Industrial IoT Data Collection & Analysis

for

Real-Time Decision-Making and Predictive Maintenance

RTInsights.com

memSQL

CHAPTER 1

The Business Value of Using Real-Time and AI/ML-Powered Decisions for Industrial IoT Applications

Across many industries—including manufacturing, oil and gas, utilities, telcos, media, and more—downtime is expensive. Operating costs can grow as staff and machinery sit idly by, and revenues may drop as services and products are not delivered.

Traditional approaches of waiting for something to go wrong are incredibly bad for business, whereas getting in front of potential problems saves time and money and makes life easier for customers. It also opens up new business potential, such as offering capabilities as a service, rather than selling equipment to customers. As a result, companies today have a great need for real-time decisions and proactive, predictive maintenance.

Other areas that can be impacted include overspending when operations are inefficient and losing out on a new opportunity by missing telltale signs of emerging customer demand or market opportunity.

Rather than reacting after the fact, insight into the root causes of the problems that lead to downtime, defects, inefficiencies, or missed opportunities beforehand can potentially cut costs, accelerate time to market, grow revenues, and increase customer satisfaction.

Increasingly, industries are using IoT devices to garner those insights, employing the devices to get information about every aspect of their operations. For example, a manufacturer might use sensors throughout an assembly line to predict when a machine will likely break down, or an oil and gas organization might use predictive analytics on data from sensors installed across a pipeline to help identify corrosion and damage.

Organizations can leverage the data derived from billions of connected devices to drive never-before-seen efficiencies in internal operations and create entirely new value streams that transform their relationship with their markets.

Data Explosion

Industrial IoT brings together connected intelligent devices and analytics in a way that allows organizations to monitor, collect, exchange, analyze, and deliver valuable new insights about their systems and processes. These insights can help drive smarter, faster business decisions.

But one problem most companies encounter is that IoT devices produce very large volumes of data. One industry report estimated that by the end of 2019, IoT devices will generate more than 500 zettabytes per year in data (a zettabyte is a billion terabytes). That number is expected to grow exponentially in the years that follow.[1]

500 zettabytes

THE AMOUNT OF DATA GENERATED PER YEAR BY IoT DEVICES[1]

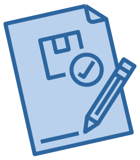
[1] https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html

Making use of, and efficiently managing and analyzing such large volumes of data, which are continuously generated, requires new thinking. Many companies are complementing traditional analysis with real-time, artificial intelligence (AI) and machine learning (ML) algorithms to get decision-making information out of the data in a timeframe that will allow a company to take proactive actions.

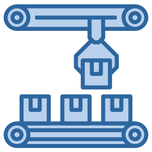
Specifically, industrial uses of AI and ML applied to IoT data are seen as a great way to supercharge and automate maintenance forecasting and other aspects of operations. The reason: AI and ML can more accurately and quickly help automate insights and decisions about complex systems. Using these more sophisticated analysis techniques, predictive maintenance can reduce machine downtime, thus lowering costs and improving margins.

Areas where AI and ML could help industrial operations include:

Equipment uptime and availability:
Telemetrics about the state of a device or machine combined with historical performance data can provide insights that foreshadow preventable failures.



Quality assurance (QA):
Better optimized and automated QA processes not only reduce costs but also enhance product quality.



Assembly line optimization:
IoT data can be used to identify inefficiencies which, once addressed, can increase output.

Other ML use cases that can help industrial operations include supply and demand forecasting, data security, and supply chain logistics and management.

CHAPTER 2

Barriers to Success

The combination of large volumes of streaming data coming from IoT devices, plus the use of sophisticated AI and ML models, creates the need for a new kind of data infrastructure that combines operational analytics with new data processing techniques to sense problems and respond to them before issues arise.

The idea of using IoT devices to track machine readiness and automate predictive maintenance in industrial applications is widely appealing. However, most organizations do not have the data infrastructure required to take full advantage of these potential benefits. And there are many challenges that must be overcome to achieve success.

Identifying the Challenges

First, there is the issue of **data preparation**. To harness the value of IoT and deliver innovation-driving insights, organizations must quickly prepare and standardize great volumes of disparate, unstructured data. In the industrial world, large industrial organizations can typically collect billions of data sets from machines, sensors, and internal business applications.

Data preparation takes up to 80 percent of the time and resources in a data analysis operation.[2] As organizations move to new IoT initiatives, it's important to consider new technologies and processes that will keep up with this huge influx of data.

Another issue is **data complexity**. Many IoT devices may generate data with timestamps or geotagged data. But this data must be combined with other structured sources to be of use. The complexity is only multiplied as more data is generated. Common data preparation tools like Excel can't handle this complexity, which leaves skilled analysts locked out of working with the data.

