Whitepaper

Implementing Data Mesh

Six Ways That Can Improve the Odds of Your Success

Author:

Ranganath Ramakrishna, LTIMindtree

LTIMindtree

Contents

Winning with Analytical Data

Data Mesh is a socio-technical paradigm that can help organizations to fully exploit the value of their analytical data. Data Mesh paradigm enables organizations to transform analytical data into building blocks called data products, which can be combined in a myriad of ways to deliver use cases, to differentiate their products, and services. Data Mesh paradigm, if understood and implemented well, can deliver the vision of a data-driven enterprise – an enterprise which deploys data and analytics to innovate and optimize every aspect of its business to deliver outstanding customer experiences.

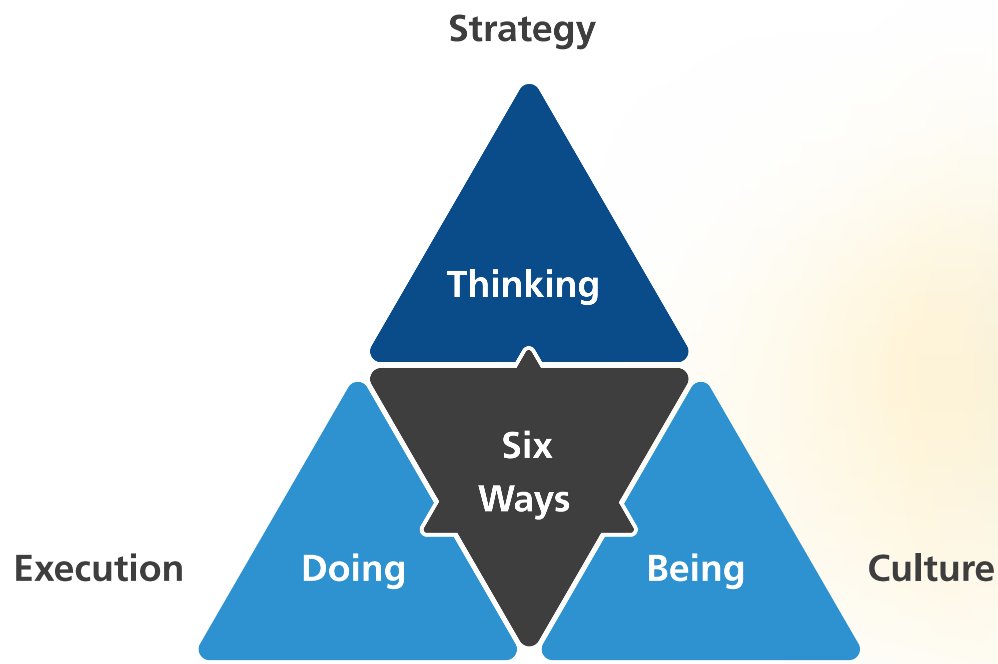
In a digital economy – software, powered by Data and AI – has become the core differentiator or probably the only differentiator. Customer experience, and market share are increasingly determined by the quality of software. Success in online banking, e-Commerce, consumer electronics, streaming, smart homes, or any other sector in economy is driven by software. Differentiating software itself, is built on the foundation of data and analytics.

Deep understanding of markets, customers and products, high value decisions, proactive risk management, and operations optimization, are all consequences of fully exploiting signals embedded within data. Today, it is not an overstatement to claim that an organization is only as good as its ability to instrument, capture, analyze, predict, and differentiate with data. How organizations leverage data will determine their success and survival in the digital age.

With data mesh, we move away from the notion that data is a tool for internal decision making. We stop treating data just as a utility and byproduct of running business processes. We no longer obsess about the volume, velocity, and variety of data. We fully embrace the idea of data as a product, which delivers value to end consumers. We consider data as a foundational building block essential for executing business strategy. Data becomes a first-class citizen, a key factor of production with direct impact on customer experience, revenue streams, and profitability. This shift in perspective can help organizations to win with data.

This paper is not about describing the concept of data mesh. It assumes you are already a convert. It assumes you are familiar with the concept of data mesh and the four principles. Zhamak Dehghani, creator of Data Mesh covers data mesh elaborately in her seminal book, Data Mesh - Delivering Data-Driven Value at Scale. It is an essential read if you are looking to understand data mesh comprehensively.

This paper synthesises and describes **six ways**, which can improve your odds of success in the adoption of data mesh. This is based on the collective experience of practitioners working on data mesh and data products across LTIMindtree. The six ways emphasise the shift required in thinking, doing, and being. The paper is for data practitioners. It is about breaking old habits and picking up new ones. It is about framing things differently and looking at analytical data with a fresh set of eyes.



Introducing the Six Ways

The diagram below shows the evolution of the data infrastructure and architecture models broadly adopted by organizations over four decades. On infrastructure, we have moved from Symmetric Multi-Processing (SMP) to Massively Parallel Processing (MPP) systems, open source-based distributed systems (Hadoop), and currently to cloud-based data platforms. On architecture, we evolved from Enterprise Data Warehouses (EDW), data marts, data lakes, and data lakehouses.

The progress in technology and architecture represented a step change in our capabilities. However, we have not been able to translate the technological capabilities into concomitant business value and outcomes. Organizations are realizing that investments in data infrastructure and architecture alone, disconnected from operating and business models, won't deliver on the vision of a data-driven enterprise.

In this context; data mesh offers a new perspective to solve the problems of lack on return on data investments and business transformation. It addresses previously overlooked aspects of domain, organization, and product thinking. It focuses equally on both social and technical aspects. This paper looks at how data mesh thinking can inform all the four major axes of an enterprise.

