TerminusX

Data is not for servers. It's for people.

Putting clean, versioned data in the hands of the people that matter: your business units.

A TERMINUSX WHITEPAPER

Contents

Beyond the Data Lake

We've all heard about the explosion of data. We've heard about how AI/ML training requires instant access to massive amounts of data, especially historical. And we've seen the proliferation of data warehouses and lakes to address these challenges.

Certainly, data warehouses and lakes have advantages over the legacy siloed transactional databases that were only understood and available to a small number of people. Yet, most companies still struggle with access to their data. While they've physically integrated data into a single warehouse or lake, users still have enormous difficulty simply finding what they need. According to ThoughtWorks, Enterprise Data Warehouse initiatives have a higher than 50% failure rate.[1] And a recent IBM report suggested that the average factory analyzes less than 1% of its data in real time. [2]

Not a lot of production for such an expensive infrastructure (A typical data lake, according to Amorphic Data, costs anywhere between $200K–$1M, depending on complexity and features). [3]

The problem is that even though the data is in one place, it's still difficult to extract information from individual siloed locations into a central database. The result is huge backlogs. There's also the all-too-common complication that if there are any issues requiring cleaning in the warehouse, it's an extremely long process, delaying everything further.

But perhaps most important is the fact that these central IT teams, the ones overloaded with huge backlogs, really don't have the best knowledge of the data content. That knowledge belongs to individual domain leaders.

If you're in sales and marketing, you need data around product performance, revenue scale, revenue per account, customer acquisition efficiency, customer lifecycle, sales efficiency, margins, pricing and channel analysis—a host of metrics that you understand a lot more than the central IT team.

The problem is that even though the data is in one place, it's still difficult to extract information from individual siloed locations into a central database.

If you are in Ops, Finance, or HR, the data, the content, and the narrative will all be equally granular and look completely different from other departments. Each domain team understands its own data much better than anyone. And those domain leaders should have access and views for all of their data. After all, it's their data!

With this access, they have the ability to craft data into a narrative, pulling metrics from other departments if necessary, and publishing a coherent analysis of their team's status.

These domain leaders—the ones who answer business questions—need to come first. What is needed is a distributed and decentralized structure that empowers these teams to employ their data to deliver the best outcomes.

Leveraging the Data Mesh

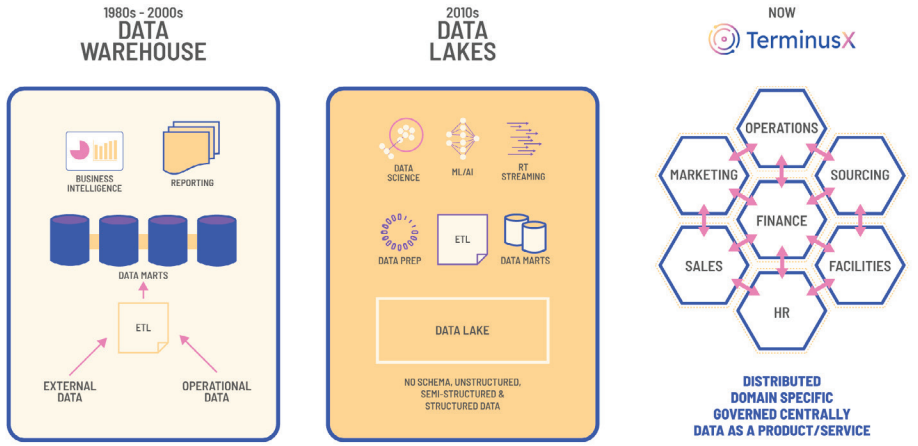
To accomplish a distributed and decentralized structure requires moving from a centralized paradigm of a lake or warehouse to a modern architecture: the data mesh. According to Barr Moses, a data mesh "is a type of data platform architecture that embraces the ubiquity of data in the enterprise by leveraging a domain-oriented, self-serve design." [4]

Zhamak Dehghani, Thoughtworks Director of Emerging Technologies, the original author of the 'data mesh' concept, in her seminal blog, described the building blocks of a ubiquitous data mesh as a platform as "distributed data products oriented around domains and owned by independent cross-functional teams who have embedded data engineers and data product owners, using common data infrastructure as a platform to host, prep and serve their data assets." [5]

The key is this domain-oriented data ownership and architecture—with the data in the hands of the business units (domains). This is especially important in extracting value from analytical data and historical facts at scale.

