Sight Machine: The Data Foundation for Manufacturing

Making Plant Data Continuously Useful for Operations, IT, Data Science

SIGHT MACHINE

Contents

One Data Set, One Foundation

Manufacturing plant floors are rich in data. And as manufacturers invest in Digital Transformation, both plant teams and internal technology teams increasingly seek to use that data. There are often two broad, sometimes overlapping initiatives:

1. **Analysis** to improve operational performance in plants.
2. **Data science, AI, and ML** for prediction. Areas of strategic opportunity include sustainability, supply chain resilience, process improvement, and, ultimately, new business models.

Manufacturing transformation requires:

- Pipes to acquire data
- Math and pictures to visualize it
- Sense-making to make it useful

Plant data is not like virtual data at all: it's voluminous, varied, missing, late, and out of order

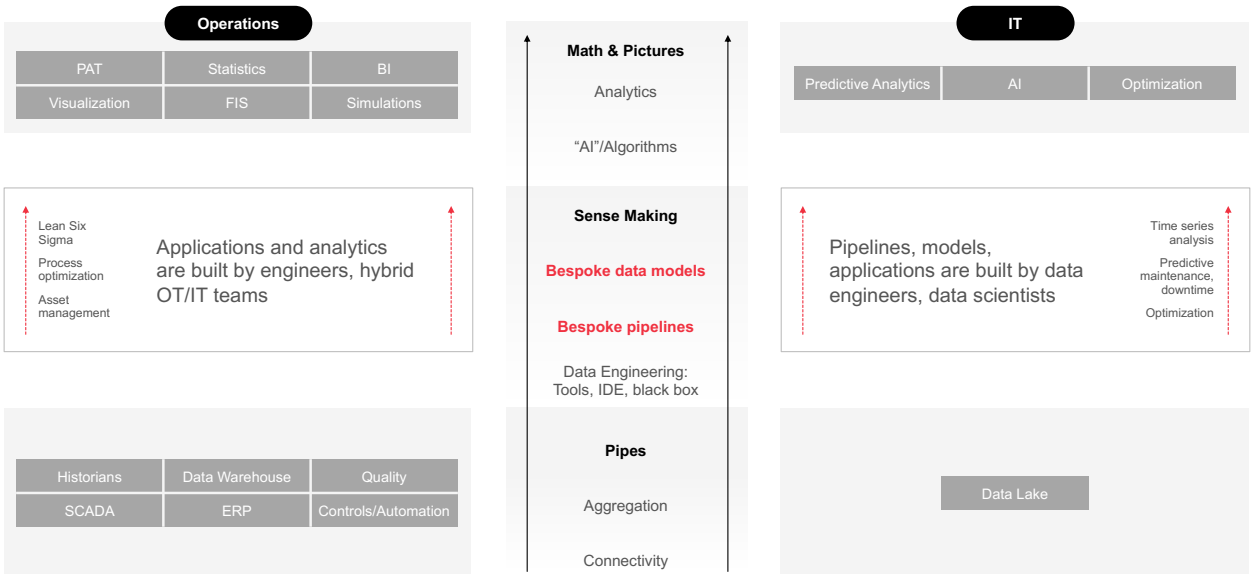
Once those challenges are overcome, plant data becomes extraordinarily useful

Plant data is abundant, raw, extraordinarily varied, and among the most difficult data in industry to work with. As a result, companies have struggled to achieve any expansive impact on either the plant floor itself or at loftier strategic levels. Scale has remained elusive.

From years of hard lessons, manufacturers know they must address the underlying data problem. It's now clear that to achieve digital transformation, three levels of technology are needed:

1. First, **pipes**. Pipes get data out of data sources, collected, and moving. The need for good pipes has been long understood, and with edge, connectivity, aggregation, and cloud, innovation in this area continues, especially as 5G comes to the market.
2. Next, there is deep appreciation for **analysis**: this level includes tools like BI, data science, and AI/ML. Significant time, money, and skills have been devoted to this level, but scale has been a terrible challenge. Efforts stall because plant data is so varied and difficult.
3. And then there's a layer no one really talked about for many years and now is increasingly recognized as necessary, right in the middle, between pipes and analysis. This level is **sense-making**. Technologies here help turn data into usable information. For some fields, horizontal solutions and DIY tools are feasible, and IT and operations teams should do the work. In others with especially complex data environments, domain-oriented products provide significant acceleration.

