RAMP-UP TO ANALYTICS:

Building a Data Model for Industrial Transformation

Industrial Transformation provides a path for companies to make dramatic improvements in business operations. With these undertakings, organizations are planning for 10% or more boost in performance. This degree of gain is highly attractive to companies where long-established continuous improvement programs can no longer generate even .05-1% annual improvement.

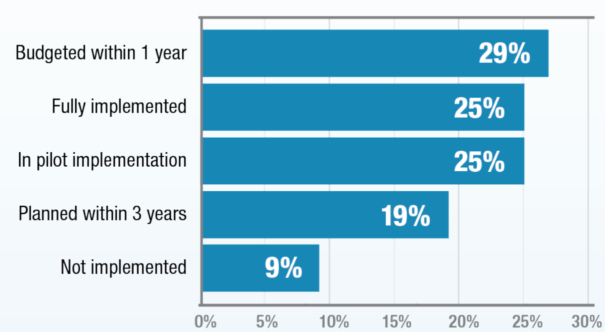
One of the major driving factors for Industrial Transformation is recognition of new ways to use the data generated in factories and plants, empowering new business insights that were previously unreachable. However, data challenges can limit the impact of transformational technologies like Industrial Internet of Things (IIoT) platforms and industrial analytics. Entrenched data silos often exist throughout organizations. Since the silos have usually been locally optimized for productivity, quality, reliability, safety, or some other purpose, driving change to the teams that built them is extremely difficult.

Consider this: most existing data architectures and use cases were built solely for IT systems and neglected the significance of data generated in plants and factories (and the benefit it can deliver for and by plant operations teams). The distributed nature and different formats of plant data are especially challenging for IT teams not accustomed to working with the unique data formats, volumes, and speeds that come with industrial operations and processes. This disparity is both one of the challenges and one of the driving factors of IT-OT (operations technology) convergence, which can be a stepping-stone in Industrial Transformation.

> **DATA CHALLENGES CAN LIMIT THE IMPACT** of transformational technologies like Industrial Internet of Things (IIoT) platforms and industrial analytics. The distributed nature and different formats of plant data are especially challenging for IT teams not accustomed to working with the unique data formats, volumes, and speeds that come with industrial operations and processes.
>
> **—PATRICK FETTERMAN**
> *Research Analyst*

Our research shows that the use of industrial analytics as an enabler of Industrial Transformation is rapidly growing. Fifty percent of companies surveyed say they have analytics deployed or in a pilot phase, with another 48% planning on deployment within one to three years.

Current state of Industrial Analytics Program

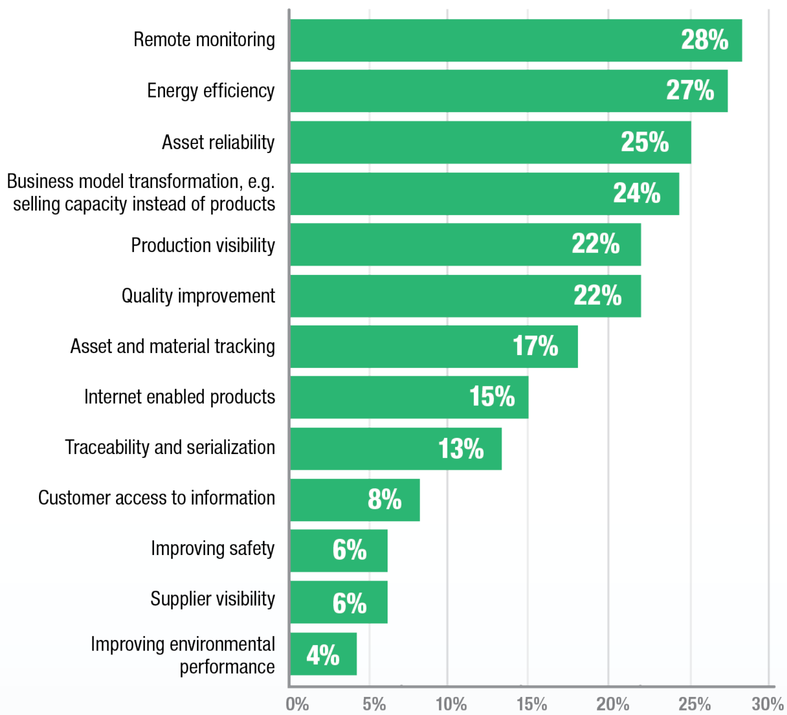


To achieve the benefits of Industrial Transformation — especially when it comes to analytics — companies need a data model that spans assets, processes, and systems to optimize performance holistically. Organizations that fail to define a common data model for transformation projects are likely to end up in "pilot purgatory" — that dreaded state of being stuck in an ill-defined project that never delivers promised results and never scales. Manufacturers that achieve their end goals, on the other hand, can gain new competitive advantages in the markets they serve and may uncover entirely new business models empowered by data.