And scale is not just volume, explains Zhamak: "Scale [means] a constant change of data landscape, proliferation of both sources of

The key is domain-oriented data ownership and architecture—with the data in the hands of the business units (domains). This is especially important in extracting value from analytical data and historical facts at scale.



Data Structure Evolution

data and consumers, diversity of transformation and processing that use cases require, speed of response to change."

Once data has been distributed to domain leaders, concerns around accessibility, usability and harmonization arise. This, says Zhamak, is where product thinking is critical: "Domain data teams must apply product thinking with similar rigor to the datasets that they provide; considering their data assets as their products and the rest of the organization's domains, data scientists, ML and data engineers as their customers."

An imperative facet of data mesh is to ensure that data can be shared across domains to be used in innovative and insightful ways. It is key to have an overarching global governance strategy that encompasses access control, metadata, harmonization rules, and much more. In this context, Zhamak outlines what the data product must look like to be a successful product. It starts with making sure the product is in a centralized data catalog for easy discoverability. Once you can find it, it should have a unique address following a global convention that helps its users to programmatically access it.

The next step is trust, says Zhamak: "A fundamental shift requires the owners of the data products to provide an acceptable Service Level Objective around the truthfulness of the data, and how closely it reflects the reality of the events that have occurred or the high probability of the truthfulness of the insights that have been generated."

When business units begin treating data as a product, the reward is "a distributed data architecture, under centralized governance and standardization for interoperability, enabled by a shared and harmonized self-serve data infrastructure."

She also emphasizes that the data product, like any good product, must be secure, with governance, and be cross-functional for all teams. When business units begin treating data as a product, the reward is "a distributed data architecture, under centralized governance and standardization for interoperability, enabled by a shared and harmonized self-serve data infrastructure."

Departmental Data Access

TerminusX is a self-service data platform that allows you to build, deploy, execute, monitor, and share versioned data products. TerminusX versions both data and schema, allowing your team to deliver a consistent product—while improving and innovating. Your data is decentralized and can be utilized by those who need it.

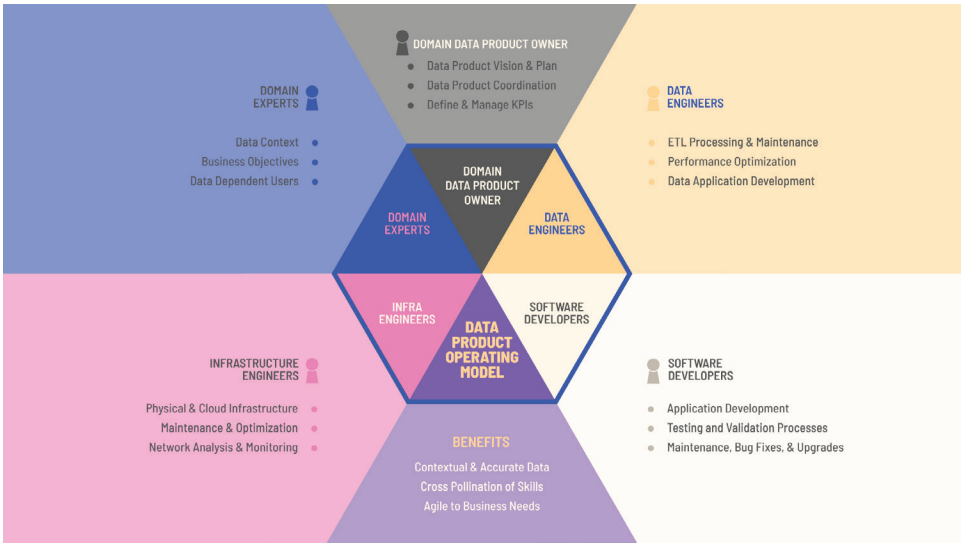
With a decentralized platform, data moves from overburdened IT to the departments that need it the most. Product, sales, marketing, ops, HR—all have their own view. And these domain experts have the speed and agility to use their data to make decisions and improve applications. No longer is marketing waiting for historical campaign numbers from an overworked IT team. No longer is the new campaign on hold until they get their data.

To eliminate these bottlenecks, TerminusX enables the marketing team to build self-service data products, including all current and historical data, as well as relevant data from other departments.

The purpose of TerminusX is to make complexity and friction a thing of the past; your domain teams can now manage the full lifecycle of your data products. And this self-serve platform contains powerful tools to provide developers with a workflow to create, maintain and run data products.

With the amount of data in enterprises today, collaboration is key. Building new customer applications, improving internal processes, increasing predictive analytics all require flexibility and access not available on centralized monolithic platforms.