In the last five years, recognition of the need for sense-making has progressed steadily. In early stages of digital transformation, as Big Data technologies matured and Internet of Things investments took hold, manufacturing enthusiasm for IoT led to massive investment in pipes. A focus on AI followed, accompanied by a widespread belief that pipes and AI combined would radically benefit plants. But then, as companies increasingly engaged with huge amounts of plant data, the need for sense-making became evident. Companies use operations engineers, data engineers, and data scientists to understand plant data. These workflows, represented in the figure below, are challenging, and, for the most part, the job of sense-making is left to people who build models by hand.



Turning data into useful information is no easy feat. If it were, given all the money and time that's been invested, sense-making would have been achieved long ago. Here are just some of the challenges that data engineers and data scientists must overcome with plant data.

- → Extraordinary heterogeneity
- → OT data does not flow through mature IT infrastructure
- → Data takes days or months to gather, model, and analyze — but to influence and improve production, analytic insight needs to be real-time
- → Critical data is often late, missing, or out of order

Sight Machine offers a unique solution aimed directly at the gap in this map. The offering is end-to-end. It includes everything from pipes to AI and BI, but it works with and bolts into existing technologies. The heart of the offering is the sense-making layer — a streaming pipeline that continuously analyzes all plant data. The product transforms and analyzes data from any manufacturing environment in process and discrete industries, and plants small and large. It complements and accelerates existing investments, and provides OT and IT leaders with a single layer of trusted information.

Once data challenges are overcome, plant data becomes extraordinarily useful. Benefits include:

- **Scale:** one common Data Foundation for all models — both operational and analytical
- **Flexibility:** handling changing data environments
- **Operational agility:** stream-processing technologies provide continuous, real-time understanding

Below, this paper briefly describes how sense-making is performed. It details two product innovations: Common Data Models and Pipeline as a Service.

Common Data Models and Pipeline as a Service

Sight Machine captures manufacturing reality with Common Data Models populated by our robust streaming data pipeline, or Pipeline as a Service. This approach handles changing data and high complexity and is applicable to every manufacturing process. Once data is expressed in standardized forms, everything in the system becomes relatable.

COMMON DATA MODELS

Consider the way most of us think about manufacturing. If we were to try to analyze a plant, where would we start?

Most of us would probably be inclined to start by analyzing machines. We would do this because, to our intuitive way of thinking, plants are built from machines, so the right building block for more complex models must be machines.

Common Data Models allow Sight Machine to model work done by machines rather than the machines themselves

Pipeline as a Service continuously stream processes and analyzes all data that can be associated with production

But this approach leads to big problems. Let's assume, for example, we just want to understand the production of a car. A car has 30,000 parts, and if we're going to model all those parts being made, we would then have to model tens of thousands of machines too. There's no way all these models can be organized and understood — but that hasn't stopped companies from trying. Machines have generally been the building block of choice.

Sight Machine, in contrast, uses only about ten Common Data Models. One of the most important of these is the Production Model. As you would expect from the name, the Production Model allows clients to continuously understand production activity at different levels: machines, lines, and plants.

But Sight Machine's Production Model doesn't begin by modeling machines. *It begins by modeling units of work done by machines.*

Let's state that again: the core of the Common Data Model for production is a unit of work. To illustrate, we can go back to all those machines needed to make a car. A unit of work can be the repetitive act of die casting an engine block. Applying a paint or coating. Injecting plastic into a mold for a dashboard piece. It can represent work done in a discrete process or in continuous production And it can take a second or hours. Types of work are limitless, but the idea of the unit of work is common, almost elemental. A unit of work is just the repeated cycle of activity by a machine.

As a matter of organizing data, here's how this Common Data Model works. Every time a machine does a unit of work, a row of information is generated from all the data associated with that work. The unit of work is described with data from sensors on the machines, quality systems, MES, historians, ERP, and even ambient data like temperature and humidity, or other data about raw material characteristics.