Industrial Analytics: Use Cases Drive Data Model Decisions

When we view industrial analytics as a critical part of an Industrial Transformation business model, the best practice is to start by building a use case. For all the grandiose claims about what analytics can do, the most common use cases to date are fairly humdrum; remote monitoring (28% of companies surveyed), energy efficiency (27%), and asset reliability (25%) top the list.

Top IIoT Use Cases



© LNS Research. All Rights Reserved.

The top use cases have two distinct things in common: they're simple to explain, and the financial impact is easy to quantify. However, we expect use cases to rapidly evolve as the deployment of analytics continues to grow and teams start to dream up new uses for advanced analytics. All these new use cases, however, will also be dependent on companies deploying a data model that empowers analytics and paves the way for transformation — the more complex the use of analytics, the more sophisticated the data model must be.

> **FOR ALL THE GRANDIOSE CLAIMS** about what analytics can do, the most common use cases to date are fairly humdrum. However, we expect use cases to rapidly evolve as the deployment of analytics continues to grow and teams start to dream up new uses for advanced analytics. All these new use cases, however, will also be dependent on companies deploying a data model that empowers analytics and paves the way for transformation.
>
> **—PATRICK FETTERMAN**
> *Research Analyst*

The impact of uses cases on data models is a two-way street: as a company defines new use cases, the team must examine the data models to ensure that the required data is available to power the use case. At the same time, the organization must examine any changes to the data model to understand the impact on current deployments and use cases.

The (Many) Challenges of Industrial Data

A data model defines the relationships between disparate data entities within an organization and provides a perspective from which to examine the data. Data models play an essential role in any IIoT solution and are especially important for industrial analytics to support a transformation project. A robust and common data model is one that spans the enterprise's applications and data sources — in this case, both IT and plant data sources. It defines all the data relationships and meanings that exist within an organization. While it's a straightforward concept, industrial companies have distinct challenges when it comes to data management and modeling.

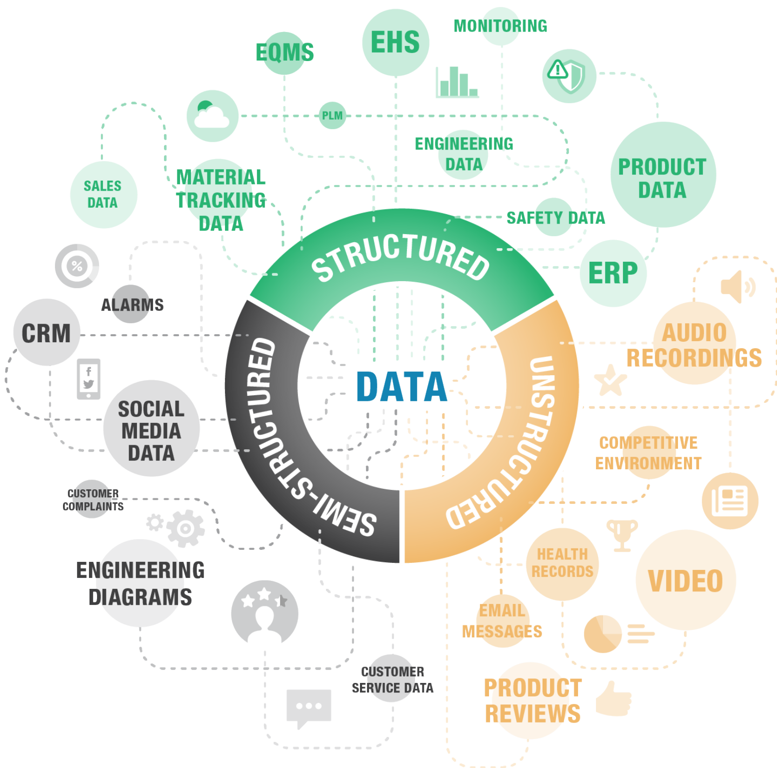
> **A ROBUST AND COMMON DATA MODEL** is one that spans the enterprise's applications and data sources — in this case, both IT and plant data sources.
>
> **—PATRICK FETTERMAN**
> *Research Analyst*

THE SIX V'S OF INDUSTRIAL DATA

When the industrial sector first started to recognize Big Data, a common mnemonic for data challenges was the "Three V's" of data: volume, variety, and velocity. While these remain fundamental challenges when building an industrial data model, there are additional factors that companies must consider.