Enterprise

01
Business model

Customer value proposition and profit formula

02
Operating model

Ways of working embedded in business processes

03
Architecture model

Structure, design, and implementation of applications and data

04
Infrastructure model

Technical components like server, storage, network, and databases

Generation 1	Generation 2	Generation 3	Generation 4
Bespoke	SMP & MPP	Hadoop	Cloud
Custom architecture	EDW, dimensional	Data vault, data lake	Data lakehouse
Fixed cost, capex	Fixed cost, capex	Fixed cost, capex	Variable cost, opex
Support function	Reporting	Reporting and analytics	Competitive differentiator
Top executives	Executives & managers	Executives & managers	Democratized, all people
Custom & silo	Centralized, monolith	Centralized, monolith	Decentralized, data products

- Who owns the data in the datalake
- Central team is slow to meet needs
- I don't trust the dashboard
- I am not sure about lineage
- We don't have a data culture
- Not able to monetize data
- Data is not a differentiator

Technology | Data Architecture | Cost Model | Expectation | Personas | Architecture Style

Brief description of Six Ways

The diagram below captures the six ways and a brief description of each of the way is provided in this section. Each of the way is detailed in subsequent sections. The emphasis is on what shift is required, in terms of thinking, and ways to become effective in deploying analytical data for business transformation. Also, many of the ways intertwine social and technical aspects, which is a given with data mesh. Data mesh is essentially a socio-technical paradigm.

The Six Ways

Business Model	Operating Model	Architecture & Infrastructure
1. Frame the opportunity right	3. Organize for value delivery – ownership and clarity	5. Use architecture to build value not infrastructure
2. Embrace experimentation and product thinking	4. Adopt dataops	6. Leverage cloud native capabilities for building data mesh

1 Frame the opportunity right

Data analytics is not about providing the right data, at the right time, to the right people, to make decisions. This was a valid assumption in an industrial era, but can be fatal in fast changing digital world. This assumption makes the entire data analytics process linear – left to right, with lack of focus on business value, no feedback loops, and with multiple points of value leakage and breakage.

2 Embrace experimentation and product thinking

Data practitioners, have been brought upon the staple diet of "data is an asset". This assumption is not cerebral, but visceral. This results in value being placed on collecting, hoarding, and governing data. The goal is to manage data and avoid risks. The value latent in data is never exposed through innovative and differentiating use cases.

3 Adopt dataOps

In the industrial age, the assembly line revolutionized the manufacturing process. The assembly line improved the productivity, velocity, and quality of the products. DataOps practices and tools can provide similar capability to data engineers and data scientists. If we have to deliver data products at a velocity and scale to meet the needs of consumers, we need to adopt DataOps.

4 Organize for value

This is a byproduct of considering data as a utility and not as a differentiator. Data teams are not directly connected with an end-to-end value stream. There is no "paying consumer" at the end point of most use cases. Most data teams generate reports and dashboards for business/internal consumption. In many cases, the value generated is notional, with no direct measurement of it in terms of revenue generated or bottom-line contribution, possible.

5 Use architecture to build value not infrastructure

Information architects generally think in stacks, tools, and platforms. It is very tempting to start applying legacy thinking to stand up yet more tools and platforms in the name of data mesh – data catalogue, data marketplace, self-service data wrangling, etc. Architects should focus on delivering data products and incrementally build the required platforms once the value is proven.

6 Use cloud-native capabilities to build data mesh

Cloud provides capabilities which align perfectly with the needs of a data mesh architecture. Though in theory, data mesh is platform-agnostic and can be built on-premise, the cloud makes the implementation much simpler and faster.

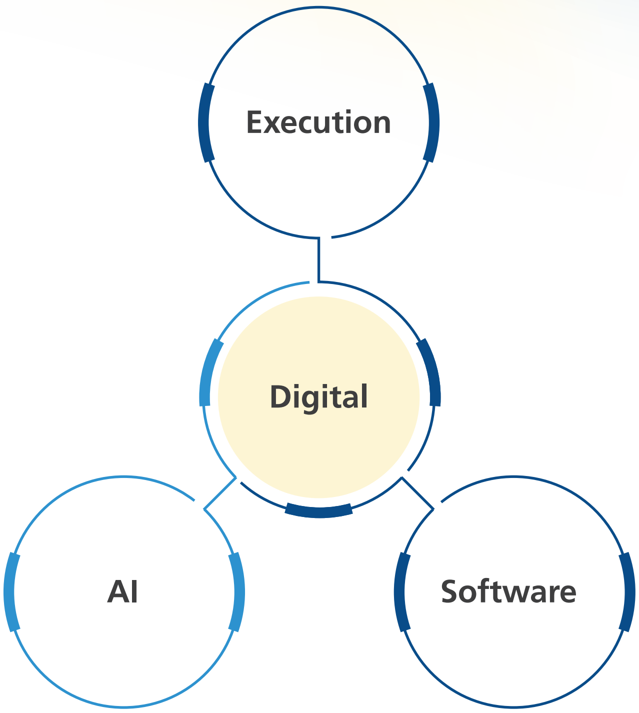
As with any new paradigm, understanding the essence becomes critical. One can adopt the rituals and forgo the value. There is a danger of calling old things with new names and declaring victory. For examples, we have seen instances where data marts got relabelled as data products and subject areas as domains.

These six ways can help you to embrace data mesh in its full context.

Way 01

Frame the Opportunity Right

Digital business is all about Data, AI, and Software. These represent the molecules with which physical-digital products and services are built and delivered to the customers. The user experience of the product or service is primarily moving to software components, powered by data analytics. The physical platform itself is a commodity and is primarily differentiated by the software – think laptops.



Intelligence embedded right into the product or service, at the point of contact with customer, determines the user experience. Predictive analytics to proactively address maintenance issues, recommendation engines to help customers choose right products, and credit card fraud analytics to approve or reject a payment in real-time are few examples. The ability to innovate at the intersection of data, AI, and software becomes core to winning in a digital world. This shift in the context needs to be factored in by data practitioners.

