Competitive landscape and fitment – Data Management

Data Management

The Modern Data Landscape is a challenging field. With data and technology advances, the complexities of data architecture have radically changed. Patching new components to the dynamically changing architectures needs to be more sustainable and adaptive to future changes. But as with any approach to have effective results, we need a planned layout that will fit the current architectures of organizations. The following approach will guide us in solutions of data management architectures.

Understand Business Problem – No two data architectures are the same. The first responsibility of a solution architect even before doing the solutions of architecture is to understand the business problem at hand.

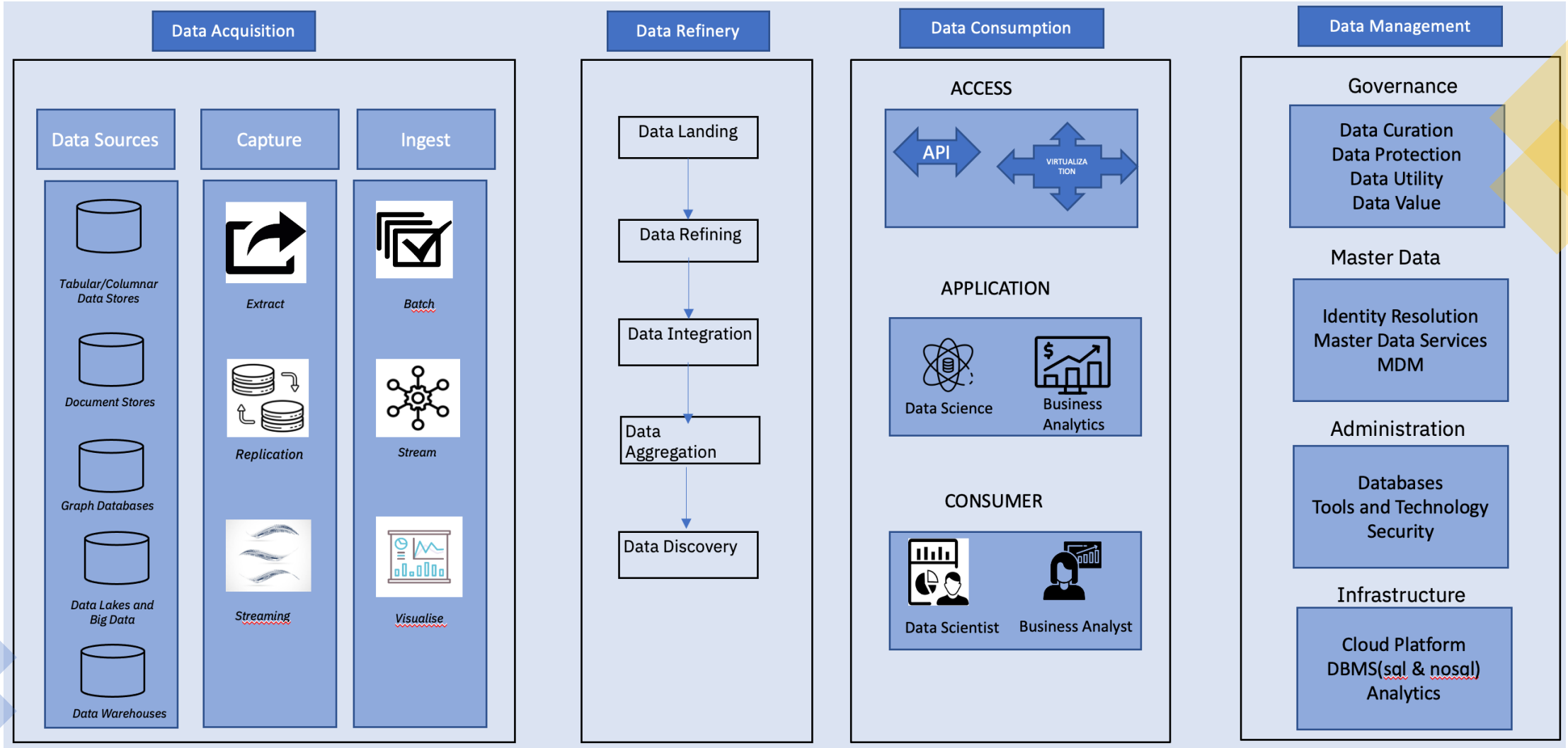
Explore Requirements – Good architectures always meets Business requirements. Having detailed business requirements is part of the architectural definition. The analysis of requirements will validate and refine the list of business capabilities.