1. **VOLUME** | A large factory can create a terabyte or more of data in a single day, while it's not uncommon for an IIoT-ready oil drilling rig to generate seven or eight terabytes daily.
2. **VARIETY** | Industrial enterprises produce and can utilize three types of data:
   - **STRUCTURED**: easily ordered and analyzed using common database technologies; often from financial, maintenance, and production systems.
   - **UNSTRUCTURED**: does not fit into a spreadsheet or database; may include text logs, multimedia content (i.e., video, images, sound files), emails, and others.
   - **SEMI-STRUCTURED**: a form of structured data that doesn't fit into relational databases, such as XML and JSON.

DATA VARIETY



A typical challenge for industrial companies is time-series data: a collection of data obtained through repeated measurements over time. Sensors and machines produce time-series data like temperature and flow, vibration readings, and many others. In fact, the majority of data generated by plants and factories is in time-series format because plants contain a profusion of sensors and intelligent machines. Many common relational databases cannot handle time-series data; specialized databases and data analysis tools manage and analyze time-series data and are often used by industrial companies.

3. **VELOCITY** | Even IT organizations that are used to millions of transactions per day in their financial systems are frequently shocked by the speed of data streaming from a plant. It's not uncommon for a plant to have thousands of sensors sampling data multiple times per second that create billions or trillions of data points every single day. When constructing its data model, an organization must evaluate the speed of each data source plus the timing of the use of that data. For example, machine monitoring for predictive maintenance requires near real-time processing of data (which also has implications for the location of the data and the

location of the analytics) to deliver valuable insights. On the other hand, combining machine data with internal and external data sources for business analytics doesn't call for the same stringent timing requirements.

4. **VALUE** | What is the relative worth or significance of each type of data and each data source? A database that has little current value (e.g., historical data for a division or plant that is no longer operational, or a process that's no longer current) may be excluded from the data model. Conversely, external data sources (e.g., weather, commodities prices, and others) may be extremely valuable; projected ROI from applying in analytics may justify the cost and use.
5. **VERACITY** (or validity and integrity) | Do you consider the data trustworthy and accurate? Nothing spoils analytics results more thoroughly than introducing invalid data to the analytics systems — as the adage goes, "garbage in, garbage out." Integrity and validity are two components that make up veracity. If the data is unchanged from its source, we say that it has integrity. If we know the data to be true and accurate, it has validity.
6. **VARIABILITY** | Is the data stream regular and dependable, even in conditions that vary unpredictably? This consideration is especially impactful for time-series data, where systems often mistakenly assume the time series will be interrupted and steady.

WE'VE HEARD MULTIPLE STORIES about problems with identifying data locations when building a data model. In one case, a large industrial company was unable to associate sensor data to originating machines in as many as 60% of cases. Naturally, this reduces or eliminates the value of insights provided by analytics.

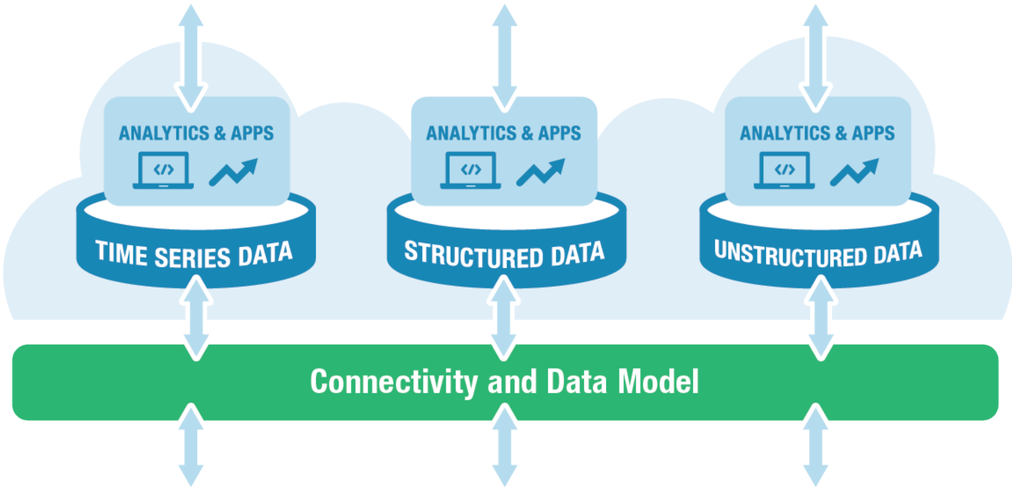
—PATRICK FETTERMAN
Research Analyst

LOCATION: DATA AND DATA PROCESSING (EDGE VERSUS CLOUD)