Data is not (just) for decision making

The paradigm that data is primarily for decision making is both dated and dangerous. It forces organizations to frame the data opportunity in a suboptimal way, and degrades data as a support or a utility function, which provides signals for decision making. IT teams focus on collecting data and governing it. The discussion is about volume, velocity, and variety, but very little about customer value, user experience, or jobs to be done with data. The goal is misrepresented as building central monolithic platforms, like datawarehouse or data lakes with vanity metrics like petabytes of data collected.

Key Idea: Think data as product

Using data for decision making is a necessary, but not sufficient to win in the digital world. Consider data as a key factor of production to create awesome customer experiences through differentiated digital products and services. AI, powered by data and delivered through the channel of software, can optimize current business models and create new ones. Build data products, not just data warehouse or data lakes. Apply design thinking, jobs to be done theory, and product thinking to data. Use your imagination to come up with new set of use cases, which can deliver an entirely new value propositions to customers.

Old Data Thinking	New Data Product Thinking
Focus on output	Focus on outcomes
On time, on budget	On value, on time, on budget
Process and internal focused	Customer and market focused
Suitable for industrial age	Necessary for digital age – complex & rapid change
Values stability	Values agility
We will build datawarehouse, data lake, or data LakeHouse	We will win with data and build what is required to achieve it
Volume, velocity, and variety	Value
Data is for decision making	Data is for delivering superior user experience
We will build reports and dashboards with data	We will build differentiatiation with data

Way 02

Embrace Data-Driven Experimentation

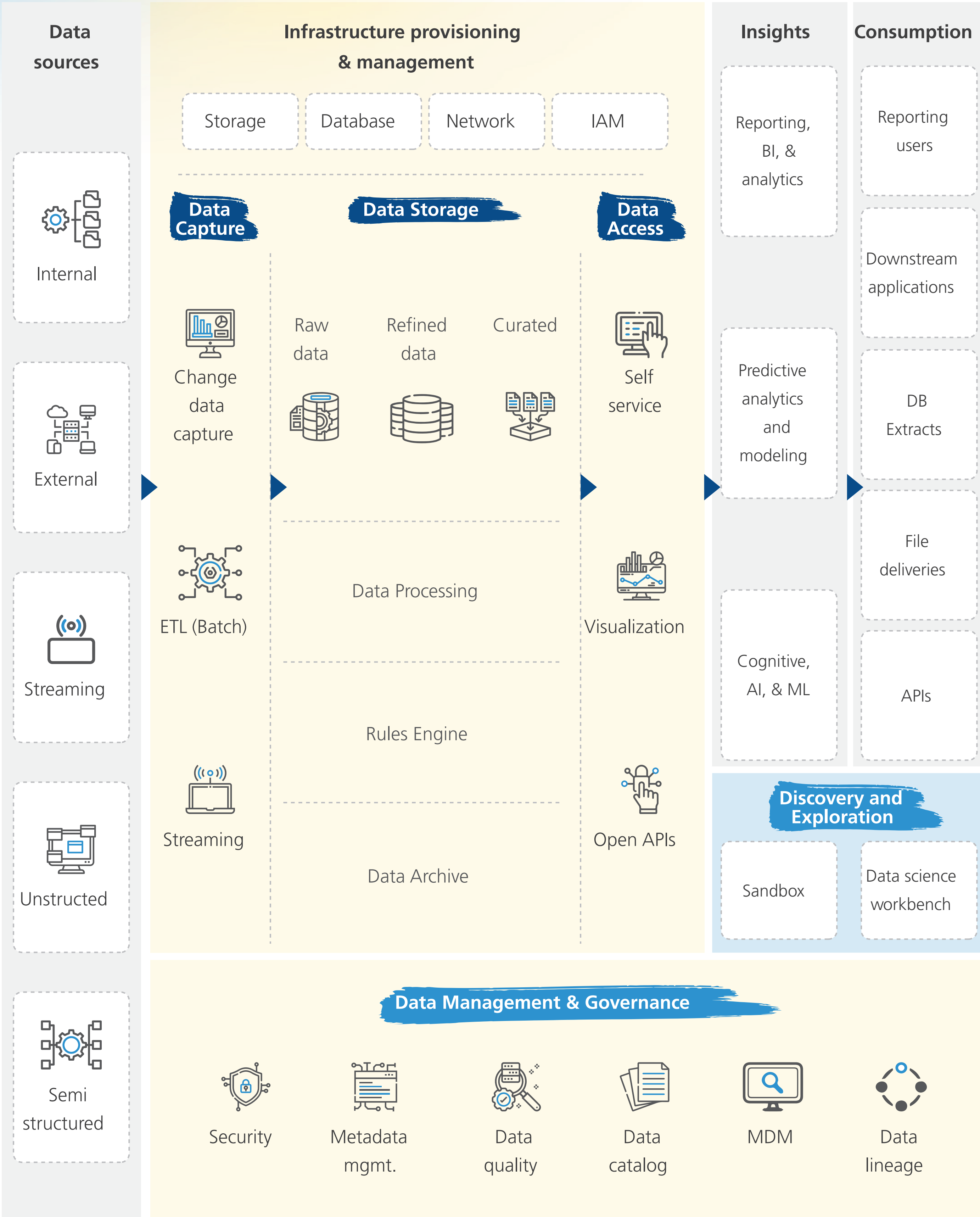
To be successful in a digital world, an organization needs to be empirical. The current environment is characterized by rapid change, volatility, uncertainty, causal ambiguity, and complexity. The model of strategize, plan, and execute over long range, no longer works. There is a need for shorter feedback cycles and continuous adaptation of business and product strategy to respond to market changes. Most products find it hard to achieve the product-market fit and fail. This points to the inability to anticipate and adjust to increasing customer needs. This is a problem data can solve, if applied properly.

Data platforms and lack of feedback loops

A typical architecture diagram for data analytics is depicted below. The data flows from left to right, from source systems to a data repository, like a data lake or a data warehouse, and finally gets consumed. The flow is not mapped to any particular value stream. There is no paying consumer at the end, so the whole process is insulated from market forces and consumer feedback. Data is considered a cost center, a necessary evil to run the business.

