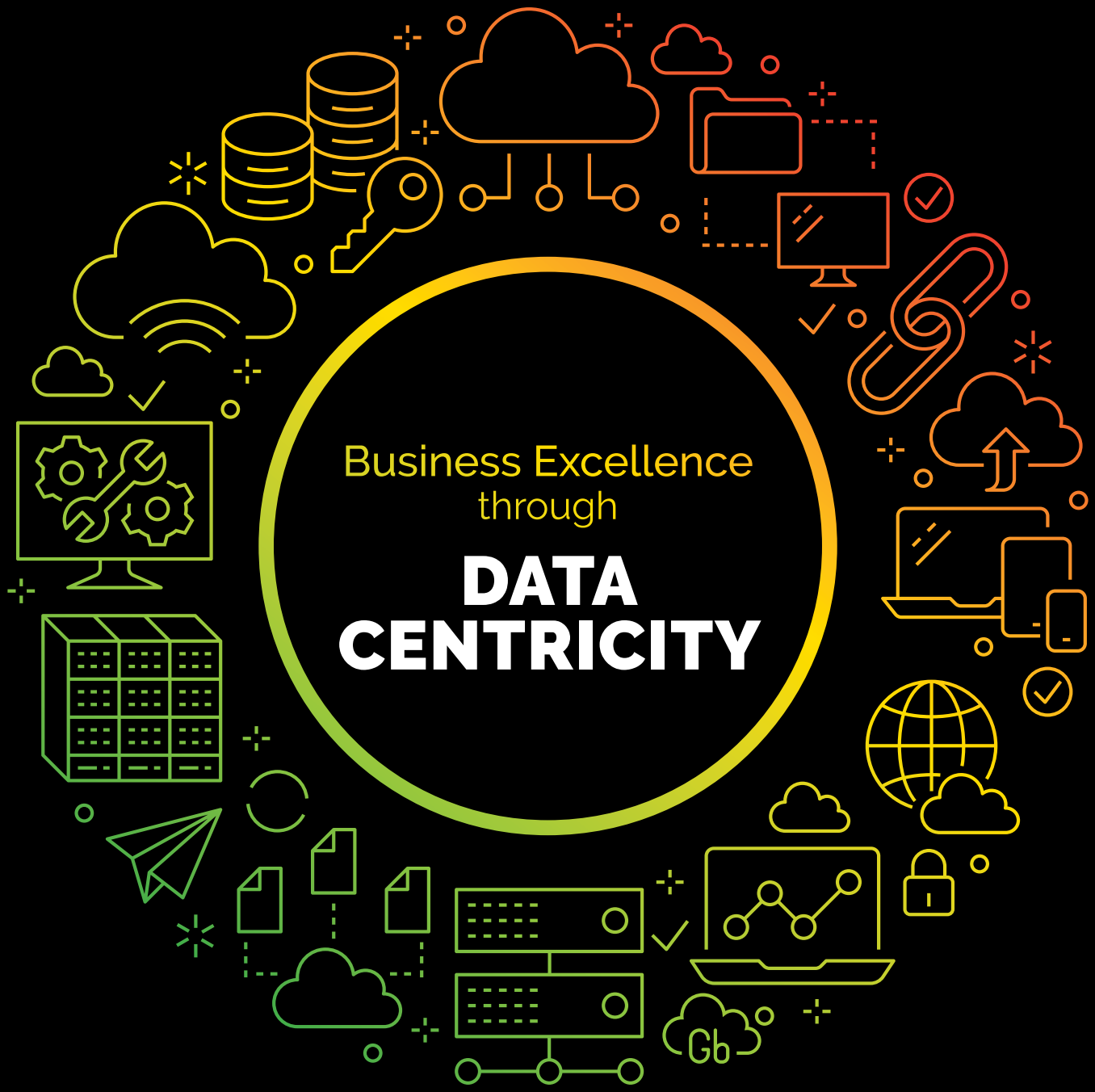
TATA BUSINESS EXCELLENCE GROUP

TATA



Based on

TCS DATOM™ FRAMEWORK

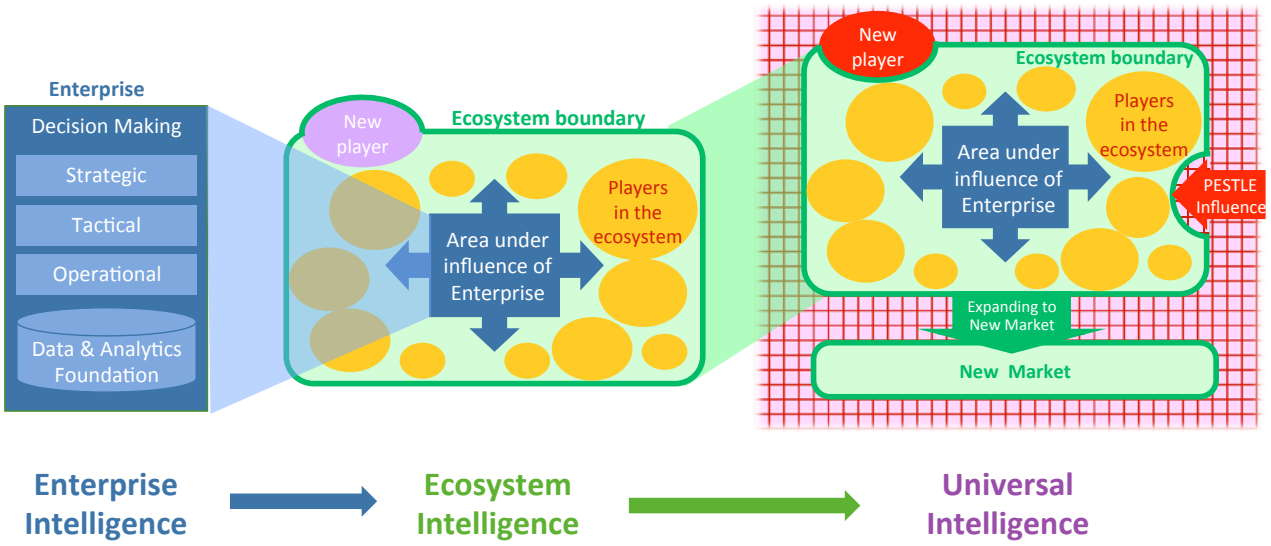
DATOM™

Contents

1 Data and Analytics – Journey so far

Today, every organization is interested in working with and utilizing data. The focus is on building enterprise level capabilities, increasing access and availability, adopting analytics, and ensuring that data and analytics solutions gain credibility and acceptance across the organization thereby improving business performance.

The following section describes **last three decades journey** of data and analytics from 'good-to-have' to 'business-critical' applications and the way data driven business capabilities were built in various organizations across the world.



▲ ***FIGURE1:*** *Evolution of data and analytics context in business continuum*

Till 1990s, mostly **mainframe systems** supported critical business applications. These were robust and reliable but **had many challenges in terms of making data available** for analysis. Simple analysis of financial or sales data used to be a major task. This required building of tactical capabilities to acquire data and then use it for analysis. This was the era of data marts and organizations were trying to learn the ways to capture data, match it with other enterprise data, use it and so on.

During 2000 and 2001, the **concept of balance scorecard** evolved. Now that organizations could acquire business function specific data, the intention was to get those consolidated at the enterprise level to get a holistic view in terms of an organization's performance and then drill down to function specific details. This led to the emergence of **Enterprise Data Warehouse** to enable strategic decision making capabilities. Enterprise data started getting integrated centrally. Organizations started focusing on the ability to analyze KPIs, measure behaviours and provide strategic insights to business. (The definition of a data warehouse was subject oriented, integrated, time-variant and non-volatile data sets used for decision support).

Around 2004, the definition started evolving again. The **need to know the current state of affairs** with minimum latency, became more prevalent. For example, the telecom industry wanted to know about the network faults as and when they occur. Call Centres wanted to know about their customers as soon as they receive a call and so on. Organizations wanted to **build capabilities to use data for supporting operational decision making**. The focus was on using internal data for decision making at various levels.

In 2008 and 2009, the world witnessed a huge recession. More regulatory bodies were established. Organizations realized that it is not just them **but also the entire ecosystem** including suppliers, customers, regulatory bodies and other affiliates, **influence the overall success**. A car manufacturer who had all its suppliers in China was affected by the tsunami that occurred in Japan. This happened because the suppliers in China were importing sub-components from Japan. This ecosystem was not just vast but also multileveled. Thus emerged the need for organizations to build capabilities to understand this multi-level ecosystem, exchange information and leverage them. Capabilities to capture, understand and analyze **ecosystem data evolved**.

After 2012, the **era of Artificial Intelligence** (AI) began, many new players started disrupting the market place. Apple, a known technology giant, introduced Apple Pay and started venturing into financial services. Amazon brought in a major disruption in the retail industry by driving customers towards online sales and enhancing the customer experience with never so thought of delivery cycles and a huge searchable product catalogue.

On the other hand, organizations also started to feel the need to extend their boundaries: either in a new geography or with a new service model or in a completely disruptive path. The **need to know more about the unknown** became stronger than ever. This obviously meant processing of very large data sets that often originated beyond the existing ecosystem. It could be data of any form, from any source. Organizations wanted to build capabilities to extract useful insights and infer from all possible data and data sources available in the universe. Reaching to this maturity level of Universal Intelligence is a long journey.

As more and more C Suite executives acknowledge that mature Data & Analytics (D&A) capabilities are key contributors to market success today, disconnected operations across People, Processes, Technology and Data continue to be a major hurdle to such success. Multiple generations of technology and constant flux of processes and techniques add further complexity. It is time that **organizations recognize** the **interdependencies** between key capability components such as **Governance, Roadmaps, Service Models,** and **Competency** and **orchestrate simplified yet holistic operating models** for their **D&A initiatives.**

TBExG Journey into Data Driven Excellence

An increasing number of Tata companies today leverage data insights to create competitive advantage as a strategic pathway for growth. A data-centric approach goes beyond operational excellence and requires more than just installing the right tools and applications. It entails putting data and analysis at the core of everyday business processes.

As the key driver of the business excellence movement at the Tata group, TBExG would like to support the Tata Companies journey in dealing with the fast-emerging paradigms of the digital world and the associated phenomenon of 'all-pervasive real-time data'.

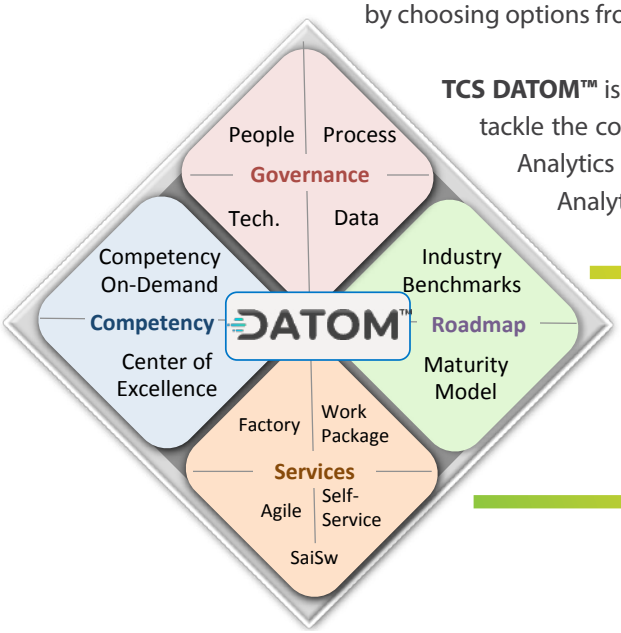
TBExG studied several frameworks available for assessing data maturity, using filters such as ease of getting assessed, comprehensiveness of the framework and the costs of training and assessing for Tata companies. The TCS's proprietary framework DATOM™, an outcome of number of years of TCS' experience in conducting data maturity assessments for its customers through its Global Data Analytics practice -- was selected by TBExG as the most suitable data maturity assessment framework for the Tata group.

TBExG partnered with Tata Consultancy Service (TCS) to adapt their framework as a diagnostic instrument for usage by Tata Companies, in helping them discover the current state of data maturity in the companies, and thereby help define a desired state for the future. DATOM™ has gone through continuous evolution in sync with evolving industry trends and TCS is committed to maintaining this evolutionary journey, thereby ensuring that the TBExG-TCS partnership will have continued access to contemporary trends and practices.

The Data Maturity assessment process using the reference manual is intended to help companies understand their maturity level towards evolving into a data-driven company. It involves diagnostics of the four core DATOM™ dimensions of People, Technology, Data management and Process aspects of the organization. The DATOM™ assessment results in a comprehensive analysis of the company's current situation and related challenges across 23 result areas under the four dimensions.

2 TCS DATOM™ Introduction

TCS's Data and Analytics Target Operating Model (DATOM™) helps organizations assess the maturity of their data and analytics capabilities and establish an effective **Operating Model** and **set up data and analytics programs** to fulfill their business objectives and goals. TCS DATOM™ provides a wide collection of best practices and solution frameworks that offers unparalleled visibility into potential pitfalls and their solutions. However, it is to be noted, that it prescribes "**no silver bullet**" that can resolve all the organizational issues. Solutions should be established by choosing options from a fabric of capabilities, depending on a particular context.



TCS DATOM™ is **an exhaustive collection of solution approaches** that can tackle the complexities and challenges spread across the entire Data and Analytics value chain. The **maturity of an organization** in Data and Analytics can be defined based on **four "key result areas"**.

These are:

- Data
- Technology
- Process
- People

▲ ***FIGURE 2:*** *TCS DATOM™*

TCS DATOM™ also has four focus areas:

- **Governance** – Establish high-level of co-ordination across functions to deliver according to the enterprise strategy across the key results areas of People, Process, Technology and Data.
- **Roadmap** – Define a business capability roadmap and identify the type of data that is required to achieve the business capability, systems required to support the data, the type of foundation that will have to be build, and the technology and people required for this activity.
- **Service Models** – Based on an organization's business needs, it can adopt a range of pattern based best of breed delivery models, ranging from Factory model, Waterfall, Agile, Self Service, etc., to deliver data and analytics initiatives using the most appropriate delivery model for a particular type of initiative. This will ensure business effectiveness and operational efficiency and thus will accelerate business value creation.
- **Competency** – Identify, deploy and retain key skill sets, reduce costs with creation of centres of excellence and development, and contain the risk of knowledge loss.

TCS DATOM™ can be used by customers in the following ways:

- **Evaluate existing ecosystem:** To assess maturity and identify areas of improvement from data and analytics perspective.
- **Establish a roadmap and an operating model:** Identify and prioritize key data and analytics initiatives/programs in an organized manner.

Why DATOM™?

The key ambition of today's organizations is optimized and effective use of data and analytics to deliver stakeholder value and achieve value chain excellence. TCS DATOM™ helps organizations to carry out an **objective evaluation** of their existing data and analytics capabilities and **establish a data and analytics operating model** which will enable an organization to **plan, adopt, govern, use and evolve key capabilities** across Data, Technology, Process and People to fulfil its ambition and meet its business objectives and goals.

Essentially it helps an organization in its journey of data driven excellence, by bringing in synergy within the enterprise and its ecosystem, instilling discipline and driving a culture of data driven decision making. The generated data and insights bring in effectiveness and efficiency in business operations and provide competitive advantage.

Some of the potential impacts are as follows:

- **Vision Alignment:** Synchronized roadmap across business capability build up and technology adoption that helps in the execution of roadmap validation and readiness.
- **TCO Reduction**: Opportunities to rationalize application, technology and services footprint, and simplify information management estate.
- **Consistent Analytics Capability:** Enterprise analytics to enable functional capability reuse across business units, lines of business, regions and so on.
- **Improved Productivity:** Contextualized service delivery routines, process and organizational alignment and standardized delivery processes.
- **Credibility for Data Solutions:** Improved data controls and governance through process maturity and digitization.
- **Enhanced Ideation:** Improved access and reach for data, enabling ideation and enhancement of business capability.
- **Visibility into Program Health:** Digitized program governance for multi modal IT environment.
- **Competency Alignment:** Future proofing competency in line with Data and Analytics Roadmap.
- **Avoid potential pitfalls:** Avoid potential pitfalls across service delivery, governance, roadmap and competency.

This reference manual is intended to provide an understanding of the TCS DATOM™ Assessment - The dimensions/Key Result Areas (KRAs), sub key result areas (sub-KRAs), various maturity levels, the overall process and the benefits that an organization can achieve at each of these levels.

3 Data Maturity Assessment – Dimensions

An organization needs to build **foundational Enterprise Intelligence** capabilities to evolve from merely managing tactical needs to realizing strategic potential while handling operational excellence. After strengthening its self-performance, an organization needs to continue learning about the ecosystem and its players and optimize the relationships. By bringing performance intelligence and relationship optimization together, **ecosystem intelligence** can be established and expanded for the benefit of the organization as well as its ecosystem. However, unknown forces of risk can influence the area of functioning of this ecosystem and an organization needs to strive to identify, understand and mitigate these risks as far as possible. Right balance between human centric and machine centric decision making can help in institutionalizing innovation, which is a key to managing self-performance, ecosystem strength and universal risks. The fusion of governance of performance, relationship and risk can help an organization to achieve the state of **self-optimization** with a strong backbone of Digital Intelligence. This can be established using the **five maturity levels (5ml) of Digital Intelligence**. *(For further details on Digital Intelligence, please refer to Appendix C.)*

TCS DATOM™'s – Data Maturity Assessment helps to understand where an organization stands currently with respect to its Data & Analytics maturity. **It evaluates an organization's data maturity** based on five levels: **siloed, simplified, scaled, synergized, and self-optimized**. This also provides a benchmark to measure progress on the data and analytics journey.

In terms of the assessment, it focuses on **four primary dimensions or Key Result Areas (KRAs): Data, Technology, Process and People**. Each of these KRAs are further divided into Sub-KRAs. This is done to ensure that a comprehensive coverage in terms of breadth and depth is available.

Data is the core element in decision making. If data is incorrect, the decisions taken will not be reliable.

Technology is the next key result area. An organization owning huge datasets, but not having the right technology to be able to capture, process, govern and exploit the same, will not be able to utilize its own data assets effectively. (For e.g. by deploying a Manufacturing Execution System (MES), a manufacturing organization may have automated operational tasks and have effected a step jump in terms of operational efficiency. However, capturing and analyzing the generated datasets can open up plethora of new opportunities for process improvement and optimization and ability to take real time decisions. These cannot be achieved unless appropriate technologies to capture and process real-time data and take decisions are deployed.

Similarly, if a bank is not able to get a holistic view of transactions with a customer across its different Lines of Businesses, there is high probability that it may be vulnerable in terms of its financial exposure. On the other hand, it may have also lost opportunities by not understanding the customer's lifetime value.)

Process as a key result area helps to establish repeated and consistent performance thereby reducing chaos. Right processes bring in agility as well and reduce time to market.

People is the next key result area where the focus is on the ability of the organization to continuously improve competency in terms of data driven decision making and manage knowledge to improve and sustain its competitive advantage.

With this overview, the **following section further deep dives** and aims to explain in detail the meaning of maturity in each of the key result areas and the various focus areas (sub-KRAs) within each of them. **The details cover some of the typical questions** that need to be asked to assess the maturity **and examples of expectations and ways in which different organizations have addressed these.**

1.0.0 KEY RESULT AREA - DATA

In the world of data abundance, every organization needs to deploy effective mechanisms to collect and classify data and effectively channelize the same for decision making. The **Data KRA** focuses on assessing the journey of data from **planning, to acquisition, to management through its entire lifecycle**.

SILOED – 1ml	• Data exists in silos in Departments *(e.g. HR, Finance, Procurement, Operations)* or with individuals • Incomplete history • Support tactical decisions • Integration issues requiring manual intervention for resolution • Very limited Data Governance
SIMPLIFIED – 2ml	• Integrated enterprise data store • Complete historical data • Well established Information Management Functions *(Data Architecture & Modeling, Metadata Management, Master Data Management, etc.)* • Established Corporate Performance Management *(covering Planning & Budgeting, Financial Consolidation, Balanced Scorecard)*
SCALED – 3ml	• Low latency/real time operational data for decision making at enterprise scale • Comprehensive Business Performance Management • Fulfillment of all data needs for Enterprise Intelligence across all management levels • Matured Operational Analytics • Command Center Analytics
SYNERGIZED – 4ml	• Capture and use Ecosystem data • Synergize ecosystem data with enterprise data during decision making • Process semi-structured, unstructured, high volume data • Established business/domain specific analytical capabilities • Complex event processing intelligent engines
SELF-OPTIMIZED – 5ml	• Identify & understand unknown risks beyond ecosystem boundary • Usage of Orthogonal, PESTLE Data • Advanced analytical, AI capabilities • High Degree of intelligent automation in decision making • State of self-optimization

Typically, at **1ml**, one will observe that **data exists in separate archives** in **departments or functions** such as HR, Finance, Procurement, Operations or with individuals. There will be many historical data available but it may not all be complete. It may include only what was required to support the individual business operations. Looking at such data sets may help to take tactical decisions. However, if one tries to connect it with other enterprise data, the issues start surfacing and one ends up with disconnected data sets.

At **2ml**, organizations will typically have an **integrated data store at an enterprise level** and not just at department levels. This store would be hosting the historical data as well. **Information Management functions** such as Data Architecture and Modeling, Metadata Management, Master and Reference Data Management (MDM/RDM), Data Quality Management, Data Security Management, Information Lifecycle Management (ILM), Data Governance are well established. **Corporate Performance Management (CPM)** is well established covering various aspects such as *Planning and Budgeting, Financial Consolidation and Balanced Scorecards*.

At **3ml**, very **low latency operational data is used at enterprise scale** for decision making. Wherever justified, real time data usage is also established at this level. Without establishing the information management framework (as explained in 2ml), it is not possible to establish and sustain the scale at which the data is required to be captured, processed, governed and utilized for decision management across the enterprise, at this level of maturity. There will be a need to enable Operational Analytics, Command Centers Analytics and so on. At this maturity level, **all data needs for Enterprise Intelligence, covering strategic, tactical as well as operational needs within the organization are fulfilled**.

At **4ml**, organizations capture and use ecosystem data from Customers, Consumers, Suppliers, Competitors, Regulators, Affiliates, etc., in the decision making process. Data can be semi-structured (as in the case of B2B transactions using xml messages, security logs etc.) or unstructured (text, voice, images, video) as well. Organizations at this level capture and synergize this with enterprise information. At this level, one will typically find established business and domain specific **Analytical Applications, Complex Event Processing** intelligent engines and other decision support capabilities with data sourced from varied sources within the ecosystem.

There is a huge amount of data even beyond the boundaries of ecosystem, such as PESTLE data. This enables us to **understand the unknown risks beyond the ecosystem boundaries**. At **5ml**, one starts looking at all these data sets. Organizations strives to take into account different data sets which helps to identify external risks. Often **advanced analytical capabilities are deployed** with large number of variables in datasets and connected datasets. An organization will use data, analytics and AI to attain **high degree of intelligent automation** to enable effective strategic decision making.

The maturity characteristics of the **Data KRA** is summarized in the adjacent table. *(A **consolidated view** of the maturity characteristics spanning all the KRAs has been provided in **Appendix B.**)*

The assessment focus is on the following sub-KRAs while assessing data:

- **Data Planning**
- **Data Governance**
- **Master Data,**
- **Data Security**
- **Data Architecture & Information Life Cycle Management**
- **Data Quality**
- **Metadata**
- **Decision Management**

1.0.1 Sub-KRA: Data Planning

Who manages Data and Analytics? Moreover, how effective is the planning process?

The Data Planning sub-KRA **focuses on measuring the ability of the organization to get the right processes to manage enterprise data assets**. Making right data available at the right time involves significant planning. The focus here is to understand the maturity of the organization's processes for managing data and analytics assets. The following aspects are what one would try to understand:

- Does the organization follow **a defined process**, which ensures that current and future data requirements are identified proactively?
- Do these processes cover **requirements for data outside of the data generated by the enterprise** itself (such as Competitive Intelligence, PESTLE analysis, and so on.)?
- Is the **accountability of the entire data planning exercise established**, that is, are there defined teams or roles who are responsible for the same?

The other important area covered here is the valuation of the data that is held by the organization. It tries to understand aspects like:

- Is the **cost-value analysis** carried out for existing and prospective future data assets? Is the financial accountability shared with data consumers?
- Does the organization **understand the monetary value** of the data it holds?
- Is the organization planning to **put up controls to avoid legal or regulatory litigations** and fines due to bad data quality?

Matured data organizations would have established Data Office in place that would be responsible for data planning and valuation.

1.0.2 Sub-KRA: Data Architecture & Information Life Cycle Management

How is data managed through its life cycle from creation to archival or deletion? How good is the data availability for decision support?

The focus here will be to **understand what data is available for analysis to the business users and how they are made available**. Do they just look at their business function data or they have access to enterprise wide data sets and can do cross-functional analysis. Do they have ability to

merge even external or orthogonal (statistically independent) data? For example, an insurance company can use telematics data from sensors embedded in cars and smartphones to get insight into an individual's actual driving behaviors and patterns. This information can then be used to reward lower-risk drivers with discounts or rebates while providing education and real-time feedback to help improve the risk profile of higher-risk drivers.

While establishing the overall data architecture of an organization the following key aspects are analyzed and established:

a) **Data Component/Data Entity Catalogue**: To identify and consolidate the data components and related data entities used across the enterprise. This is analogous to the Data Dictionary for an enterprise.

b) **Data Entity – Business Function Mapping:** The purpose is to depict the relationship between data entities and business functions. This enables the following activities:

- Define system of origin, system of record, and system of reference for data entities
- Assignment of ownership of data entities to business functions
- Understand the data and information exchange requirements
- Enable development of data governance programs across the enterprise (establish data steward, develop data standards pertinent to the business function, etc.)
- Identify gaps

c) **System – Data Entity Mapping:** This helps to understand the CRUD relationship (Create, Read, Update and Delete) relationship between systems and data entities.

Generally, the data architecture is represented using a number of viewpoints each representing a specific aspect of the architecture. **Some of the important ones** are as follows:

i) **Entity Relationship** – This depicts the relationship among the critical data entities within the enterprise. (This helps the business users to understand the data organization and facilitates effective usage and insight generation from an enterprise data repository). This is also referred as a data model and can take different forms depending on the purpose.

ii) **Data Lineage** – This depicts the flow of data from source to the target applications/enterprise data repositories. This provides a comprehensive visual representation of the spread of source and targets and all the intermediate layers, explaining the purpose of each. (This is one of the most common data architecture views providing a holistic view of the enterprise data estate and enable effective auditing and establish traceability).

iii) **Data Life cycle** – This is an essential part of managing business data throughout its life cycle from conception until disposal within the constraints of a business process.

iv) **Data Security** – The purpose of this viewpoint is to depict which "actor" can access which enterprise data.

All these information together help to establish the data architecture of an enterprise.

The TCS DATOM™ assessment helps to understand and evaluate the "**As-Is**" data architecture of an enterprise and helps to identify the opportunities for improvement to establish the "**To-Be**" data architecture in alignment with the business objectives. If an organization has the right data architecture defined then it will have the ability to provide right data sets to the right users at the right time. This will essentially mean the organization would have identified critical data sets and then defined their refresh frequency for reporting and analytics. Data architecture essentially involves models, standards and best practices that defines storage, relationship, naming conventions and life cycle management of the data elements. Data models helps in the visual representation of data and enforces business rules, regulatory compliances, and government policies on the data. Data models also ensure consistency in naming conventions, default values and semantics. Typically, there are three different layers defined: **Conceptual, Logical and Physical**.

There would also be a retention policy defined post which historical data may have been aggregated or compressed and then stored on a less expensive storage.

A good Data Architecture helps to improve the following aspects of data:

- **Timeliness:** Is the data available at the right time or is it available, but not at a time when it was needed?
- **Completeness:** Is the information being delivered to the business complete or does it have missing elements that need to be stitched together manually before the final picture can be drawn up?
- **Ease of use:** Is the information available in a way that is accurate and easy to interpret. For example, if a summary statistics is needed then is the user provided with the required information or is the user provided with all constituent details to derive the summary himself. Vice versa may be true as well, the user may just be given summarized data without the details required for deeper analysis.

This Sub-KRA also measures how the different life stages of data is managed from its creation to deletion. The value of data generally decreases with its age and hence it needs different strategy at different stages.

The following graph shows the value of data with its age and indicative storage policies that could be adopted for the same.