With a decentralized platform, data moves from overburdened IT to the departments that need it the most. Product, sales, marketing, ops, HR—all have their own view.



TerminusX-Data-Product-Operating-Model

Version Control, which is at the heart of TerminusX, is the key to effective collaboration. With Version Control, you can access, share, and work on multiple versions of the same asset at the same time, all in a controlled, optimized way.

Data decentralization and distribution powers continuous change and scalability and lets your productive teams work at a fast pace, not slowed by the 'centralize everything' ethos of legacy architectures. Each domain can expose one or many operational APIs, as well as one or many analytical data endpoints. The architecture supports the autonomy of your domain teams.

Perhaps best of all, TerminusX can be tested with little to no risk. It can be implemented on discrete data sets initially, and can be quickly scaled as needed—a process simply not possible on monolithic, centralized platforms. In addition to mitigating risk, this also significantly speeds development time, and allows you to scale where you need it, when you need it.

How it Works

TerminusX is powered by TerminusDB, a database which uniquely combines the power of knowledge graphs with the simplicity of documents. This means complicated documents (with say an internal structure and sub-documents) can all be connected in custom ways.

TerminusDB is a database which uniquely combines the power of knowledge graphs with the simplicity of documents. This means complicated documents can all be connected in custom ways.

The TerminusDB owes something to software best practice. In the last 15 years, software has transitioned to a continuous development and deployment process. This was facilitated by distributed version control systems with discrete branches that could be used as a testing environment, then released. TerminusDB does the same thing for data.

Why is this important for data? Say you have an inventory system that you want users to access, but you also want to ensure that the records in it are reviewed, corrected or run through an editorial process. Today this is done in a haphazard way, where people edit in a live transactional database—something you'd never do with code. TerminusX solves this, moving data into the same development and deployment process that has revolutionized software.

Versioning and branching data has many other uses too, you can create bitemporal datasets and withhold certain information and make them available to other stakeholders, such as regulators, customers, or other teams.

This software development perspective dovetails perfectly with the new data product mindset. Now data products can be continually created and optimized—domain teams can easily build, curate, and manage individual data products, enabling them to focus their efforts on business outcomes.

Knowledge Graphs

TerminusX allows you to build a distributed knowledge graph made up of a vibrant ecosystem of interoperable data products. By harmonizing internal and external data relevant to your organization you can improve operational efficiency for the enterprise and competitive advantage for the business units.

Most tools make knowledge graphs difficult to develop, difficult to scale and difficult for business units to consume. TerminusX merges the best of document stores and knowledge graphs, keeping your teams in the JSON world they know and love.

TerminusX knowledge graphs can connect to each other in pods, enabling quick, massive scaling. This enables very large knowledge graphs for multiple domains within the company. While each has a domain specialty, everyone can see them.

TerminusX knowledge graphs include a schema, so you have definitions on the shape of the graph, and be assured of a strong structure where data customers know where data sits and its context, thanks to the metadata. This ensures data quality and is useful for automatic processing and for the release of data via an API.

TerminusX knowledge graphs can connect to each other in pods, enabling quick, massive scaling. This enables very large knowledge graphs for multiple domains within the company. While each has a domain specialty, everyone can see them.

The fundamental advantage of knowledge graphs is that they model the world as things that have properties and relationships to other things. The knowledge graphs that companies like Google and Facebook have accumulated are demonstrably commercially powerful and it is unlikely that such complex and rich datasets would have been possible without a graph. TerminusX allows you to model your data with properties and relationships to build a bigger picture to drive operational efficiency and service innovation.

