 360 AD group setup, RBAC permission assignment is considered out of scope

Git & on-prem Integration

- Git integration is supported in limited mode and enabled only through Azure DevOps. Cost and provisioning for DevOps is extra and would be charged separately
- Presently, MS Fabric doesn't support direct onprem private connectivity. The setup and configuration of the same is considered out of scope
- Any such feature which is not fully supported by Microsoft shall be considered out of scope



BI Dashboard

- Tiger shall develop upto 6 Dashboards (having 4-5 visualizations each) for the 2 identified use cases based on the 15 ingested datasets
- The dashboards shall not implement any live (streaming) datasets or reports
- Complex or ML based dashboards are considered out of scope
- Tiger team shall develop the Power BI dashboards after discussions and mutual agreements with the client



Modeling/ Cleansing

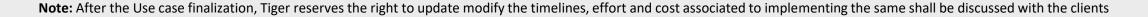
- Data modeling (Star schema, data vault, normalization etc.) is considered out of scope
- Any excessive data cleansing or deduplication work is considered out of scope for the initial use cases

Fabric SKU

- MS Fabric SKU and workspace provisioning requirements must be discussed in advance. Tenant setup, capacity purchase and cost associated is not covered in scope
- Any MS Fabric Cost Monitoring will be done by Customer's IT team







Deliverables in 50 days

Tiger shall provide the following 5 key deliverables as part of the overall implementation:

Use case & Requirement Analysis document

Document detailing out the 3 prioritized use cases and requirements in detail which would also be the scope for implementation

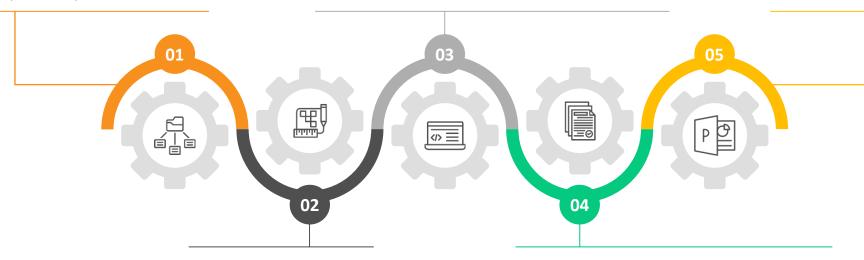
All code, artifacts and Power BI reports

All the code/templates/test cases etc. developed as part of implementation of the use cases including pipelines, frameworks, metadata,

Power BI reports

PowerPoint Deck

Supporting deck for final use case demonstrations



High Level Architecture

High level architecture diagram with relevant Microsoft Fabric Experience and featured tools.

Run books & documentation

Application run-book with key configurations/parameters, procedure to onboard new datasets (in future) and troubleshoot commonly faced issues while development

Client Responsibilities





MS Fabric license and capacity on MS Azure

The client would have acquired Microsoft Fabric licenses and capacity to support all 3 environments (Dev, QA and Prod). Tiger team shall be provided with required access on Microsoft Fabric in the development environment to perform the required Use case development



Source Data

Data sources will be made accessible within the 1st week of the start of the engagement in order to perform necessary exploratory analysis.

02



Source System Owners

Source system owners shall be available for discussions and provide data or make it accessible for the Tiger team to access and ingest in the required schedule, frequency and format.

Power BI licenses

The client shall provide Tiger team the right licenses/resources to develop the required Power BI reports and dashboards.

05



Client Project Manager

The client shall designate a Project Manager to act as the point person for all matters for the Tiger Team. The client Project manager shall also be the authorized signatory for all the deliverables.



Stakeholders

Relevant client IT and Business stakeholders and point-of-contacts shall be identified and communicated to the Tiger team. These stakeholders should make themselves available for discussions and working sessions for which meeting invites shall be sent out at least 1 business day prior



Policies and Existing Documentation

The client shall make available to Tiger team relevant policies, architecture documents and other details of its existing/current system

07

Key Assumptions

MS Fabric licenses and access to use the relevant workspaces for development shall be made available before the project starts. All the licensing and capacity cost on MS Fabric shall be borne by the client.

Timelines outlined are contingent on availability and active involvement of client's resources in the project in the due course and as requested by the project team staff.

Data to be ingested will not require any significant data cleansing by Tiger. Overall objective of the initial phase would be to "enable" the right foundation for bringing in the required datasets to power the identified use cases.

Client will provide licenses for Power BI etc. – for Tiger to perform the necessary development activities. Client will provide all necessary documentation (like Data Dictionary) that Tiger needs to be aware of to perform the required tasks at the start of the engagement.

No special data handling for GDPR compliance is assumed. Sensitive data will be masked, and all PII data shall be anonymized before it is made available in landing zone for the development team.