Data integration is yet another obstacle to success. Most analysis infrastructures are not designed to exchange or process the vast amounts of complex information pulled from sensors and connected devices. It's difficult to quickly integrate and enrich machine-generated data with data from business applications and data repositories. Greatly adding to this problem is the fact that much of the data is siloed.

Most data scientists spend only 20% of their time on data analysis.

80% of their time is spent finding, cleaning, and reorganizing data.

[2] https://www.infoworld.com/article/3228245/the-80-20-data-science-dilemma.html

Traditional analytics processing would first collect and store multiple streams of data and then put that data into a relational format for real-time or historical analysis, which is commonly executed using slow batch processing. These tasks can stress traditional analytics infrastructures, resulting in significant analysis delays or an inability to run workloads. This precludes the instant discovery of anomalies and rapid generation of predictions.

Scalable stream ingest also must be addressed since most traditional systems either struggle to collect every event, or store data in an inefficient and costly way. Users are then forced to do all sorts of tricks to move or delete data onto less costly storage media.

Naturally, **usability** can be an obstacle to success with any operational analytics endeavor. Many times, a solution will be used that requires deep data science expertise to craft queries and conduct investigations. This greatly limits the use of sophisticated analytics within the organization and incurs delays as business and operational staff wait for the data scientists to perform their duties.
Finally, there is the issue of concurrency. In the rush to solve for the challenges of the volumes of IoT data and large numbers of concurrent users, some solutions abandoned the core features of databases that make them highly performant and easy to use.

Adopting New Approaches to Analytics

Simply put, these challenges are too much for current infrastructures. They cannot perform the needed operations to deliver decision-making intelligence in a timely enough manner. Not only does a great amount of data need to be analyzed, but the true value also comes from a system that can continuously monitor, identify, and respond to opportunities or issues in real time.

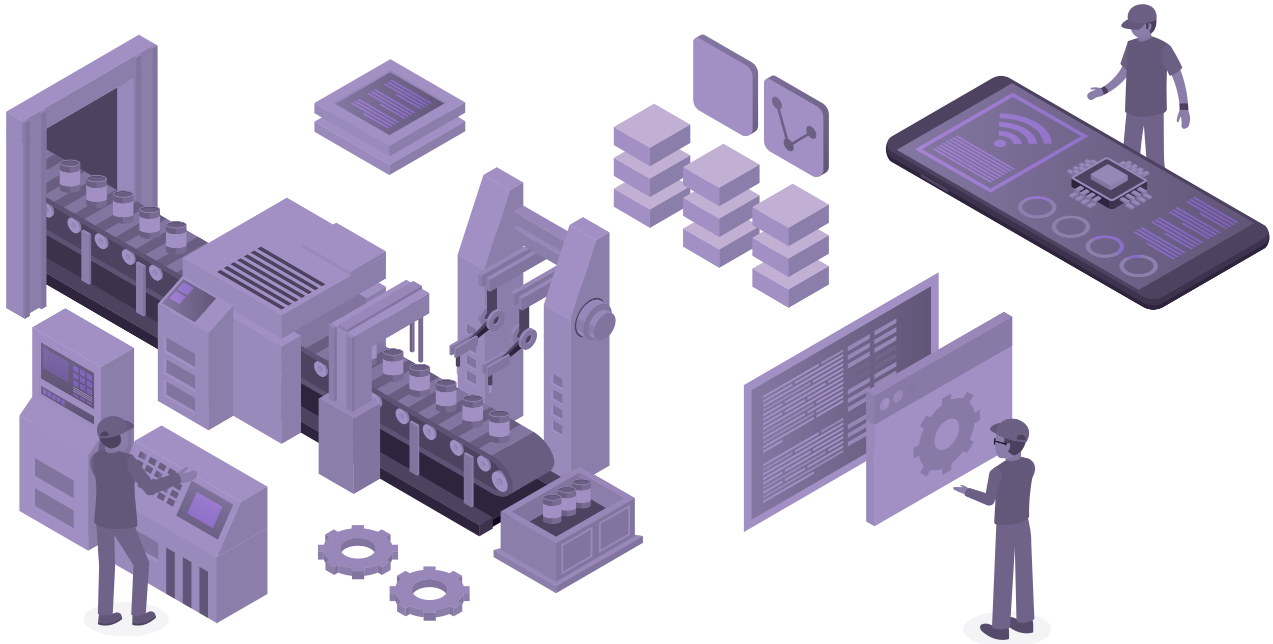
This situation is leading organizations to adopt new approaches to their traditional analytics efforts. Specifically, organizations are looking to leverage the power of AI and ML to spot critical information in real-time data. This is causing operational strains as siloed data and slow analytics prevent the power of AI and ML from being brought to bear.