Adapt to Reference Architecture – Rethinking architectures to embrace the state of art methods and patch the legacy data management with the ability to integrate the new data management concepts is a must adapt to the complex digital transformation of business.

Having to design or propose a solution to the expressed architecture from the clients can be a daunting task if starting with a blank page. The Better approach is to support and cater to them using a reference architecture with the needed capabilities. The reference architecture will serve as a template that will represent the best practices followed for Data Management architectures.

The architectural model showcases the steps involved in designing data management architectures. Remember, the reference architecture is your template, it will help you kickstart and prepare you for your solution journey. You might and will need to add or remove the capabilities as per your data management architectures and placements of existing components around the periphery.

Modern Data Management Orchestration

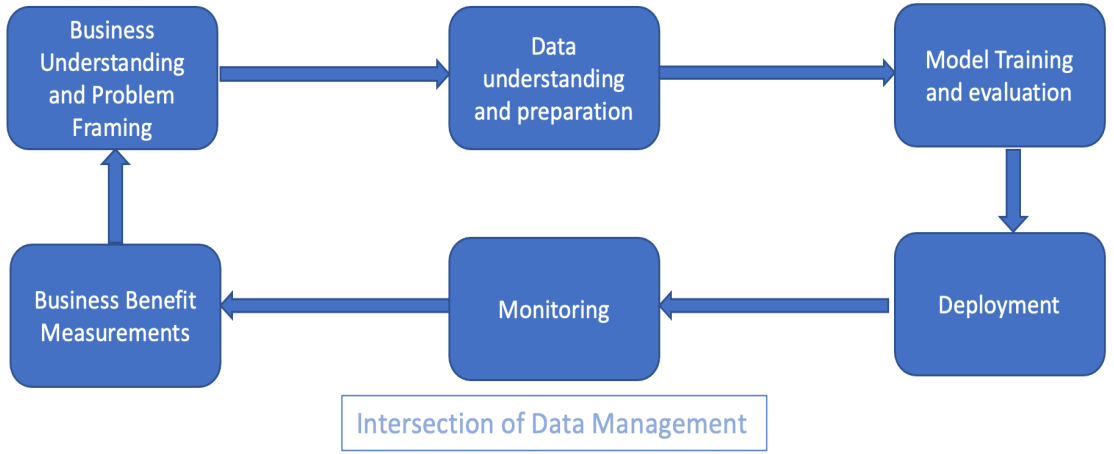


❖ When and Why WKC

- Often, we have our partners or customers sitting on an existing architecture using dominant competitor services.
- We will be primarily focussing on governance fitment in the existing architecture. It is important to understand having governance co-exist in the competitor-heavy space we will need to reach a consensus on how to determine what to determine.
- First and foremost, we will go with the metrics the partners or clients are looking for in the governance space. It should all depend on the features the Business would want to focus on. When we begin with this approach, we will stick to the problem to be solved. Let's understand it via example.
- We have a partner whose architecture setup was on a private cloud. They were looking for data governance with data quality. The contender we were competing against here was Collibra.
- With Collibra as a contender, we knew it would not be a direct hit if we position WKC for governance, as it's a pure governance solution. But the partner was also looking at the data quality feature which is a missing key with collibra even after their recent acquisitions of OwIDQ. Thus, with the competitor's knowledge in hand, we could position WKC which is a more complete data governance including superior data quality.
- Hence, Positioning WKC should rely on pure business requirements of the existing landscape and how we can help clients have more robust, up-market, and enhanced solutions.

Knowing the architectural Landscape

- Any architecture's focus area is to develop a broad understanding of the business domains, workflows, and relevant data.
- In today's digital age, it's imperative to understand the client landscape which often involves multiple clouds or open-source services to stay relevant and adapt to the market changes.
- By now we are well-versed in terms of objection handling which we touched upon in L3 content in proposing data governance, lineage, and management in various facets like client starting their data governance journey or using a feature or two of the complete set.
- But in the run-time business, we do have challenges when the client is interested in an asset but is in a stable space of its existing architecture.
- At higher-level data management intersects with the ML life cycle as shown in the figure below. With the growing adoption of ML solutions, there are new business and technological considerations for data management platforms. It's important to design an architecture that can easily integrate with services that are developing models using the enriched and quality data being processed.
- Let's understand the idealization of real-world solutions where we have positioned the Watson Knowledge Catalog which integrates well within the existing structure.