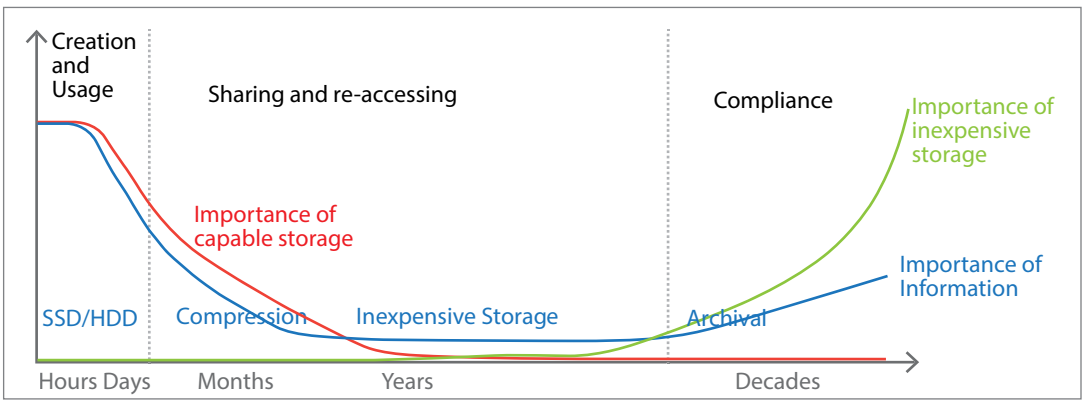
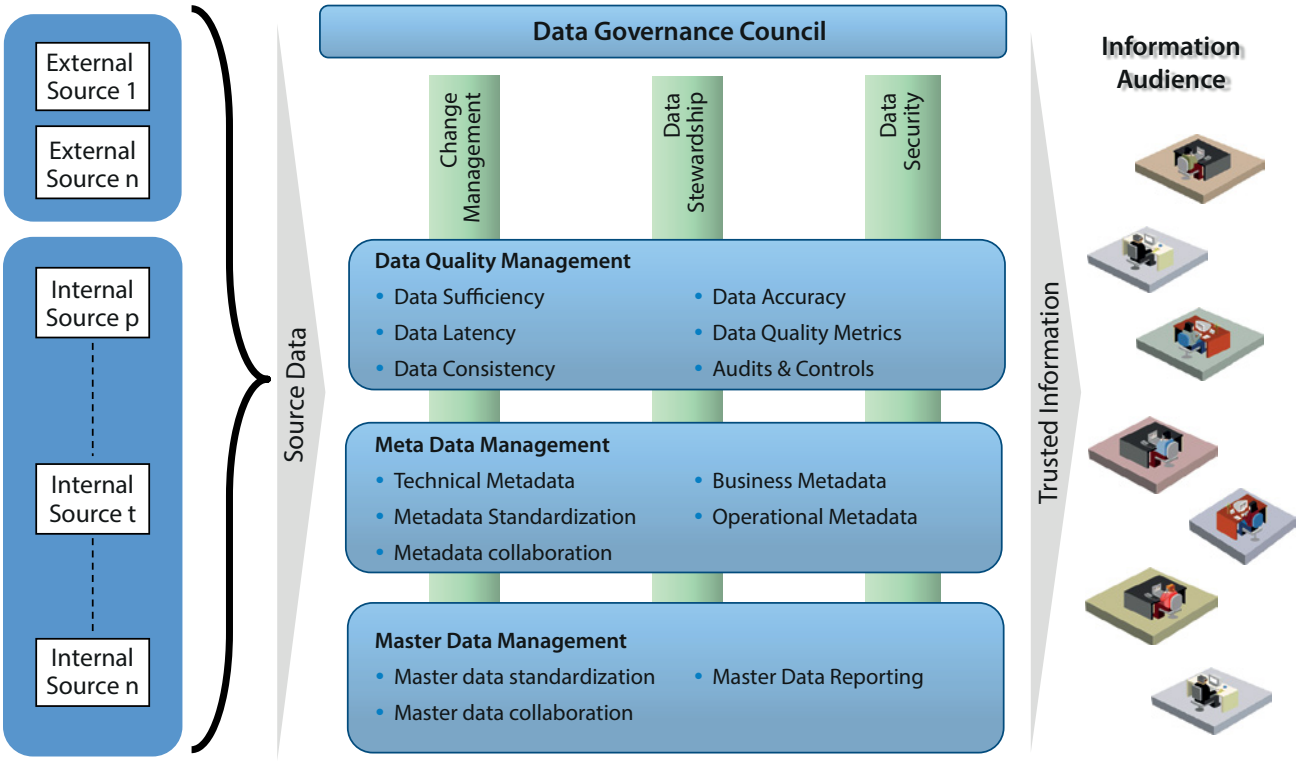


FIGURE 3: *Information Life cycle Management*

1.0.3 Sub-KRA: Data Governance

Is there a governing body to oversee and manage data?

TCS DATOM™ prescribes using data as an asset. This essentially calls for governing many aspects of data, which would mean establishing methods and structures with clear responsibilities and processes to guarantee understandable, correct, complete, trustworthy, secure and discoverable data, covering, captured and analyzed enterprise as well as external data.



▲ ***FIGURE 4*** *: Data Governance Framework*

Data governance refers to the overall management of **the availability, usability, integrity, and security of the data employed** in an enterprise. It **encompasses the people, processes and procedures** required to create a consistent, enterprise view of a company's data.

The key goals of data governance are to:

- **Minimize risks**
- **Implement compliance requirements**
- **Increase the value of data**
- **Establish internal rules for data access and use**
- **Improve internal and external communication**
- **Reduce cost of data management**

Enterprise Data Governance helps an organization to define the structure, roles and processes that provide avenues for interaction and communication paths for gathering appropriate input, making decisions, identifying and resolving issues, escalating when necessary, implementing changes and communicating actions. It will increase the accuracy of the data and help to make the organization act and plan more efficiently.

Implementing Data Governance will provide **a common set of reference data, reduced data duplication and a common vocabulary** for all of the organizations' information assets. It will also result in **increased user trust** in data stored within the organizations information management systems.

A **Data Governance Team** should ideally include:

- **Executive Sponsors** from both senior management and the business units
- **A Data Governance Management Committee**
- **Data Governance Lead, wh**o will establish Data Governance in the organization and have overall responsibility of managing it, will be responsible for resolving disputes and be the single point of accountability
- **Data Stewards**, the caretakers of the organization data assets
- **Data Owners**, responsible for the creation of the data and enforcement of organization business rules

In terms of the assessment, one needs to evaluate if the organization has a governing body to manage data. If the governing body is present then the scope of the activities it undertakes needs to be evaluated.

1.0.4 Sub-KRA: Data Quality

Is the organization aware of the quality of available data? Is there an established mechanism to regularly measure the same?

Generally, **data quality is the ability of a given data set to serve an intended purpose**. In the context of data and analytics, this intended purpose is generally **deriving insights for business decision making**. As data takes the center stage in decision making, it becomes essential to ensure that the data is certified to be fit for use; in essence, it should be of good data quality. Organizations typically face data quality issues for various reasons, ranging from system integration issues, silo processes, manual interventions causing data errors and so on.

Data quality is not a one-time process or issue. An organization may have good quality data one day and may lose it the next day. This is because new data, processes are getting added to the organizational landscape. **Data quality is also context or use case based**. What an organization perceives as high-quality data in a particular context may not be 'fit for use' in a different context. Poor quality data lacks in terms of one or more of the following factors: ***consistency, accuracy, sufficiency, latency, uniqueness, reliability, availability and usability*** while fulfilling the decision making needs in a particular context. The impact is inaccurate, improper and delayed decision making.

As an example, a leading US based health care organization had a very poor campaign effectiveness as they were not able to target the right health care professionals (HCPs) and health care organizations (HCOs) during their drug campaigns. This was due to incomplete and inaccurate Customer data (data related to HCPs and HCOs). This impacted their business significantly.

This Sub-KRA evaluates the **initiatives undertaken by an organization to establish and maintain quality of data**. It finds out what roles, teams, programs and processes have been defined to maintain usable quality of the data. The same needs to be socialized with consumers to impart confidence in using this data for business decision making. The **organization should enable the business users to identify data stewards and have the stewards then define certification rules** based on which the data gets certified regularly before reports are spooled out. Normally these rules are embedded in specific programs which are used to validate and identify quality data during data ingestion into an enterprise data repository. For data with identified quality issues, there are remediation mechanisms set. **Data issues are assigned to the stewards and then it is the steward's responsibility to ensure that the issue gets fixed at the source to avoid recurrence**. In matured organizations there are common reusable components built to support such activities. There are dashboards available through which business users can themselves assess the quality of the data they are using.

In situations where the operational processes are system enabled, digitized and automated with appropriate data validation at the data entry points and the data ingestion process into the enterprise data repository is also automated and seamlessly integrated, data quality issues introduced due to manual data handling can be avoided to a great extent. **Issues like inconsistency, inaccuracy and incompleteness can be handled to a great extent**. However, whether the data is fit for purpose or not is dependent on the exact information need or decision making requirement.

(More information on Data Quality Management and Data Quality have been provided in the glossary).

1.0.5 Sub-KRA: Master Data

Has the organization identified its Master Data?

Master Data is that subset of data that represents core business entities in an organization for example Customer, Product, Vendor, Chart of Accounts (COA), and so on. It is essential that an organization is able to identify and establish a single set of master data across its various operating entities.

Master data management aims at standardizing the processes for collection, aggregation, matching, consolidation, quality-assurance, persistence and distribution of master data throughout an organization. This ensures a common understanding, consistency, accuracy and control in the ongoing maintenance and use of the data.

Poor-quality master data is an enterprise liability. It causes unnecessary confusion throughout the organization and it dramatically increases the IT cost burden with ceaseless and fruitless spending for incomplete error remediation projects. *Some examples have been known where incorrect product master data was recorded on product labels for consumer products which resulted in the rejection of*

a whole shipment resulting in considerable financial and reputational losses. In a different instance, improperly maintained location information resulted in logistical errors and longer order to cash cycles for a large telecommunication company thereby impacting customer satisfaction adversely. Inaccurate and multiple versions of customer data maintained across a health-care organization resulted in dispatching of drug promotion information to wrong specialists.

Besides healthcare, pharmaceutical or food & beverage companies which are regulated by health and safety standards may have significant legal implications and can even lose their licenses in case of incorrect master records on expiration dates, product composition, storage locations, recording of ingredients, etc.

The benefits of master data management increase as the number and diversity of organizational departments, worker roles and computing applications expand because it acts as the pivotal IT process and provides a uniform, consistent and accurate data across the enterprise.

(More information with respect to Master Data management has been provided in the glossary.)

1.0.6 Sub-KRA: Metadata

Does the organization have enough insights on the data it uses?

Metadata, commonly known as 'data about the data', is the data, which describes other data. This can be understood with a small example.

'102250 Richard King' can have many interpretations and a few of them have been given below.

- 10:22:50 EST appointment with Richard King
- 1022 is Order# and 50 is Line Item# of a shipment delivery to Richard King
- 10,2250C Temperature of a QUASARS called Richard-King

To know about which of these interpretations is correct some more information will be needed and that is what Metadata is.

Metadata is the data-describing context, content and structure of data and their management through lifetime. In short, Metadata is 'the context of the data'. If the right context is not available, data may be interpreted in different ways and that can hamper the decision making capabilities of the organization. Consider an organization that claims that its on-time delivery metric is always above 95% and the customers still complaining that it takes them long to receive the delivery. The **RCA** discloses that the expected delivery date is an **updatable** field, which is updated every time a negotiation happens with the customer. There is a change in delivery schedule irrespective of whether that change is acceptable to the customer or not. The fact that this field is **updatable** is a **metadata information**. Similarly, the **size, type, valid values,** etc., are all examples of metadata for a particular data field.

Another key aspect that needs to be addressed here is the **definition of a measure or a KPI** and how well it is understood by the organization. The choice of measures can influence and drive the organizational behavior (desirable or undesirable). One of the largest automotive companies wanted to motivate its employees and bring in innovation. In that attempt, it selected the measure "number of engineering changes introduced". The measure also had an incentive associated with it without really understanding that it might lead to huge inventory obsolescence.

So it is important to know the following attributes for a KPI.

- **Measure/KPI Name:** Should be self-explanatory
- **Purpose:** The intention behind defining the measure
- **Objective Supported:** The business vision/goal that the measure is going to help achieve.
- **Target Values:** The level of performance that is expected.
- **Formula and Data Mapping:** The actual calculation of how you need to compute the measure. This should be accompanied with appropriate data mapping, indicating which attribute needs to be sourced from which data source to compute the same.
- **Frequency:** The frequency at which the measure needs to be computed
- **Owner:** The person or function who is responsible to act on the results of the measure.

Such details should be available in the **Enterprise Data Dictionary** and should be accessible by each individual through a seamless interface.

Hence, the focus of this KRA is to understand the following:

- If importance of metadata is known?
- Are practices and process defined to ensure right metadata is managed and maintained?
- Is there any governance set-up to manage the metadata?
- Are there any technology investments available for better knowledge representation and reasoning?

Depending on the size and complexity of the organization one may choose a tool (like Collibra, IBM - IGC, and so on) for managing metadata or it may be managed as part of the project documentation process. Another important factor to be assessed is that managing metadata is not a one-time activity and has to be updated as and when things change or there are chances of this documentation getting disharmonious.

1.0.7 Sub-KRA: Data Security

How secure is the Enterprise data?

Data security is a broad topic that includes **preventing unauthorized access, use, disclosure, disruption, modification, inspection, recording or destruction of information**. It aims to set basic guidance and policies on password, antivirus software, firewall, encryption software, legal liability and so on.

In the context of TCS DATOM™, it primarily evaluates the **governance aspect of data security** like access authorization process, access audits, ownership management of the data and rules around how to use and share the data with respective stakeholders.

It needs to be ensured that all data access happens through a secure and authorized channel. Data exchange over emails and sharing excel sheets are a big threat to security since there is no traceability as to who has access to the data and what is being done with it. The aspects covered under this Sub-KRA would be:

- Availability of secure channel (authentication and authorization) to access data within and outside the ecosystem
- Ability to trace changes happening to both data and the structures it is stored in. Also the ability to trace who accessed the data and whether it was an authorized source.

Apart from these, assessment of whether the organization has adopted the right data privacy principles and whether it is ready for open data policy are also carried out. Some of the key aspects are as follows:

- Are critical data elements identified?
- Is there a security classification available?
- Are there datasets that the organization can make available in the public domain?
- Are the right processes available to make such data public?
- Has the organization identified PII related data that they store and whether it is about their employees, ecosystem partners and affiliates or their customers? If they have captured such information, have they taken the consent from the associated party to capture and store such data elements?

*In this context it might be worthwhile to introduce **some of the prevalent information security standards/certifications** available for enterprises.*

a) **ISO/IEC 27001:** It specifies a management system that is intended to **bring information security under explicit management control**. It adopts an overarching management process to ensure that the information security controls continue to meet the organization's information security needs on an on-going basis.

b) **HIPAA** is the acronym for the Health Insurance Portability and Accountability Act. It mandates industry-wide standards for health care information on electronic billing and other processes and requires the protection and confidential handling of protected health information.

c) The Payment Card Industry Data Security Standard **(PCI DSS)** is a widely accepted set of policies and procedures **intended to optimize the security of credit, debit and cash card transactions and protect cardholders against misuse** of their personal information.

d) The General Data Protection Regulation (GDPR) is a new EU regulation which has come into force on May 2018. Its aim is to improve privacy and give greater control to customers and citizens over their personal information and how it is used.

e) **FIPS (Federal Information Processing Standard) 140-2** is the benchmark for validating the effectiveness of cryptographic hardware. Although FIPS 140-2 is a U.S./Canadian Federal standard, FIPS 140-2 compliance has been widely adopted around the world in both governmental and non-governmental sectors as a practical security benchmark and realistic best practice.

f) **ISO 22301:2012** specifies requirements to plan, establish, implement, operate, monitor, review, maintain and continually improve a documented management system to protect against, reduce the likelihood of occurrence, prepare for, respond to, and recover from disruptive incidents when they arise.]

1.0.8 Sub-KRA: Decision Management

Is the organization driving all its decisions based on data?

Decision management, also known as enterprise decision management (EDM) entails all aspects of an organization's decision making process and the enabling systems that it uses.

This Sub-KRA measures the data centricity of the organization, which means the culture that puts data or information at the center of all enterprise decisions, whether strategic or operational. Data centricity brings objectivity within an organization by bringing in data or facts at the center of a decision making process and eliminates the sentiment or emotion part of it, making the entire process more rational or factual.

The second aspect is, as the organization becomes data centric, **whether the infrastructure is appropriate to help foster this culture**. Is the organization trying to deep dive in the data space, working in the areas of predictive and prescriptive analytics, and looking into optimization problems wherever applicable? It is also essential to identify if data centricity culture is adopted throughout the organization or is it isolated in some of the business functions only.

This sub-KRA essentially **assesses the decision making ecosystem of an organization** which can span across descriptive, diagnostic, predictive, prescriptive, and pre-cognitive/cognitive capabilities. It evaluates the extent to which the organization's decision management is data driven and leverage available capabilities spanning simple reporting, data discovery and visualization, dimensional analysis, machine learning, natural language processing and cognitive computing, to improve decision effectiveness.

Computerization and availability of Data & Analytics capabilities have changed the way organizations are approaching their decision making today. More and more decisions are getting automated and are based on analysis of historical behavioral data, prior decisions, and their outcomes, thus improving reliability, consistency and performance of organizational processes.

Typically in an organization **the focus initially is on descriptive reporting**, to measure and report on past events, aggregate trends that can be used for strategic decision making and investigate special situations. Gradually, the **focus starts encompassing use of advanced analytics for**

predicting future events and take proactive actions. Then **data and analytics suites gets integrated with operational processes** for prescriptive analytics to use analytical models to automatically identify and execute the next based action.

Enterprises across a wide spectrum of industries are prioritizing and leveraging advanced decision management capabilities to improve customer experience, reduce time to market, unlock process efficiency, enhance resource utilization and so on to improve overall organizational performance.

As examples of matured decision management capabilities, a very large company in the insurance domain is currently using chatbots to underwrite certain type of insurances, resulting in insurance sales without any manual intervention. Another company from the same industry is using artificial intelligence to process claims resulting in substantial saving in terms of working hours and claim processing time.

In the telecom industry artificial intelligence has the capability to inform a contact center agent that a telecom customer who has called for some billing clarifications, can opt for an international roaming pack to save money, as he travels abroad frequently.

2.0.0 KEY RESULT AREA - TECHNOLOGY

Technology is the next key result area in terms of the assessment process. Technology is continuously evolving and moving from generalized offerings towards specific customized offerings. Hence, it is **essential to build upon a foundation**, which allows quick adaptation of a new technology or changing/upgrading the existing one to reap desired benefits. **Agility in adopting "the new" is the key**. Hence **TCS DATOM™ prescribes a fabric-based solution**, which enables utilization of different technology components in a plug-in plug-out manner, and quick adaption with respect to an organization's specific needs. Technology components in context of TCS DATOM™ are **geared to enable a data and analytics environment**.

The way to assess maturity in this Sub-KRA will be to understand the existing technology landscape and assess various aspects like *practices for evaluation of latest technology options and their appropriate adoption and usage with respect to business needs, the extent of usage of legacy applications, the connect with technology vendors and so on. It is also important to understand the process of managing and maintaining the established IT infrastructure hosting/supporting the D&A capabilities.*

Typically, organizations **at 1ml will mostly have traditional Business Intelligence (BI) Technologies** involving ETL, MIS, OLAP reporting, dashboards and so on. Typically, data is stored in business function specific data repositories. Data storage in a Data Warehouse or in a Data Mart or in an Operational Data Store (ODS) is not uncommon. There would be cases where even multi-dimensional databases (MDDB) may exist. Advanced Analytics (Data Mining) capabilities may exist within some of the business functions. Generally, the prevalent tendency is to carry out **analytical activities on desktops**, although instances of carrying out data mining activities in sandboxes do exist.

SILOED – 1ml	• Traditional BI Technologies involving ETL, MIS, Reporting & Dashboards. • Business function specific data repositories • Data may be in a Data Warehouse, Data Mart, Operational Data Store (ODS) ; sometimes even in MDDBs • Advanced Analytics (Data Mining) capabilities may exist within a few business functions • Data Mining/analytics mostly on desktop
SIMPLIFIED – 2ml	• ETL suites to integrate different type of data across enterprise in batch mode • Enterprise Data Warehouse and mostly EDW Appliances (MPP architecture) are used • Technologies for Information Management are well established • MPP or Clusters or Grid are used for future scale • Specific tools needed for CPM are used
SCALED – 3ml	• For low latency/real time data ingestion, appropriate technologies are looked at *[e.g. CDC, Middleware (EAI, ESB, B2B), Data Services, APIs, Micro Services, EII]* • Active Data Warehouse (ADW) appliances • Usage of automation technologies such as BPM, Rules Engines, Business Activity Monitoring (BAM) • IT Delivery & Operations automation – handling scale & operational excellence • Information Management is digitized
SYNERGIZED – 4ml	• Data ingestion for semi-structured and unstructured information • NLP, Computer Vision, Case Based Reasoning, CEP engines • Advanced analytical capabilities across enterprise • Big Data for high volume data & multi-structured storages • Cognitive capabilities, (Semantic) Search Engines and interactive chat/voice bots • Cloudification to scale on demand • Data Virtualization; Data Privacy & Security
SELF-OPTIMIZED – 5ml	• Right balance of Human centric and Machine centric decision making • Institutionalization of different disciplines of AI *(Automated Forecasting & Control, Knowledge representation, Learning, Reasoning, Perception, NLP/G);* Self-Healing • Scaling compute and storage to much higher levels • Enable humans to take decisions faster, using visual analytics, immersive analytics (MR, VR, voice enablement)

At **2ml**, one would see the usage of **ETL suites to integrate different types of data across enterprise**, mainly in batch mode. **Enterprise Data Warehouse** along with **technologies for Information Management** would have been well established for Data Office. One would see mostly **EDW Appliances** (MPP architecture), Clusters, or Grid being used for future scale. Specific tools needed for CPM would have been established.

At **3ml**, to support **low latency/real time data ingestion**, some of the following technologies may be established: Change Data Capture (CDC), Middleware (EAI, ESB, and B2B), Data Services, APIs, Micro Services, EII, In-Memory databases and so on. The **data store would be an Active Data Warehouse (ADW) appliance** since it needs to handle the mixed workload. **Automation technologies** such as BPM, Rules Engines, and Business Activity Monitoring (BAM) would have been established. Automation of IT Delivery & Operations to handle scale and ensure operational excellence, **Information Management program** would have been already **digitized**.

At **4ml**, **data ingestion capabilities for semi-structured and unstructured information** (files or large messages) is needed. There is also a **need for NLP, Computer Vision, and Case Based Reasoning, CEP engines to process the raw data** and extract information. To support this, technologies such as Big Data (Hadoop and NoSQL) are available. (Semantic) Search Engines are available. **Advanced analytical capabilities** support business functions across the enterprise. **Cloudification** strategy would be in place to be able to scale on demand, along with right management of Data Privacy and Security needs. **Interactive chat or voice bots** and even **cognitive capabilities** may be available to support IT Operations. Some organizations may as well be using a **data virtualization layer** to access different types and sources of data sets.

At **5ml**, establishing the **right balance of human centric and machine centric decision making** is the key. Matured cognitive capabilities for automated decision making are integrated within the business processes. One would see the usage of **different disciplines of AI**, such as automated forecasting and control, knowledge representation, learning, reasoning, perception, NLP or NLG institutionalized. Usage of Graph DB for advanced analytical capabilities are common. Cloudification and even Quantum Computing are planned to scale the compute and storage to a much higher degree. To enable humans to take decisions faster, **visual analytics, immersive analytics (MR, VR, voice enablement) technologies** would have been established.

The maturity characteristics of the **Technology KRA** is summarized in the adjacent table. *(A **consolidated view** of the maturity characteristics spanning all the KRAs has been provided in **Appendix B**.)*

During the assessment, one would hence focus on the following Sub-KRAs to assess Technology maturity:

- **Technology Landscape**
- **Technology Centricity**
- **Non-functional Requirements**
- **Estate Modernization**
- **Technology Partnership**

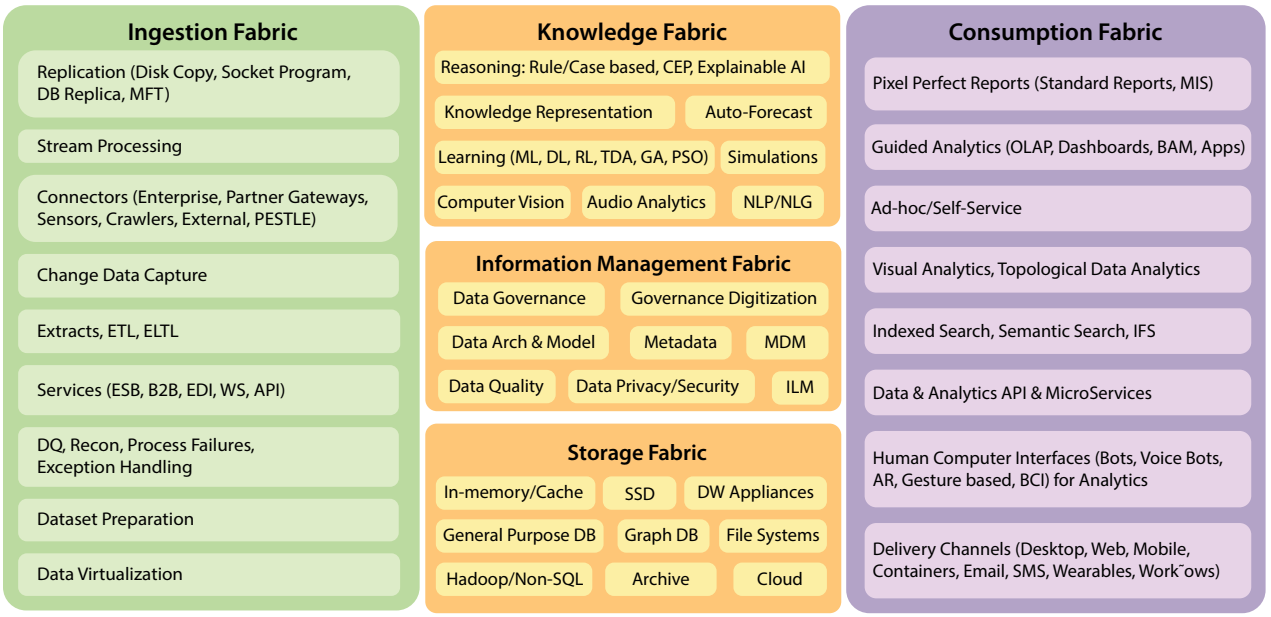
2.0.1 Sub-KRA: Technology Landscape

What technology components are available to support data and analytics?

Technology landscape encompasses all the key technology components required for **efficiently collecting, integrating, storing and making use of data for decision making**. This sub-KRA assesses the degree of maturity of the data and analytics technology landscape of an organization.

TCS DATOM™ provides effective guidance in terms of technology selection and technology adoption with reference to a comprehensive **solution fabric** established as part of the framework. The solution fabric provides a **holistic view of the different solution options** available for data integration, knowledge management and delivery of data and insights for business consumption. **It acts as a reference during assessments** and also helps to evaluate and establish the most suitable data and analytics technology landscape for an enterprise during roadmap creation.

The following diagram depicts the various technology components or solution fabrics that together make up a technology landscape:



▲ ***FIGURE 5:*** *Technology Reference Architecture*

The data that an organization works with, gets generated and resides in a number of separate data sources both internal and external. **Integration fabric** is a set of technology options that are used to **capture, validate, process, combine and store** all these data sets from different sources into an appropriate data and analytics repository to enable data driven decision making. For e.g. real time data processing needs can be handled using "Stream Processing", and requirements to capture and process incremental changes can be addressed through "Change Data Capture". (Informatica, IBM, Oracle, SAP, SAS provide some of the cutting edge data integration tools. Cloud service providers like AWS and Azure also provide comprehensive data integration capabilities. Dell Boomi & Mulesoft are examples of robust integration platforms which can enable real-time data integration as well.)

The **automation of the data integration process** cuts down on the time it takes to manually gather and combine datasets for decision making, significantly. Additionally, it also eliminates manual errors thereby increasing the reliability of data used for decision making. As data gets integrated into an enterprise repository, quality issues are identified and necessary improvements are implemented thereby enhancing data accuracy.

Storage Fabric describes the **various ways data can be stored** and is also dependent on the various consumption use cases. Thus, small to medium size structured data with normal response time requirements may be stored and managed using "General Purpose DB" technology options, while real time reporting needs will ideally require "In-Memory" storage. Effectiveness of storage and use of ontology based relationships may be enhanced through usage of an appropriate "Graph-DB". (HANA and Spark are few examples of capabilities using "In-Memory" storage, whereas, Oracle, DB2, MS SQL Server are few examples of "General Purpose DB"; Neo4j is an example of a Graph-DB.)