The focus is on delivering data for decision making. However, it is difficult to ascertain what decisions were taken, who took them, what was the quality of those decisions, and how they generated positive business outcomes. It is a push process based on information requirements. There is a lack of coupling with end consumer, hence there is no scope for improvement. In large data estates, typically there are 10-15 percent reports or dashboards which are not accessed over 12 months and with an equal percentage accessed quite rarely. So, there is a persistent question on RoI of data investments as metrics connecting to the business brass tacks are not available.

LTIMindtree



Key Idea: leverage data products for business experimentation

Organizations need to adopt the scientific method, which is the essence of a data-driven organization. It is necessary for data teams shift from output-based project orientation to outcome-based product orientation. The diagram below shows a virtuous cycle, which can be setup around a data product to deliver a key business outcome.

The process is circular, with bi-directional feedback loops across levels and end-to-end. We can build a data product to solve a business problem, starting with a hypothesis. We can identify data signals needed to design the experiment. In case data is not available, we setup the instrumentation to get the signals.

We can build analytical models, generate insight, and make the interventions to check if they work. Based on the real-world evidence, we can scale the intervention, pivot, or simply stop the work if no value is found. We can start with a new hypothesis. We can run multiple such experiments across lines of business or product lines. We can link those data products to generate higher order insights and value.

The investments are scaled based on the achieved business outcomes. Technology is applied to solve the business problem. We don't start with building large platforms and wait to on-board use cases. We can build a cross-functional team, which can autonomously solve the business problem.

Illustrative example – Churn Preventor data product

Let's assume a telecom provider is suffering from high rates of customer churn. There is no clear understanding on why this is the case. There are several opinions: ranging from poor network quality, high call drops, higher subscription prices, roaming charges, difficult signup process, competition, macro-economic factors, lack of variety in subscription plans, and others. Let's apply the data product thinking to address this issue.

We will build a data product – Churn Preventor – to identify and solve the churn problem. Note that we are not sure what is the solution, nor are we interested in a particular tool or technology. The goal is to solve a business problem. The process is cyclical, with an initial business hypothesis, say churn is due to high roaming charges. This can pivot as we learn more through the iterative process and feedback loops.

Data product name	Churn Preventor
Product vision	Reduce customer churn by 10% in the next 3 months to prevent revenue loss of 10 million
Domain	Customer domain
Sub domain	Customer retention
Product goals	This data product will: • evaluate the identified driving parameters for customer churn • find the root cause for customer churn with probabilities • identify any missing parameters impacting churn • recommend and implement interventions to reduce the churn
Product personas	• Customer retention team • Marketing team
Success criteria	% of churn reduced post the product deployment
Data sources	Call data records, billing information, roaming charges, customer survey, tower logs, etc

With the above approach, the value is obvious and focused on the business outcome. It is easy to measure the RoI in business terms.

Way 03

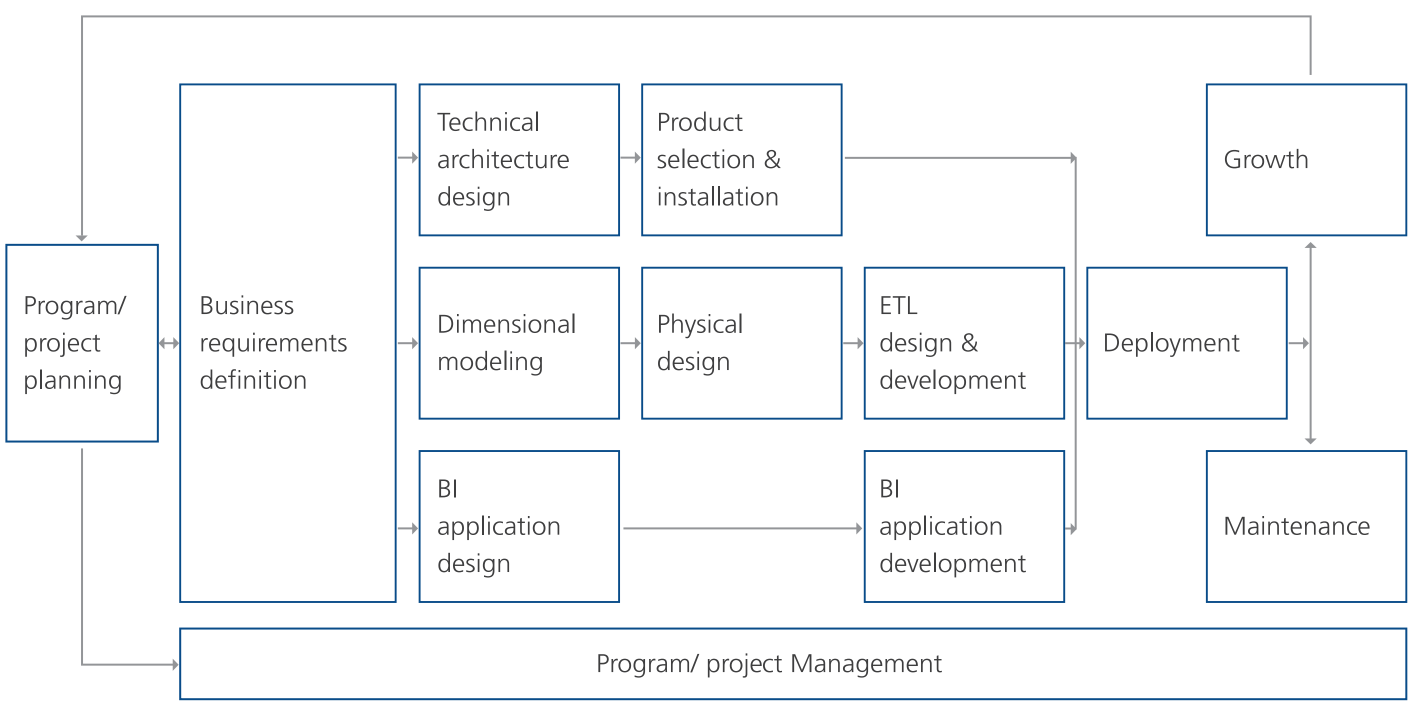
Adopt DataOps

Data practioners are adept at building data warehouses and data lakes for over decades. The process used to build these platforms resembles the Kimball Life cycle approach, which is illustrated in diagram2.