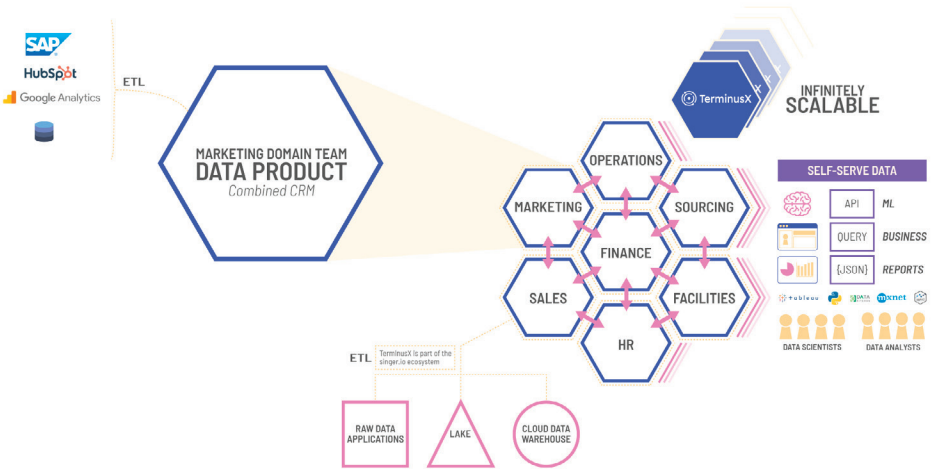
Commit Graph

The TerminusX database records and tracks all versions of data, enabling live query at every version, at any point. This covers the whole data lifecycle–you can audit any change with a full commit history (Even automated changes done by bots are tracked).

Every version can be accessed and optimized, with the original always intact. And the versioning is not only linear—you can almost instantly branch to another database and create a new timeline that shares the timeline of the new database. The commit graph, which clearly shows who updated what and when, can be valuable to management, and can assist with compliance requirements.

With the Commit Graph, TerminusX also significantly reduces compliance and audit costs. This is due to the ability to quickly access every version of any audit/compliance document.

With the Commit Graph, TerminusX also significantly reduces compliance and audit costs. This is due to the ability to quickly access every version of any audit/compliance document.



TerminusX in an ecosystem of data products

Use Cases

Tamper-evident data audit
Audits are laborious, expensive and subject to rigorous legislation and standards from regulatory bodies such as HMRC and Companies House. And as standards and compliance continue to increase, so do costs. Strategic Finance notes that in a survey, 83 public companies reported average audit fees of $9.8 million and a median fee of $3.7 million— up 4.2% from the previous year. [6]

The requirement is for extremely detailed documentation over time, a full accounting of records, reports, operating practices, and documentation. TerminusX records all changes to your data, giving you a full, immutable audit log. In your existing data infrastructure, you can add new versions to existing records, but the database never changes or deletes records. This lets you store critical data without fear of tampering and provides proof of data inclusion and historical consistency in real time.

Regulatory and compliance demands
A new DeLoitte article points out that "the pace of regulatory change, convergence in global regulation, and competition from new market entrants...have created a complex environment for compliance leaders across all industries." [7]

The requirement is for a data mesh architecture, which allows for clean, governed data and "frictionless access and sharing of data in a distributed network environment," according to Gartner.

Compliance failures, which result in reputational damage and significant financial penalties, are a moving target with requirements that are constantly evolving — and increasing. "Organizations are realizing that business and operational value," continues DeLoitte, "can be derived from anticipating risks and meeting regulatory requirements, making compliance an increasingly integrated part of the business investment strategy. New regulations demand visibility of enterprise and market information, and require provisioning data to regulators." TerminusX helps you to meet compliance challenges by providing detailed data lineage, flexible data modeling, and performance capabilities required to satisfy regulators at an enterprise scale. As an immutable store, you can now give regulators a snapshot of the exact situation at any point in time.

Data management

In a recent whitepaper, KPMG points to incorrect data being a huge data management problem plaguing organizations today: "Businesses lose massive amounts of money each year by basing their decisions on bad data," states the report. "Therefore, it is essential to have appropriate data quality processes in place to ensure decisions that are based on complete and accurate data." Equally important, they say, is data governance: "Bad data governance can lead to non-compliance with regulatory mandates, uncertainty about the definitions of data and about who is responsible for what data." [8]

The requirement is for a data mesh architecture, which allows for clean, governed data and "frictionless access and sharing of data in a distributed network environment," according to Gartner. TerminusX delivers just this: now your domain teams can leverage product thinking to improve your data management, provide a holistic customer view, and manage risk across multiple domains.

360-degree customer view

Gartner recently predicted that "in a few years from now, 89% of businesses will compete mostly on customer experience." Which means, says Forrester, "that customer service organizations must accelerate their adoption of AI-powered self-service technologies for frictionless service." As we all know, AI means lots of current—and historical—data easily available to domain teams. [9]

By harmonizing internal and external data relevant to your organization you can quickly identify any weak links in your supply chain.

To genuinely improve the customer experience, AI is critical to develop valuable context across customer data. TerminusX provides the ability to seamlessly employ multiple data products from current and historical trends, to build a customer view with superior context and quality.

Supply chain analysis

Recently, as we have all painfully seen, a secure, reliable supply chain is a critical differentiator in any industry. Supply chain risk management (SCRM) is the discipline of identifying, assessing, and mitigating supply chain risks. Identifying risk in the supply chain is complicated. It requires understanding how contingencies and dependencies across the network can lead to specific links or businesses representing risk.