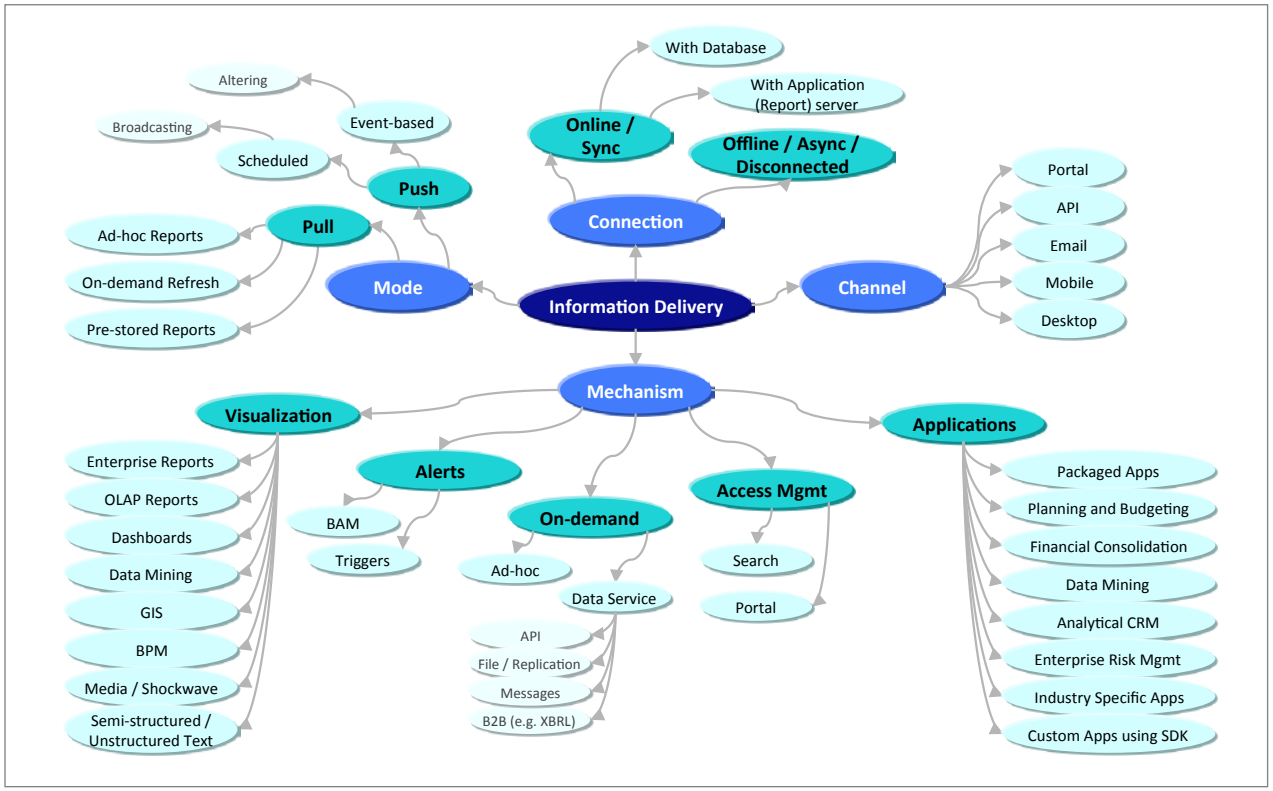
Information Management Fabric brings together the **governance related capabilities**. Depending on an organization's needs, one or more of these capabilities may need to be selected and enabling technologies identified to establish the overall technology landscape. (IBM Data Governance, Informatica, Collibra are some of the leading Data Governance capabilities.)

Decision Fabric provides the capability options for **the decision making ecosystem** and include descriptive, diagnostic, predictive, prescriptive, and pre-cognitive/cognitive capabilities. Capabilities are selected considering business needs and availability of required quality of input data. *Thus, to establish a Customer 360º view, "Dimensional Modeling & Analysis", a descriptive capability, may be used. On the other hand to predict Customer Churn, "Machine Learning", a cognitive capability, may need to be deployed. Appropriate technology options for the chosen capabilities are decided to establish the overall technology landscape for the enterprise. (Oracle, Business Objects and MicroStrategy are few examples which can enable dimensional Modeling. SAS, Data Bricks, TIBCO, R, etc., provide machine learning capabilities.)*

Delivery Fabric essentially provides the capability choices for the **information consumption layer** and includes various ways to disseminate information internally and externally. The information delivery solution is a resultant of choices primarily with respect to the **channel of delivery, the mode of delivery, the delivery mechanism and type of connection** with the information consumer. *For e.g. a) An organization might implement a real time, API based, Master Data Reference capability within the enterprise to enable all enterprise applications to relate to a single version of truth. (Informatica MDM Hub can be considered as a solution option in this situation). b) It might decide to deliver alerts to mobile devices of relevant stakeholders, whenever the waiting time of a priority issue exceeds the defined SLA. (Possible usage of SAP HANA based solution). c) The CEO's dashboard, might be a high quality GUI available in the Enterprise Portal and accessible to authorized users. (OBIEE, Tableau, Power-BI can be some of the solution options).*

Depending on the different information delivery needs for an organization appropriate solution options are selected from the delivery fabric to help establish the overall Data & Analytics technology landscape for the enterprise.

As an example of a solution fabric, the detailed information delivery fabric is depicted below:



▲ ***FIGURE 6:*** *Information Delivery Fabric*

2.0.2 Sub-KRA: Estate Modernization

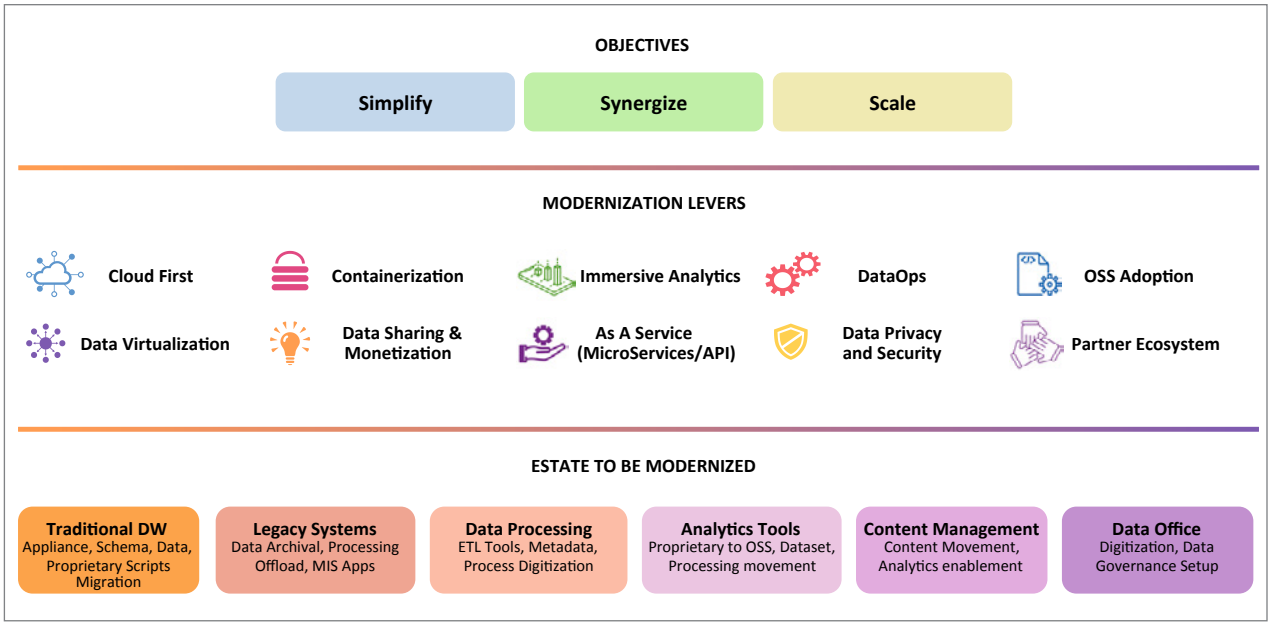
Is the organization able to harness the benefits of the latest technology options to enhance business or operational efficiencies?

It is important to **understand the relevance of modern technology options and use them effectively** to ensure competitive advantage and bring benefits to business. The technologies being deployed may improve business efficiency and effectiveness, may improve system availability and performance or can ensure standardization and reduce technical debt.

*(**Technical debt** is a concept in software development that reflects the implied cost of additional rework/necessary future enhancements, caused by choosing an easy/tactical solution, instead of using a better approach that would take longer and/or may be costlier.)*

Estate modernization evaluates effective use of emerging trends and technologies in the data management space to drive business agility, ease of use, as well as optimize costs. The objective here is to simplify the landscape, synergize the ecosystem and build the ability to scale as the organization grows.

For e.g., It evaluates whether an organization has the **capability of using large volume, wide variety and high velocity data** (wherever justified), internal and external, in order to generate effective insights. This is the realm of Big-data technologies.



▲ ***FIGURE 7:*** *Modernization levers*

It also checks **whether cloud-based solution options have been evaluated** to reduce infrastructure cost, improve agility and development cycle times, improve system availability and performance, improve governance, enable technology standardization, and reduce IT operational costs.

It **assesses propensity to use advanced capabilities** such as data virtualization, in-memory computing, Iota, real-time decision making and so on. It also looks at the **proliferation of legacy technologies** within the organization and **architectural blueprints to replace them**.

To improve delivery reliability and reduce release cycle times, **adoption and maturity of DevOps methodology** is also evaluated. In this scope, the focus is more on the technology enablement and end-to-end automation of IT Delivery and IT Operations. On a broader context, there are ways and means available today to **support automation** and hence it is important to ensure that there is minimum manual intervention to support either a business or operational process. One would typically deep dive and understand all the areas that have manual interventions and see if an automation solution is worth the investment.

The modernization aspect looks at the legacy technologies that the organization is still dealing with and identify and prioritize areas that can be explored. These can be:

- **DW Appliances** that are still part of the Ecosystem (Schema, Data, Proprietary Scripts Migration)
- **Mainframe** - Data Archival, Processing Offload, MIS Applications.
- **Data Processing Tools** - ETL Tools, Metadata, Process Digitization
- **Analytics Tools** - Proprietary to Open Source migration, Dataset, Processing movement
- **Content Management Tools** - Content Movement, Analytics enablement
- **Data Office** - Digitization, Data Governance Setup

As new technologies are planned, **appropriate guidelines should be formulated by the architecture team to ensure standardization and prevent compatibility issues** within the enterprise.

2.0.3 Sub-KRA: Technology Centricity

Does the organizational value system promote the usage and enrichment of technology and identifies it as a central factor to promote business efficiency?

Technology Centricity **measures the culture in the organization to promote new technology adoption** and technology debt management.

It evaluates how focused the organization is in terms **of continuously evolving its technology landscape for the betterment of the business**. This includes availability of a reference architecture, which can clearly lay out the use case and guidelines for each technology. What kind of investments are done to support the technology evolution? How is technology promoted? Is there a focus team who is responsible for evaluating newer technology options and how they can be best leveraged? Does the organization participate in vendor events or themselves sponsor ones, are key factors contributing towards technology centricity.

Technology centricity **also evaluates the actual utilization of the technology** with respect to business scenarios. How much of the available technology is really utilized and exploited? Is it tracked as a KPI? Are corrective steps defined for low usage and so on?

2.0.4 Sub-KRA: Technology Partnership

What partnerships the organization has in the Data and Analytics space and how is it benefitted from the same?

Technology partnership **assesses the relationships that the organization has in data and analytics space with external stakeholders**. These can be product vendors, service providers, academic institutes, training institutes and so on. It also assess' the mode of management of these relationships along with accompanied risks. Do partners provide regular trainings? Does the organization get early access to new product releases for exploration? How much does the organization contribute and to what extent does it influence the vendors?

[In this context it might be worth mentioning that in many organizations, the delivery and maintenance of analytical capabilities might be outsourced. (There can be different types of arrangements and nature of contracts with vendors in such scenarios. The arrangements can span from contracting for Insights-As-A-Service to complete outsourcing of data & analytics service delivery. Sometimes the infrastructure is owned by the parent organization and vendors log in remotely to deliver and manage capabilities and sometimes the infrastructure might also be owned by the vendor). In such situations, before awarding the contract itself, it needs to be ensured that the vendor is capable to fulfill the legal, regulatory, techno-functional and non-functional requirements and obligations of the parent

company. Pre-contract assessments and post-contract audits are carried out. Sometimes, especially for process maturity and data security, established industry certifications and third party audits are mandated. Service Level Agreements are agreed both for service delivery and for analytical capability performance which are monitored and managed. In case of data exchange requirements appropriate (secured) interfaces are designed, deployed, monitored and audited. The IT governance function of the parent company oversee the vendor operations and performance on a regular basis.]

2.0.5 Sub-KRA: Non-functional

How stable, efficient and usable are the existing Data and Analytics applications and systems?

Non-functional parameters are an important aspect to gauge the efficiency of the already established landscape in terms of performance, ease of maintenance, scalability and so on. Following criteria are important non-functional parameters to be considered.

Performance: This measures how well the deployed systems are performing, whether on the data integration side or the report consumption and analytics side. Are the performance SLA's being tracked? In addition, are they are in line with the business or end user expectations?

Availability: Availability is about the application/system availability measured as percentage of time a system is in operable condition. It can be for reporting applications or the database or data store itself for ad-hoc analysis.

Scalability: Can the system scale up and scale out to meet business growth targets? Are there any known issues with the underlying technology/infrastructure?

Usability: This measures how easily an end user can utilize the data and analytics application. Whether it is the reporting application, self-service application, analytics outcomes and so on. Is it easy for an end user to navigate through the application and find what he wants to find?

There are other aspects also which are looked at based on organizational contexts. These are **flexibility, agility, reliability, maintainability, extensibility, complexity, portability and security**. Since security have been already covered under Data security, they are not covered as part of this Sub-KRA.

3.0.0 KEY RESULT AREA – PROCESS

The focus of the Process KRA is to **understand the various aspects that are essential to keep the data and analytics landscape intact, aligned to business needs and continuously improving**. The focus is to understand the service delivery mechanisms and program management aspects of an organization. Essentially it encompasses everything from requirements to delivery to training and usage. Besides it also evaluates extent of business engagement and branding, degree of digitization and functioning of Data Office.

SILOED – 1ml	• Business Functions drive D&A requirements independently • Program Management processes at initial stages of maturity • Basic SDLC for project execution • Projects are short lived and aimed at delivering reporting/dashboard applications for specific use cases with limited scope
SIMPLIFIED – 2ml	• Business process maps created and linked with data architecture and consumption layer, to enable business consumption • Enterprise Scale Program management & Requirement Management capabilities • Factory processes deployed to bring scale in developing enterprise scale D&A capabilities • Established Data Office processes using Information Management framework • Established Center of Excellences • Enterprise scale D&A Initiatives
SCALED – 3ml	• Multi-modal delivery capability established • Rapid Application Development (RAD) used for operational analytics based process improvement initiatives • End-to-End Business Engagement • Competency Programs for Business Users • Digitization (/automation) across all Information Management processes including Requirements Management, Program Management, Knowledge Management, Competency Management, and so on
SYNERGIZED – 4ml	• Matured Agile processes to deliver Ecosystem Intelligence • Institutionalized Innovation processes. Processes such as Digital Re-imagination, Small Group Activity established • Reuse program extended to Decision Management itself • Stringent Data Privacy & Security processes
SELF-OPTIMIZED – 5ml	• The intelligent automation using AI would need processes of "Service as intelligent Software" to be established • AI Governance • Comprehensive Self-service capabilities • Self-healing capabilities for systems • Competency build on AI democratization, AI governance as well as self service analytics

Typically, at **1ml**, business functions drive data and analytics initiatives independently. **Program management processes are at initial stages of maturity**. Basic SDLC is used for data and analytics related project execution. Projects are generally of smaller scale and short lived (6-9 months) and aimed at delivering reports and dashboards for specific use cases with limited scope.

At **2ml**, **business process maps are available** which are linked with data architecture and consumption layer. This simplifies the data landscape for business consumption. Comprehensive **program management and requirements management processes are established** at enterprise level. Factory processes deployed to bring in scale in developing D&A capabilities *(ETL suites, reports, dashboards etc.)*. **Data Office processes are established** using Information Management framework. **Center of Excellences are set up** to develop capabilities, establish best practices, foster re-use, build knowledge repositories and provide design assurance. Transformational enterprise scale data and analytics initiatives are common.

At **3ml, multi-modal delivery capabilities are established. Rapid Application Development (RAD)** used for large number of operational analytics based process improvement initiatives. Business engagement spans the entire delivery life cycle. **Competency programs are planned and delivered** for large number of business users and all supporting processes are established. **Digitization and automation across all Information Management processes** including Requirements Management, Program Management, Knowledge Management, Competency Management, and so on.

At **4ml**, a fair amount of experimentation is needed while dealing with ecosystem and hence **agile processes are well established. Innovation is institutionalized**. Processes such as Digital Re-imagination and Small Group Activity are practiced. To improve agility and speed of execution with higher precision, reuse program is extended to decision management itself. **Stringent Data Privacy & Security processes** are put in place as more and more external data from ecosystem are captured and used.

At **5ml**, the **intelligent automation solutions using AI** would need processes of "service as intelligent software" to be established. As AI institutionalization grows, the **AI Governance becomes more critical** to channelize the benefits of AI instead of getting into ruptured situations. For humans to take decisions with speed, matured self-service capabilities and methods are established. Self-healing processes and capabilities for IT systems gets deployed. The competency build process on AI democratization, AI governance as well as self-service analytics gets higher focus.

The maturity characteristics of the **Process KRA** is summarized in the adjacent table. *(A **consolidated view** of the maturity characteristics spanning all the KRAs has been provided in **Appendix B**.)*

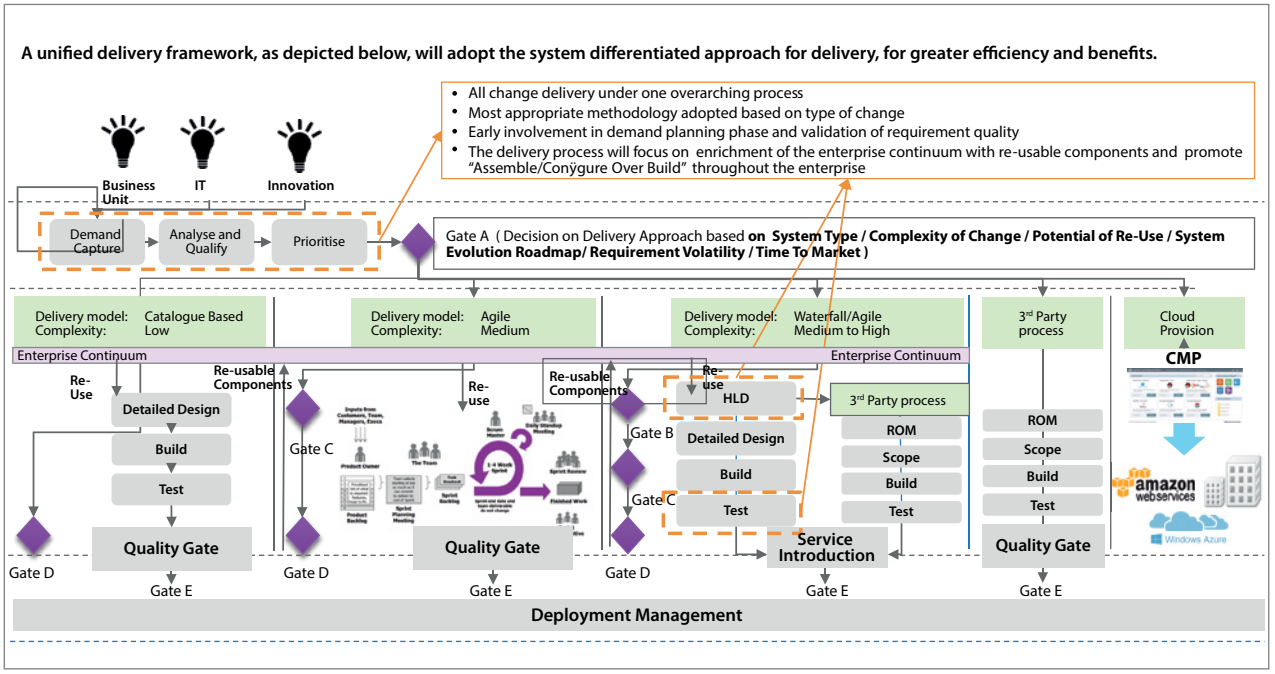
For the assessment, from process perspective, the following Sub-KRAs are assessed to determine the maturity:

- **Service model**
- **Program Management**
- **Digitization**
- **Business engagement**
- **Branding**
- **Data Office**

3.0.1 Sub-KRA: Service Model

How are Data and Analytics projects and programs delivered?

Service delivery is a set of processes and standards that govern the selection, execution and governance of a delivery model for data and analytics initiatives. It also defines the practices to be adopted for release management, change management and configuration management. Organizations **adopt a range of delivery models** ranging from **Factory model, Waterfall, Agile, and so on** to deliver the data and analytics initiatives. *An example of a unified/composite delivery process with multi-modal delivery capability with decision point on delivery approach is depicted in the diagram below for reference.*



▲ ***FIGURE 8:*** *Unified delivery process for multi-modal development*

For this sub-KRA, **the assessment is about an organization's ability to adopt the right delivery approach** to ensure optimal outcome with respect to business needs or end user requirements. Normal practice is to follow a **waterfall model** in which there are a sequence of steps from requirements gathering to implementation. This leads to longer delivery and change management cycles. In the face of changing requirements and priorities this approach is rendered ineffective. **Agile delivery methodology** handles this situation effectively and improves the "time to market" and provides early and continuous visibility to business users in terms of what they need. However, this works wells only if the right governance is in place. **Catalogue based (Factory)** delivery approach can be deployed for low complexity and repetitive tasks.

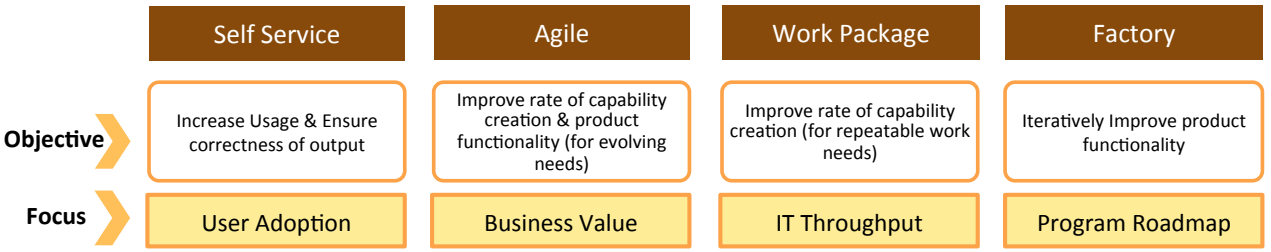
TCS DATOM™ prescribes that the organization should have processes and procedures defined to choose the right delivery model based on solution contexts, efficiently. As explained above (Fig. 8) it may so happen that for larger project or program deliveries, different models may be adopted. The following table depicts what are the different delivery models that are available and what they can be used for:

Service Model	Description	Accelerator	Productivity	Flexibility in team changes	Required proficiency	Cost of service
Factory based	Support, bug fixes, small enhancements and repetitive work	Productivity	High	High	Low	Low
Work-package based	• Clear scope of delivery • Max. 3 months of duration of projects • Any change in scope is delivered in next phase • Strict release cycles	Solution/ Framework	Medium	Medium	Medium	Medium
Agile based	• Unclear scope and hence iterative (or agile) delivery • Tight Business and IT interaction with Show & Tell approach • POC/Sandbox Env. • Production release with work-package approach	Innovation	Low	Low	High	High
Self-service	• Users want to handle end-to-end development without intervention	Time to Market	High	Low	High	Medium
Service as an Intelligent Software	• Software handles operations without human intervention	Cognitive Computing	Too High	Not Applicable	Not Applicable	Too Low

▲ ***FIGURE 9:*** *Service models*

This Sub-KRA would assess **if the organization is aware of various delivery models**. If yes, have they tried enhancing the business user experience by adopting the right composite? If they have not done that so far, where do they stand in terms of taking this aspect forward? In many organizations, the service delivery is completely outsourced. In such a situation, if the outsourced vendor is able to meet the business aspects of agility and time to market then they are likely to adopt the right practices. If not, a brief assessment of the practices being followed by the vendor should be carried out.

The next area of assessment is to **understand if the specific objective of adopting a service delivery model is being met or not**. Each of the service delivery models has a focus area and an objective to be accomplished. It is essential to understand and assess if the same is being delivered and that there are no path corrections needed. The following diagram depicts the expectations from each of the service delivery models:



▲ ***FIGURE 10:*** *Delivery Models Objective & Focus*

One needs to assess **if the organization has established processes and identified appropriate tools and capabilities that enable it to measure the success** of each of the delivery methods deployed.

The other important aspect of service model maturity is **the automation and/or digitization deployed as part of the service model**. This needs the following evaluation:

- Is the requirement management process digitized? To what extent is it integrated with delivery and release management process?
- Is the end-to-end traceability, from requirements to product features to test cases established?
- How automated are the development, testing and deployment processes?
- Does the deployment require the application to be taken down or can it be done without disrupting business as usual?
- How easy is it to rollback a deployment and will it be part of the quality gate checklist?
- Are all the defects captured correctly? Is the root cause analysis done?
- To what extent are delivery performance analysis and defect capture and analysis automated and digitized?

3.0.2 Sub-KRA: Business Engagement

How does the Business Team get involved through the data and analytics project or program delivery?

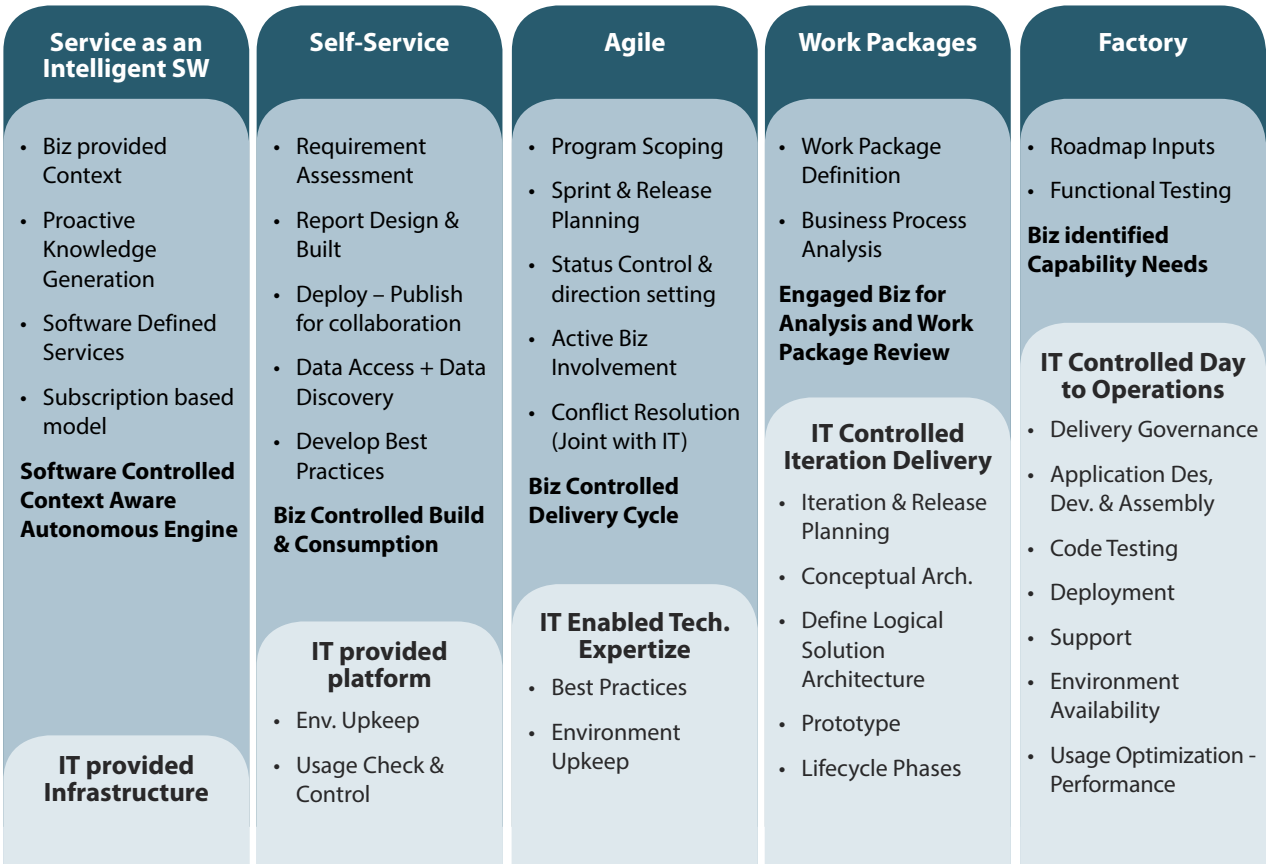
Business Engagement Sub-KRA measures the **level of participation of business users in data and analytics initiatives**. It is generally seen that the efficacy of data and analytics initiatives increases in proportion to the level of participation from business units. Activities that business users typically participate in are:

- **Roadmap Creation:** High-level roadmap for data and analytics function maintaining close alignment with business roadmaps and plans.