We start with requirements across three tracks – the data modelling track, the ETL track, and the reporting track. We integrate the tracks and deploy the code. We create a set of documentation, including dimensional models, ETL specifications, test cases, deployment documents, and others. The primary focus is on extracting, transforming, curating, and storing data based on business requirements.

This roughly follows a waterfall approach. It does not cover steps related to data science, as the focus is only on BI use cases. This approach is still dominant, though we have recently sprinkled few agile terms over this process model.

If we must scale data-driven innovation, we need to evolve a new set of processes, practices, and tools for the digital era. We need to prioritize agility, learning, and speed over stability and conformance to requirements. We need to adopt the DataOps way of working. The DataOps represents a new set of practices and tools which can help data engineers and data scientists to deliver value in a consistent and productive way.



Key Idea: adopt DataOps practices

→ Move from GUI tools to code

Data practioners are used to GUI-based tools for implementing the business logic. While application developers look at their work as coding data, practitioners look at development as working on tools. The visual tools, though easier to use, do not integrate well with DataOps practices and tools. They are difficult to automate and parameterize. Consider using code-based approaches – SQL scripts, stored procedures, Spark framework, DBT or a visual tool, which can generate code. This will improve the quality and velocity over time.

→ Leverage CI/CD pipelines

Establish end-to-end assembly line for your data products by leveraging CI/CD tools. Version management, automated testing, and CI/CD pipeline together over time eliminate inefficiencies and deliver higher quality code. They provide a baseline setup, which can be continuously tuned and improved. Enhancements like static code analysis for SQL or Spark can ensure enterprise standards are followed and code is of consistent quality.

→ Adopt good testing practices

One of the persistent problems in data projects is the lack of well-defined testing approach and test cases integrated with code. This is one of the side effects, of using GUI-based tools for development. Adopt coding approach, where testing can be integrated and automated as part of code. Adopt tools like DBT and Great Expectations, which allow tests to be automated. Features like cloning provided by Snowflake can enable full scale performance testing in a stage environment without duplicating data.

→ Use version control

Use version management tools like GIT. This improves development practices. Modern practice is to consider not just the code, but even infrastructure, configuration, governance policies and data as versioned artifacts. A well-defined branch and merge strategy can enable faster and parallel development for data products.

→ Adopt infrastructure-as-a-code practices

Use templates or APIs to spin up on-demand infrastructure for building, testing, and deploying data products. This is one of the greatest benefits of moving to cloud. The platform teams can provide hooks or templates, which can be used by the domain teams to self-serve their infrastructure needs. Apply cost governance controls from day one to keep costs within limit.