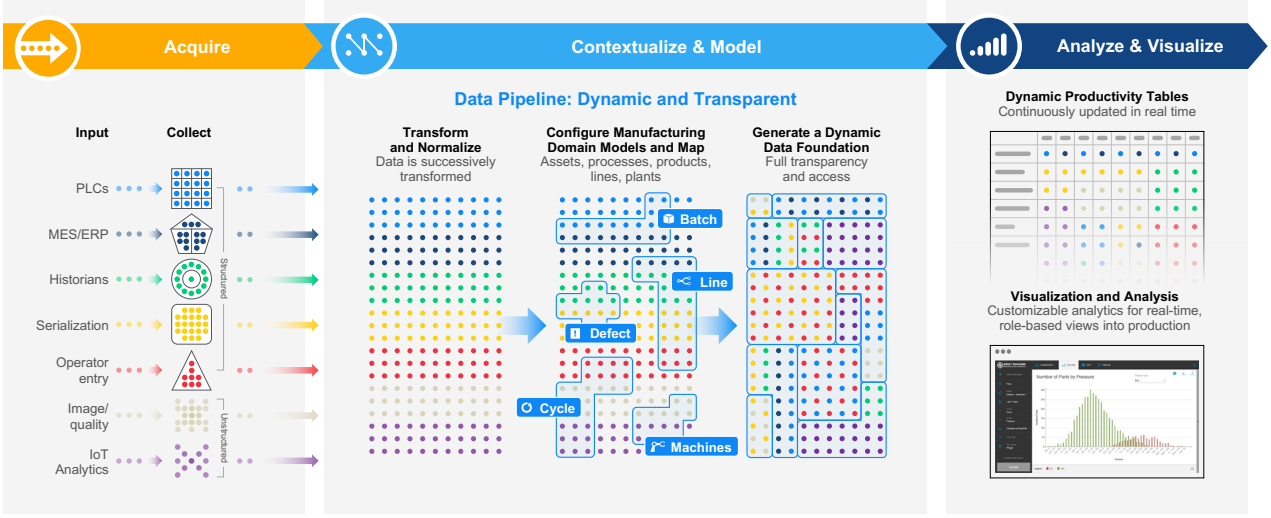
There is no minimum level of data required to characterize a unit of work, and no maximum. Information about the unit of work can "flex" with the level of ingested data. In data table terms, the row of information can be modified to add parameters in new columns as new data becomes available. Information about the unit of work also changes as data sources change or late-arriving data or missing data is incorporated. This kind of flexible recalculation, not possible until the advent of robust stream processing tools, is essential for factory data.

Once data is transformed, these are the kinds of things we can know about every unit of work:

- **How long did each unit of work take?**
- **What were the characteristics of the material being worked on at the beginning of the unit of work and at the end?** What was its weight, density, or temperature when it entered and when it left?
- **What happened as the work was being done?** For each parameter we want to measure — and the number can be thousands if useful — what are the values by parameter during that unit of work? Is pressure just right? Too high? How does it trend from one unit of work to another? For example, from the unit of work at 10:00:30 a.m. to the next unit at 10:01:15 a.m.? Or how about between machines? Or different plants?

PIPELINE AS A SERVICE

From almost a decade of streaming and analyzing data from plant environments, Sight Machine has developed a second innovation: Pipeline as a Service for manufacturing. The pipeline includes hundreds of techniques to map, model, and distill data so raw plant data is continuously mapped accurately into Common Data Models, and in turn elevated into analytics. Pipeline as a Service employs as many as 50 automated transformations. It's browser-based, configurable, and transparent: data engineers and data scientists can see both raw and transformed data throughout, and they can use it themselves to generate information. Once configured, Pipeline as a Service continuously generates useful information, and adjusts in the background to changing data.