- **Business Architecture:** Putting up the big picture that depicts high-level business functions and their sub functions, inter-relationships between them and interfaces with external actors. It also describes the key business processes, maps them to the business functions and highlights the KPIs or measures being computed for each of them. Business architecture provides the foundation to establish the data requirements for each department.
- **Requirement Management:** Defining business requirements and business rules, ascertaining business benefits or ROI and driving prioritization and release management for these requirements.
- **Design Reviews:** Design is generally owned by the technical team but certain components of the design needs business involvement such as user interface design, data design, use of appropriate data, defining right archival policies, defining right rules for required KPIs and so on.
- **Functional & User Acceptance Testing:** To ensure functionality being delivered is in line with the original requirement and is fit for use.
- **Data Governance:** Business owns the data and it should be governed as the most expensive asset of the organization. Business must have an active role in all phases of data journey right from data creation to its destruction.

The **level of business engagement is different in each of the service delivery models** as well. Business Engagement Sub-KRA **measures the level of participation of business users in data and analytics initiatives**. The following diagram depicts how IT and Business teams will be involved in the different delivery models:

▼ ***FIGURE 11:*** *The Business-IT Conundrum*



One would have to take a call based on the organization structure to assess if the involvement from both IT and Business teams is right or if there are areas of improvements that can be identified.

3.0.3 Sub-KRA: Program Management

How efficient are the processes around Program Management of the Data and Analytics Environment?

Program management is the broad umbrella that encompasses Requirements Management, Estimation & Planning, Maintaining Standards & Guidelines, Communication, Change Management, Metrics Reporting, Assurance, Innovation, Value Measurement, and so on.

Program management office plays a vital role in ensuring the overall success of data and analytics initiatives. *For example, in the absence of a common estimation template to be used across the enterprise, it will be difficult to compare the ROI of all these initiatives and take the right business decision in terms of prioritizing one over another.*

Communication is another important pillar of good program management principle. How are the organization level announcements for data and analytics initiatives made to flow through the organization? Is there a defined process to manage the same? How does the user community get to know the initiatives that are being taken up and how they can contribute towards the same? These are important factors for achieving the desired scale in data and analytics initiatives.

Change management is one of the core and important processes to be managed by the Program Management office and covers other sub areas within it. It is necessary t to understand the changes needed by the business and have the right processes in place to ensure that the changes flows through cohesively within the enterprise. Along with this, the **Assurance** function needs to ensure that there are right processes in place to ensure that business happens as usual without any disruptions.

Program management office is also responsible for **setting up a measurement system in place for data initiatives**. The measurement system aims to define key KPIs to be monitored for each program, the frequency at which they should be measured and whom each KPI has to be reported. This helps track the success of the project through its life cycle and enables targeted executive communications.

The **measurement system should also look at evaluating the ROI of the project** and should be able to tie it back to the original business case.

Other things that need to be fostered is an innovative culture. A culture that will help each individual to contribute creatively towards generating new ideas in their field of influence. This can be done systematically where individuals can log ideas, visualize which of their ideas were selected. A system will also help implement a process to selectively curate ideas and take them forward.

3.0.4 Sub-KRA: Branding

How is the data and analytics landscape being promoted in the organization?

Branding aims to understand how an organization encourages data driven decision making and promotes it through usage of various analytical capabilities. When the top most executives of the organization promote data as the driver of any decision, the maturity journey of data and analytics in that organization gains momentum.

Building a data driven organization is a big cultural change that one needs to bring in and it requires awareness to be brought in. Internal marketing is an important step to make users aware of the benefits of data driven decision making. Using various platforms and forums such as knowledge magazines, internal campaigns, innovation days and so on, one needs to encourage user participation. All these require focused attention. **Usually, the failure of a data and analytics ecosystem is blamed on system features**, but in most of the cases, it is actually due to low acceptability of the system. The system usually gets a low visibility and is considered a non-mission critical system. Business can continue as is, without it being in place. **This is where branding and promoting the data and analytics applications help**.

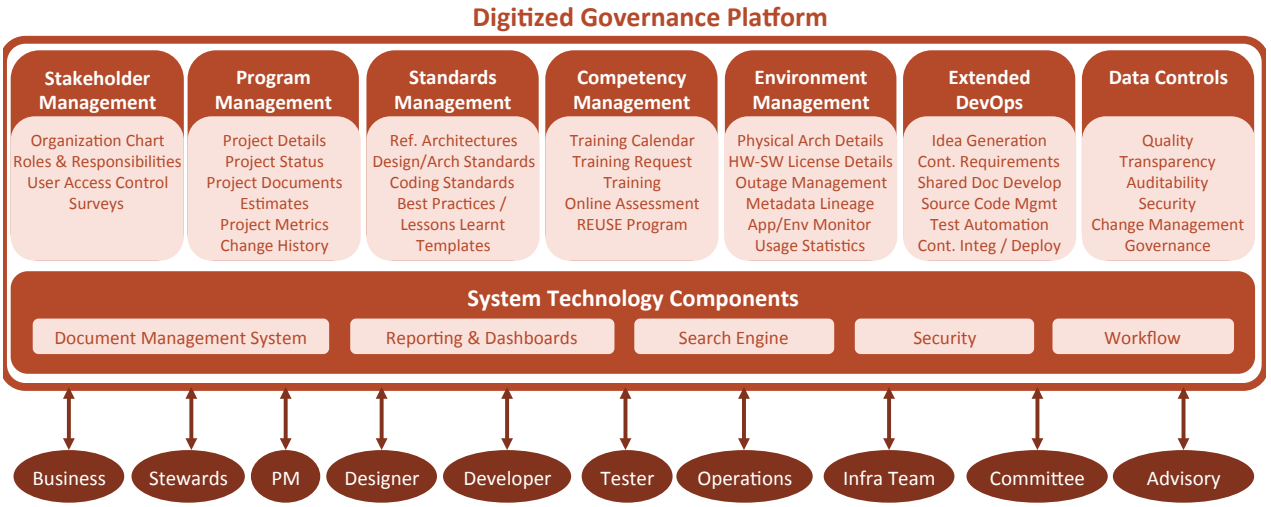
Apart from internal branding, one should **encourage participation in external events**. Once you see top executive teams promoting data at external events, the importance and acceptability within the organization as well as the avenues of building a healthier ecosystem of partners and affiliates increase.

3.0.5 Sub-KRA: Digitization

What is the level at which the data and analytics related process are digitized?

The whole focus of digitization is to get better visibility and hence better control. In the context of TCS DATOM™, digitization includes digitization of processes like delivery process monitoring and control, data quality monitoring, workflow management for business glossary or Metadata update, information life cycle management, usage statistics monitoring of dashboards and reports to evaluate performance, requirement management and traceability of the same.

The following picture depicts the indicative capabilities and processes that should be digitized in an organization:



▲ ***FIGURE 12:*** *Digitized Governance Platform*

3.0.6 Sub-KRA: Data Office

Is there an established Data Office to support data and analytics initiatives?

Data is omnipresent in an organization. A typical IT landscape will involve multiple systems and applications who are either generators or consumers of data in the organization.

While the ownership of quality, completeness, availability, interoperability and so on, can be maintained at the system level, data and analytics initiatives gain from setting up a centralized data office that oversees these activities across the enterprise. **Data Office helps define cohesive standards and guidelines** that can be followed across the organization for data management, data sharing, and data security and so on. **These cohesive processes enable better governance** of an enterprise's data assets compared to when it is done in a silo mode. As the maturity in the data and analytics area enhances, the need for establishing an organization that is responsible for maintaining and managing all the data assets become very important to ensure availability of quality data, on time for effective decision making.

4.0.0 KEY RESULT AREA – PEOPLE

The next important aspect of Data Maturity is the people supporting these initiatives. It is important that **data initiatives be supported at all levels across the hierarchy**. The business and top executive level driving data maturity helps in gathering a lot of momentum and is a critical factor in supporting a sustainable maturity. In this KRA, the team compositions in both Business and IT teams, the competency enablement programs available and knowledge management are assessed. A brief assessment of adoption of Data and Analytics within the user community, their feedback and involvement in this area is also done.

SILOED – 1ml	• Mainly IT driven D&A initiatives • Business Users, IT teams does not have much exposure to analytical capabilities. Statisticians and Data Scientists may be used sometimes • Very limited business involvement. Mostly during requirement capture and acceptance testing
SIMPLIFIED – 2ml	• Involvement of Business Users & Sponsors throughout a D&A program • Enterprise Architects develop end-to-end solution architecture; Chief Data Officer handle Data Office and enforce data practices • Program Managers & Visionary Advisory available for large scale program execution • Comprehensive knowledge management • Competency management initiated
SCALED – 3ml	• People with skills to handle low latency/real time data interfaces • Digitization and automation experts to establish IM Digitization • Competency enablement of Business users (e.g. how to define business rules in a given technology; Self-service reporting) • Trainers with ability to train operational users • Operations Research skills for optimization at scale
SYNERGIZED – 4ml	• Agile program executioners to execute transformational programs • Active involvement of Data Scientists • Strong legal advisors to understand legal & regulatory implications of ecosystem data. Enables data monetization platforms • User Experience skills are important • Matured Enterprise Continuum
SELF-OPTIMIZED – 5ml	• AI skills across different branches to build AI programs and agents • Besides Data Scientists, neuro-scientists and psychologists needed for AI • Visual Analytics and self-service analytics skills for business users • Human Computer Interface skills to build Immersive Analytics interactions • Quantum Computing skills to manage scale

At **1ml**, one observes mainly **IT driven data and analytics initiatives**. Business users do not have much exposure to analytical capabilities. Statisticians and Data Scientists may be doing data mining. **Business involvement is at a minimal**, possibly only to provide requirements and test deliverables.

At **2ml**, one observes the **involvement of business users and sponsors throughout the program**. Enterprise Architects are there to ensure business architecture is mapped to

Information Architecture and Application Architecture, and then to Technology and Infrastructure. Chief Data Officers handle the Data Office and enforce data practices at the enterprise level. For large-scale program execution, Program Managers and Visionary Advisory are available. Comprehensive knowledge management processes at an enterprise level has been established. Competency management has been initiated.

At **3ml**, people with **skills to handle low latency/real time data interfaces** are necessary. Digitization and automation experts are needed to establish Information Management digitization. **Business users will be technology savvy** and will know how to define business rules in a given technology. Comprehensive competency management plans for business users across all layers. Trainers with ability to train operational users on technology will be available. Operations Research skills are necessary to bring process optimization at scale.

At **4ml**, **agile program executioners** are seen executing business transformation programs. Active involvement of **Data Scientists** who become essential at this stage. Strong legal advisors needed to understand the implications (specifically regulatory), of ecosystem data. This also helps in terms of building data marketplace or data monetization platforms. **User experience skills become important** and are available. In terms of knowledge management, a matured enterprise continuum gets established.

At **5ml, AI skills across different branches** are essential to build or configure AI programs and agents. Along with data scientist skills, specialized skills such as neuro-scientists and psychologists will be important to handle AI. Visual analytics and self-service analytics skills are expected from business users. Human computer interface skills to build immersive analytics interactions will also be needed. To improvise the scale, Quantum Computing skills may be essential at this level.

The maturity characteristics of the **People KRA** is summarized in the adjacent table. *(A **consolidated view** of the maturity characteristics spanning all the KRAs has been provided in **Appendix B**.)*

To assess the maturity of the People KRA, the following Sub-KRAs will be analyzed:

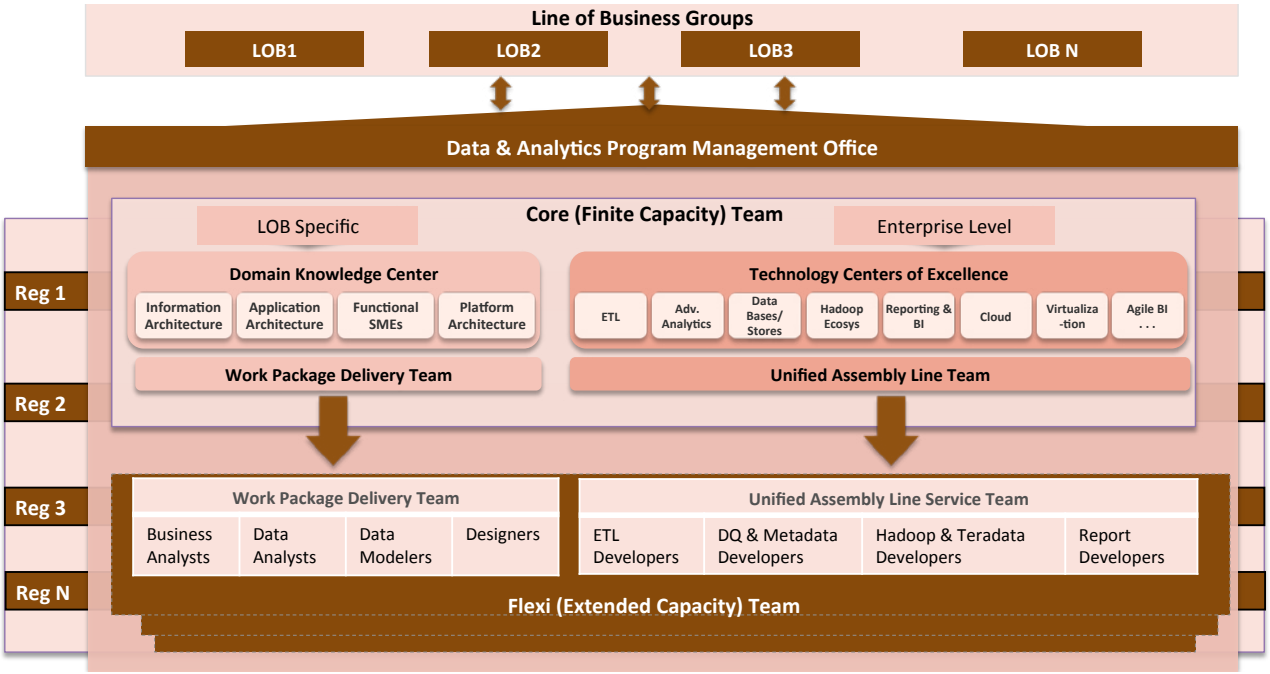
- **Team Composition**
- **Knowledge Management**
- **Competency Enablement**
- **User Maturity**

4.0.1 Sub-KRA: Team Composition

How are the teams involved in the data and analytics landscape organized?

Success of a data and analytics organization rests heavily on the **presence of right teams and right roles** with defined responsibilities. An effective data and analytics function will need various roles like data stewards, business analysts, data architects, data scientists, program managers, solution architects, data officer, developers and so on. Common team organizations are Data Office, Architecture Board, Program Management Office, D&A Delivery, Center of Excellence etc. These teams will need an **optimum mix of people**, with appropriate roles, from IT as well as Business units. There are certain roles, which should be owned by business while some needs a presence from both business and IT. Besides, typically there are finite capacity "core teams" and extended capacity "flex teams" (mostly in delivery organizations). This helps to scale up and scale down in terms of manpower capacity, depending on business needs.

*The following diagram depicts **a sample operating model** involving data & analytics team organizations.*



▲ ***FIGURE 13:*** *Global organization operating model - sample*

It is worth mentioning here that **each role in the diagram does not need a separate person**. A single role can be enacted by more than one person and vice versa depending on a particular situation and the size of the organization.

4.0.2 Sub-KRA: Competency

Is there a structured route via which people can build upon data and analytics capabilities?

By definition, competency is a combination of observable and measurable knowledge, skills, abilities and personal attributes that contribute to enhanced employee performance and ultimately result in organizational success.

In the context of TCS DATOM™, **competency mainly refers to competencies relevant to data and analytics space**, but is not limited to IT teams since business users play a significant role in the data and analytics maturity of an organization.

Competency management process typically covers phases like:

- **Competency base lining** - Building competency maps, setting up competency assessment frameworks to establish competency gaps etc.
- **Competency building** - Identifying the required competencies, defining the content and mode of trainings such as, instructor-led, virtual, self-paced learning programs.
- **Continuous monitoring and reporting** - Tracking progress by setting up competency dashboards, monthly or quarterly status reports and reviewing competency gaps to drive closure.

TCS DATOM™ recommends a Competency on Demand Program to achieve economy of scale in competency building. Employees should be provided with learning labs along with training and mentoring platforms to build their skills on different data and analytics competencies. Assessment is centered to certify and track achievements and a collaboration platform to exchange learnings.

This is the area where one would assess how the mentioned aspects on competency are addressed by the organization.

4.0.3 Sub-KRA: Knowledge Management

How easily is knowledge available?

Knowledge Management platform is **ideally a digitized enterprise repository where all artifacts that contribute towards an organization's knowledge are stored**. These artifacts may include design patterns, best practices, re-usable components, standards and checklists, data architectures and models, business process maps, design documents, architectural policies, KPI lists, data dictionary, training modules and so on. *(With reference to TOGAF parlance, in a matured state, the repository takes the form of an Enterprise Continuum.)*

Knowledge management is not a onetime process and hence **a comprehensive governance process enabled by an appropriate tool/technology needs to be established**. This will streamline processes such as how these artifacts are managed, who owns the artefact, who can update, who needs to review before any change is accepted and so on.

Organization of the artifacts should consider **ease of retrieval** and content should be indexed and searchable.

4.0.4 Sub-KRA: User Maturity

User Maturity denotes **the maturity of business users in using data and insights to improve business performance**. For an organization to become data centric, it is imperative that business users are aware of the value that data and analytics can bring to their business functions or processes. They should also be aware of the broad use cases or patterns on how data and analytics is being leveraged in their industry. Users should be conversant with the modern visualization platforms and basic tools to adopt self-service for faster response to business demands.

Some examples of the aspects assessed within this sub-KRA are as follows: Are the users aware of the organizational data assets? Do they use automated reports or do they prepare reports manually? Can they independently access and utilize data for analysis whenever required, using relevant data discovery, visualization or other analytical tools? Are they using the reports and analysis outputs for their decision making?

Essentially, the focus is to ascertain whether the business users only produce data for consumption of higher ups or are they capable to take decisions at their own levels using established data and analytics capabilities in the organization and operate effectively as **data consumers**.

CONCLUSION

The KRAs & Sub-KRAs described above **form the grain of TCS DATOM™ framework**, since each one of these measures a distinct capability of an organization in the data and analytics space. They are the central theme in an organization's data & analytics maturity assessment, and can assure an onward journey with certainty. The learnings from the assessment are leveraged to scale up and accommodate evolutions in an organization's data & analytics journey. Further, these enable organizations to achieve their business goals by simplifying the operating models, and setting up frameworks to interconnect Data, Technology, Process and People.

4 TCS DATOM™ Assessment Process

TCS DATOM™ assessment helps organizations to reflect on the importance of data and to evaluate degree of data usage in the decision making process. It empowers organizations to better orchestrate their data initiatives, and cascade data-driven advantages throughout the enterprise to make smarter decisions for driving business growth and staying ahead of the curve.

It assesses, to what extent each of the capabilities related to the sub-KRAs within each KRA (described in section 3) **are established and industrialized as enterprise capabilities**. How are these working in close co-ordination with each other and have enabled the organization to go beyond and learn about the ecosystem and its players and optimize the relationships. Further, it also evaluates whether they help the organization to innovate and strive to identify, understand and mitigate unknown external risks as far as possible. *For example, along with the aspect of data quality itself, **it is important to establish a matured enterprise process** as well, to ensure quality of data repeatedly and consistently to enable effective decision management. Similarly, an established program management capability is necessary to ensure that right data and analytics capabilities are identified and prioritized at an enterprise level and delivered efficiently using an appropriate service model to enable business functions.*

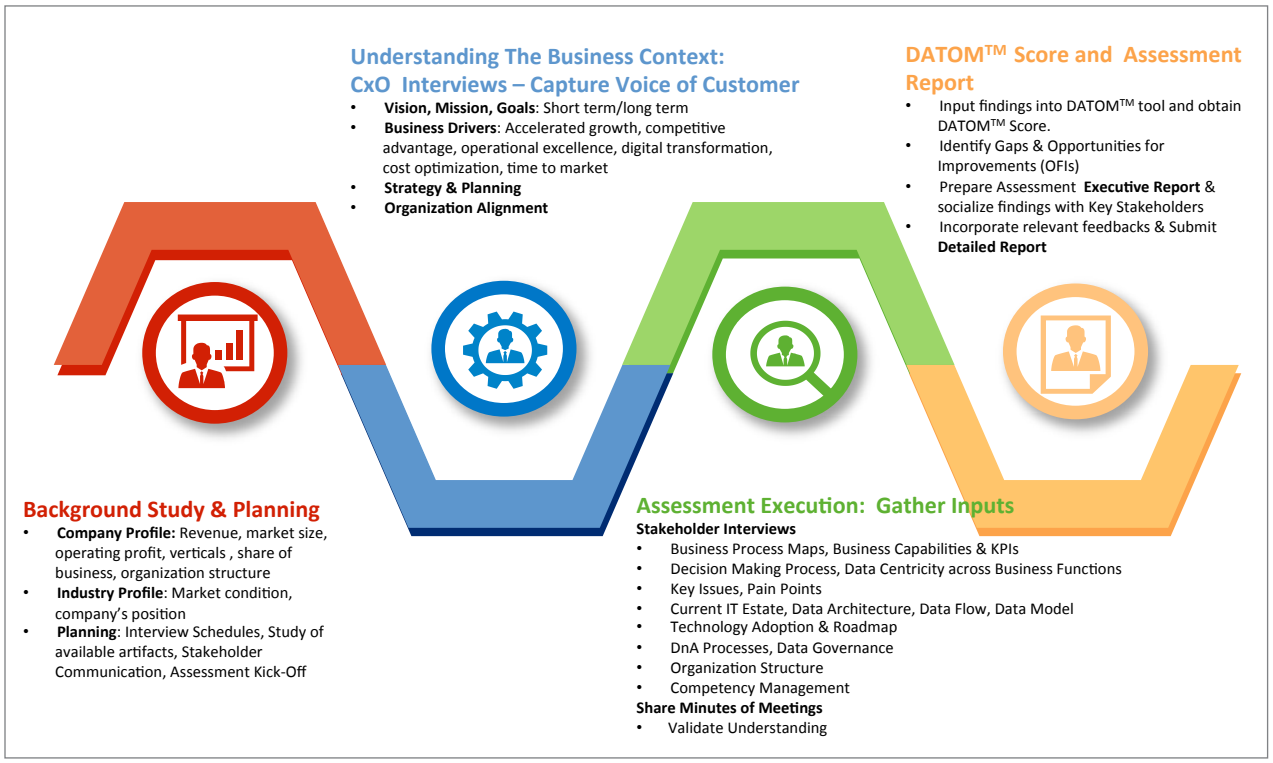
The assessment process has been established and is conducted in accordance with the **TCS proprietary TCS DATOM™ framework** which is based on decades of TCS experience in this domain. It looks at data and analytics initiatives holistically and in a structured manner and assesses an enterprise's data maturity level by how its data is managed and used across the organization. It identifies gaps in the existing data and analytics programs and suggests specific approaches for plugging these gaps.

The **process is enabled by a comprehensive TCS DATOM™ assessment tool**, which incorporates the entire assessment questionnaire across all the KRAs and sub-KRAs and facilitates capture of responses through an intuitive GUI. Based on the responses, the tool has the capability of scoring an organizations maturity at the sub-KRA, KRA and at the aggregated/ organizational level. Besides, it has the ability to benchmark an organization's maturity, relative to other organizations in similar industries, who have already undergone this assessment. *(More details on the assessment tool have been provided in **Appendix E**.)*

Apart from establishing the maturity level, the assessment process is designed to identify **opportunities for improvements** as well, across the KRAs and sub-KRAs. These findings enable organizations to build comprehensive data and analytics strategies in alignment with their business objectives and goals and use data and analytics as a strategic capability. *Implementation outcomes of these strategies are simplification of operating models, setting up of data governance models, identification of correct technology patterns and architectures and identification and development of required analytical capabilities to ensure data and analytics programs truly drive business growth.*

As part of the overall assessment process, there can be a **self-assessment** whereby an organization is provided direct access to the assessment tool and the authorized stakeholders of the organization carry out a self-assessment by responding to the questions based on their understanding. This provides an organization a first-hand experience in terms of the assessment process. The self-assessment score, can give an indication of the Data & Analytics maturity and can provide an early visibility in terms of the strengths and weaknesses.

TCS DATOM™ assessment is carried out **independent** to the self-assessment and the overall process is depicted below:



▲ ***FIGURE 14:*** *TCS DATOM™ Assessment Process*

Broadly, the assessment process has the following phases:

A. Background Study & Planning:

This is essentially **a pre-assessment phase where the assesse company's profile** and the industry profile in which it is operating are studied in reasonable depth to understand the business environment, its current size, products and services, financial performance, organization structure, market position, key competitors, challenges and risks. Additionally any other input providing information with respect to the organization's operations, information management capabilities, technology landscape, ongoing initiatives, are studied to gather as much information as possible.

Besides, **detailed assessment planning** in terms of overall assessment plan (schedule, resource and governance), stakeholder identification, stakeholder communication, preparation and finalization of interview schedules and assessment kick-off are carried out in this phase. The list of standard documents and artifacts required for the assessment are also shared with the relevant stakeholders.

B. Understanding the Business Context:

This phase involves **discussions with the organization's leadership team** to understand the key business drivers, current business strategy, short term and long term business objectives and plans of the company in the context of its Vision, Mission and Goals. The leadership's observations and views in terms of the degree of data usage in the decision making process across the organization, the current challenges & pain points and the organizations ongoing initiatives and future plans towards a higher maturity are also elicited.

C. Assessment Execution:

Interviews are conducted with each of the identified business functions to understand their scope of operations, key processes, systems and technology landscape, current data and analytics capabilities and degree of data usage during decision making across the operational, tactical and strategic layers. Current challenges and pain points, any ongoing data & analytics initiatives and future roadmaps are also discussed. **Key inputs required for the assessment across the four KRAs and sub-KRAs and relevant for the specific business function are elicited**. Representatives from all the organizational layers namely strategic, management control and operational are interviewed during the assessment to have **a holistic view of the as-is state** of the organization.

Apart from the business functions, **meetings are conducted with the IT Organization** to understand how and to what extent it is enabling the business functions towards data driven decision making. As-Is Enterprise level Information Management capabilities across People, Process, Technology and Data KRAs, ongoing initiatives and future roadmaps are discussed.

Detailed Minutes of Meetings (MoMs) are prepared and shared with all relevant stakeholders after each session to validate the captured inputs and to avoid any communication gap. Feedbacks are discussed and incorporated. Potential Opportunities for Improvement are identified. **The inputs captured in the MoMs form the basis of responses to the questions in the TCS DATOM™ assessment tool**. Also, artifacts required as evidences are requested from the relevant stakeholders during the assessment meetings.

D. Establishing TCS DATOM™ Score & Sharing Assessment Report:

Once the interviews/meetings are completed, the **assessors analyze the collected inputs and received artifacts and collectively discusses and agrees on the responses to each of the questions in the TCS DATOM™ assessment tool**. The Opportunities for Improvements in the context of the specific organization are also consolidated, discussed and finalized. The inputs are fed into the tool, relevant evidences are uploaded and necessary comments documented. Subsequently, the entire assessment is reviewed by a review panel, which includes the Lead Assessor, and finally approved, once all the comments are closed.

The assessment score, findings and opportunities for improvement are then socialized with the key stakeholders of the assesse organization. Any relevant feedback is incorporated.

Subsequently, the **DATOM™ Assessment Executive Report** is prepared and presented to the organization's Board Members/Leadership Team.

Finally, the **TCS DATOM™ Assessment Detailed Report** is submitted after incorporating all relevant feedbacks from the Executive Meeting.

5 Interpreting the TCS DATOM™ Score

TCS DATOM™ score indicates the maturity of an organization and ranges from zero to five. The following section provides an explanation in terms of how to interpret the score.