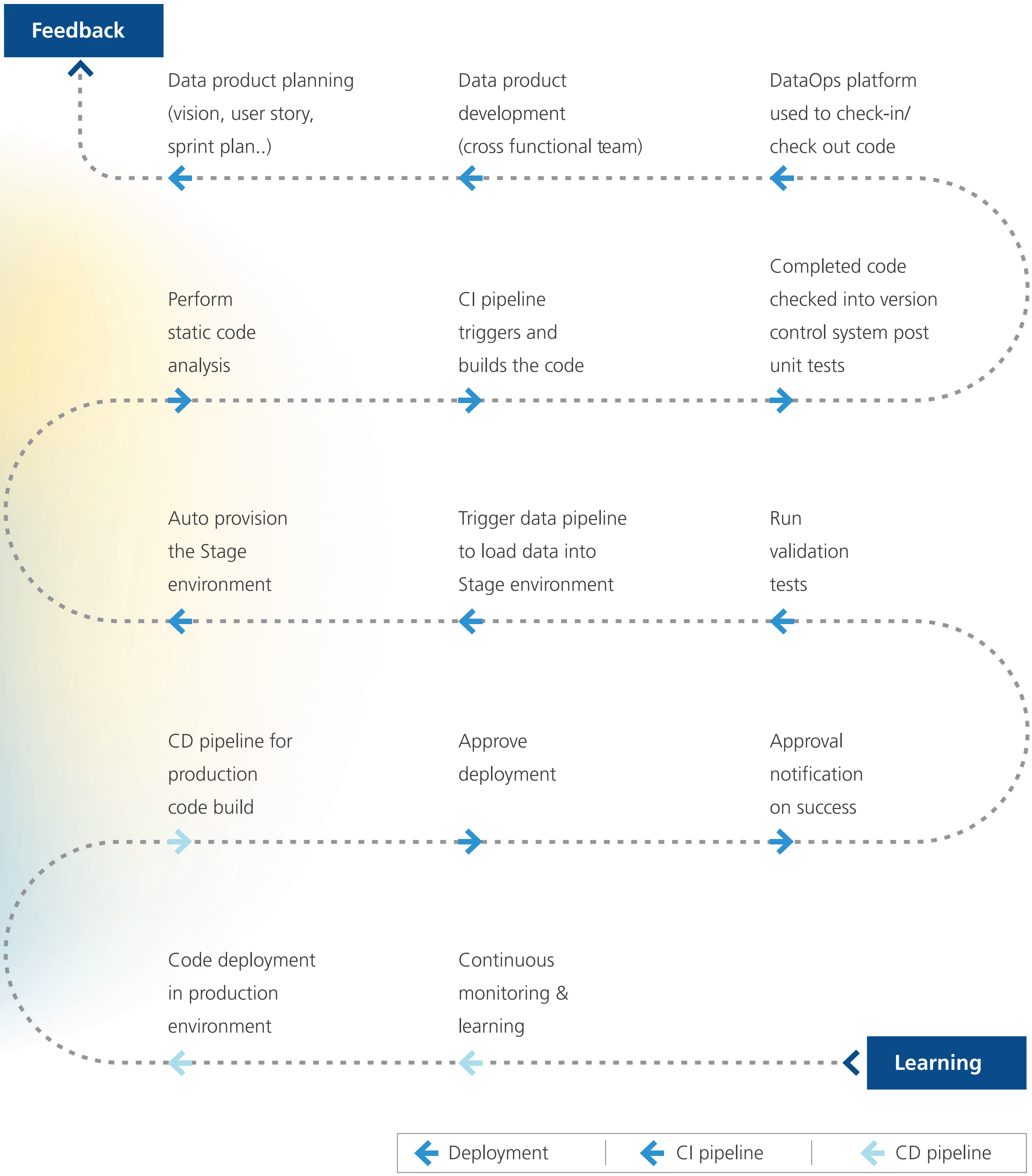
→ Think policy-as-a-code

Define governance policies-as-a-code, which can be implemented using infrastructure-as-a-code tools like Terraform. New tools like Open Policy Agent can be used to decouple the policy from code.

→ Automate everywhere

For speed and scale automation must be deployed both in depth and breadth. Data practitioners generally think about the automation of data pipeline – scheduling, notification, data quality rules, etc. However, for successful data mesh, we need to expand that scope. Infrastructure provisioning, code deployment, and data governance all aspects need to be automated.

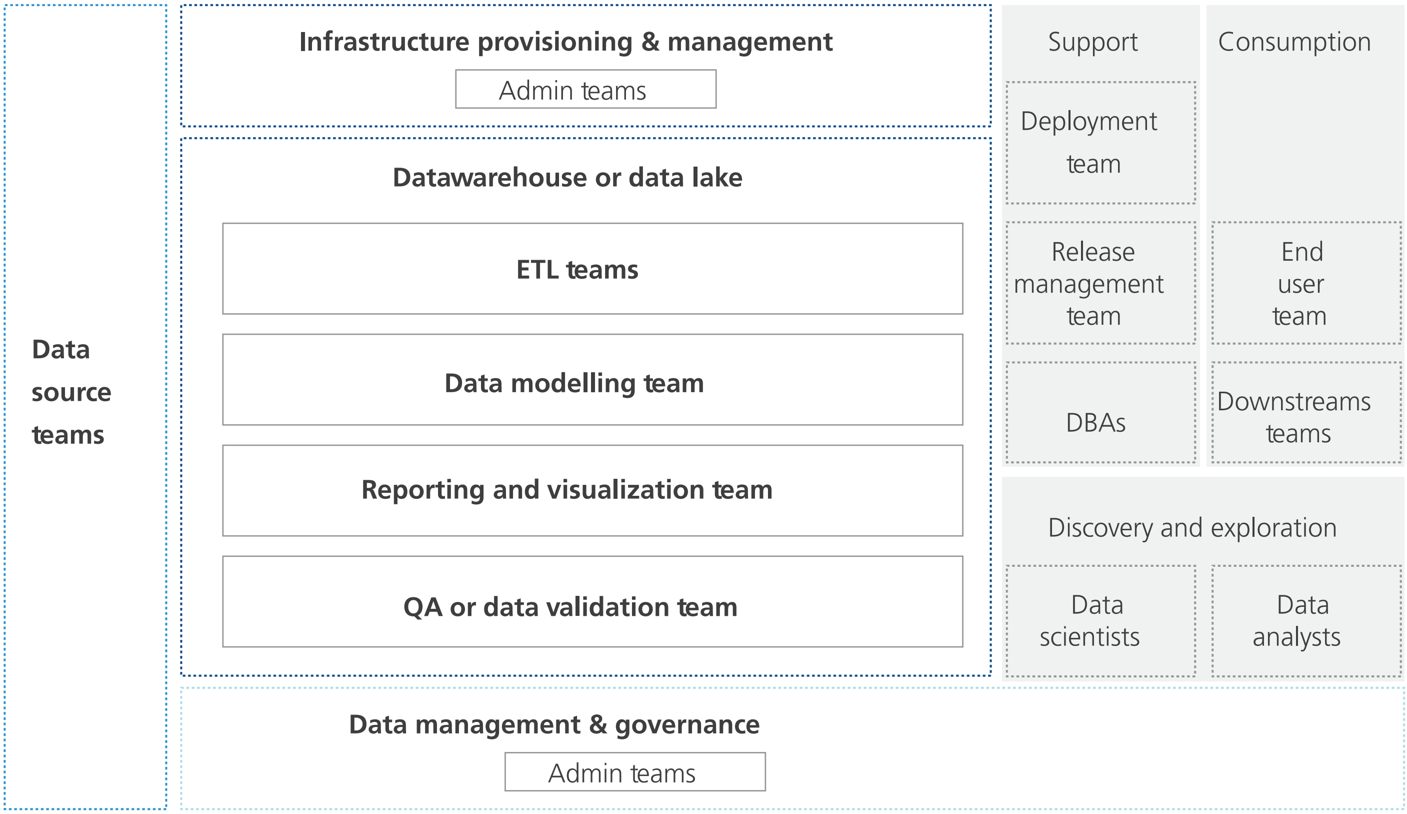
Illustrative DataOps pipeline



Way 04

Organize for Value Delivery

Traditional data teams are organized for activities. There is an ETL team for building data pipeline, data modeling team for building the data models, database administrators doing the physical tuning, scheduling team for creating schedules, and visualization team to build reports and dashboards. Generally, teams work in their area of competence with well-defined hand-offs. There is an episodic collaboration. For example, poor report performance may require visualization experts, DBAs, and data modelers to come together to modify the data model and create indexes. The diagram below captures the traditional team structure of an organization.