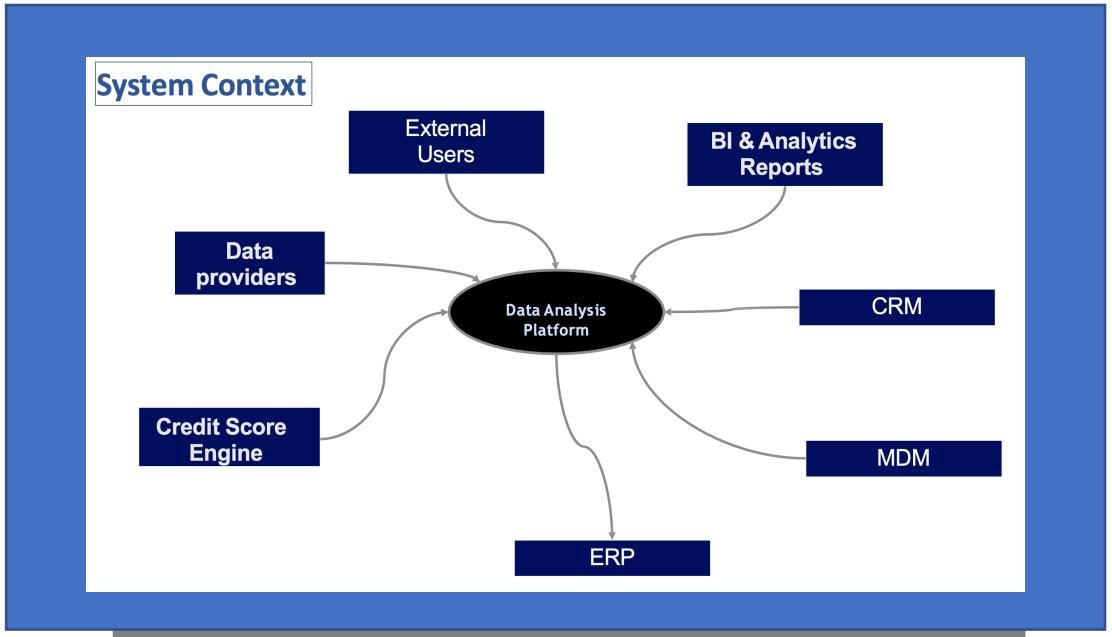


Business Problem

Client Proposition

Financial Services is one of the most technologically savvy industries and a front-runner in Data/AI investments and adoption. One of our clients was solving the problem of credit score rating which is highly dependent on data. Whether it is about the availability of that data, its quality, governance, analytics is done on them, or the whole data lifecycle, a modern Data and AI platform is required for such business problems.