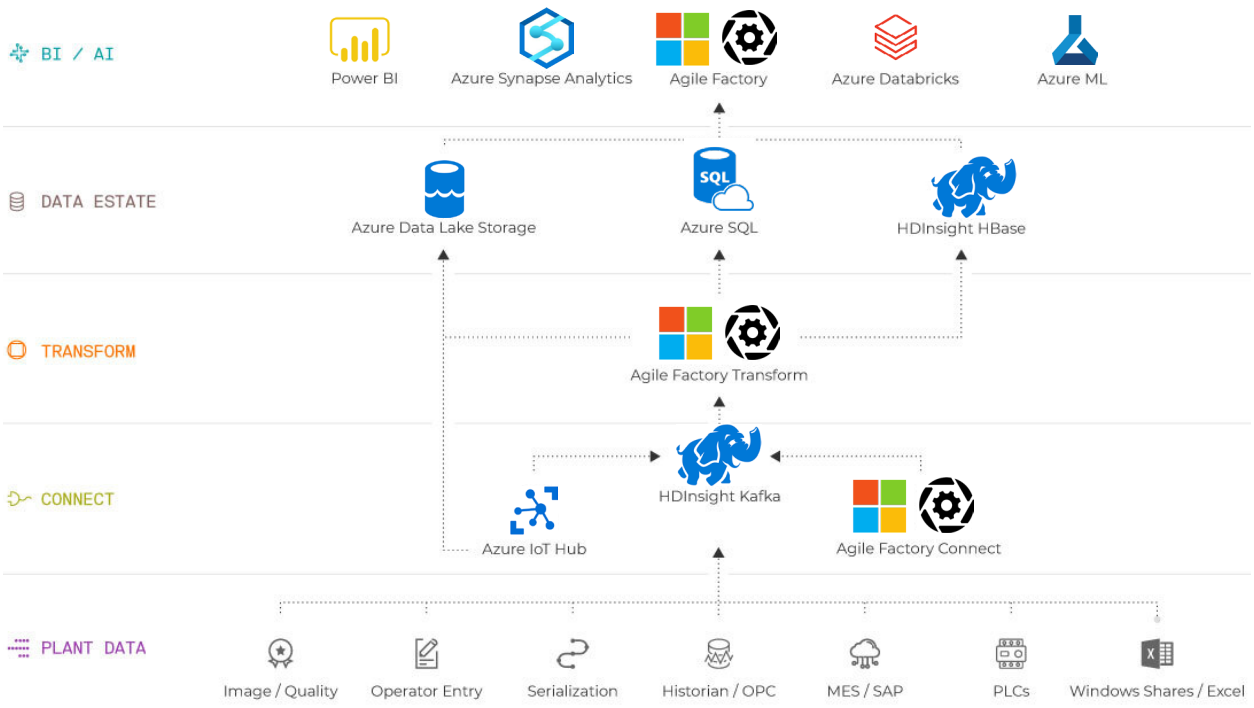


Reference Architectures: An Integrated Solution for Manufacturing

With Data Foundation, companies can now accelerate and unify their efforts for all work streams — operations, IT, data science — around established technologies.

Sight Machine supports and integrates with technologies from all major edge, cloud, and analytics providers. The solution has been implemented regularly in AWS, GCP, and Azure.

Below is an example of a reference architecture of Sight Machine in a full Azure environment: ADS, Data Bricks for Azure, Synapse, Azure ML, Azure Purview, Azure Edge, Azure Cloud, and all of Microsoft's manufacturing and supply chain partners. Together, these technologies enable fast insight, scale, and impressive business impact.



Foundations Are for Building

Converting torrents of data to useful information is the breakout move.

It's a move enabled by products, and for that reason the move is itself more standardized and goes faster than many expect. Plants can be up and running and pay back their investment with hard, quantifiable results in well under a year.

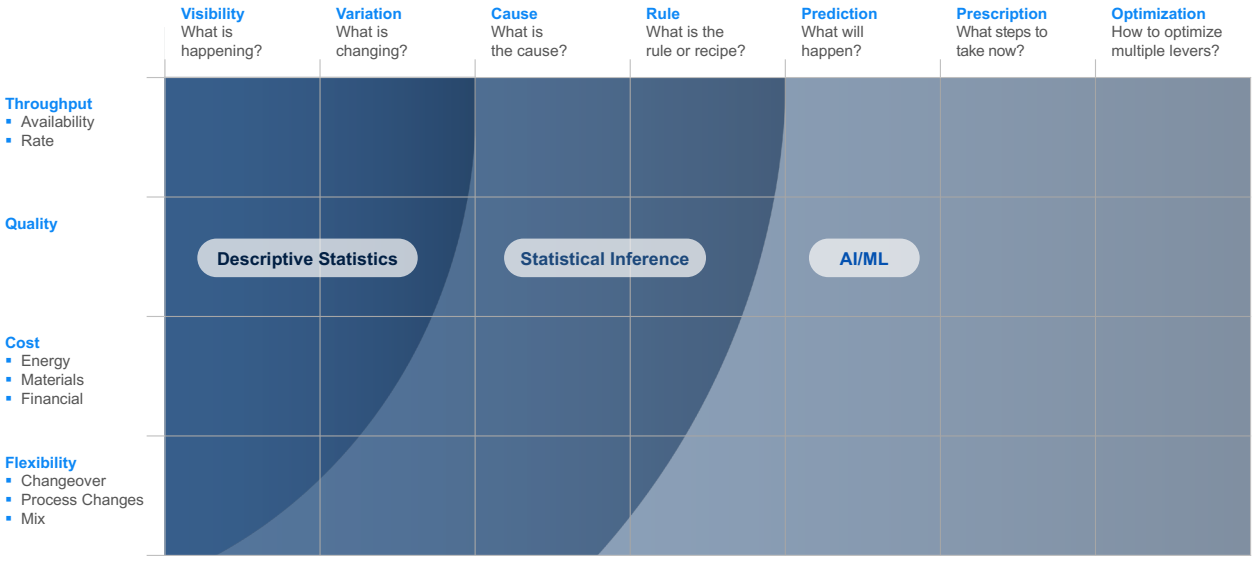
The next moves are in the hands of technology and operations leaders. Companies may want to develop their own algorithms or combine Contextualized Information about production with other information in unique, proprietary ways. This is the domain of tools and the abundant number of good visualization, analytic, and AI frameworks that are available. With the right foundation in place, both operations and technology teams can soar.