Any changes to scope may have impact on schedule, effort and cost. The same shall be discussed and agreed upon on mutual terms.

Customer's Point of Contact will ensure availability of their SMEs for review and sign-off.

QA & UAT will be done on the same environment.

Commercials and FAQs

Commercials



 5 Layered Foundational Architecture (Landing -> Bronze -> Silver -> Gold -> Reports)

Indicative

- ELT -->Spark, Gold --> Relational Tables --> Consumption. No Real-time or Near Real Time processing
- 15 Source data table, 4-8 Gold Layer Tables. No Streaming Datasets.
- Transformations between the Silver to Gold layers are simple
- Reusable and metadata driven Frameworks for Ingestion, Transformations and Data Quality.
- Up to 6 Dashboards. No ML Use Case.
- Each dashboard has 4 to 5 sections/Charts
- Tenure: 2.5 months (10 Weeks)
- Est Effort: 8 person months
- Est Price: \$45,000



FAQ's

	?	
1	Can all existing Azure components and functionality be migrated and replace with MS Fabric?	No. It depends. Only few features and key Azure functionalities can be migrated to MS Fabric 'as-is'.
2	How is MS Fabric different from Azure platform?	MS Fabric is a SaaS offering providing integrated data and analytics capabilities without overheads of platform management. The traditional Azure platform is a cloud platform offering several PaaS, laaS services and tools which need to be setup, managed, and administered which are additional overheads.
3	How is MS Fabric different from Azure Databricks?	In short MS Fabric is a SaaS offering while Azure Databricks is a PaaS offering. Azure Databricks has available for a few years while MS Fabric is very new and continues to evolve.
4	How is MS Fabric different from Azure Synapse Analytics warehouse?	MS Fabric offers all the functionalities including Data factory pipelines, Spark pools (serverless), serverless SQL analysis (Lakehouse), and real time ingestion (KQL/Data explorer pools). Synapse dedicated SQL pool has been rearchitected inside MS Fabric as a serverless Warehouse with storage on One-Lake in Delta-parquet format (open source).
5	What happens to existing Azure Synapse after MS Fabric rollout?	Azure Synapse will continue to exist separately as a PaaS solution offering.
6	Is there a migration path available for migration from Synapse to MS Fabric?	Microsoft has created some migration toolkits/knowledge artifacts around this. However, a comprehensive packaged migration solution is not available as of today
7	How can I migrate existing data from on-prem databases/other cloud databases into Fabric?	Some in-built mechanisms such as Mirroring (pvt. preview), data factory pipelines, AZcopy (ADLS storage data migration) can be adopted for this.
8	What use cases are best suited for MS Fabric?	Almost all the modern data platform architectural patterns such as batch data integration/orchestration, Data quality, Medallion Lakehouse, Data mesh, Streaming/Change data capture use cases, Time series analysis, data replication patterns are supported.
9	What is Tiger Analytics' point of view on MS Fabric?	We have a detailed PoV which can be shared on request
10	What are the key advantages of using MS Fabric?	SaaS solution, easy to setup and work with, no infrastructure setup or management, single open storage format repository, serverless, auto-scalable and many more. (refer to slide 36 in Appendix)
11	Can MS Fabric be fully integrated with Git?	MS Fabric is now generally available. While most of the features can be integrated with Git, there are some features like pipelines etc. which are not fully supported currently. This is being worked on by Microsoft and shall be made available soon.
12	How are costs managed and tracked with MS Fabric?	Costs are tracked through usage of Fabric capacity allocated to a workspace. A granular level view of capacity usage can be obtained through MS Cost Management app as well as through MS Fabric Capacity metrics dashboard.
13	Does MS Fabric need a license? Can exiting Power BI license be used?	Yes, a license is needed. Existing PowerBI license can be used if they are equivalent to Premium tier. Specific Fabric capacity tier can be bought separately as well.
14	Have more questions?	Feel free to reach out and we shall be happy to help!

Secure On Prem Connectivity to Microsoft Fabric

On prem connectivity can be secured through On-Prem Data Gateways. Support for Secure user access to Fabric is planned.





About Us

Tiger Analytics is pioneering what AI and analytics can do to solve some of the toughest problems faced by organizations globally. We develop bespoke solutions powered by data and technology for several Fortune 500 companies. We have offices in multiple cities across the US, UK, India, and Singapore, and a substantial remote global workforce.

We have received multiple awards ranging from being recognized as a Leader by Forrester Research to being ranked among the fastest-growing tech companies by Inc. and Financial Times. We consistently feature in prestigious 'Best Analytics Firms' lists.

www.tigeranalytics.com