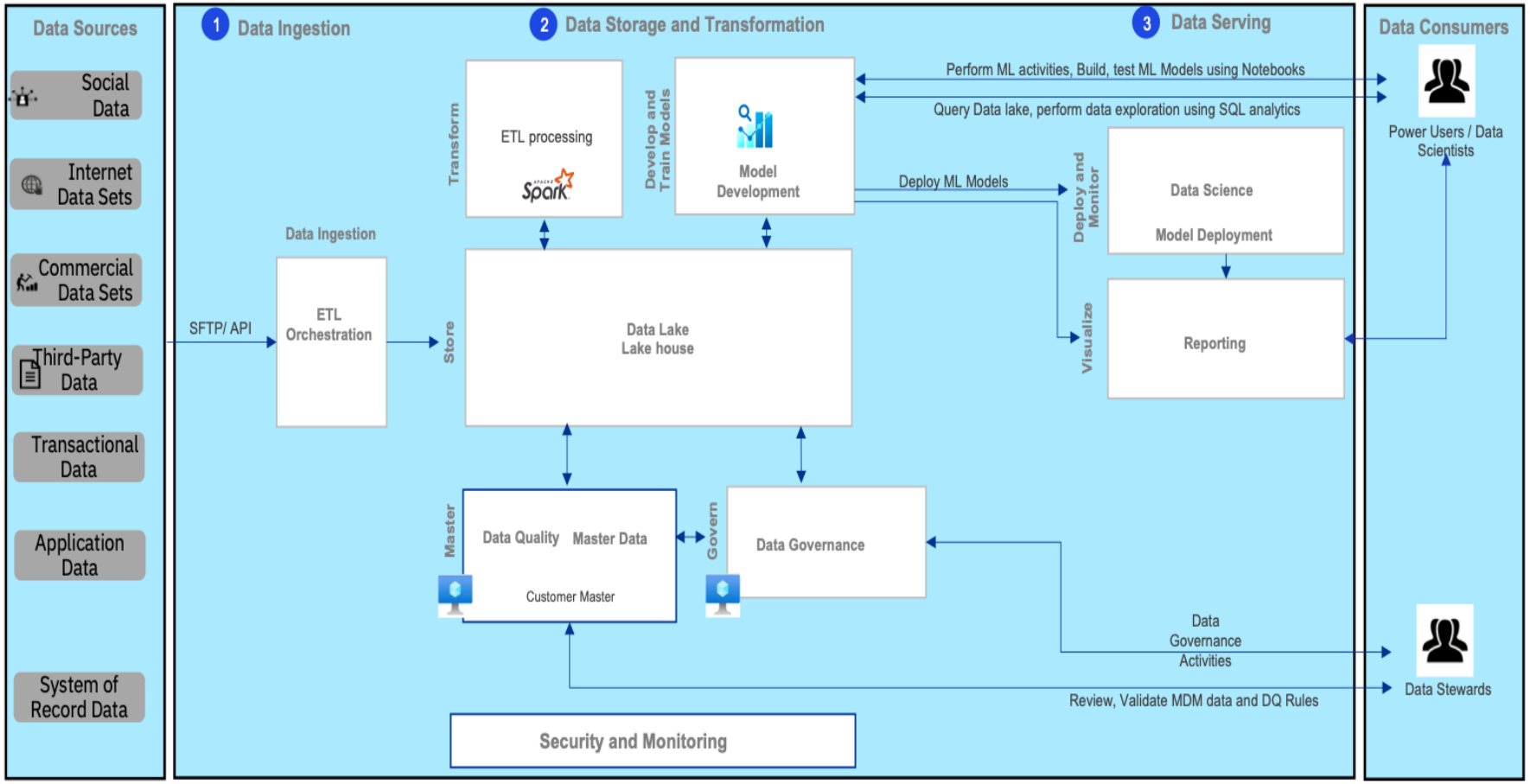
A client brought us the existing architecture and wants the use cases that need to be addressed in different areas in the current proposition. The Primary use case they proposed to look into from Data and AI potential is Credit Score Engine as represented in the flow diagram.



Use cases to be served by the Data & AI platform

Architectural Requirement

Developing a credit score engine includes the integration of data and ai services. The client did provide its Existing reference architecture with the services they want to change and the services it would like to stick to. Below is the proposed solution architecture which shows what components of the data management pipeline are to be incorporated into the final architecture.