Manufacturing is a single discipline with a deep structure

- While industries are different, data unlocks value in the same ways for all
- All plants work to improve the same levers: throughput, quality, cost, flexibility
- Contextualized Information supports improvement of each lever

But, as every transformation leader knows, the acid test for success is not just success at the data level, but ultimately improvement in operations.

Below is the value map for manufacturing. On a single page, this map lists the four levers of manufacturing performance — levers that apply to every manufacturer regardless of what product the manufacturing firm may make. Columns in the table list key insights for each lever: what is happening? What is changing? What is the best setting? What will happen?



Applied to Contextualized Information, statistics and other data science techniques drive improvement across the map.

Examples include:

Operational Visibility

- Finding and fixing the root causes of quality issues
- Identifying the causes of micro-stops in high-speed automated production
- Developing best practices across sites
- Minute-by-minute visibility of key production KPI's and changes across the manufacturing enterprise
- Visibility of production across the Intelligent Supply Chain

Production Operations

- Dynamic recipes for optimizing process control. Recipes take account of variations in raw materials, production conditions, goals to be optimized

Asset Productivity

- Reducing energy use in heat-intensive industries through improved quality and use of assets
- Optimizing asset productivity and scheduling of process steps in processes involving varying demand and supply

Traceability and Supply Chain

- Joining insight about upstream and downstream production processes
- Sharing quality and throughput data between tiers
- Visibility and analysis of distributed assets: useful to providers of assets, plants, and enterprises who seek to manage production facilities as an integrated fleet

Use cases are limitless. With the benefit of a real-time Data Foundation, manufacturers generate feedback and predictive insight for every stakeholder.

Data Alone Is Not Enough

It is now widely accepted that for almost every type of manufacturing, Digital Transformation will create radical improvements in productivity and sustainability. Almost a decade into transformation, most believe the path to those improvements is built on data.

It is not hard to see why. Manufacturing generates more data each year than any other sector, and uses almost none of it effectively. Morgan Stanley, for example, estimates Data Era advances in manufacturing will generate a 5% improvement in global GDP and a 20% improvement in manufacturing EBITDA.

However, while the vision is compelling and the efforts to achieve it considerable, impact still remains elusive. This is the paradox for manufacturing: factories present the most opportunity and the most data of any sector — and the most difficulty. In the buildup to and evangelism of transformation, most experts looked at the amount of data in plants, its obvious value, and the explosion of new technologies for understanding other types of data, and assumed plant data could be put to use with the traditional data processing technologies.

In other sectors, that assumption was warranted. Relatively clean, structured data (or at least data that is suitably structured for effortful analysis) is abundant in sales and marketing, product development, energy generation, smart buildings, smart cities, and, in the industrial realm, product design and supply chain logistics. But of all the parts of industry to be touched by transformation, perhaps none remains more rich in data and more difficult and dark than the factory floor.

As one prominent Silicon Valley firm wrote recently, data alone is not enough. There is data, and then there is the ability to use it. They are not the same.

Almost ten years into Digital Transformation, there's good news for manufacturers. The ability to use data is finally here.