A TerminusX knowledge graph enables this analysis, as complicated documents can all be connected in custom ways. Now you can build a distributed knowledge graph made from a vibrant ecosystem of interoperable data products. By harmonizing internal and external data relevant to your organization you can quickly identify any weak links in your supply chain.

TerminusX knowledge graphs can connect to each other in pods, enabling quick, massive scaling. This enables very large knowledge graphs for multiple domains within the company. While each has a domain specialty, everyone can see them.

Traceable data lineage

TerminusX is an immutable data store, tracking every version of every piece of data. And the versioning is not only linear—you can branch to another database and create a new timeline that shares the timeline of the new database.

The TerminusX commit graph delivers the ability to monitor inventory, parts, cost control, and KPIs related to audits across the supply chain. You can access data remotely, maintain stakeholder data, and unify related concepts using your data model. This new agility enables you to rapidly respond to business changes and new requirements, easily including new data sources.

TerminusX is a fully managed platform, designed to get you up and running in a matter of minutes. Instant access and flexible management of data for all of your domain leads.

Predictive analytics
The predictive analytics market is projected to reach $10.95 billion this year, growing at a compound annual growth rate (CAGR) of about 21%, according to Zion Market Research.[10] Whether you're using predictive analytics for marketing or ops optimization, fraud detection or risk reduction, TerminusX allows you to leverage graph analytics to improve the quality of the data going into your predictions. Equally important, our version control allows you to capitalize on all your valuable historical data to make the best decisions. TerminusX's end-to-end prediction pipeline accurately identifies the likelihood of future outcomes—based on all of your data.

Conclusion

Department heads screaming for metrics. Overworked IT lacking context about the data they're working with. Delays and more delays. And with all of this, you need to innovate, you need to scale.

In a Forbes survey, they found that "Data scientists spend 60% of their time on cleaning and organizing data. Collecting data sets comes second at 19% of their time, meaning data scientists spend around 80% of their time on preparing and managing data for analysis." [11]

Talk about a bottleneck. TerminusX is a fully managed platform, designed to get you up and running in a matter of minutes. Instant access and flexible management of data for all of your domain leads.

With TerminusX, you reinvent your teams' ability to leverage data. Now your organization is:

- **Agile**, with domain teams working in interoperable but containerized ways, deploying when they need to.
- **Scalable**, so you can create new data products when new data appears.
- **Accurate**, since data product responsibility lies with the producers and experts.
- **Self-serve**, allowing business teams to create fully end-to-end, automated deliveries.
- **In Control**, with metadata, provenance, and audit history with data lineage.

Data scientists spend around 80% of their time on preparing and managing data for analysis."

The TerminusX dashboard allows domain leaders to query, explore and manage data products as linked documents. It includes time-travel and version control, enabling you to mix and match multiple data products, advance collections at their own pace, query and perform transactional ACID management across data products, and access metadata to track data lineage and data provenance.

With TerminusX you are fully utilizing your data–in the hands of the people who know what to do with it. Now your business units are able to innovate with data products, employing improved service, applications, processes, reporting—and lower costs.

Best of all, it all happens in just a few clicks.

Sources

[1] https://www.thoughtworks.com/radar/platforms/enterprise-data-warehouse
[2] https://www.ibm.com/topics/industry-4-0
[3] https://www.amorphicdata.com/resources/how-much-does-it-cost-to-build-and-manage-a-data-lake
[4] https://towardsdatascience.com/what-is-a-data-mesh-and-how-not-to-mesh-it-up-210710bb41e0
[5] https://martinfowler.com/articles/data-monolith-to-mesh.html
[6] https://sfmagazine.com/post-entry/december-2019-lower-the-costs-of-your-external-audit/
[7] https://www2.deloitte.com/gz/en/pages/about-deloitte/articles/transform-saudi-arabia/changing-role-compliance.html
[8] https://www.kpmgsofy.com/blog/data-management/blog-data-management-blog-data-management-common-data-management-issues/
[9] https://learning.callminer.com/c/whitepaper-forrester-2021-customer-service-megatrends?x=CFl8z6&lx=amFxJO
[10] https://www.globenewswire.com/news-release/2018/03/02/1414176/0/en/Trends-in-Predictive-Analytics-Market-Size-Share-will-Reach-10-95-Billion-by-2022.html
[11] https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=372900656f63