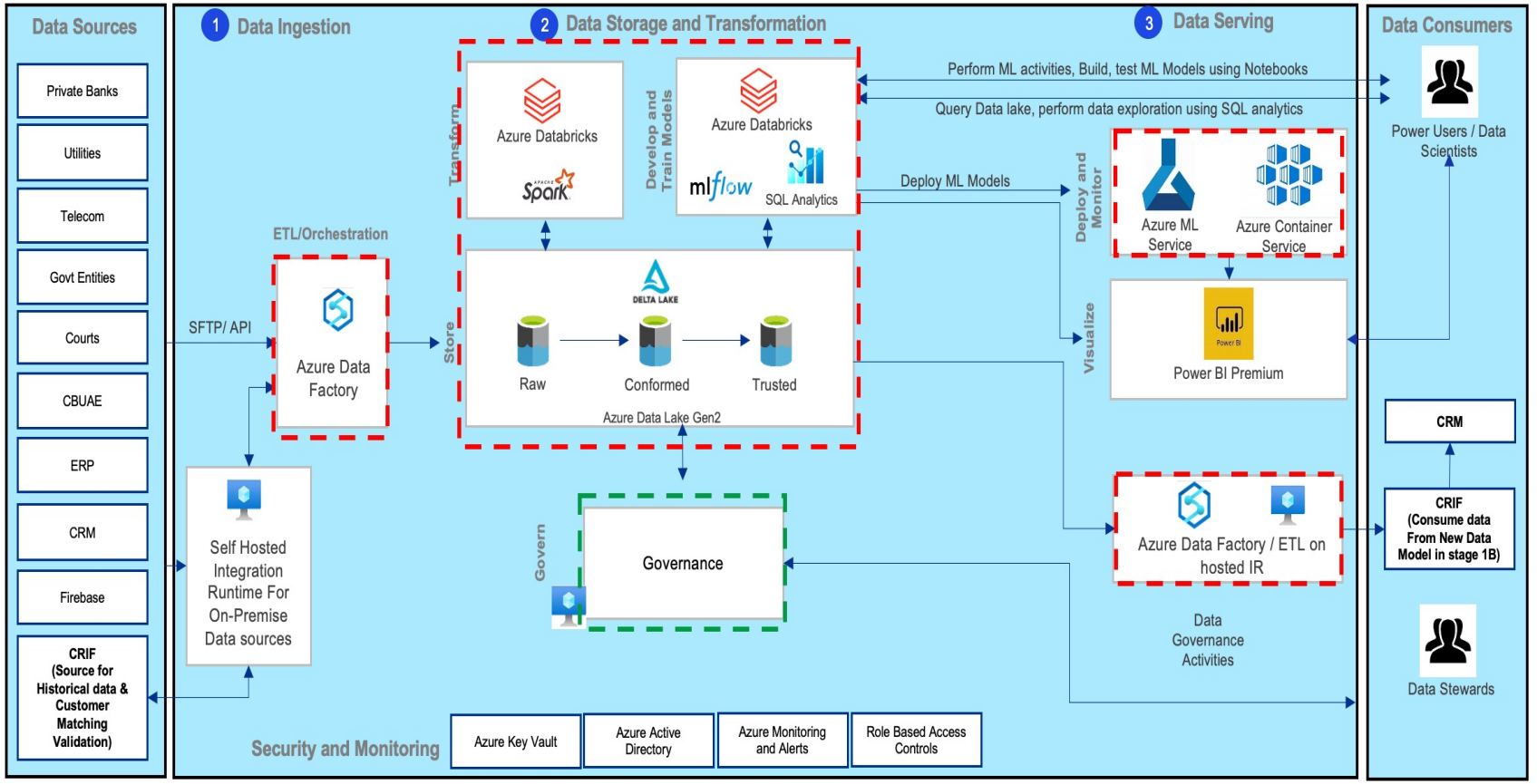
Customer Architecture Requirement

Existing Client Landscape

From the solution architecture, it is visible that the customer is highly vested in the azure cloud (Hybrid). The data ingestion, data storage, and security involve hyper scaler (Azure) footprints.
Let's go through the architectural design and understand the complexities to address.

Azure Native Services

Fitment Service



Cloud based modern reference data architecture

Complexities in existing environment

Data onboarding is a time-consuming and complex operation, especially when new data providers are onboarded.

In the architecture we see data sources with varied formats and velocity are being onboarded from providers, these sources are driving significant data growth and the platform requires to have trusted source of data to deliver effective analytical and business insights.

Data quality activities are manual and time-consuming.

Data Governance policies are not established as defined in the existing target operating model. The Model requires review to address all data management domains and ensure that governance committees are operational.

Scope – Part I

- From the complexities, we know the pain areas of an existing architecture. This gives us the idea to focus on the primary concerns of the client. Now that we know the client is interested in the quality, governance, and golden truth version of data, we have our initial starter.
- We know WKC can be positioned here for Governance and data quality, but it is not that straightforward to conclude our decision. If we dig deep into the architectural landscape, we see that the whole architecture is mainly on the azure shop. Hence before we go and propose it's very essential to understand the azure services for governance and its market consumption.
- The correct method to understand the positions of contenders is to check Gartner Magic Quadrant, Gartner Critical Capabilities, and Forrester Wave Guide to get an understanding of the feature capabilities of the services. This helps buyers choose the purchase options in the technology marketplace.
- We know Microsoft has a service called Azure Purview for Data governance. But azure purview has significant governance and quality capability deficiencies. It has very poor data privacy and limited capability in the Business Glossary.
- We thus have a very strong ground to position WKC which is bagged by the Gartner Forrester scores in respective categories from the below table.