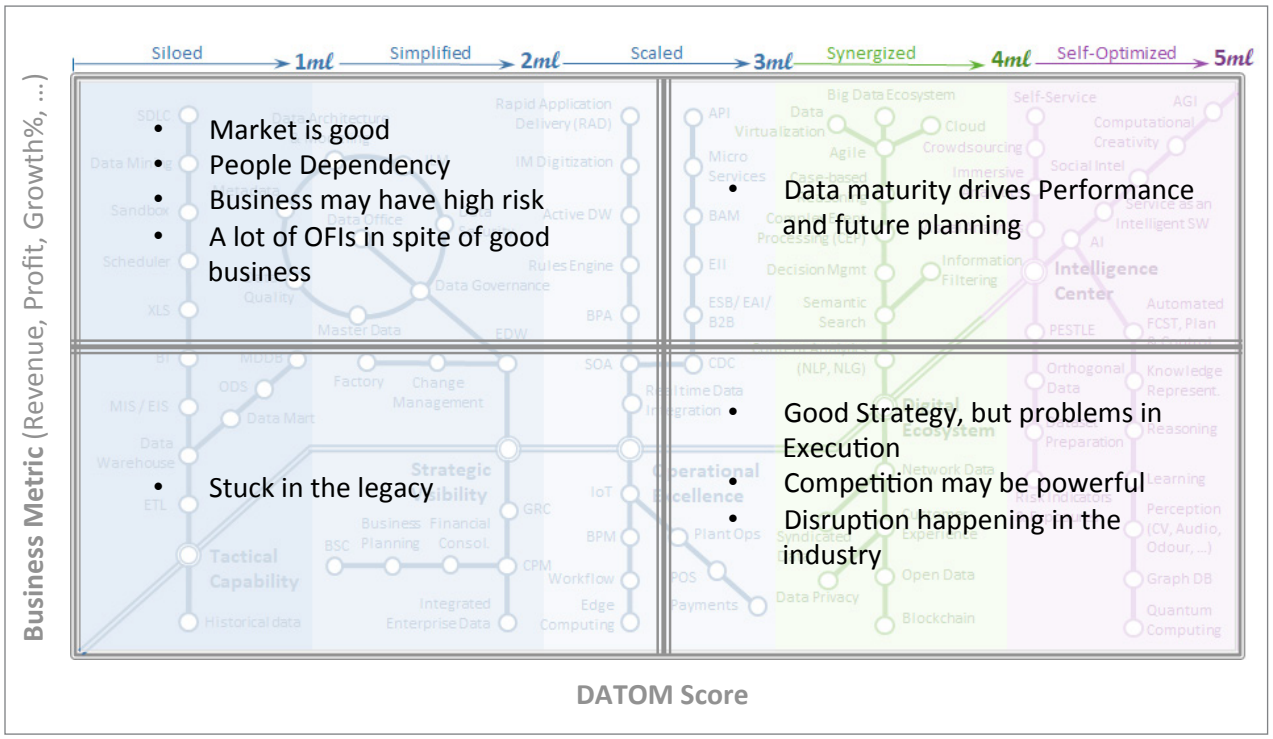
- **Score 1: Siloed**: Essentially reflects on an **Independent and Tactical capability** confined within Business Units/Lobs/Area Business Offices or Functions
- **Score 2: Simplified**: Represents the **Standardized Capabilities** at the enterprise level which has been or can be deployed across Business Units, thus providing strategic visibility for the Enterprise – though at simplified level this is still basic. *For instance, let us say a data ingestion framework is used consistently across different projects/initiatives, but metric definitions might still be different for different BUs. Which means there is some standardization but it is still too early to say that the organization has reached scaled level.*
- **Score 3: Scaled**: Represents a more **mature Enterprise level capability**, which enabled **Operational Excellence**, *for e.g. real time decisions at enterprise level, or consistent reporting and analytics with a proactive control on data issues. So, in effect, more mature and comprehensive enterprise level capability – the highest that can be achieved within the boundaries of the enterprise.*
- **Score 4: Synergized**: Represents a Scaled Enterprise Capability + Established Digital ecosystem enabling complex event analytics across an enterprise and its affiliates. Therefore, this is where organizations venture into an area of analytics which provides a view on a wider ecosystem while already having established enterprise level capabilities. At this level, the data and analytics function is expected to be managed with digitized data governance and processes.
- **Score 5: Optimized**: An Optimized organization represents Intelligent Automation driven balance between human centric and machine centric decision making, that can allow broader ecosystem and PESTLE level analytics. At this level, Data, BI., Analytics and AI all work in a systemic manner and use of different disciplines of AI (Automated Forecasting & Control, Knowledge representation, Learning, Reasoning, Perception, NLP/G) are expected to be institutionalized in terms of user adoption, technology and governance.

TCS DATOM™ Index when analyzed with the KRA and sub-KRA indexes **provides deep insights** to an organization and helps to identify **specific areas of improvements**. On similar lines, more insights can be derived by analyzing the TCS DATOM™ Index in multiple different contexts as mentioned below:

- Where does the organization stand vis-à-vis its own vision/mission/goals to be data driven high performance enterprise?
- How does the organization score vis-à-vis other organizations involved in the same business or related business?
- What are different Industries doing? Can best practices be adopted from other industries to innovate?
- What is the co-relation between the TCS DATOM™ Index and Business growth metrics indicating?

It may be possible that due to the unique nature and complexity of an Industry the ability to exploit data is restricted. Hence, **a low score may not always be interpreted as a bad score**. It needs to be analyzed using the above context before concluding on the organization performance.

One way to understand the significance of the TCS DATOM™ score is to **plot it with the growth metrics**. Growth metrics cannot be the absolute numbers. One can looks at metrics like percentage growth in EBITA and plot them along with the TCS DATOM™ Index.



▲ ***FIGURE 15:*** *TCS DATOM™ Score Vs Business Performance*

If the **plot falls in the bottom left quadrant** then that essentially indicates that the **business is not growing at the expected pace**. The lower TCS DATOM™ score here also indicates that the **organization is not effectively utilizing its data for decision making**. Decade old process are being used for business operations and is stuck in legacy. Data is not being used as a lever to build efficiencies, to understand customer needs and market trends.

If one has **excellent business metrics and a low TCS DATOM™ score** then there may be a couple of reasons for the same. The **market may be overall good** for the business and there are very few players, and hence growth targets are easily achievable. The other important aspect could be that **the business is supported by excellent talent pool**. This situation has its own pros and cons; unless the tacit knowledge is converted to processes, the organization carries a huge risk for people dependency. Hence, though the organization is already doing well **there would be still a lot of areas of improvements** to be worked upon.

The next quadrant is that **TCS DATOM™ score is high but the business metrics are not as good**. This means that the organization already considers data to be the key focus area but may be because of a strong competition or a disruption in the industry, it is not able to do well. It might

also be because the execution or operations are weak and hence in-spite of a good strategy the time to market is possibly not good.

If the organization's TCS DATOM™ score and business metrics are both high then definitely it is on the right track and Data Maturity is driving its performance.

The **opportunities of improvements** are derived based on the way the leading organizations in the same industry are performing, the best practices adopted and where the industry as a whole is approaching. It is essential to keep a tab on the way an organization is progressing with its maturity level. **Unlike the process world, the data world is continuously evolving as technology enables to do more with data**.

It may not be always viable for an organization to reach the highest maturity level. In addition, it may not be a sequential journey for an organization to reach the levels one after the other since there can be a lot of diversity within the various departments or functions within the organization as well. The key is to be aware of the current state of the organization and compare its position within the same industry and across different industry verticals, in terms of the best practices adopted. **It is important to understand the Industry trends and work towards reaching the optimum maturity level** in line with the organization's strategic objectives and vision, mission and goals. A strategy needs to be built and a roadmap to be laid out for its maturity progression. A governance mechanism to ensure that the roadmap is being followed, is of utmost importance.

Appendix A

Glossary

Anonymization

The sanitization of personally or financially identifiable data for the purpose of privacy protection, so that the sanitized data can be used for archiving, other kinds of analysis and testing. **In this process the Personally Identifiable information (PII) are converted to non-identifying data**. The contents of personally identifiable fields are replaced in a way, such that it cannot be associated with specific individual, project or company. For example, IP addresses, usernames, SSN, zip codes.

Data Encryption is a popular anonymization technique, which replaces sensitive data with encrypted data. This process provides effective data confidentiality, but also transforms data into an unreadable format. It is often not very suitable for practical purpose where the anonymized data need to be used further for analysis or testing.

Substitution consists of replacing the contents of a database column with data from a predefined list of fictitious but similar data types so it cannot be traced to the original subject. ***Shuffling*** is similar to substitution, except the anonymized data is derived from the column itself.

Number and date variance are useful data anonymization techniques for numeric and date columns. The algorithm involves modifying each value in a column by some random percentage of its real value to significantly alter the data to an untraceable point.

Nulling out consists of simply removing sensitive data by deleting it from the shared data set. However, this might pose challenges while doing analysis post anonymization.

All of these techniques provide an effective process to protect personal information and assist in preserving privacy.

Business Architecture

Business Architecture is the term used to depict the artifacts that visually represents the high-level business functions and their sub-functions, inter-relationships between them and interfaces with external actors. This can be for a system being built or for an enterprise as well. It also describes the key business processes and the detailed sub-processes (L0 to L4), maps them to the business functions and describes the KPIs used to measure each of them and how they map back to the organization's vision, mission and goals. Business architecture provides the foundation to establish the application, data and technical architectures. This live document is updated as the organization evolves and adopts or changes the various business processes it follows.

Business Metadata

Within a data store, be it a source application or a data mart or Data warehouse or a data lake, the term business metadata represents the **functional or business description of any subject area, entity, element**. It may include information such as the business owner, definition, scope, business rules, required for its derivation, applicable constraints, allowable domain values, security classification and relation to other entities and attributes.

It is usually defined at three levels of abstraction:

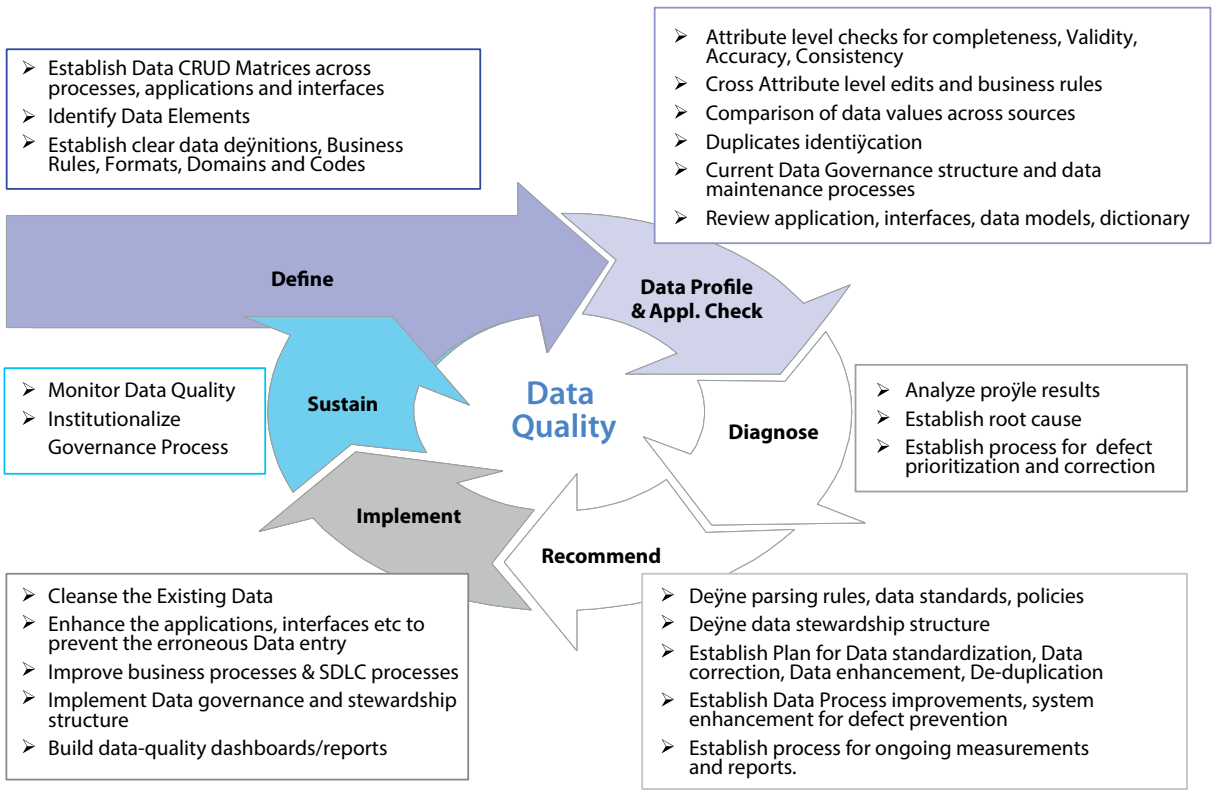
- **Subject Areas' or 'Concepts'** at the highest level of abstraction. HR, CRM, billing and payment are the examples of subject areas of a business, which are defined at the time of gathering business requirements.
- **Each subject area can be further broken down to Business Entities or Business Transactions or Enterprise Business Objects.** Customers, Vendors, Partners, and Business Transactions like Order Management etc., form the Business Entities.
- **Business elements or attributes** form the lowest level of abstraction of Business Metadata. For Business Entities like Customer, the Business Elements will be Customer Identifier, Customer Name.

Data Quality Management

Data quality management is a set of practices that aim at **maintaining the quality of information** on a regular basis. It involves maintaining the **reliability[2], effectiveness and usability** of data. Typically this includes standardization, enhancement, enrichment and de-duplication of data. Effective data quality management is recognized as essential to any consistent data analysis, as the quality of data is crucial to derive actionable and – more importantly – accurate insights from available data. Following are some of the key focus areas of the data quality management process:

- **Business Rules, Metrics and Targets**
- **Monitoring and Reporting**
- **Governance, Best Practices & Engagement**
- **Issue Detection and Management**
- **Remediation Process**

The following diagram depicts the typical data quality management approach and methodology:



Data Quality

The quantification, monitoring and management against business and technical expectations, of the quality of data loaded into an enterprise data store in terms of ***Latency, Uniqueness, Consistency, Accuracy & Sufficiency (LUCAS)***. This is followed by the initiation of root cause analysis in case of any unexpected deviations and application of the identified remedial measures.

The **essential DQ process steps are Profiling, Analyzing & Cleansing**. These steps are iterative in nature and hence may not be executed sequentially. In a scenario where many data issues are already reported, analysis would yield corrective rules and DQ metrics. Hence, the process would start with analyzing, lead to data cleansing and then data would be profiled to monitor quality of data on an ongoing basis.

As part of the Data Profiling step, basic rules are applied to data to understand the content, quality, and structure of the data. It is an ongoing activity, which enables discovery of any new quality violations as well as helps to assess the implementation of already identified issues. As part of the Data Analyzing step, these profiles are used to define and formulate metrics and set the threshold for further data quality monitoring and plan the corrective actions. As part of data cleansing step, standardization, de-duplication and setting up processes for handling data latency issues are performed.

Encryption

The encoding (and subsequent decoding) of data (plaintext), in motion and at rest, in such a way (cipher text) that only authorized parties can read it.

The primary purpose of encryption is to **protect the confidentiality of digital data** stored on computer systems or transmitted over the Internet or other computer networks. The encryption algorithms not only play a vital role in assuring the security of any system by providing confidentiality but also provide authentication, integrity and non-repudiation of the data.

Data, often referred to as plaintext, is encrypted using an encryption algorithm and an encryption key. This process generates cipher text that can only be viewed in its original form if decrypted with the correct key. Decryption is simply the inverse of encryption, following the same steps but reversing the order in which the keys are applied.

Encryption algorithms are divided in two categories – Symmetric and Asymmetric or public key cryptography. Advanced Encryption Standard or AES is the most widely used symmetric key cipher.

The use of public and private key asymmetric algorithms, also known as public key cryptography, are the most popular method of achieving data encryption. This also uses digital signatures to achieve integrity, authenticity and non-reputability of data in rest or motion. Cryptographic hash function is a next generation algorithm widely used in data encryption processes.

Enterprise data architecture – the conceptual, logical and physical layers

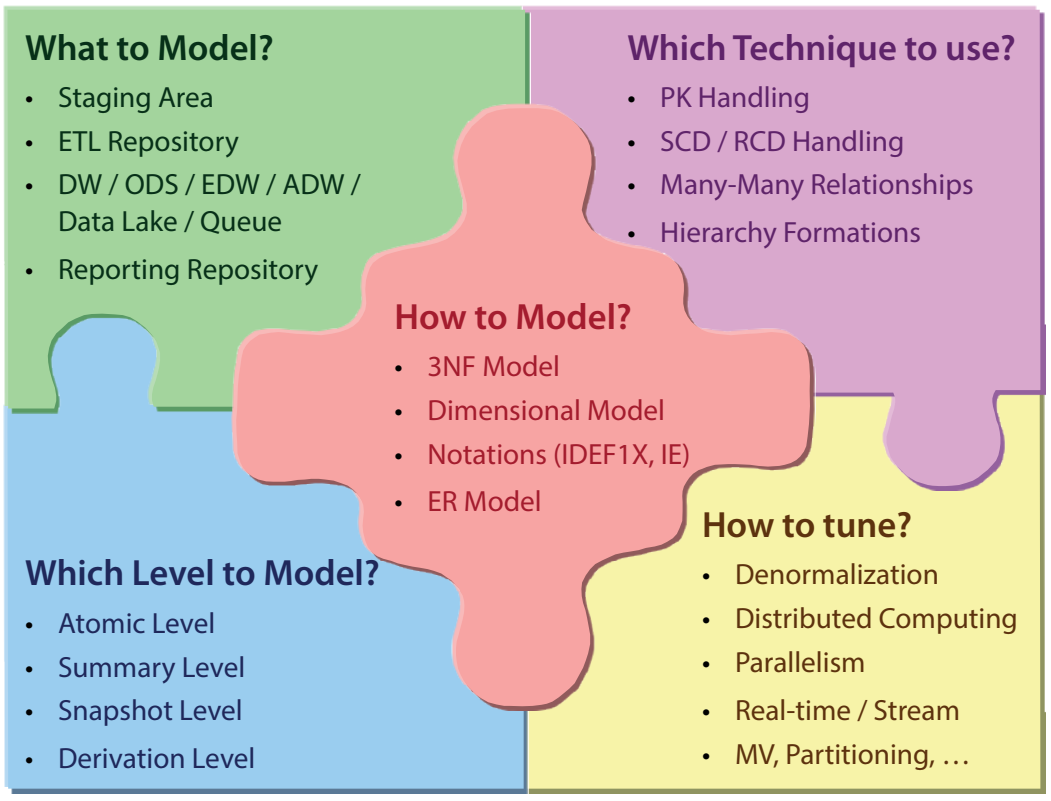
Based on the Business Architecture an enterprise then defines its data architecture. An enterprise data architecture consists of three different layers:

Conceptual: This layer defines **WHAT the system contains**. Business stakeholders and data typically create these models architects. The purpose is to organize scope and define business concepts and rules.

Logical: Defines **HOW the system should be implemented regardless of the technology platform**. Data architects and business analysts typically create this model. The purpose is to develop a visual map of rules and data structures.

Physical: This layer describes **HOW the system will be implemented using a specific technology or storage platform**. DBA and developers typically create this model. The purpose is the actual implementation of the models keeping choice of technology in mind.

Proper data models have to consider multiple aspects as depicted in the following figure:



Exception Handling

The definition, categorization, monitoring, capturing, storing and initiation of remedial action of or for deviations from acceptable behavior in the system. This includes errors for which the data integration or provisioning processes are brought to a halt and warnings for which the processes

are allowed to proceed once the exception is logged. *One typical difference of exceptions with Data Quality issues is in the fact that the former focuses on deviations from acceptable platform, application or data behavior but the latter focuses on deviations from the expected data behavior. However, there may be some data scenarios, which qualify as both.*

Exceptions can be classified as:

- **Process level exceptions:** E.g. exceptions due to failures, which affects the load process.
- **Data exceptions:** Data exceptions are the errors due to incorrect data format, incorrect value, unmet business rules and incomplete data from the source system.
- **Infrastructure related exceptions:** Infrastructure exceptions are caused because of issues in the network, the database and the operating system. *Common Infrastructure exceptions are FTP failure, database connectivity failure, file system full and so on.*

Logging

The recording of all required information for all kinds of routine operations and deviations in the system. Logging can be done in different areas. Following are some of the important ones:

- **Data loading** – Details such as load process-start-date, end-date, details of operations on records like records read, deleted, rejected, loaded, process status could be tracked under this category.
- **Application access** – Application access by downstream systems such as access time, access status, type of request, downstream system details accessing the application can be tracked.
- **Exception management** – Data exceptions such as adherence to LTV constraint, referential integrity violations can be tracked.
- **Data reconciliation** – Data reconciliation issues such as inconsistencies in data between systems can be tracked.
- **Data quality issue** – Data quality issues encountered during data quality analysis such as data inconsistency, duplicate records can be tracked.
- **Application management** – Application level maintenance details as activities performed on different servers, failure details can be tracked.

Master and Reference Data Management

An integrated set of data management processes, a governance framework and a technical solution that consistently defines and manages the critical master and reference data of the organization and provide **a single point of reference**.

Master Data Management (MDM) provides organizations with a reliable process for collecting, aggregating and distributing high-quality data throughout an enterprise. The process must ensure consistency and control for ongoing maintenance and use of such information. **MDM solutions**

safeguard an organization against the use of multiple (potentially inconsistent) versions of critical data in different areas of its operations.

Master Data is information essential to the operation of a business and **includes data about the business entities** like customers, products, employees, vendors, materials and so on. This type of information tends to be non-transactional. Product codes, material codes, list of valid values are some examples of **Reference Data** which generally does not change over time unless there are changes in the business rules. **Single version of truth** is created by cleaning, de-duplicating, consolidating and validating the original master data. Thus, MDM solutions typically contain the following architectural components:

- Technologies to capture, profile, consolidate, establish, synchronize and publish master data across the enterprise. This may include capabilities to manage, cleanse, and enrich the unstructured master data as well.
- Technology to enable comprehensive master data governance with rich UI interfaces and workflow capabilities.

MDM initiatives must consider an enterprise in its entirety.

This may be achieved through various mechanisms. Either there is a single Master Data Management System upstream (master entity is created here and then propagated to various business applications) or downstream (master entity is synchronized and a single view formed for reporting).

For instance, building a single view of customer across various lines of business helps build a better customer profile. A bank can sanction a personal loan for its customer who is regularly paying back the existing home loan or credit card bills. It can understand the risk profile much better when all the fixed deposits or trading transactions the customer does are brought together.

The following figure depicts the different master data architecture options:

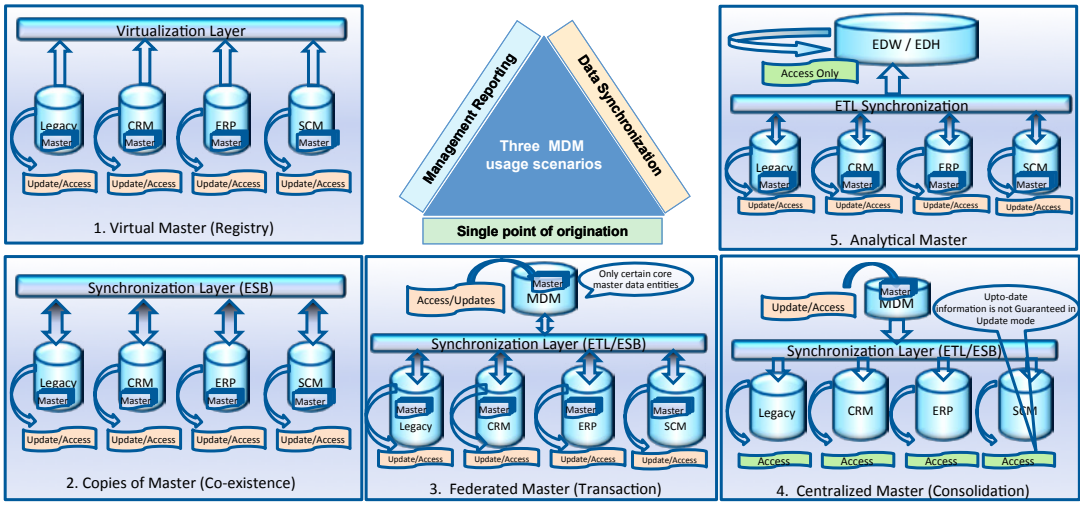


FIGURE 18: *Master Data Architectures*

Organization Structure

The hierarchy of roles and responsibilities, which is an integral part of all the frameworks, and policies, defined. This is a cross-functional group, with representatives from business and technical communities. Often the Chief Data Officer (CDO) will lead this organization. Some of the key roles of this structure are as follows:

- **Data Owners**
- **Data Steward**
- **Super User**
- **Information Security Officer**
- **Operations Manager (responsible for Incident Mgmt. & BCP)**
- **Compliance Officers**

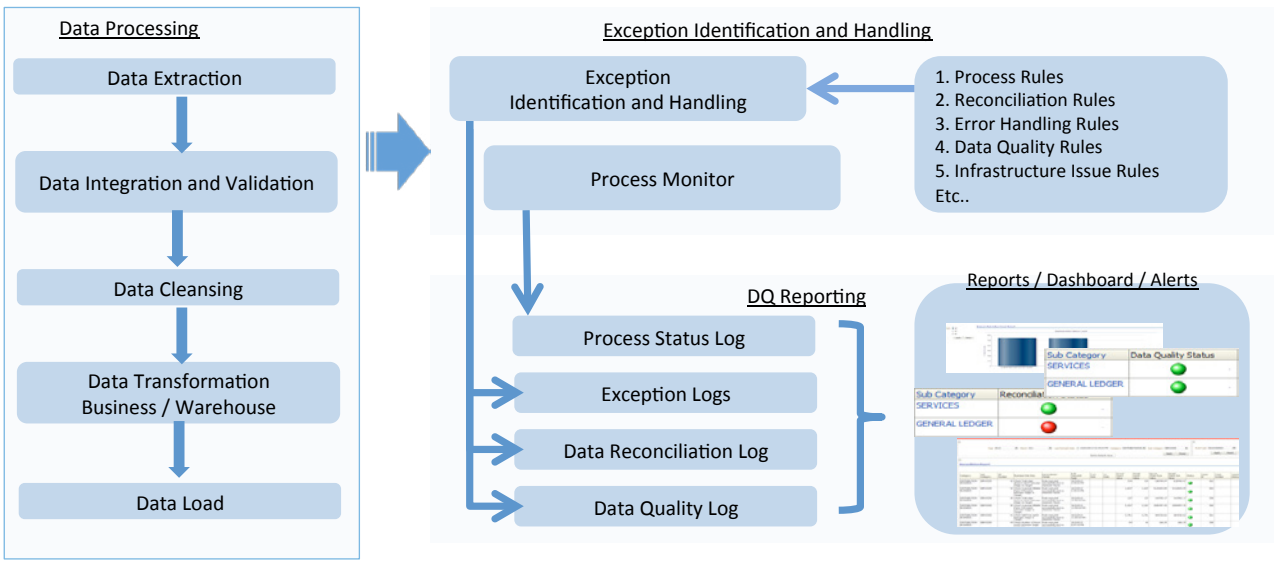
Organizational Charts helps to define the governance framework. It describes ownership, roles and responsibilities of each of the stakeholders involved.

Quality Trends

The monitoring of data quality, data reconciliation or system exceptions over a period to identify any consistent improvement and degradation in system quality, with a view to optimize Data Quality, Reconciliation and Exception handling processes. Typically done postmortem and in a batch-mode at monthly frequencies and **often employs statistical techniques such as Statistical Process Control (SPC)**.

▼ ***FIGURE 19:*** *Data Quality Management – Quality Trend Analysis*

A typical Quality Trend analysis process would cover aspects as depicted in the following figure:



Query Governance

Rapid increase in data volumes has intensified the need for high processing power of databases and data mining applications. Researchers have actively sought to design and develop new architectures for improving the performance. Recent research shows that the **performance can be significantly improved** not only by using high computational power and memory bandwidth of the hardware but **also by making effective utilization of architectural features and memory hierarchies of the applications**. One of the common method of effective usage of the architectural features and memory hierarchies is by setting **proper query governance** in the applications. With more and more use of processing using the same resources, this governance has become critical in present day systems. **This governance ensures that one code, user, group or application does not affect any other code, user, group or application in the ecosystem**.

Database limits are the most common method of query governance where a limit can be exercised to govern the number of rows retrieved from a select query and the number of records on which DML operations are performed in a single transaction. **Profiles** can be set on users, roles and at the application level. **OS profiles** also can be used to set database limit.

Execution limits are in place to **help control runaway processes or limit execution times** which can be greatly increased by things like Web Service or HTTP callouts. Some of the more common limits of this type include total number of executed script statements per transaction and total number of future calls allowed per transaction.

Application level Limits are also set using the license of a particular application to restrict the maximum number of concurrent users or user requests.

These limits can be set on a **particular period of a day or even on a permanent basis** based on the usage pattern and overall resource availability.

RASIC

A Responsibility Assignment Matrix describes the participation by various roles in the Organization Structure. It identifies the roles, which are **Responsible, Accountable, Supports** and have to be **Informed** and **Consulted**, for decision making and carrying out tasks.