This way of organizing by stacks, by activities, and by technology specialization is the very antithesis of how we want to organize for data mesh. This not only limits collaboration, it also completely ignores business domain, as there is no alignment with business outcome or business value. The best-case scenario is to complete activities on time and within the budget.

Key Idea: organize for ownership and clarity

With data mesh, the goal is to build data products which deliver value to a consumer. This essentially requires the data team to be able to deliver the end-to-end value stream. This will require a multiplicity of skills, with domain understanding being the key. There are four core teams required for successfully building data products. This structure is derived from Zhamak Dehghani' s book – Data Mesh - Delivering Data-Driven Value at Scale

1 Data product team

Responsible for delivering the data product. It is cross-functional, autonomous, and self-sufficient in skills to deliver the data product. Can consist of data engineers, data scientists, modelers, analysts, or other technology SMEs. However, it should have domain experts or business SMEs, so that the right data product can be built, which will be able to deliver value to the customer.

2 Platform team

Responsible for providing various technical platforms used by the data product teams in a as-a-service model. These may include database platforms, data science platforms, observability tools, etc.

3 Cross-product governance

This team is responsible for ensuring that data products don't become data siloes. They define the boundaries, scope, and overlap of data products, resolve any conflicts and ensure global optimization of the investment.

4 Special purpose teams

They provide specialist knowledge not available with data product teams. This might be related to areas like compliance, legal, security standards, etc. They are engaged by data product teams on need basis.

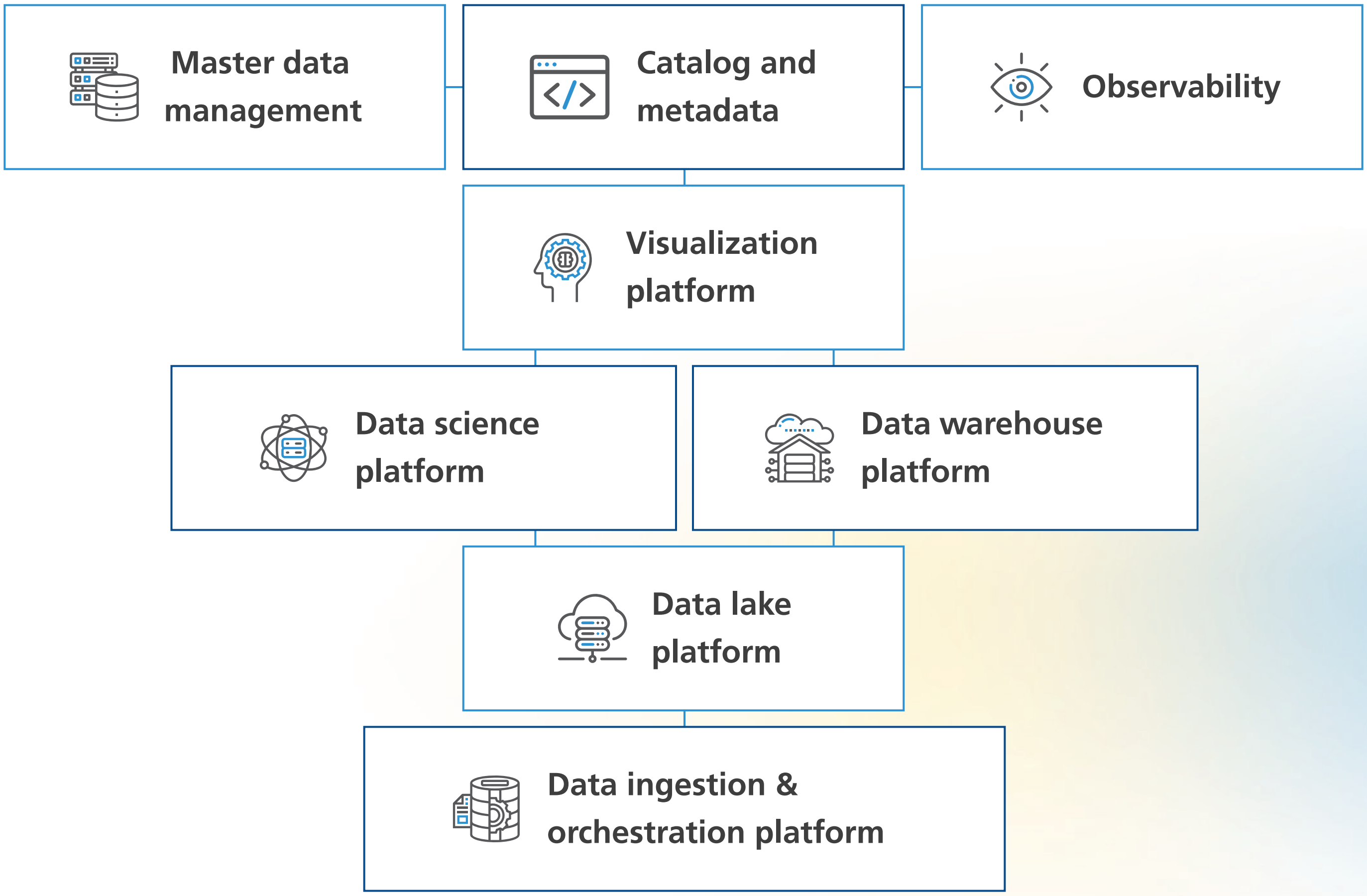
Key Idea: elevate the domain, it's your differentiator

Data teams are experts at creating frameworks. Data ingestion frameworks, exception handling, audit and logging, data quality, and others. However, these are domain-anaemic. They address architectural concerns and reusability. Embedding domain experts and business SMEs with data engineers and data scientists will add value to both sides. It can result in innovative use cases, which either of teams will not be able to come up with independently. It is necessary that data products teams focus on delivering consumer-focused, domain-rich, and differentiating data analytic applications. Platform teams can build and support generic frameworks.