	Automation and Augment.	Analytics and Data Science	Data Eng.	Data and Analytics Gov.	Master Data Man.	Ops. / Trans. Data Quality	AI and ML	Catalog	Data Intelligence	Data Gov.	Connect. & Interop.	Data aaS	Data Man.	Stream and discover	Data process. Persist., & events	Data access, search, and delivery	Deploy. and admin.	Total
IBM	3.5	4.02	4.05	4.12	4.15	4.03	3.94	3	3	2.4	3	2.32	5	3.6	4.2	3	3.4	3.57
Microsoft								3	3	1.4	3	3						2.68

Scope – Part II

- The next step is to scope down to the right services required with the current and the needed.
  - Implement the defined Data governance strategy with supporting technology to work in accordance with the defined target operating model, policies ,and procedures.
  - Implement an analytics environment that leverages data from the data layer and supports data analysis and insights needed for reporting and operations. The environment should contain subcomponents and technology such as the following but not be limited to:
    - Data transformation, analytics, scorecards, machine learning and mining
    - Data Visualization via BI tools, reports and dashboards
    - Sandbox environment for self-service analytics
  - Implement a data quality framework, including technology, to monitor data quality over existing and new components of the data architecture
- To keep the scope for the expansion of the framework
  - To meet the growing needs of business it is important to realize that we develop an architecture that’s elastic and scalable and resilient to the changing demands