Task	PMO	CC Gov.	Leadership	LOB IT	App Dev	Users	QA	Other Prog	Operation
Demand Management	R	C	C	C, A		S			
Contract & Financial Management	R	C, I	C	C, A	S	S			
Project Planning & Tracking	R	I	I	C		A			
Resource Management	R	C, A		C, I					
Issue & Risk Management	C	R	C, A	C	S	S	S	I, C	S
Governance & Communication Mgmt.	C	R	I, A	C, I	S, I	S, I	S, I	I	S, I
Reporting Management	R	C, I	I	I		I			
Knowledge Management	C	A		S	R	S	R	S	R
Change Management	R	A	I	A				C	S
Quality Management	I	S	I	O	C	A	R		I
Operations Management	I	I		A	S	I		I	R
Release Management	R	A	I	S	C	I	I	I	S
Deliverable Acceptance Management	R	S		C	S	A	R		
Benefits Realization Management	S	S	A	R	I	C	I		I
Integration Management	S	R	R, A	S	I	I	I	I	I

▲ ***FIGURE 20:*** *Program Management : RASIC*

Reconciliation

This is the periodic monitoring of data loaded into the data stores against their upstream and original sources, to establish that the internal data plumbing (ETL) processes have not modified the volume or content of data unexpectedly.

The reconciliation can be done at different levels:

- **Technical or Batch Reconciliation:** Technical Reconciliation is concerned with ensuring sets of data (batches) are moved from source to target systems, through an Integration Platform, with the data being always accounted for.
- **Business or Item Reconciliation:** Business Reconciliation is concerned with ensuring business logic is satisfied within the Integration processing.

Restartability

The tolerance of the system towards any data quality, data reconciliation, application or system exceptions. Includes the resumption of normal operations during warnings, restarting of normal operations in case of errors, once the required remedial action are carried out. It also covers recovery of the system in case of system failures. Matured systems have restartability capability from the point of failures. This requires Rollback and Restartable Updates to be included as system design considerations.

Restartability is required due to failures at different levels such as Application failure, Database failure, Load Process Failure and Hardware failure. Appropriate strategies should be formulated and designed at the application level.

Root Cause Analysis

The definition, publishing and management of standard operating procedures that are to be followed in case of any data quality, data reconciliation or system exceptions. **The objective is to identify the root cause and required mitigation and prevention strategies and track them to closure**. Often techniques such as Fishbone diagram, Scatter Plot Diagram, 5 WHYs, FMEA etc., are used for this analysis. Includes continuous communication to all the stakeholders during the process.

Service Delivery Models

Organizations adopt a range of delivery models ranging from Factory model, Waterfall, Agile, etc. to deliver the data and analytics initiatives. TCS DATOM™ prescribes that the organization should have processes and procedures defined to choose the right delivery model based on solution contexts, efficiently.

The following diagram depicts in brief what each of the service delivery model may be used for:

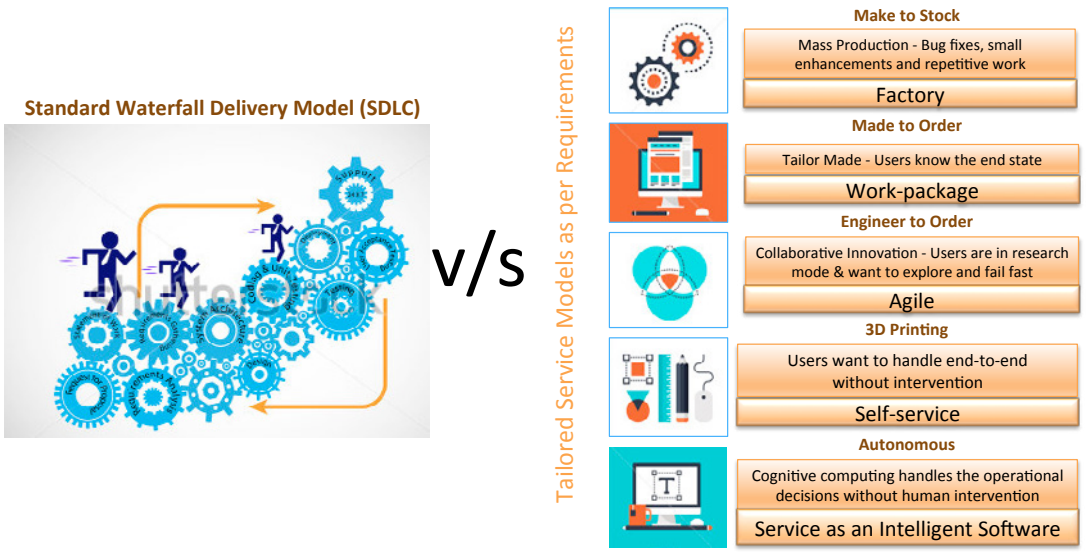


FIGURE 21: *Service Delivery Models*

TCO

Total Cost of Ownership (TCO) is a model that is used to identify and analyze the overall costs associated with ownership of Information Technology (IT) assets, **from cradle to grave**. Reducing the cost of ownership of IT investments results in higher Return On Investment (ROI).

It is essentially the purchase price of an asset plus the costs of operation. Assessing the total cost of ownership represents taking a big picture look at what the product is and what its value is over time. When choosing among alternatives in a purchasing decision, buyers should look not just at an item's purchase price, but also at its long-term price. The item with the lower total cost of ownership is the better value in the long run.

Technical Metadata

It is the **technical (IT) description of any system, entity or element**. It takes different forms starting from the SORs, Applications, Database Schemas, Tables, Data Flows, Physical Data Models, ETL Mappings, Reports, Dashboards, Columns, Datatype, Format, Transformations, Unit-of-Measurement (UoM), Access Control List (ACL), Horizontal Lineage (as it passes through the systems), Vertical Lineage (uses/used-by table/columns), Infrastructure details (hardware, software, system capacity, SLA) and Audit Information (timestamps and user accounts).

It is usually defined at three levels of abstraction, but there can be multiple hierarchies representing different IT technologies and domains such as SOR, Data Integration (ETL, Middleware), Data Stores, Information Delivery (Data Services, File dumps, Reports), Infrastructure (hardware, software).

- **Corresponding to business areas**, technical systems or applications are developed to meet the requirements of each subject area. Example: Oracle HRMS can be developed for HR subject area and SIEBEL Systems may be implemented for CRM. These form the 'Systems' of the IT/Technical Metadata.
- **Corresponding to each Business Entity**, there will be Technical Objects storing the details of these entities such as tables, reports, or processing the data such as mappings.
- The lowest level in IT or Technical group is **Technical Elements**. Any detailed information, which exists at element level like columns, fields, or any transformations form Technical Elements.

Usage Monitoring

Usage monitoring **keeps track of the overall health** of any application. Proper tracking of applications will ensure significant short and long term benefits. Usage monitoring can be achieved by **monitoring database activities, CPU usages, memory usages and even unused data in a data store**.

- **Database Usage** – Database usage monitoring is a critical step for security and compliance. This enables compliance controls, operations monitoring and data protection and does so without interfering with business processes. This monitoring gives the ability to analyze use of objects, user behavior, volume of data, source and destination systems, various applications and contents. Thus, compliance and security policies can be applied in a very granular and precise way. Recording database activities helps to track unauthorized database changes helping in audit compliance.
- Along with the security the other important aspect of database usage monitoring is **dormant data**. This data resides in the database but is not accessed by any user or applications. Thus, it affects the performance and cost of databases negatively.
- **CPU Usage** – CPU consumptions need to be monitored regularly to avoid performance issues in the applications. This also helps to identify the idleness of the CPU, the peak hours of consumptions, the consumption patterns and the overall utilization of the physical or virtual servers. With the data gathered during the monitoring process, proper benchmarks can be created and CPU intensive processes can be handled to improve the overall application performance.
- **Memory Usage** – Monitoring memory usage helps to detect network overloads, before they result in downtimes or data loss. In addition, it helps to identify underused servers, and redistribute loads accordingly.

Versioning

Data Versioning

Once a change goes LIVE, newer incremental changes start coming in. It is important and often required that the versions of data before and after be presented together in a consistent view. This is to preserve the traceability of information over a period, irrespective of technological changes underneath.

Data versioning in order to track the history of events can be achieved by implementing **Slowly Changing Dimensions (SCD) Type 2** for dimensions and fact records can be derived based on change in event.

Data versioning to track the change in natural (business) key values can be achieved by a map strategy to map the old and new values.

For example, Credit Facility Ids may be different in the old and new systems. In order to provide a historical view consistently, it is important that the traceability between the old and new id are preserved and published. This may even be a required for audit and legal purposes.

Metadata Versioning

Most changes, apart from plain software version upgrades, include some change in the business or technical data. Just like Data Versioning, Metadata Versioning may be required to preserve the traceability of information over a period, irrespective of technological changes underneath.

Appendix B

Maturity Characteristics: Consolidated View

KRAs / LEVELS	Siloed – 1ml	Simplified – 2ml	Scaled – 3ml	Synergized – 4ml	Self-Optimized – 5ml
DATA	• Data exists in silos in Departments (e.g. HR, Finance, Procurement, Operations) or with individuals. • Incomplete history • Support tactical decisions • Integration issues requiring manual intervention for resolution • Very limited Data Governance	• Integrated enterprise data store • Complete historical data • Well established Information Management Functions *(Data Architecture & Modeling, Metadata Management, Master Data Management etc.)* • Established Corporate Performance Management *(covering Planning & Budgeting, Financial Consolidation, Balanced Scorecard)*	• Low latency/real time operational data for decision making at enterprise scale • Comprehensive Business Performance Management • Fulfillment of all data needs for Enterprise Intelligence across all management levels • Matured Operational Analytics • Command Center Analytics	• Capture and use Ecosystem data • Synergize ecosystem data with enterprise data during decision making • Process semi-structured, unstructured, high volume data • Established business/domain specific analytical capabilities • Complex event processing intelligent engines	• Identify & understand unknown risks beyond ecosystem boundary. • Usage of Orthogonal, PESTLE Data • Advanced analytical, AI capabilities • High Degree of intelligent automation in decision making • State of self-optimization.
TECHNOLOGY	• Traditional BI Technologies involving ETL, MIS, Reporting & Dashboards • Business function specific data repositories • Data may be in a Data Warehouse, Data Mart, Operational Data Store (ODS); sometimes even in MDDBs • Advanced Analytics (Data Mining) capabilities may exist within a few business functions • Data Mining/analytics mostly on desktop	• ETL suites to integrate different type of data across enterprise in batch mode • Enterprise Data Warehouse and mostly EDW Appliances (MPP architecture) are used • Technologies for Information Management are well established • MPP or Clusters or Grid are used for future scale • Specific tools needed for CPM are used	• For low latency/real time data ingestion, appropriate technologies are looked at [e.g. CDC, Middleware (EAI, ESB, B2B), Data Services, APIs, Micro Services, EII] • Active Data Warehouse (ADW) appliances • Usage of automation technologies such as BPM, Rules Engines, Business Activity Monitoring (BAM) • IT Delivery & Operations automation – handling scale & operational excellence • Information Management is digitized	• Data ingestion for semi-structured and unstructured information • NLP, Computer Vision, Case Based Reasoning, CEP engines • Advanced analytical capabilities across enterprise • Big Data for high volume data & multi-structured storage. • Cognitive capabilities, (Semantic) Search Engines and interactive chat/voice bots • Cloudification to scale on demand • Data Virtualization; Data Privacy & Security	• Right balance of Human centric and Machine centric decision making • Institutionalization of different disciplines of AI *(Automated Forecasting & Control, Knowledge representation, Learning, Reasoning, Perception, NLP/G)*; Self-Healing • Scaling compute and storage to much higher levels • Enable humans to take decisions faster, using visual analytics, immersive analytics (MR, VR, voice enablement)
PROCESS	• Business Functions drive D&A requirements independently • Program Management processes at initial stages of maturity • Basic SDLC for project execution • Projects are short lived and aimed at delivering reporting/dashboard applications for specific use cases with limited scope.	• Business process maps created and linked with data architecture and consumption layer, to enable business consumption • Enterprise Scale Program management & Requirement Management capabilities • Factory processes deployed to bring scale in developing enterprise scale D&A capabilities • Established Data Office processes using Information Management framework. • Established Center of Excellences • Enterprise scale D&A Initiatives.	• Multi-modal delivery capability established. • Rapid Application Development (RAD) used for operational analytics based process improvement initiatives. • End-to-End Business Engagement. • Competency Programs for Business Users • Digitization (/automation) across all Information Management processes including Requirements Management, Program Management, Knowledge Management, Competency Management, and so on.	• Matured Agile processes to deliver Ecosystem Intelligence • Institutionalized Innovation processes. Processes such as Digital Re-imagination, Small Group Activity established. • Reuse program extended to Decision Management itself • Stringent Data Privacy & Security processes.	• The intelligent automation using AI would need processes of "Service as intelligent Software" to be established • AI Governance • Comprehensive Self-service capabilities • Self-Healing capabilities for systems • Competency build on AI democratization, AI governance as well as self service analytics
PEOPLE	• Mainly IT driven D&A initiatives • Business Users, IT teams does not have much exposure to analytical capabilities. Statisticians and Data Scientists may be used sometimes. • Very limited business involvement. Mostly during requirement capture and acceptance testing.	• Involvement of Business Users & Sponsors throughout a D&A program • Enterprise Architects develop end-to-end solution architecture; Chief Data Officer handle Data Office and enforce data practices • Program Managers & Visionary Advisory available for large scale program execution • Comprehensive knowledge management • Competency management initiated	• People with skills to handle low latency/real time data interfaces • Digitization and automation experts to establish IM Digitization • Competency enablement of Business users (e.g. how to define business rules in a given technology ; Self-Service reporting) • Trainers with ability to train operational users • Operations Research skills for optimization at scale	• Agile program executioners to execute transformational programs. • Active involvement of Data Scientists • Strong legal advisors to understand legal & regulatory implications of ecosystem data. Enables data monetization platforms • User Experience skills are important • Matured Enterprise Continuum	• AI skills across different branches to build AI programs and agents. • Besides Data Scientists, neuro-scientists and psychologists needed for AI • Visual Analytics and self-service analytics skills for business users • Human Computer Interface skills to build Immersive Analytics interactions • Quantum Computing skills to manage scale

▲ ***FIGURE 22:*** *Consolidated View of Maturity Characteristics of TCS DATOM KRAs*

Appendix C

5ml of Digital Intelligence

Business involves many activities and there is always a decision making involved in these activities. People, based on the information available in hand, perform analysis and take decisions. This is sometimes accompanied by gut feeling as well. For more than two decades, enterprises have been leveraging the Analytics and Information Management platforms with the sole goal of '**management by facts**' for business benefits. Over this period, the **technology landscape has changed**. And this change is not just limited to the corporate world but also to the social world.

Analytics (and Information Management) is definitely one of the most important business initiatives, which has shown **positive impacts on the health of organizations**. But, as everyone is aware, this impact has never happened by taking a big dose of Analytics, rather it usually resulted in a major failure. Organizations have achieved good health and differentiating growth with carefully prescribed doses based on '**Think big, but build step-by-step**' philosophy. A lot of questions have been raised regarding the definition and the scope of these doses to get clarity on the maturity of the Analytics initiative. After obtaining the desired benefits from the Analytics initiative, the usual question has been 'What next?'

There is always a benchmark and a model to evaluate the maturity of a business initiative, which has (/can have) a potential to impact business performance. The **5ml of Digital Intelligence** (5 maturity levels of digital intelligence in enterprise using systems such as Analytics, Business Intelligence, Big Data, iBPMS/IBO and Information Management) has been devised by considering various scenarios and dimensions involved in the development of intelligence in an enterprise. This model provides a concept to comprehend the maturity level of Analytics and Information Management initiatives and gives the perceived technology backbone to yield business benefits with certainty. The model attempts to build a correlation between the development of human intelligence over the lifespan and the evolution of Digital Intelligence in an enterprise.

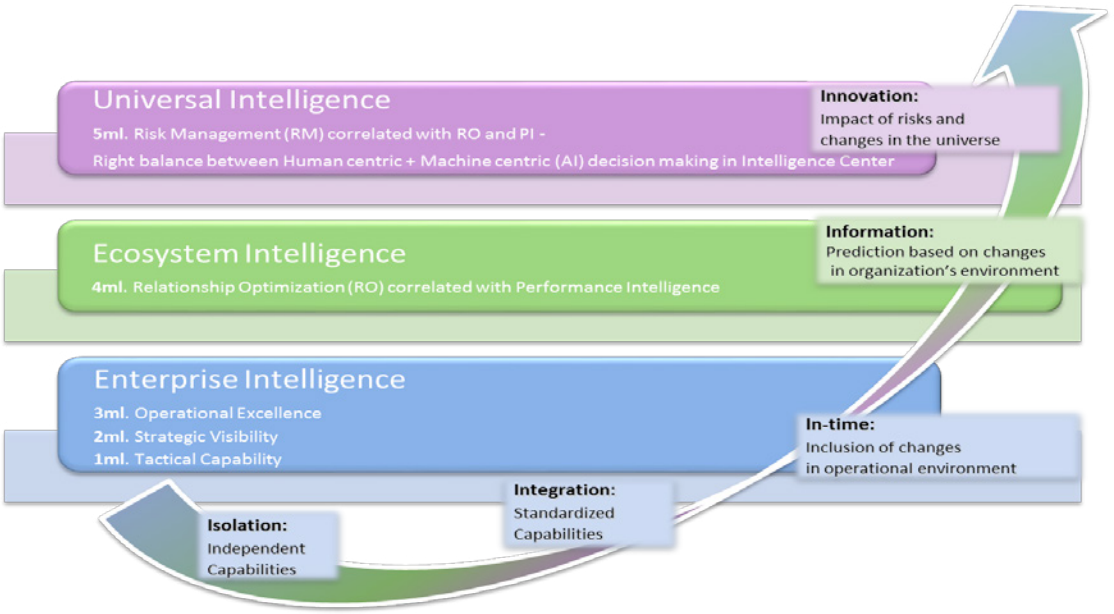


FIGURE 23: *"5ml of Digital Intelligence" Model*

Enterprise Intelligence

1 ml: Tactical Capabilities in Enterprise Intelligence

A newborn child maximizes its abilities in isolated limbs. It tries to use its different limbs which have different capabilities and strengths. The legs are not strong enough to walk or run, the hands are not strong enough to hold, but voice is good enough to get its mother's attention when it is hungry. **The unified approach towards the self is missing**.

In Sigmund Freud's "Structural Model of the Human Psyche", a new born child is impulsive and responds directly and immediately to the instincts instead of long term gains. This personality is termed as "id". Same is the case with an enterprise, where different departments or divisions work in isolation. They have tactical analytical needs and end in building analytical systems accordingly. These departments do achieve intelligence maturities of different level, but miss the bigger and long term picture.

There are obviously a few parts of the body, which by default function well (such as heart) to ensure survival of the child. Similarly, a few departments do well to ensure that the enterprise survives. Finance department may be strong in analyzing the information and giving push to the marketing division for reducing the expenditure, but marketing department may not be mature enough to optimize the spend. Services division may be having a bad time in handling the customers against the cost as they do not know the complete picture of the customer's lifetime value and interactions across all divisions or product lines. In short, the enterprise departments or divisions are working in isolation and managing it tactically. The vision and focus is limited to the department and hence the benefits drawn are restricted for the enterprise.

This is the **first level of maturity (1ml)**, where the **tactical capabilities are built within departments** and there are "**pockets of excellence**". The Analytics and BI initiatives are departmental in nature. Information sharing across the departments is minimal. Most of the times it involves a political struggle to get the information from other departments. Different departments build their Operational Data Stores (ODS) or Data Marts (DM) or Data Warehouses (DW) or Big Data Solutions with historical data and try to minimize the system performance impact of analysis on the OLTP (source) systems. Usually, the data volume is not so large and hence general purpose databases are used to build these data stores. For data coming from sensors, social media and high volume transaction systems, the data volume is very large and Big Data (Hadoop, NoSQL and various flavors) cost effective stores can be used. MIS, Dashboards, Standard, OLAP, Ad-hoc reports and analytical applications are built on these stores. This usually leads to proliferation of the tools and databases. Individuals in the department do follow the same culture and find it easy to dump the necessary data into Excel and MS Access and build their own reports. There are multiple copies of the same data across the enterprise leading to high cost, less manageability and eventually low trust on the data. In addition to proliferation of data, uncontrolled as well as unstructured (reports in PowerPoint or Word or Excel) versions of data outputs are generated and distributed as well. The obvious outcome is the individual versions of the truth.

Data reconciliation is one of the major issues faced at this stage. When different individuals and department try to produce reports in a specific area, the numbers often do not match and it is difficult to know which data is correct. Hence, the trust in the correctness of the data at enterprise level is minimal. The main challenge to rise to the next level is in establishing the governance especially Data Governance. The lack of such governance is one of the key reasons why most

enterprises struggle to reach the next level. That is why the regulatory authorities across different verticals impose strict norms on how to govern the data.

2 ml: Strategic View with Enterprise Intelligence

The child grows and starts coordinating its different limbs and thoughts in a more governed manner to perceive itself as a whole individual.

As per Sigmund Freud, the direct influence of the external real world modifies a part of "id" into "ego". Now, the child operates to the reality principle and uses reasoning to devise realistic strategy instead of just satisfying tactical needs. The unified governance of the body enables it to plan and control the behavior. The analogy is the "id" is the horse while the "ego" is the rider providing the governance. It starts using the body as a whole instead of pockets of excellence. If a specific limb is not functioning well, then it can strategically compensate using other limbs. It is a very difficult stage as self-governance is a big milestone that needs to be achieved.

Enterprise needs Governance to operate as a unified enterprise and build a strategic view. Integration of data to facilitate unified enterprise perspective is a giant step. The second maturity level (**2ml**) is all about the **strategic visibility with analysis of the integrated information**. Usually, this is achieved with the backbone of an Enterprise Data Warehouse (EDW). It **needs a strong governance spanning across the Key Focus Areas, namely Data, Technology, Process and People** and hence most of the enterprises struggle to build it. The following are the glimpses of the complexity involved in governing these Key Focus Areas:

- **Data:** Enterprise Data Management (EDM) covering seven areas, namely Data Architecture and Modeling, Metadata Management, Master Data Management, Data Quality Management, Data Security, Information Lifecycle Management and Data Governance Council.
- **Technology:** Establishing technology and infrastructure for the integrated information with scalability, sustainability and low TCO. The enterprise data can be so large that appliances or big data may need to be considered. The tools proliferation needs to be controlled while keeping in mind the issues with 'one size fits all' solution.
- **Process:** Defining the policies and procedures, delivery operations models (Work-package, Agile or Factory based) and value (benefit) measurements.
- **People:** Setting the Competency Center based on the enterprise topology (Distributed, Centralized or Federated), Business-IT alignment and user maturity.

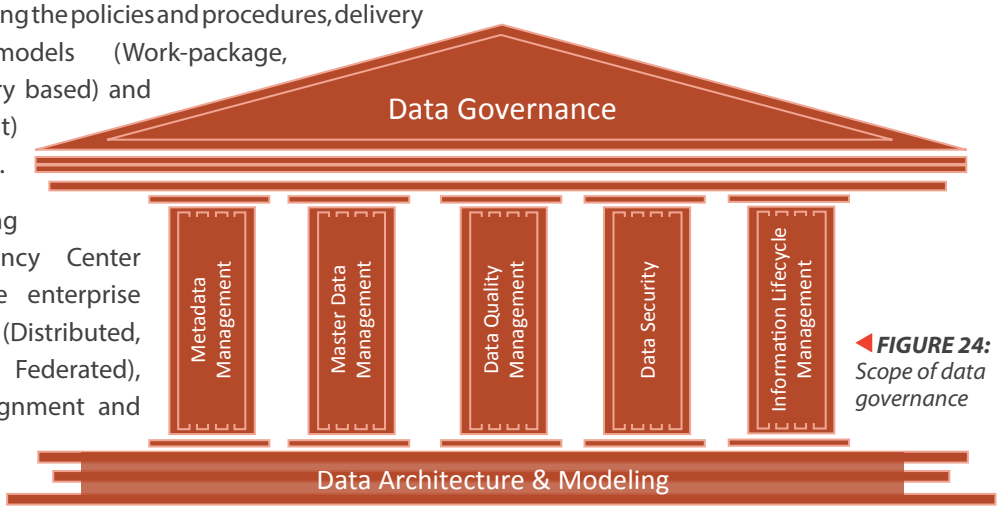


FIGURE 24: *Scope of data governance*

If any of these aspects are missed, the governance fails. So if Data Quality is not managed, the 'Garbage in – garbage out' philosophy prevails, if centralized metadata is not defined, the 'On Time Delivery' may mean delivery against customer requested date for one unit and against promised date for another unit. This failure results in the decline of the maturity back to 1ml.

The EDW or Data Lake (the Data Lake is sometimes considered as a dumping ground of different type of datasets in the enterprise, but here the Data Lake refers to a governed storage of data in Big Data environment) facilitates the "**authorized source of truth**" with huge breadth (across the enterprise functions) and depth (granularity of details) of the historical information to enhance the analytical capabilities such as ad-hoc analysis and data mining. 1ml also facilitates the ad-hoc analysis and predictive analysis capabilities, but the information span area is limited and hence holistic perspective is missed. There have been supportive talks on the use of a semantic layer which can hide the isolated storages and give a single view such as Enterprise Information Integration (EII) technology. But the advantages of a physical EDW or Data Lake are far greater than virtual integration achieved through the semantic layer.

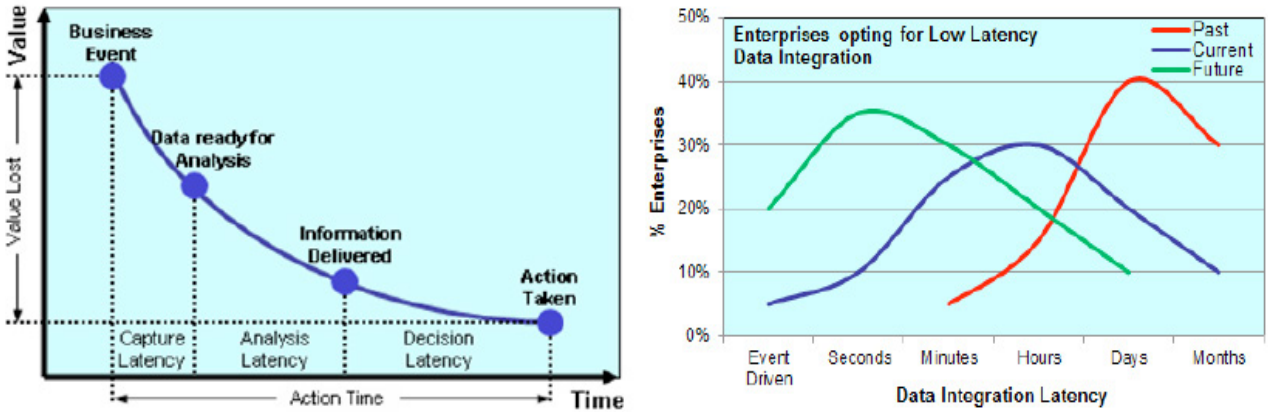
The CIO always has the aim to make the right information available to all the stakeholders. **The EDW or Data Lake has been found to be the most useful in making it achievable**. Following are a few examples that help substantiate this are:

- **Walmart** started their EDW journey in 1989 and was the first to reach 1TB in 1992 gaining a huge competitive advantage. Since then their integrated data set not only enabled internal organization, but also the buyers and the suppliers.
- **GE Rail Car** built EDW in 2002. It gave an advantage to keep their operating costs low in holistic manner and maintain high service quality and efficiency, making them the advisors to Association of American Railroads.
- **ICICI Bank** built EDW in 2001 and gained end-to-end understanding of customers-households and financial portfolios to lead the change in modern Indian banking.
- In 2002, **Reserve Bank of India** built Nations Economy Database connecting the datasets across Macro Economics, Financial Market, Liquidity-Debt, Financial Sector Stability, Currency Management, etc., with a power to discover butterfly effects in Indian economy.

Around 2010, many enterprises started consolidating their data in Data Lake on Big Data platform. Initially, co-existence of Data Lake and EDW dominated the market, but now enterprises especially Banks and Retailers are moving onto the enterprise wide Data Lake with Data Governance Controls enabled on it.

3 ml: Operational Excellence with Enterprise Intelligence

Initially, the ego is that part of the id which has been modified by the direct influence of the external world. The ego develops in order to mediate between the unrealistic id (responding to tactical needs) and the external real world (strategic movements). It works by reason at an operational level spanning across a set of psychic functions such as judgment, tolerance, reality testing, control, planning, defense, synthesis of information and intellectual functioning. Day to day, minute by minute the behavior needs to be managed in relation to the tactical capability and also the strategic goals to be achieved in future. **The complete traceability from Strategic to Tactical to Operational is the essence to get the end-to-end Enterprise Intelligence**.