A fundamental step in building a data model is mapping the locations of data and data sources, which can be incredibly complicated for industrial companies. This step includes agreeing on common terminology; any misunderstanding about what constitutes "Edge" will lead to confusion and unnecessary complexities. For example, to an IT person, everything in the plant might be considered Edge; the operations professional might view only remote plants and machines as Edge. Regardless of your company's definition of Edge, it should be a key source in your data model.

LNS Research has heard multiple stories about problems with identifying data locations when building a data model. In one case, a large industrial company was unable to associate sensor data to originating machines in as many as 60% of cases. Naturally, this reduces or eliminates the value of insights provided by analytics.

Some data that can be extremely valuable for business analytics may be from sources external to the company. For example, economic forecasts, weather patterns, and transportation costs may all be extremely valuable but are likely dependent on external data sources. Each organization should consider the value of each potential external data source when building a data model, and when building use cases for analytics.



Additionally, deciding what to do with data that resides in different locations across the enterprise is essential. One question that comes up often is: Leave large databases in remote locations for processing and only provide pointers in the data model, or move the data into a corporate data lake? The decision will depend on the intended use of the data, among other things. For example, Edge computing devices may be necessary for real-time processing of machine data in remote locations (for monitoring and predictive maintenance). However, this approach isn't sufficient for business analytics that will combine data from plant systems and business systems. Analytics running on large combined data lakes may be more suited for cloud computing. Research findings from LNS show that the industrial enterprise will perform a wide distribution of analytics at multiple locations.

WHO OWNS THE DATA?

Keep in mind, lots of industrial organizations have inadvertently created "data silos" with a specific use in mind. Most likely, they optimized the databases for a singular objective, such as productivity, quality, reliability, safety, or some other purpose. Getting access to data silos can be complicated; breaking down the silos, so the data flows seamlessly into an analytics system can be even more problematic. A related concern is that the systems may have security policies in place that prevent the data from being exported or accessed via APIs.

Our recommendation for bridging these silos is to include silo owners on the Industrial Transformation team. Bringing the data owner onto the team provides an insider's view about why each data silo was created and how it's used; the insight and collaboration simplifies integrating them into the corporate data model.

DATABASES, DATA HISTORIANS, AND DATA LAKES

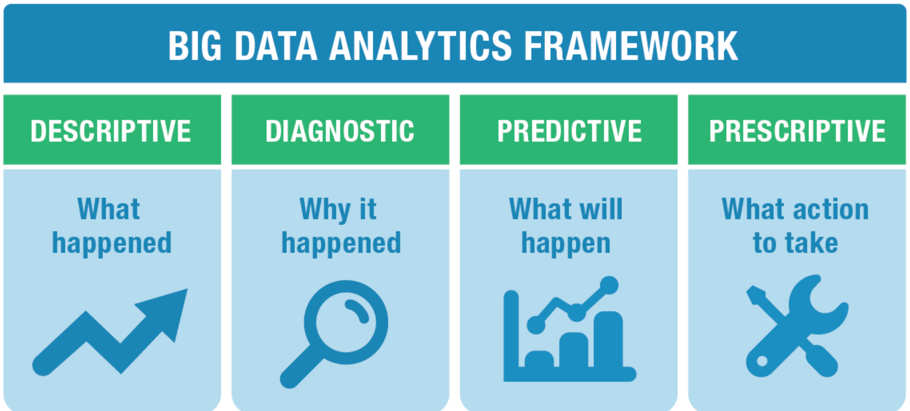
Industrial companies have been collecting data in different formats for decades. Recall that time-series data is particularly abundant in industrial companies; over the years this tendency led to the rise of data historians and other specialized databases designed specifically to manage time-series data. When combining time-series data from specialized systems with other types of data to power analytics, specific challenges arise from trying to mix disparate data types in a single database. In some cases, companies solved the problem with a "data lake" — a storage repository that holds enormous amounts of raw data in its native format. More typically, however, data lakes are used to contain Web, image, and sensor data. Organizations should incorporate all data lakes into the data model as a significant data source. The ability to combine these disparate data sources and analyze the results enables companies to align production and maintenance processes with financial and customer requirements — in many cases, for the first time.