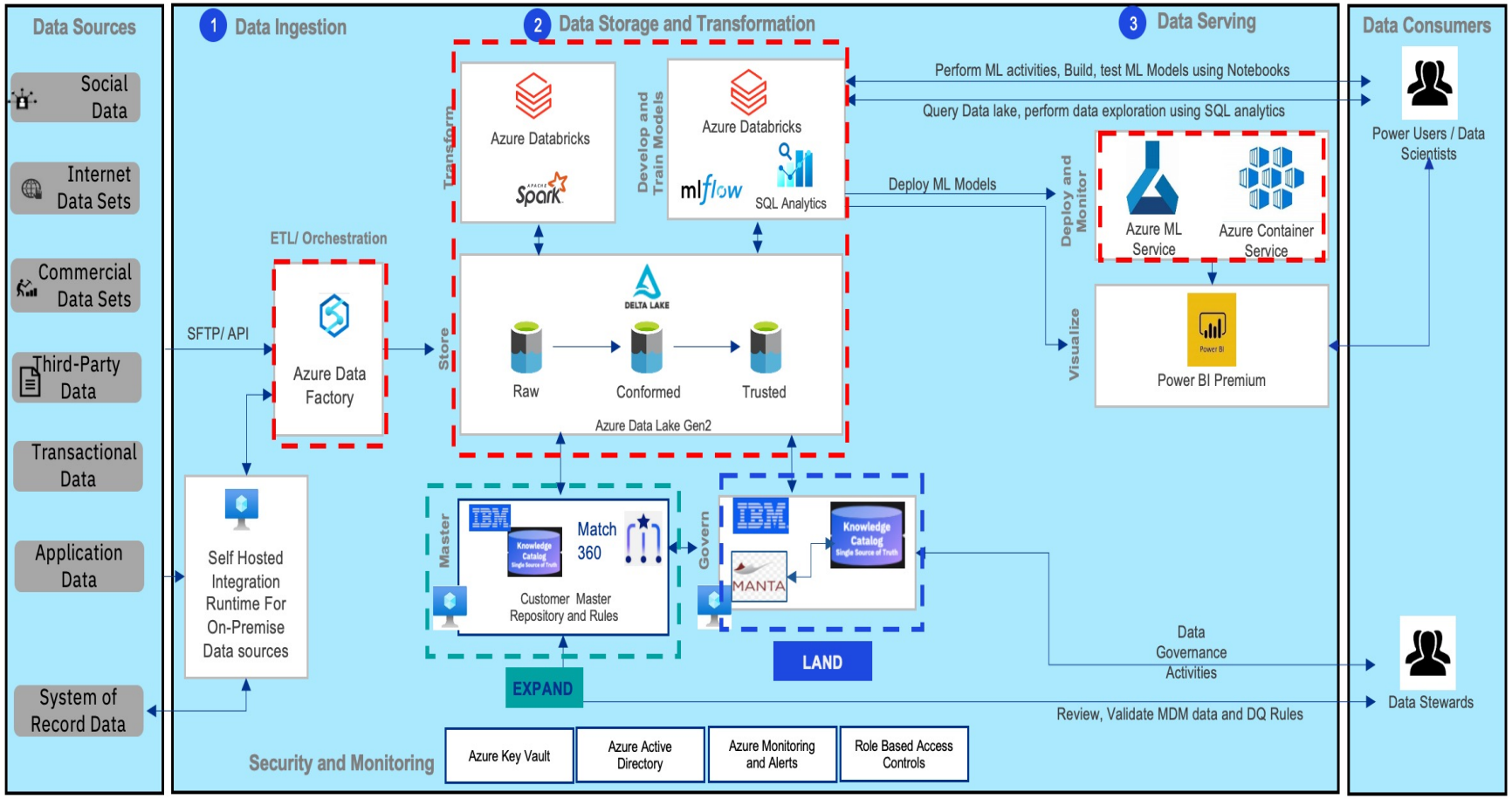
Architectural Strategy

- The architectural strategy stage is our final step when we are sure of the problem statement, pain areas to focus on, and services to fit into the landscape.
- Deep dive into the existing architecture.
  - In the existing architecture the data ingestion layer gets the data from different sources and is integrated using azure data factory. Azure Data Factory is a managed ETL cloud service that brings together raw data to operationalize and refines to use it for actionable business insights . The data from azure Data Factory is then pushed to azure storage.
  - In the data transformation and analytics layer then the data from storage is used to build the ML models along with the data from the spark cluster which is using azure databricks for high compute-intensive data transformations and pushes the data for the model pipeline.
  - In the data serving layer the model built is deployed using native azure ml service and further using this data, the analytics dashboards are built on Power BI.
  - Finally, the Azure security layer is responsible for all the security-related operations of authentication, access control, confidentiality, and integrity in order to reduce security breaches.

  STRENGTHS
  - We see here that the architectural landscape is very concise.
  - All the services are well integrated addressing each layer of the ML lifecycle.

  WEAKNESS
  - The Data Governance framework is not defined.
  - Data Quality activities are manual.
- Solution
  - The thumb rule to design architectural strategy is never to disrupt the existing but make it smoother.
  - Using Watson knowledge Catalog we will be able to rapidly build , run and manage the high-quality and trusted data to have authentic results from data and ml pipeline.

Land and Expand Architecture



Target Data Architecture

Azure Native Services

IBM Native Services

Expansion Services

Land and Expand benefits

- In Our proposed Target architecture, we are not disrupting the architecture of customers but integrating our services to stitch with the existing setup and reduce the complexity and time consumed to onboard all necessary data for Credit Bureau Engine.
- In the Land section, we help the data stewards to easily integrate data from the azure data factory in the Watson knowledge catalog ensuring to discover, classify and manage information that meets the obligations enforced by regulatory and corporate mandates. Thus, enabling immediate access to trusted high-quality data.
- Our Expand tab helps scale the current architecture further by introducing Manta and Match 360. Manta will give the transparent data pipeline execution view of the dynamic incoming data in runtime thus enabling collaborators to keep a check on the data changes. Further Match 360 can enhance and generate a customizable data model to give tailored personalized business results.
- The trusted data is further used by Data scientists to develop, train and deploy different ML models and build a concise ml pipeline.

Value Proposition

With any existing architecture, the client starts with, it's always essential to understand the growth the architecture can lead toward. Since there would be a plausibility that the architecture will be scalable in the future with the ever-growing business demands.. Which then must be seen through a 360-degree view of how the setup can be more robust, durable, and flexible. With our client in mind, we will focus on the value proposition around areas of data governance.

We have IBM ramping up its fabric offering aggressively to become more competitive in the cloud. And with this in focus, all the services in the Data Fabric offering are being driven to provide more customizability, breadth of function, and product roadmap visibility than any other competitor.

- IBM offers Privacy and AI Governance capabilities with the AI Factsheets component of Watson Knowledge Catalog. AI Factsheets are the tool that will enable Data scientists to capture the model metadata thus empowering more transparency through the complete model life cycle from development, deployment, testing, validation, and operation.
- One of the critical capabilities of Data Fabric is Data Lineage. Data lineage is used to track how your data is moved, transformed, and consumed throughout your data ops journey. Thus, helping Data engineers and business users to have more trust in data with a comprehensive understanding.
- MANTA is an add-on provided with Watson Knowledge Catalog. It initiates lineage scans through metadata import and automates the discovery and analysis of data flows.
- Also, in the Expand journey, we have the strong usage of Match 360 which will help customers with master data management. Implementing a Customer Matching Engine through Master Data Management that supports improved master data of consumers by utilizing inbound feeds and credit bureau data.
- MDM drives significant savings in cost and time for onboarding new data sources. Provides no-code approach with the auto profiling, auto-classification, auto-term assignments, and auto-mapping to onboard additional data sources for modern MDM use cases.