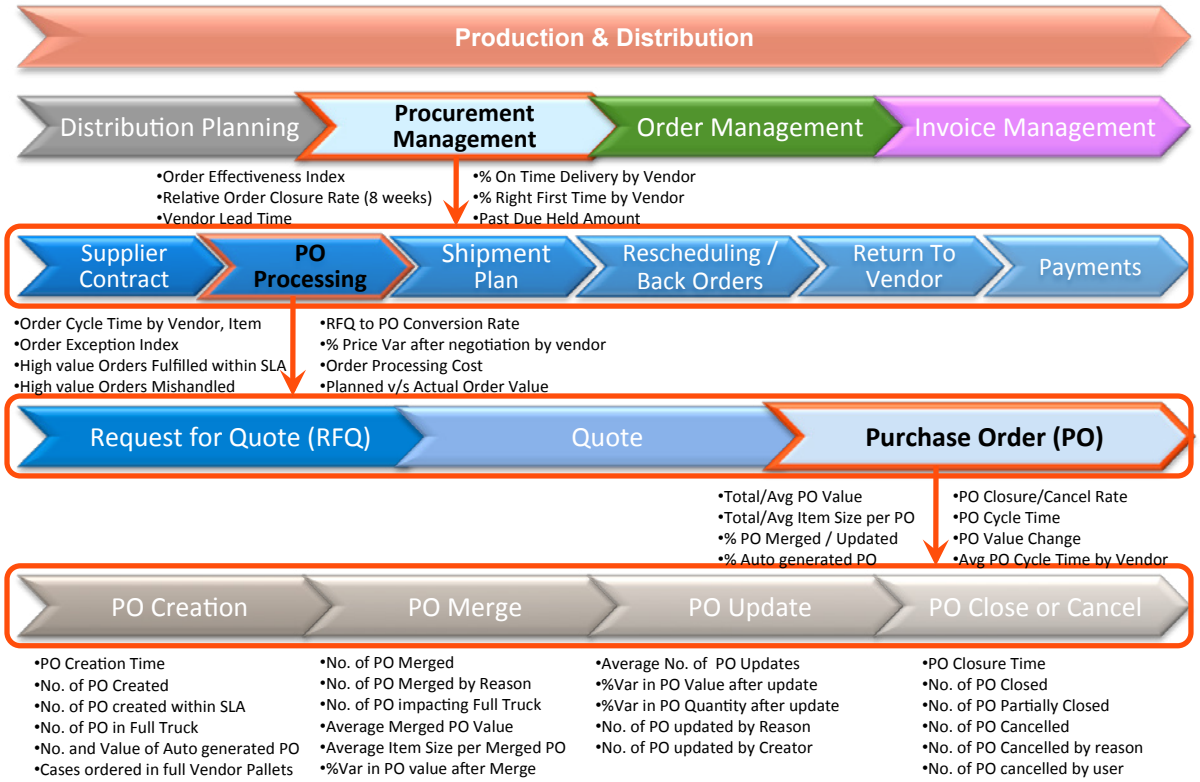
▲ ***FIGURE 25:*** *a) Action Time vs Value Lost b) Data Integration Latency Trends*

The philosophy is to have '**right information available to right person at the right time**'. As per the Time-Value Curve for Decision Making, the **business value decays with time** and the definition of the 'right time' depends purely on the optimization need in the business decision making cycle. If decisions are not taken in the right time then the insights generated from the event become "**Perished Insights**". Most of the enterprises are shifting towards lower latency information capture to enable faster decision making and hence ensure higher business value of an event.

If a call center person needs to perform the 'Best Action' while interacting with the customer, they need to know the as-of-now customer profile (customer information along multiple focus areas such as Account, Subscription, Affinity towards Products, Campaigns sent and Faults reported). In the absence of that, improper action may be taken resulting in the customer churn. In this case, the as-of-now information can be with 5-10 minutes latency. But in case of Circuit Breakers in the surveillance unit of a Stock Exchange, the decisions need to be taken truly on real time basis. **The third maturity level (3ml) is about operational excellence through strategic visibility and tactical capabilities with in-time information analytics**.

Operational excellence loses meaning when it is not linked with the tactical and strategic goals. Consider a global manufacturing company in which a plant operations manager is focusing on downtime sensitivity of the assembly lines of critical items, whereas the head of LOB is thinking strategically to outsource that task to another company. The investments demanded by the operations manager to handle the downtime of those assembly lines may not be entertained by the head of the LOB. Hence, taking a top-down approach for connecting the strategic goals to tactical and operational goals is necessary. This is easily represented in terms of the **business process decomposition**. The following diagram shows a sample decomposition of supply chain process in a retail company:

There are measurements done at the L1 level in terms of Key Performance Indicators (KPI), which are at high level and are usually tracked and analyzed by executives and senior management. For e.g., On time Delivery and Right First Time KPIs at supplier level can help them to negotiate with the suppliers during the contract. Decomposing the Procurement process from L1 to L2 to L3 level and further levels gives more granular measurements, which are usually done at operational level. **This linking of strategic KPIs to tactical KPIs to operational KPIs establishes a strong and transparent performance driven culture**. For the deeper reach of performance measurements in process decomposition (at operational activity or workflow level), **it is necessary to automate the processes**, as the time needed for capturing measurement decreases and demands lower



▲ ***FIGURE 26:*** *Decomposition of Supply Chain Process - A Sample*

latency information capture. It is not possible with a manual measurement approach. Unless there is automation in the information capture at that level, the measurement capture process itself becomes an overhead. *Hence, technologies such as Operational BI, Business Activity Monitoring (BAM), Business Rules Engine (BRE), Streaming and Business Process Analytics (BPA) play an important role in this area.*

Processes give the end-to-end perspective to business, making it necessary **to perform enterprise-wide information integration (2ml) first and then achieve 'right time' and 'in-time' information integration**. As the organization starts venturing in real time integration, the volume of data to be dealt with at low latencies is humongous making it essential to have data governance established. If data quality or master data etc. processes, technologies and involved people are not trained on higher latencies beforehand, then at these low latencies, the data will keep pouring in and one will not get sufficient time to correct it. Eventually, in spite of having great infrastructure to handle the real time data integration, it will lead to inappropriate decisions resulting in uncertainty in business operations.

The operational decision making demands analysis of information (measurements) captured at lower latency which is near real time or real time. **But there is no hard and fast rule that the information capture needs to be done at lower latencies. Depending on the business model and scenario the information capture latency can be defined and accordingly specific technologies can be deployed**. If the operational data needs to be analyzed only three or four times a day, then ETL technology can still work, but for near real time or real time decision making (e.g. telecom network fault identification in black spot analysis) a messaging or service based (EAI, ESB, EII or B2B) or replication based (DB SQL or log replication, disk to disk replication) technologies can be considered. For a high volume, ultra-latency, the technology landscape can be changed to ultra-messaging or socket programming.

The current data can be analyzed by looking into the OLTP systems themselves, but that may lead to performance impact on the business operations and hence it is not accepted by the business. In such cases, Operational Data Stores (ODS) are created to store the recent data (or replica of the OLTP system). However, knowing only the current situation without comparing with history does not add any value. The information needs to be integrated on real or near-real time basis along with the history and the choice of backbone appears to be the **Active Data Warehouse (ADW)**. Electronic Control Modules have been installed in vehicles in order to diagnose the problem on real time basis and through transceivers the sensor information is sent to a central diagnostic system. But in order to diagnose, normal, abnormal, noise patterns need to be compared. This is possible only if there is historical information available along with the latest data coming in from the sensors. Analyzing sensor data on a real-time basis is necessary for business operations. Example: Assembly line sensor data analysis or telecom network components data to prevent downtime of operations. ADW enables operational decision making and can also be extended to form the **Enterprise Data Hub (EDH)** for the downstream or external applications. The main challenge in ADW or EDH or streaming data processing is the mixed workload, which involves simultaneous READ/WRITE operations along with higher user concurrency. Newer appliances are enabling the mixed workloads with the availability of SSD and in-memory analytics features (of course with higher price per performance). But performance is sustainable only if ADW data model is robust. There are specific design practices that need to be followed in modeling the ADW or EDH such as right ER model with clear definitions of canonical entities to enable information exchange across applications. Restricting to specific school of thoughts usually ends-up in huge rework or system performance issues. Right blend of granular information and aggregations give huge performance, scalability and flexibility benefits.

IT may have questions on how to get the necessary buy-in from the business to invest in such an infrastructural capability, which enables real time integration and manages high data volume and concurrency. Since the decision making is done at process, activity, workflow level, the decision making system is no more just a non-mission critical system. It becomes a part of the mission critical environment as it gives **360 degree decision feedback mechanisms** into the OLTP systems.

The Enterprise, similar to an individual person, needs to improvise its performance from tactical capabilities to strategic potential and also to operational excellence. Strategic thinking, governance, learning and control over minute activities is the key to the superlative performance of an enterprise. **The first three levels are about building the foundation to measure, analyze and govern the performance with the help of Enterprise Intelligence**.

Ecosystem Intelligence

4 ml: Relationship Optimization along with Performance Intelligence

With "ego" the child gets control over its activities and knows what it has to do in short to mid-term to achieve long term goals. It knows the "**self**", but it has an entire community around it which is not yet fully explored and managed. "**Superego**" **persuades the ego** to turn to moralistic goals rather than simply realistic ones and to strive for perfection. The superego consists of two systems, namely the conscience and the ideal self. **The conscience governs the ego in the community**. The ideal self builds future aspirations, relationship with other people and how to be a member

▲ ***FIGURE 27:*** *Ecosystem*

of society. Based on its acceptance at the community level, its growth as well as existence can get influenced. Hence, the child needs to perform well at the community level and manage and optimize peripherals for sustained existence and growth. First, the child needs to understand what this community is made up of and then start building the relations with various individuals and groups through influence and then optimize the relations.

"In a free enterprise, the community is not just another stakeholder in the business, but is in fact the very purpose of its existence", Jamsetji Tata.

Enterprises are a part of larger ecosystem and they need to acquire the ecosystem intelligence by managing and then optimizing the relations with various entities in the ecosystem. These entities can be customers, suppliers, competitors, regulatory bodies, sister-companies, subsidiaries and also individuals such as consumers and employees. How will the enterprise be able to understand and build knowledge of such a wide spectrum of these entities? How will it maintain its knowledge about these entities with employee attrition scenarios? How will it manage and optimize the relation with them on on-going basis? How will it facilitate the sustenance and growth of the ecosystem made up of these entities? Even the child gets the similar questions when it finds itself in the middle of a community made up of siblings, parents, neighbors, teachers, friends and foes. But it gets community consciousness over a period of time through interactions, news, books, readable and unreadable expressions or gestures, correlation of events and knowledge building. This is not just dependent of the performance intelligence build on top of structured data and statistical analysis, but rather it is highly dependent on unstructured information sensing, responding, learning, collaborating, effecting change and reusing. This process is used in enterprises for relationship management and optimization.

Relationship with different entities in the ecosystem is handled in different manner, but the process is similar in nature. **First know the ecosystem player, then manage the relation and then optimize it**. How a **Customer** entity can be dealt with has been outlined in the following table:

Table 1: Dealing With a Customer Entity

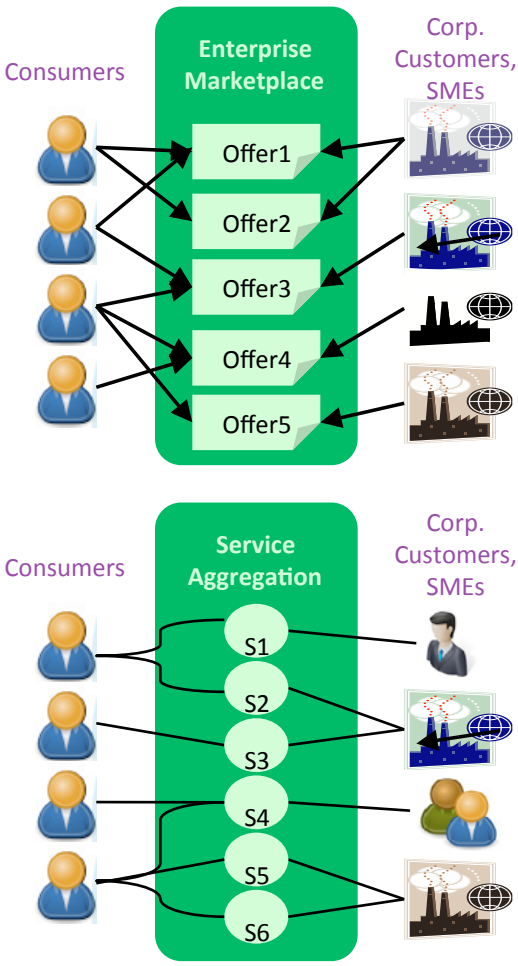
Step	Sub-steps	Technology
Know Your Customer & its network (KYC-C)	• Customer attribution • Customer data collection across internal customer data, semi-traceable, non-traceable customer data, external data services, market research data, social data • Customer master formation • Customer experience measurements • Customer segmentation • Customer Life Time Value (LTV)	• Metadata management • Integration (ETL, file transfer, replication, middleware, API, ultra-messaging), big data (+document/content management), EDW/ADW, standard reports, dashboards • (Analytical) master data management • KPIs, sentiment analysis, modeling • Data/text mining, visual analytics • Statistical modeling
Relationship Management	• Predict next behavior • Broadcasted/Targeted/ Personalized Campaigns • Next best action	• Big data, ADW, data/text mining • Big data, ADW, campaign management analytics, iBPM (especially Business Rules Engines, Complex Event Processing) • Big data, ADW/EDH, case based reasoning, data/text/voice mining, BRE, CEP
Relationship Optimization	• Multi-channel integration • Internal processes & external efficiency optimization (analogous to Heisenberg's uncertainty principle) • Customer 3600feedback to business processes	• Data/text/voice/video mining, BRE, CEP, • Statistical Process Control (SPC) systems, iBPM (Business Process Analytics, Business Activity Monitoring), Planning • All the above

Similarly, **Supplier** can be dealt by understanding the supplier and supplier network to build the relation, and then manage the relation. Eventually, for optimizing the supplier relation, processes and technologies can be used for supplier process integration, vendor managed inventory, information driven negotiations, transaction data or feedback sharing, resilient supplier networks. Similarly, competitors, employees, affiliates need to be understood well to play a major role in the ecosystem, and increase the area of influence.

As mentioned in the beginning, the Analytics and Information landscape has changed over a decade. The general trend is moving towards encompassing more of **unstructured and semi-structured data** in the analytical arena. The demand on English query on structured and semi-structured as well as unstructured data is increasing. Data Warehousing has been augmented with the power of Big Data technologies to handle semi-structure and unstructured realm of information.

For example, a credit card company matching the credit cardholder information with the Facebook or Twitter based unstructured information to understand the likes and dislikes of the individual and then create personalized campaigns to improve card usage. The speech synthesis software, used over tapped phone lines by Intelligence Agencies to identify terrorists or probability of real threats, is finding its use in the commercial market to understand customer communication as a part of Voice of Customer analytics (VoCa). Voice transcript analysis provides insight into customer communication, especially sentiments. This helps in taking appropriate decisions for improving customer experience as well as internal efficiency across different customer channels. Call center voice analytics used for improving customer satisfaction instead of just meeting the SLAs of call and complaint management. This helps to obtain answers to questions such as which call center agent handles calls well, which agent addresses complaints well in less time, what are the best practices for the same and which agents need to be trained better.

▼ ***FIGURE 28:*** *Enterprise Marketplace & Service Aggregationm*



- In addition to this analysis, the results themselves can be stored to build the **Knowledge Repository**. Using Case Based Reasoning, relevant cases and solutions can be retrieved quickly for faster resolution of the matter. Helpdesk using case based reasoning to get solution for the similar problems raised by the customers.

- Reducing the time taken to diagnose the issue with vehicle by changing the standard operating procedures of testing with adaptive testing built from the knowledge of the maintenance job sheets.

This also reduces the need to have skilled labor for maintenance and hence lowers the cost. To meet customer performance requirements, a Hi-Tech manufacturing company needs to create a specification, price it and place it in a proposal. For a similar proposal, they can perform text mining/search on the knowledge database and fetch an appropriate proposal which will act as the baseline and reduce the time to create a proposal.

The ecosystem data is usually not in a structured format. It takes finite time to get the **semi/unstructured data** transformed in to structured form to enable reports or alerts.' Why is the sale of Engine ABC in USA dropping, while the market is showing a stable demand?' If such questions are not answered immediately, the business will face sales as well as inventory

management issues. The answer may be in the definition of obsolescence of the engine or engine assembly based on engineering documentation. The existing inventory of assembly may be getting discarded based on the engineering definition of obsolescence or the production may be on hold due to unavailability of the replacement parts. Now, such critical issues involve correlating structured demand/supply data with unstructured engineering documentation in the 'right time'. Just to add complexity, consider demand driven supply chain for a retailer. The retailer will need to understand and correlate the product specific consumer behavior, competitive product pricing, campaign effectiveness, social trends, economy drifts and influence of global-local events and effect of weather. This is in addition to the correlation with POS and on-hand inventory data. If this needs to be done at right-time then technologies such as Complex Event Processing along with **in-memory and Big Data** based knowledge repository are needed.

Ecosystem is all about different players in value chains coming together for business(s). An enterprise can rethink about the existing value chains and build a Marketplace by bringing multiple players together in a business.

A financial services company bringing its Corporate customers (primarily Small Medium Enterprises - SME) to a marketplace with their offerings (such as new books in store, new movie releases in theater, new franchise items related to movie) and informing the appropriate consumers about these offerings to maximize sales through personalized campaigns. Similarly, a retailer bringing SMEs in CPG industry to an eCommerce marketplace (along with other standard CPG companies) and informing the appropriate consumers about the offerings through multi-channel personalized campaigns. This involves understanding consumers with the help of enterprise data as well as social data and then recommending specific offerings to specific individuals. The next step of marketplace is definitely the Service Aggregator role played by an enterprise. For a consumer, all the service or commercial activities are finally settled through financial transactions. Hence, Enterprises (primarily from financial services) can play the role of service aggregator by brining multiple services together with the backbone of financial payment or settlement hub.

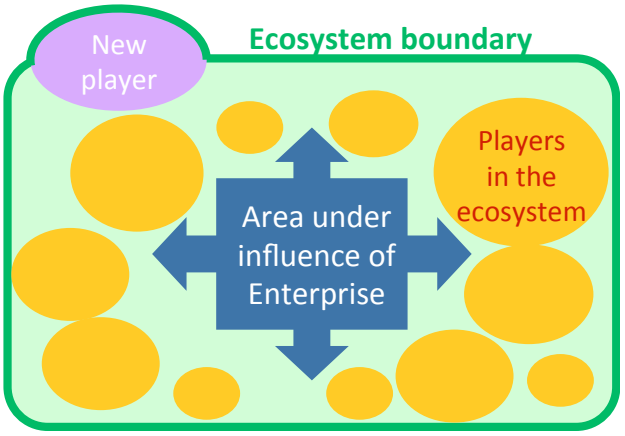
The other type of marketplace is the **Data Marketplace**. This is more known as Data Monetization area with different service models for sharing data as shown in the following table:

Table 2: Service Models: Sharing Data and Insights

Service Model	Description	Examples
Data Services	The data is supplied in different forms to the data consumers. Based on the form the service charges differ. • Raw data dumps • Data Aggregation • Enriched data (with the help of measurement sciences) dumps or reports • Enriched data service APIs. Data Privacy, Legality, Cross border regulations and Security are definitely critical aspects in the consumer data sharing area.	• Retailers like Walmart share their POS data to suppliers (CPG, Manufacturer, Home Entertainment companies) • Market Research Companies like Nielsen aggregate the data and enrich it to share it with CPG, Retail companies. • Consumer data providers like Experian, Equifax supply consumer (or corporate) level details and credit ratings • Credit ratings and other details of individuals or corporates are available over API services to insurance, financial services or individuals from credit rating companies.
Analytics Services	• On-demand Analytics services are provided by Market Research companies by enriching the data and making it available through reporting or advanced visualization services.	• Nielsen provides AOD (Answers On Demand) services using OLAP Reporting tools. The insights are also provided through these analytical tools.
Advisory Services	• Instead of data or reports, advisory services are given to CPG, Retail companies through the consulting arms of the Market Research companies.	• Companies combine data across value chains e.g. TV set top box, mobile, internet browsing, POS and research data is combined by Nielsen to provide advisory services on advertisement effectiveness, product bundling, seasonality etc.

There are four main parameters based on which these marketplaces are valued, namely Quality of data, Measurement Sciences, Volume (history, breadth of value chain) of data and Speed of delivery of data. Based on excellence, premium rates are given for these services. In an ecosystem, **this opens up a new business model for the enterprise** and the data becomes a profitable asset instead of just a cost center for IT.

Ecosystem spans across the boundaries of the frame of reference in which the enterprise resides. The more the influence of the enterprise on the ecosystem, the more is its area of control. It can maximize the area of control by maximizing the reach across the ecosystem players and optimizing the efforts to do it. More players may get added to it and hence expanding it is the frame of reference. This may affect the area of control of the enterprise in a positive or negative manner. This ever changing boundary of ecosystem needs to be understood and managed. *Netflix used the data of TV viewership to enter into the content creation business by creating the famous show "House of Cards". This has shaken the ecosystem of content creators.*



▲ ***FIGURE 29:*** *Ecosystem Boundary*

While improvising the self with operational excellence through strategic visibility and tactical capabilities, it is necessary to sense, respond, manage and optimize the relation with various entities in the ecosystem. **The fourth maturity level (4ml) is all about bringing Performance Intelligence and Relationship Optimization together to establish ecosystem intelligence.**

Universal Intelligence

5 ml: Risk Intelligence along with Relationship Optimization & Performance Intelligence

The 2ml enables predictability and scenario modeling with the backbone of EDW and Data Mining. In addition to that, with 4ml, we can build the ecosystem knowledge and understand the Past, Present and to some extent the future, but we cannot predict with certainty 'How it should happen?' The how becomes all the more challenging to define when we consider the current dynamic market, where the rate of change is faster than yesterday and what we do today may not be relevant tomorrow.

Furby was the top sold item in the 2000 holiday season, but the next year it had negligible sales. If DELL wouldn't have innovated to shift from Make-to-Stock to Make-to-Order, they would have become a victim of home computing and the open architecture evolution. If American Airlines would not have created Frequent Flyer system by analyzing their ticket discounting patterns, they would not have re-gained the market captured by their competition.

To survive and grow in the vibrant market space, every enterprise is facing outrageous competition, business convergence, demand for profitable growth, intense pressure on cost and disruptive innovation. But what about the survival and growth of the market place itself? Is the

enterprise fully aware of the risks that exist in the market place and beyond the boundaries of the market place? In order to facilitate business excellence, **the understanding and correlation of enterprise performance, ecosystem support and risk prediction with certainty is necessary and the fifth maturity level is the lever for that**.

Donald Rumsfeld's famous quote is about understanding the risks; "Reports that say that something hasn't happened are always interesting to me, because as we know, there are **known knowns**, there are things we know, we know, We also know there are **known unknowns**, that is to say we know there are some things we do not know. But there are also **unknown unknowns** – the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tends to be the difficult one. With the understanding of **Risk Given Event (RGE)**, Enterprises build risk management systems and track the risk indicators. This is the standard part of known risks and exposures. The implementations of Information Management Controls (Transparency, Quality, Change Management, Security, Governance and Auditability) and Operational Risk type of applications at second or third levels respectively, give better control over Known Knowns. The known issues with unknown exposures and unknown issues with exposures analogous to known issues are tracked by enterprises with the understanding of the ecosystem and players and their contribution, influence or knowledge. This is where market risk applications, credit risk information from external sources, network information of entities (supplier, customer, competitor.) and their combined effect gives a lot of control over Known Unknowns and Unknown Knowns. When neither the risk is known nor the exposure, the only way forward is to get more information about unknown parameters and expand visibility beyond the ecosystem. There are different levels of tolerance based on enterprise capacity to manage the exposure, appetite to take risk and ability to stretch limits while managing the point of no return. If Ministry of Defence does not look into news across various channels (Newspaper, Radio, TV, Internet, Mobile.), tapped communications and other secret sources on daily basis and synthesize it to make more and more sense out of it, then the country cannot be safe. On the other hand, they do not create panic on certain extreme information as well. This confidence is due to the ability to predict the risk with certainty. The **precision in predicting risk gives advantage**, such as more capital available for operations as well as innovation, regulatory authorities giving relaxation on the Cash Reserve, provided the risk predictions are consistently well under the threshold.

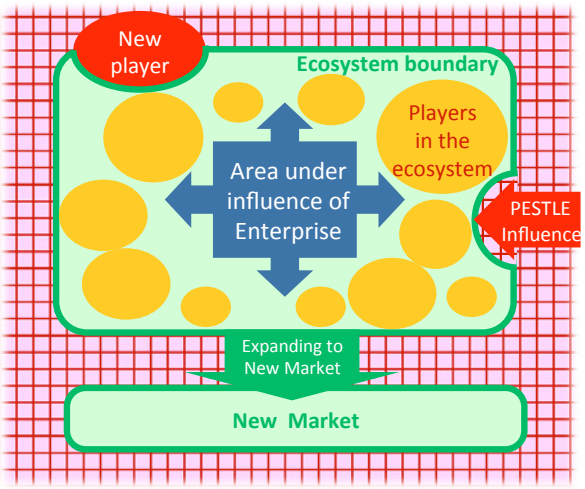
Visibility \ Reality	Known	Unknown
Known	1-3ml	4ml
Unknown	4ml	5ml

▲ ***FIGURE 30:*** *Visibility vs Reality Maturity Quadrant*

Although the topic appears to be of risk with increasing scope from known risk indicators to semi-known correlational risk to unknown risk) it should be seen in holistic perspective of enterprise performance management, relationship management and risk prediction. Risk events are not always risks. These are events which can take any form based on how enterprise treats them. Hurricanes are opportunities for retailers for extraordinary sales of packaged drinking water and items like pop-tarts and epidemic outbreaks are opportunities for medicine. Some enterprises **use these opportunities more from relationship perspective** than just from (or in addition to) business growth opportunities. Free medicine or vaccination in case of epidemic outbreak brings a huge loyalty and goodwill from the people to the pharma company and it lasts for decades. Events (or information) are like **free radicals** (an atom, group of atoms with unpaired electrons and they are highly reactive). They immediately get associated with other events forming a chain of events or what human mind interprets as cause and effect. Some theories suggest that

there is no correlation like cause and effects; events are independent and it is only human nature that associates things together to create a meaning or pattern out of it. Human nature is by default configured to create these associations among events as neurons work in the same manner. The activation of a receptor with a signal causes the effects or functions associated with the specific neuron resulting in actions. The sympathetic nervous system based on the stimulation triggers the body's fight-or-flight response. Correlation of events or objects and interpretation is our basic cognitive nature. The more variables (events or objects) we correlate, the larger the scenario we can build. The larger the scenario, the clearer is the understanding of the Universe. If we do not know about a few variables, then that does not mean they do not exist. There is a saying, "Absence of evidence is not evidence of absence." A global car manufacturer with manufacturing plant in China incurred more than a billion dollars of loss due to the Tsunami in Japan. The link that was missed in these two events was that the China plant was using dashboards from a dashboard manufacturer in China, who was using integrated circuit chips from a Japan based manufacturer and there was no alternate supplier for the dashboard or chip within the supply chain of the global car manufacturer. Risks can be opportunities and opportunities can be risks based on our intelligence span across the Enterprise-Ecosystem-Universe.

The expansion of the picture of ecosystem boundary given in 4ml will look like this. There are new players, who try to cross boundaries. The enterprise may never recognize these players as competition until it starts losing business. Such as, Brick and Mortar Retailers losing business to eCommerce Retailers, Banks losing business to Hi-tech mobile companies (ApplePay) or mobile service providers (eWallet) and Taxi services losing business to mobile app based broker system (Uber). The re-imagination of the services in Digital Consumer Economy is opening huge risks for the traditional business players and there are clear cases of business espionage. On the other hand, traditional players are also entering new business models to de-risk loss or create new opportunities for growth such as Banks, CPG and Retailers with saturated businesses in various geographies are entering international markets. Mergers and acquisitions are happening to enter new markets or wipe-off and damage competition, and companies entering services space of B2C business. These changes are so rapid and disruptive in nature that traditional analytical methods and processes cannot cope. Decision making aided by Artificial Intelligence systems used in conjunction with and agile processes is driving business in the digital era. Hence **Digital Reimagination™** of the business model, Business processes, Channels, Customer Segments, Workplaces, Products and Services is necessary to expand the area under influence and also expand the boundaries of the ecosystem.