Way 05

Architecture Must Build Value Not Infrastructure

One of the consequences of central teams running data projects has been that we have built large monolithic platforms and infrastructures. The idea was to onboard multiple applications on the central platform as requirements were essentially the same – an ability to generate reports and dashboards. It made perfect sense to aggregate all the demands and build a single platform, which can deliver economies of scale. Since data platforms were considered as utilities - stability, uniformity, lower costs through shared platforms, and services and central management were optimized. The diagram below captures the core platforms typically built upfront.



Strategic data initiatives, like adoption of big data, started with capacity planning and infrastructure – say a 100 node Hadoop production cluster with smaller Stage and Dev environments. Many organizations later discovered little or no value in those investments as anticipated use cases never turned up or took off. The organization were saddled with sunken investments and non-value adding infrastructures.

With data mesh, the emphasis is on decentralization and fast-moving domain-oriented teams building data products. So, while the data platforms and infrastructure are important, we optimize for a different set of criteria. We are competing in the context of time, hence speed, experimentation, learning, and market fit become more important. The center of gravity moves from technological platforms to consumer value. It is perfectly fine to use heterogeneous technologies, as long as data products deliver value to customers and make money for the organization.

So, focus must be on building platform and infrastructure incrementally, once the value is proven. Ultimately, a small set of platforms which fit the organization context will converge. However, organizations must desist from large upfront investments. The cloud can be leveraged to effectively address this challenge, which is the subject of next section.

Way 06

Leverage Cloud-Native Capabilities for Building Data Mesh

Cloud-native capabilities make building data mesh simpler and faster on the cloud. Cloud- native features like self-service provisioning, elastic scaling of compute and storage, data sharing, availability of CI/CD tools, containers, cost transparency, and others make it easy to operate autonomously and leverage high levels of automation. Using cloud PaaS and SaaS offerings can eliminate undifferentiated heavy lifting and enable agility and scale for building data products.

Key Idea: build on data mesh enabling capabilities

With data mesh, the goal is to build data products which deliver value to a consumer. This essentially requires the data team to be able to deliver the end-to-end value stream. This will require a multiplicity of skills, with domain understanding being the key. There are four core teams required for successfully building data products. This structure is derived from Zhamak Dehghani' s book – Data Mesh - Delivering Data-Driven Value at Scale

→ Use PaaS/SaaS offerings

As data products are built by decentralized domain teams, it is necessary to bring down the technical complexity and democratize enabling technologies required to build data products. Leveraging PaaS/SaaS cloud offerings will enable data product teams to self-service most of their needs and move fast without dependence on platform teams. Technologies like Snowflake take out most of the undifferentiated heavy lifting activities, which were needed on on-premise platforms. For example, defining indexes, partitioning, capacity planning, and resource contention management are not needed, simplifying building of data products. Also cost attribution for the resources consumed is very transparent, which makes it easy to determine the value generated by data product.

→ Build on multi-cloud/cross-cloud capability

Most organizations will have multiple public clouds and a single cloud in multiple regions. It is best to look at cross/multi-cloud capabilities between data products to ensure that we don't end up creating data silos on cloud. As decentralized teams across multiple domains build data products, a common standard for interoperating between them is needed. This will ensure data products are composable and higher order data products can be built using the existing data products.

→ Leverage data sharing

Data product consumption is multimodal, there are many ways to consume the data or insights, APIs, native connectors, views, and files are all valid access mechanisms. However, it is best to avoid creating data pipelines for extracting and sharing data with consumers of data products. This will result in same problem data teams faced with FTP processes – network issues, duplicate transfers, data latency issues, performance issues for large files, etc. Cloud-native data sharing capability will alleviate many of these problems.

→ Create data marketplace and catalog

Cloud platforms enable building of marketplaces and catalogs, which can list the data products of an organization. They can provide visibility on what data products already exist. Data marketplace can also provide description, sample data, usage statistics, cost, and other useful metrics.

LTIMindtree

Conclusion

The paper covered the six ways which can improve your odds of successfully implementing the data mesh. The ways require us to change assumptions about data, requiring us to think and act in new ways. We need to fully embrace the idea of data as a product. We need to embrace business experimentation leveraging data products. We need to organize in new ways to deliver value. We need to move beyond old engineering practices and adopt DataOps. We should take advantage of enabling technologies like cloud and resist the urge to stand up infrastructure prematurely. We need to move from a model of solving technology puzzles to solving business challenges.

Today, consumers expect and demand superior experiences. Digital is spawning nimble and agile competitors who are attacking the incumbents. Digital giants like Amazon and Google are making industry boundaries irrelevant by expanding into new industries in search of growth. Macro-environment continues to become more complex and volatile. Zeitgeist demands organizations put data at the centre of their business transformation. Data-driven innovation will separate the winners from losers in the digital age. Data mastery is pre-requisite for digital mastery, and data mesh provides a viable pathway to accomplish that goal.

Notes

1. https://future.a16z.com/software-is-eating-the-world/
2. https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dw-bi-lifecycle-method/

Abbreviations

Abbreviation	Expansion
AI	Artificial Intelligence
MVP	Minimum Viable Product
CoE	Centre of Excellence
PoC	Proof of Concept
SMP	Symmetric Multi-Processing
MPP	Massively Parallel Processing
BI	Business Intelligence
SME	Subject Matter Expert
CI/CD	Continuous Integration/Continuous Delivery
EDW	Enterprise Data Warehouse
GUI	Graphical User Interface
ETL	Extract Transform Load
API	Application Programming Interface
FTP	File Transfer Protocol
SaaS	Software-as-a-Service
PaaS	Platform-as-a-Service
DBT	Database Build Tool
Dev	Development

About Author

Ranganath Ramakrishna leads the Data Mesh CoE at LTIMindtree. He is passionate about applying data solutions to solve the business problems. He is a certified enterprise architect with more than 16 years of industry experience. His core competence lies in Data Strategy Development, Cloud Data Platforms, Data Warehousing, and Data Analytics. He has a proven track record of successfully designing, leading, and executing enterprise scale, complex data projects for Fortune 100 clients.