> **BUILDING YOUR DATA MODEL** is not a short-term project. While it may take months to establish a data model, once it is ready, implementing industrial analytics can be relatively quick.
>
> **—PATRICK FETTERMAN**
> *Research Analyst*

Building Your Data Model – and Keeping it Current

Building your data model is not a short-term project; the initial model should involve a dedicated cross-functional team and may take months to design, with hundreds of decisions along the way. It must include both IT and OT staff members, plus owners of key data sources and users of the analytics and other outcomes. Moreover, the effort must be accompanied by use cases that will drive decisions along the way.

While it may take months to establish a data model, once it is ready, implementing industrial analytics can be relatively quick — moving to predictive and prescriptive analytics rapidly, as there will be a high level of trust in the data and the analyses. Ultimately, a well-defined data model can eliminate the risk of pilot purgatory, accelerate the deployment and scaling of IIoT technology like industrial analytics, and drive the organization's Industrial Transformation. Without a data model, all of this is at risk.



At the same time, a well-defined data model also readies an industrial company for emerging technologies like artificial intelligence (AI) and machine learning (ML). With these technologies advancing rapidly, the long-term value of a data model (and the analytics it empowers) is likely to be in uses that the market has not yet envisioned, but that will emerge as companies begin to deploy these technologies on top of their data models. (And, recall that emerging use cases will have an impact on data models, which need to evolve as required.) Additionally, AI systems may be powerful enough to help build and iterate these models in the future. At LNS, we have already seen demos of IIoT systems that can "crawl" data sources throughout an enterprise and begin mapping them to a data model with little or no human intervention.

> **A WELL-DEFINED DATA MODEL** readies an industrial company for emerging technologies like artificial intelligence (AI) and machine learning (ML). The long-term value of a data model is likely to be in uses that the market has not yet envisioned.
>
> **—PATRICK FETTERMAN**
> *Research Analyst*

Recommendations

For a company gearing up for Industrial Transformation, it's clear that a well-planned data model is a fundamental preparatory step that can accelerate transformational technology projects and increase ROI from tech investments. It's not a short and simple project and can consume substantial corporate resources. However, the result is a high-value, high-fidelity data model that can yield benefits that far surpass the investment.

Companies that want the best outcome in the shortest amount of time in building a data model should take steps that encompass people, processes, and technology:

ASSEMBLE A CROSS-FUNCTIONAL TEAM of operations engineers, IT staff, data owners, and end users who can collaborate to build the right use cases. Including both IT and OT personnel ensures that experts in each type of data will be available to evaluate each data source based on the criteria mentioned throughout this report.

BUILD AND DOCUMENT USE CASES, including definitions of success and projected ROI. How will the data model power your use cases around operations, reliability, financial performance, and other areas? It's only by envisioning the use of data in analytics — and, more broadly, in Industrial Transformation — that the team can understand which types of data will be needed and how the company will use them.

GET BUY-IN FROM A CROSS-SECTION OF EXECUTIVES on how critical a data model is for the company's Industrial Transformation. This buy-in is most important for the departments and business lines that generate the data, and for those teams that will experience the greatest impact from future use of analytics.

ANALYZE AND DOCUMENT AVAILABLE DATA SOURCES, including internal and external sources, to identify formats, speed and volume, validity, usefulness, locations, and ownership of each. Be aware of data silos — including why they evolved and what their current condition and use case is. Note especially the criticality of that database for operational functioning, and any potential risk of touching or using that data. Also, document all concerns when building the data model.

SCHEDULE REGULAR REVIEWS far into the future to support the evolving nature of the industrial data model. A static data model can lead to a static business. Every company needs a well-documented process to evaluate the existing data model, continuously update it to meet emerging use cases, and take advantage of new technologies as they become commercially viable.

Presented by:

ACRONYM
QUICK REFERENCE
VIEW ON BLOG

Author:

Patrick Fetterman,
Research Analyst
patrick.fetterman@lns-global.com

LNS Research provides research and advisory services to guide companies through industrial transformations. Our research focuses on how digital technology drives industrial transformation across the value chain and offers insights into the people, processes, and technologies required for achieving Operational Excellence. Learn more at www.lnsresearch.com.

JANUARY 2019