▲ ***FIGURE 31:*** *Extended Ecosystem – Universe*

Along with the new players, **PESTLE** (Political, Economic, Social, Technological, Legal and Environmental) changes also influence the ecosystem boundary. This is the prime area where the universal intelligence to understand the unknown unknowns becomes extremely important. It is extremely difficult (assuming that nothing is impossible) for human intelligence to keep track of these frequent changes and forecast the effect on business. It is also difficult to rely on

unsupervised Artificial Intelligence to forecast and drive business decisions. Intelligence Agencies have been handling these scenarios for more than a decade. It is now that their ideas are finding its use in the commercial environment. They have always used right mix of human centric decision making (with highly skilled Analysts) and machine centric decision making (CEP and Artificial Intelligence systems) to solve complex problems of PESTLE and give stability to the countries as well as universe. Hence, it is logical to use a mechanism similar to these agencies (of course it cannot be same as these agencies as the investments are limited in any enterprise) in which the right balance of Human Intelligence and Artificial Intelligence is formed in an Agile Intelligence Center. An unbalance can cause severe damage to a business, brand, ecosystem or even existence.

While building **Intelligence Center** four things are extremely important:

- **Agile Culture:** Instead of always following a waterfall model from problem definition to solution delivery, how agility can be induced in the culture for outcome driven delivery. Prioritization of requirements, synchronized vocabulary, adaptive processes, fail-fast approach and highly motivated and co-active teams facilitate the agile culture, which is the foundation for the disruptive innovation on continuous basis. In Spotify Methodology, although there are teams like Tribes, Practices and Guilds, the members can move based on the need of the hour, but still maintaining and extending the knowledgebase of the overall group. Unlike Waterfall, the delivery cycles are shorter. There is no fear of failure in long term due to the time boxed processes. This makes the team more confident and they can take calculated risks. But not everything can be agile like technology or data migration programs. In such cases, Waterfall model needs to be there with Factory or Work package approach. Both the cultures need to exist with the right balance.
- **Human Intelligence and Competence:** The enormous flow of information, limited time (only 24 hours in a day, not a single second more) and numerous decisions to be taken in the given time. The scenario is becoming more and more difficult in the digital era. That is where a self-realization is necessary in humans. Taking all types of decisions with all types of analytics is not necessary. Prioritization needs to happen to focus energy on high impact decisions. For other decisions and actions, either rule based automation or machine centric automation is necessary. Overloading self usually happens due to human nature of not accepting failure in actions in spite of having zero bandwidth. Movement is necessary from unconscious incompetence, where humans do not realize the mistakes (unlearning process) to conscious incompetence, where they realize the mistakes to conscious competence, where they build the skill to unconscious competence, where the skill becomes their "second nature" (like refluxes of master black belt in karate). There are people who analyze the data and give a gut-feel and most of the times they are precise in answer. This is a typical example of unconscious competence, which sometimes supersedes the machines as well. Various new techniques such as Visual Analytics make it easier for humans to look at voluminous data in an easy and intuitive way. Hence, introducing a higher class of visual analytics (D3) instead of tabular reports or simple charts is necessary. It improves productivity as well.
- **Artificial Intelligence:** Investment in systems is necessary to relieve humans from mundane jobs or for synthesizing data in a short span of time so that humans do not need to do these tasks manually or by building a knowledge base instead of building a dependency

on individual humans in an enterprise. Making these systems more intelligent is the natural progression towards self-optimizing an enterprise. Artificial Intelligence spans across the branches of:

- **Deduction, reasoning, problem solving**
- **Automated planning & forecasting**
- **Natural Language Processing (NLP in communication)**
- **Perception (speech, facial, object recognition)**
- **Knowledge representation**
- **Machine learning**

There are long term goals of Social Intelligence, Creativity (Practical and Theoretical i.e. Psychological + Philosophical) and General Intelligence (self-awareness), but those will take five to ten years to come to commercial markets.

- **Governance:** While understanding the performance capability and ecosystem equations the tolerance to manage the risk exposure, appetite to take risk and ability to stretch limits need to be highly governed with different lines of defense mechanisms. When it comes to governing the point of no return, the balanced Human and Artificial Intelligence with Agility plays the lead role. Governance should not be set to as a hurdle in actions. Rather, it should be set as a protective gear in disruptive innovation.

As per John Nash's game theory, when all the players have same objective in the game, only then the outcome can be maximized. When all these four things have been directed towards the same objective of business excellence, then predictability of outcome can be certain. In case of lean manufacturing, in order to maintain the near-zero inventory, the Out-of-Stock situations across all plant warehouses can be predicted just-in-time (and neither ahead of time nor after the cut-off time) and material procured without putting the labor on hold. With a centralized command center global power generation operations can be monitored and preventive maintenance can improve generator availability and reduce loss. CPG companies can manage demand driven supply chain through a single global operations center. In the absence of same objective, Governance can block the agility and human intelligence can choose to override the artificial intelligence most of the time. The results will obviously be the ineffective Intelligence Center.

Hence, **the fifth and the final maturity level (5ml) is all about universal intelligence**, which is known as "**Turiya**" in the Indian Philosophy. **It helps to institutionalize innovation by correlating Performance, Relationship and Risk for the enterprise to achieve the state of self-optimization** with the backbone of the intelligence center.

Appendix D

Data Democratization

Data as a catalyst for exponential growth

Data is a valuable resource and mankind has a knack of producing it abundantly. In recent years disruptive technologies have permeated our everyday lives and triggered ever greater levels of data generation. Unique ways of both creating and capturing data have emerged which provide numerous possibilities for optimization, targeting, and the generation of insights and innovation. But unless this data is available to, and used by, relevant people and programs, its value is dramatically curtailed. Democratizing data – ensuring it's of, by and for the people, – is crucial to realizing its benefits. Its objective is to enable individuals and groups to use it at their convenience, where possible without the need for expert assistance.

Data boosts our understanding, optimization, and productivity since it has reciprocal effects on the technology from which it breeds. Take, for example, Tesla, and its self-driving cars. A flood of data about road conditions, maneuvering techniques and driving best practices makes the technology possible. The capture and interpretation of data in this manner advances the technology and in turn, more data is generated.

Digital-first companies with embedded Business 4.0 technologies such as Artificial Intelligence (AI), Big Data and analytics, Cloud, Internet-of-Things (IoT), have an upper hand but legacy businesses can also benefit by democratizing data. Taking advantage of this data will help them thrive through faster decision making, enterprise agility, improved customer experience (CX), operational efficiencies and new products.

Why do we need data democratization?

Data has been called the new oil. Refining it into actionable intelligence unleashes its potential. This can be achieved at three levels.

- **Within organizations,** it has the ability to fuel business projects across functions.
- **Outside of them, within the ecosystem**, it can provide invaluable insight to and about the customer- and supplier networks, and other stakeholders.
- When data is **shared into a universe of academia, data scientists and policymakers,** it acts as feedstock for new and innovative products and services.

Democratization pushes organizations to rethink how they manage, distribute, and interpret data. Traditionally, data has been considered an IT asset, but this limits its value. **When data, and knowledge, are made available to functions across an organization its latent potential is unlocked**. New collaborations and innovation take root. Technology acts as a great enabler. With a balanced combination of machine and human interventions, the decision making ability of an organization increases.

Unleashing the true potential of data

To democratize data, the effort needs to address three categories: Data literacy, data enablement and the creation of borderless ecosystems.

Literacy: The foundation stone

For data to be relevant **requires a level of data literacy and an ability to understand its provenance**, its value, its limitations, and risks. As the volume of data available increases, so do concerns around its governance and hence governance becomes a key facet of data literacy.

At the heart of the data literacy debate are questions about where the data resides, who owns it, and how good it is. Digital technologies enable answers to these. But it is also crucial that leaders understand the implications of democratization for individuals within an organization: What's in it for me? How will it shift the paradigm of decision making? What does it do for career progression? How does it elevate the role of humans in an AI world? These questions are precursors to data democratization.

Governing data democratization is essential to plan and hedge associated risks and avoid potential reputational damage. Consider the six dimensions in this checklist:

- **Ethics:** An ethical process requires understanding the purpose for which data is shared. Assessing whether the intent and basis for collection and sharing of data are published, and frequently monitored, is critical.
- **Privacy:** Adhering to privacy norms to protect personally identifiable information is key in maintaining stakeholder confidence.
- **Trust:** Globally businesses find themselves facing a trust deficit, in part because of the way data has been treated or mistreated. Issues related to trust with the entities in the data value chain, including regulators, must be identified and managed.
- **Legal:** A dedicated legal team focused on democratization activities should be appointed to oversee regulations and data related legal risks.
- **Risk:** Every business process involves data-related risks. An action plan to reduce its impact will strengthen the chances of success.
- **Valuation:** Data is a valuable, as well as expensive, asset. The democratization process should only be undertaken if it adds value and the effect on brand value, security, moral values, and principles can be monitored.

Data Enablement: Capture, management, and consumption

Digital technology makes capturing large volumes of information possible. This now includes IoT sensors embedded in products, images, videos, audios, and other paperless forms. Once the information is captured, the raw data has to be made usable which means converting it to a meaningful format.

A global bank needed to have authorized data stores to meet regulatory compliance. Data used for decision making in banking needs to be traceable and metadata tracks its origins, transformations, and approvals. Master Datasets including customer, product, and channels information, need to be

uniformly defined and managed across the organization. So, when regulators ask questions, it can now answer in 1-2 weeks, not 2-3 months. The bank's stress-test ratings are up.

Information on which decisions are based has to be trustworthy so its authorization is important. It cannot be done manually so digitization of the certification process in real-time, and its frequent audit, are pressing concerns. Controlling and authorizing ever-increasing volumes of real-time data is only possible digitally, whether it be structured, semi-structured, or unstructured. Configuration of data controls should be focused on activities such as the correlation of business-data elements, cross-verification for quality, cleanliness of masters, adherence to privacy and security norms, and records of the data life cycle. This improves the trustworthiness of the data.

Capture and management of information should be the responsibility of the business and not just of its IT function. This builds enterprise-level accountability. Data capture and management is a key enabler of democratization so that the right people in the organization can use it on their own instead of depending on help from experts. Until this phase is done properly, consumption cannot take place.

Individuals prefer different ways of looking at data. We're not all data scientists so a key aspect of democratization is the packaging of data in forms that encourage consumption. Three categories of services can be used:

- **Data Services:** These can be facilitated through mechanisms such as API, microservices, reports and dashboards, search engines and conversational agents. At the heart of these is the simplified semantic layer which is overlaid on top of different data stores. Such plug-and-play technologies encourage effortless consumption and business users find a value that can be extracted from data. These can also be extended to partners through controlled channels and new value streams can be sought from existing investments.
- **Visual analytics:** Most people understand visuals better than numbers. Illustrated data is easier to consume than a large content-heavy dataset. Visual depiction is gaining traction in the form of infographics and interactive reports. There may be a number of people in the enterprise who are good at data visualization and they should be enrolled. Self-service visualization tools that build appealing reports and help easily identify patterns are increasingly accepted in business user communities. Visual analytics is a key enabler of consumption. Augmented and virtual reality are gaining prominence with data scientists and business users.
- **Conversational systems:** Asking for the answer has never been easier. Conversational systems are changing our expectations of our interactions with machines. Whether it is a point data or a report or a document search a strong data services layout at the backend is required to fulfill such needs.

People dependency may lead to knowledge leakage so establishing digital knowledgebase, as a part of decision management exercise, is wise. A combination of policies, procedures, and technology are needed to retain knowledge even if there is employee attrition. This knowledge, when democratized, can help more people to make better decisions.

Beyond the Organization: Borderless ecosystems

Making data available outside the organization creates opportunities for suppliers, partners and other members of the business ecosystem (including competitors) to **create new value**. Beyond the organization and its partners, regulators, policymakers, academics, and the general public may also benefit greatly from the broader availability of some kinds of data and even they can reciprocate the benefits.

Opening organizational data to the ecosystem can lead to improved efficiencies, optimization of resources, and innovation. Insurance companies identifying fraudulent claims are limited by the data within their own organization. If this data were shared, the number of frauds could be reduced as the ecosystem is able to perform better vigilance and weed out offenders. Democratization here creates a win-win situation for the company, its partners and end customers.

A global retailer opened up its data to suppliers (ecosystem) via an extranet. Suppliers can now analyze their performance and importantly compare it to that of others. They can plan better to meet SLAs and strategize against the competition. All parties have access to the data which then informs negotiations and drives improvements for the retailer, its supply chain partners, and customers.

A market research organization has developed new revenue streams by monetizing the data it and its ecosystem partners gather. Platforms are emerging as data marketplaces allowing partners to decide what and how they wish to share data commercially.

Open data policies encourage the democratization of data. They provide individuals or businesses a means of accessing information to identify traits and risks and they can stimulate new business ideas. Open banking policy in Australia, and the UK's PSD22, policy are great examples. Sharing information encourages the identification of customer behavior and risks. But organizations need to consider the latent risks data policies bring with them. Open banking policies may help to stop money laundering activities, frauds, and financial crime while opening sensitive information raises concerns about data privacy.

Problems involving a large combination of data may prove to be overwhelming for a company's limited set of data scientists. Opening the data to a larger set of people including academics, institutions, and regulators could be the solution. The drug discovery process in life sciences is a case in point. Research into new molecules requires possible combinations of multiple factors such as regions, demography, ethnicities, weather conditions and reactions to other molecules. Such scale and complexity create immense challenges – often too complex to be handled by scientists within a single business. Currently, the potential to discover personalized medication is enormous and deducing patterns for dosages is an important part of that process. It's clear how life sciences can greatly benefit from the democratization of data at a universal level.

Building a secure environment on the cloud and using containerization allows data sharing and experimentation without compromising security. Proper tokenization, masking, and pseudonymization techniques will further protect data whether it's being shared within the organization, ecosystem or universally. A crowdsourcing platform with such capabilities can enable the global data scientists' community to help organizations make better decisions.

Dare to Democratize

The governance checklist lets data-rich companies evaluate if and how they can increase their data's value exponentially through democratization. If the outcome is positive, then they must develop an implementable strategy. Data security and privacy should remain in firm focus throughout.

Education and skilling–improved literacy–contribute to the ecosystem-wide adoption of democratization. Breaking organizational siloes is a necessity to make data a mutual asset so that its benefits can be shared collectively. Without **collective accountability and active support of all participants** across the **enterprise-ecosystem-universe**, democratization cannot be accomplished. An enterprise-wide culture shift will be required if data democratization is to achieve its full potential.

A global data revolution is underway. Leaders in their sectors will take advantage of their data to become optimized, innovative and agile.

Appendix E

TCS DATOM™ Assessment App

The TCS DATOM™ Assessment App is a **digitized, self-service tool for data and analytics assessment** based on TCS' DATOM™ framework. The App presents a comprehensive set of data maturity assessment questionnaire across each of the sub-KRAs related to the four KRAs (Data, Technology, Process and People) and provides rich GUI capabilities to capture responses, record notes and upload relevant artifacts to conduct the assessment process.

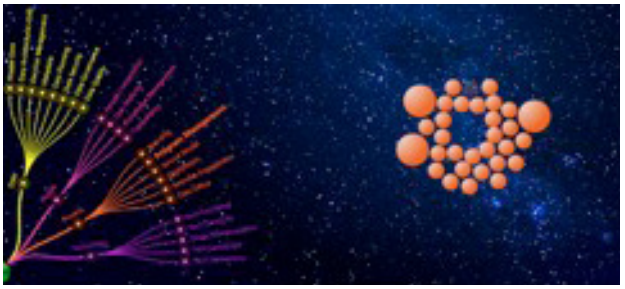
▼ ***FIGURE 32:*** *TCS DATOM™ App Login*

A customer organization can undergo assessments at various levels – Conglomerates, Enterprise or Line of Businesses or Functions or Region. Based on the inputs from different stakeholders, the tool has the capability to calculate a Data Maturity score, called the DATOM™ Index and derive the maturity of the organization based on 5ml maturity levels *(explained in Appendix C): Siloed, Simplified, Scaled, Synergized, and Self-Optimized.* Besides, it has the ability to benchmark an organization's maturity, relative to other organizations in similar industries, who have already undergone this assessment.

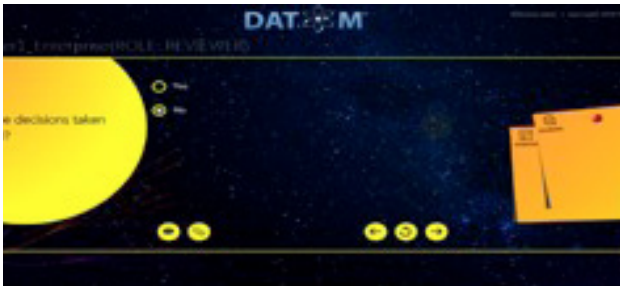
Key features of the tool are as follows:

- **Built-in Adaptability** to conduct assessments for Conglomerates, Enterprises, Line of Businesses or Functions,
- **The curated set of questionnaire** aimed at evaluating the data maturity of an organization. It considers the depth and breadth of complexities, nuances and varying scenarios of data analytics capabilities present across diverse organizations and industries.
- **Unique Scoring Algorithm** based on 5ml Maturity Model.
- **Analytics and Dashboards**, which give:
    - Comprehensive view of the organization's maturity in the Data and Analytics space, highlighting weak and strong areas at the sub-KRA grain, and aggregated at KRA and organizational level.
    - Comparative analysis of data maturity across different lob's/functions/regions for an organization.
    - 5ml based Industry and Geography specific benchmarking model for Data & Analytics giving a comparison of maturity with respect to the best within and across industries.

▲ ***FIGURE 33:*** *TCS DATOM™ App Landing Page*



▲ ***FIGURE 34:*** *Assessment Navigation & Control*

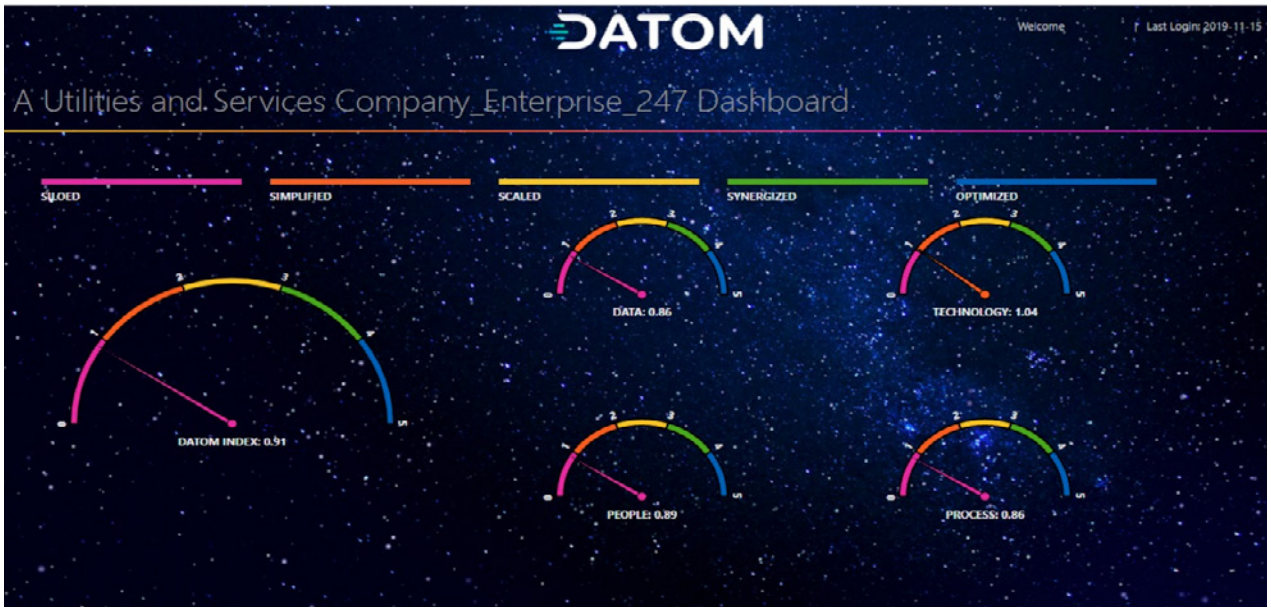


▲ ***FIGURE 35:*** *Capturing Response*

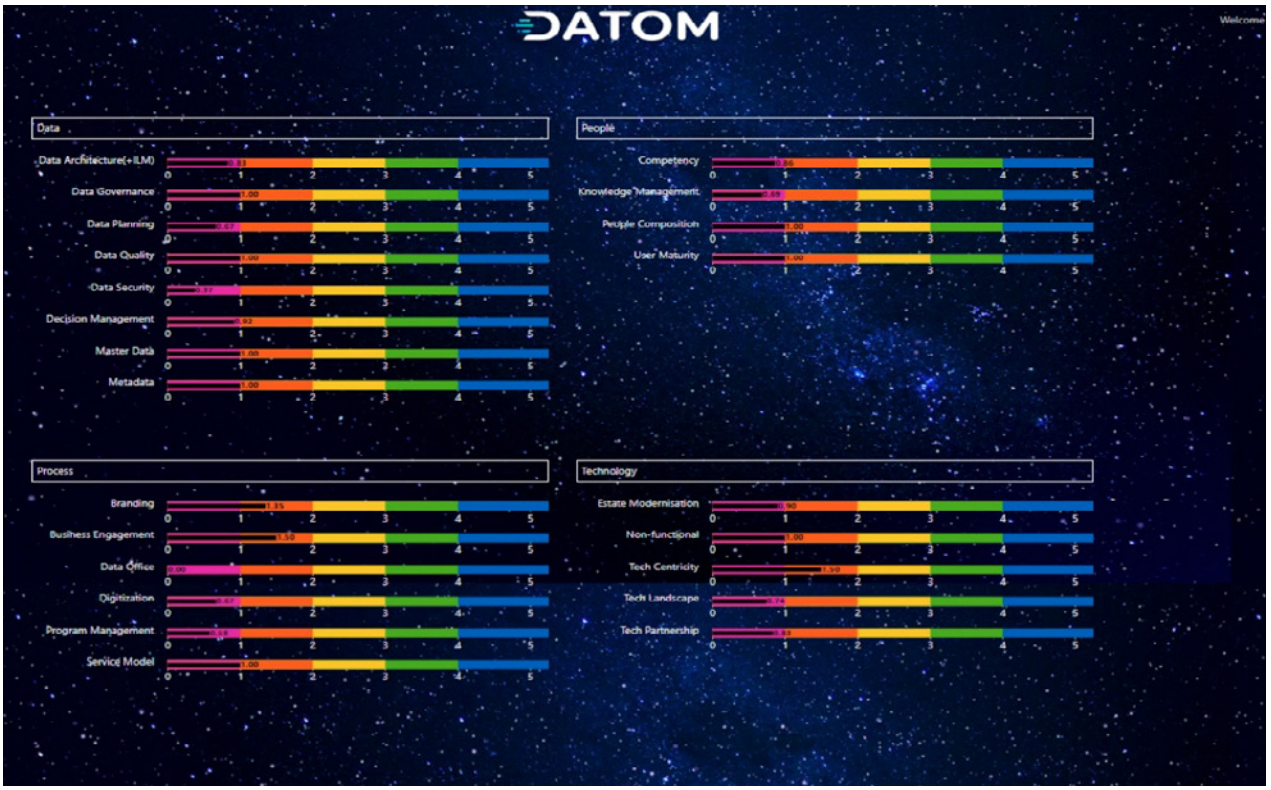
- **Contextual recommendations** and Point-of-View on Data & Analytics roadmap
- **Assessment Approval Workflow** with complete transparency in execution
- **Elevated User Experience** through creative UX/UI Design, Visual Aesthetics and Advanced Visualizations
- **Text and Voice Bot for virtual assistance.** Guided navigation to simplify the assessment & complete in optimal manner

The following are some of the sample outputs from the TCS DATOM™ Assessment App:

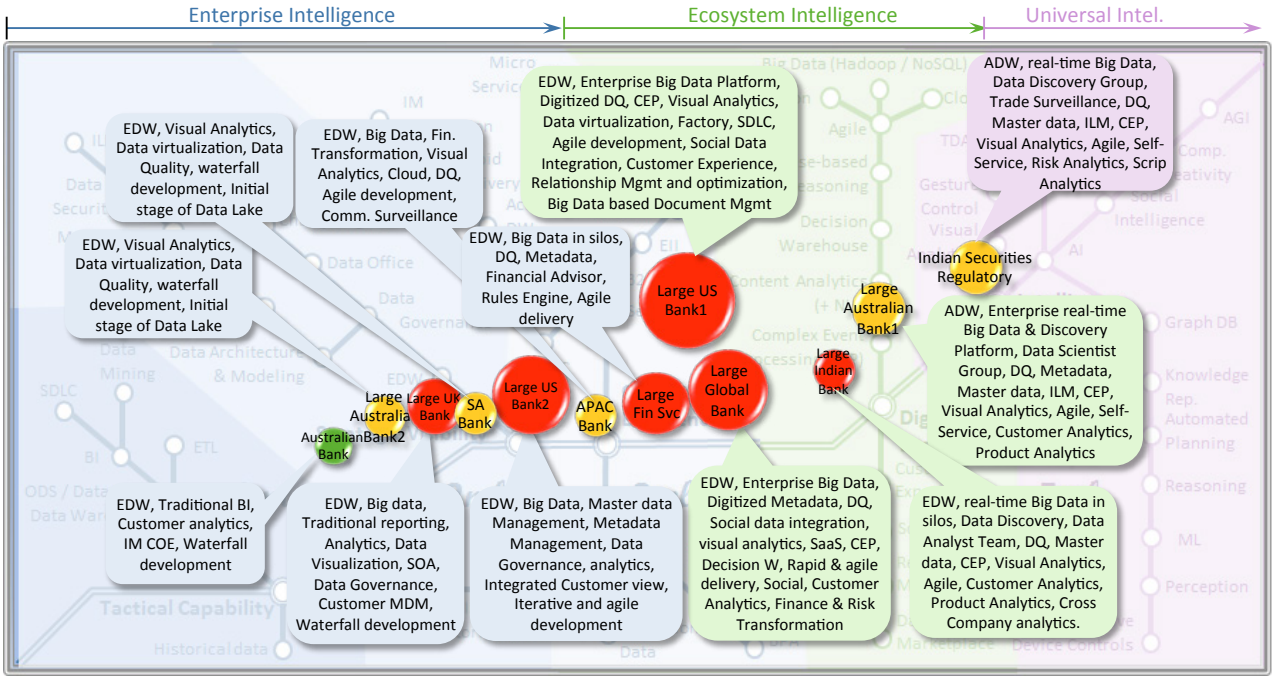
More details on the tool is available in the TCS DATOM™ Assessment App User Manual and TCS DATOM™ Assessment App Administration Manual.



▲ ***FIGURE 36:*** *TCS DATOM™ Index Dashboard - Top Level*



▲ ***FIGURE 37:*** *TCS DATOM™ Index Dashboard - Sub-KRA Level*



▲ ***FIGURE 38:*** *Benchmarks*

About TCS Analytics & Insights

TCS' Analytics & Insights (A&I) service line helps organizations view, understand, and reimagine their businesses through an intelligent data-centric approach. We design innovative solutions for superior business outcomes and help customers execute effective data-driven strategies. The A&I unit researches key emerging trends, including Artificial Intelligence, immersive analytics, mobility, cloud computing and social networking to develop innovative, practical, and powerful applications to deliver business results.

The founding principles of A&I unit is to provide Business Stakeholder Advocacy, Full Services Play, and IP-based Value Realization and the NextGen strategy is powered by:

DATOM™ - Our Data & Analytics maturity assessment, consulting and advisory framework

DAEzMo™ - Data and Analytics Estate Modernization powered by TCS' off-the-shelf offerings and solution accelerators

Decision Fabric™ powered contextual industry offerings

TCS **DATOM™** helps organizations to adopt best practices depending on the context of their data and analytics programs. By simplifying operating models, we help organizations leverage complex technologies with ease, and set up frameworks aligned with their business goals.

We offer **DAEzMO™** (Data & Analytics Estate Modernization) powered solutions to simplify, synergize, and scale an organization's data landscape and analytics portfolio to enable exponential business growth through agile transformation as technology continues to evolve at rapid pace.

Our award winning solution accelerator, **Decision Fabric™**, caters to disruptive innovation for next generation contextual industry offerings. It weaves components from different disciplines of AI such as Reasoning, Learning, NLP, Perception and Knowledge Representation, to continuously learn and take decisions for advanced automation as well as human assistance.

To help enable scaling up the execution of A&I strategy, we have created an A&I Evangelist Community Development Program. TCS **DATOM©** assessment will be one of the first goals of this community.

For more details, reach out to **Analytics.Insights@tcs.com**

List of Figures

Notes

TATA CONSULTANCY SERVICES
www.tcs.com.

TATA BUSINESS EXCELLENCE GROUP
www.tatabex.com