In response, organizations are moving to operational analytics. This is a hybrid processing approach that depends on the ability to run both transactional and analytical workloads at scale on an integrated SQL database.

The ability to run both kinds of workloads on a single platform has been given a variety of names, from **hybrid transaction/analytical processing** (HTAP), to hybrid operational-analytics processing (HOAP), or **translytical data platforms**. In this genre of analytics, in-memory processing is a key technology enabler. It delivers the performance needed to analyze streaming data. An added benefit is the simplicity of the architecture. Only one system needs to be maintained, and there is no movement of data.

CHAPTER 3

What's Needed: A Reference Architecture for Operational Analytics and Automated Decisions



Overcoming the obstacles requires a new architecture. Specifically, what's needed is an operational analytic architecture, one that plugs into existing data infrastructure and analytical tools without disruption. The challenge is how to transition to this new architecture in phases so as to not disrupt the business.

An operational analytics architecture should be based on the well-known Lambda architecture. A Lambda architecture is a data-processing architecture designed to handle massive quantities of data by taking advantage of both batch- and stream-processing methods. This approach to architecture balances latency, throughput, and fault tolerance by using batch processing to provide comprehensive and accurate views of batch data, while simultaneously using real-time stream processing to provide views of online data. The two view outputs may be joined before presentation.

The rise of Lambda architecture is correlated with the growth of big data, real-time analytics, and the drive to mitigate the latencies of map-reduce, which provides a simplified architecture for real-time, stream processing engines. The Lambda architecture is designed to model everything that happens in a complex computing system as an ordered, immutable log of events. Processing the data is completed as a series of transformations that output to new tables or streams.

Such an architecture is needed to move beyond the problems of the past caused by siloed data. This new architecture must continuously integrate all the needed data available for analysis and decision-making into one place. An advantage of this approach is that different groups within an organization can use the same data to achieve different objectives. For example, by getting IoT data from equipment on a manufacturing plant floor, service technicians can perform predictive analytics to prevent downtime, while quality control teams can use the information to help identify product defects.

Another factor to consider in the architectural choice is performance. An organization might have data coming from many thousands of IoT sensors and monitors. The platform must be capable of high-performance computing and sustained workflow throughput to enable analysis of large volumes of real-time data on the fly.

Given the business-critical use of the information derived from such an analysis, the solution selected must deliver enterprise-class reliability, security, tools, and features.

Taking these requirements together, the critical elements in an operational analytics architecture include:

- A core data platform that supports high availability, security, performance, scalability, and management
- Data integrity and consistency to deliver trusted data
- Native tools or integration with third-party vendors' tools to support data management functions
- Support for concurrent queries, transactions, reports, and data access patterns
- Deployment options for on-premises or cloud
- Access to data using standard connectivity such as SQL, ODBC/JDBC, XML, or REST

Enabling New Types of Analysis

A solution that incorporates the above critical elements would be able to perform the analysis of IoT data needed in industries today. The information derived in such an analysis could, for example, take predictive maintenance in new directions.

For example, a common way to avoid equipment downtime is to replace a component when it gets close to its Mean Time to Failure (MTTF) age. There are two issues with using such an approach. A part may fail before its MTTF age, leading to unplanned downtime. Or, a part that is fully functioning may be replaced well before it actually fails, thus making less than the maximum use of that part.

ML-based analytics run on streaming IoT data and complemented by product history information could be used to spot precursors to trouble in a part that has not reached its MTTF age. Similarly, that analysis could be used to keep a part in service for an extended period of time, knowing the analysis will tip off service staff before the failure occurs.

CHAPTER 4

Operational Analytics Processing in Action

Industrial operations require a fast and flexible data processing infrastructure that enables continuous monitoring, analysis, and actionable information to respond to the problems that can arise in complex systems. IoT operational analytics processing solutions can predict and avoid costly maintenance events at scale using a database platform that cost-effectively streams, scores, and responds to data events in real time.

There is a great need for this type of information to prevent disruptions. The impact of downtime when equipment or plants go offline is a loss of production capacity. Beyond that direct cost, downtime might jeopardize contract commitments. If a produced product misses a shipping date, a manufacturer might look for another parts supplier to ensure their production schedules are met.

These problems and costs can be avoided by analyzing IoT data in real time. To get actionable information from this data requires a platform that offers real-time dashboards with live data analysis for drill-downs and aggregations, and ML-enabled analysis that runs sophisticated algorithms in SQL or uses third-party ML frameworks such as TensorFlow.

Use Case: Maximizing Performance in the Energy Industry

Using these technologies, one energy industry company built an IoT solution that used complex scoring to track dozens of readings in real time to adjust drill operations to maximize performance and availability. The solution also predicts drill bit failure before outages occur. This can result in significant cost savings, as a day of unplanned drill downtime can cost an oil and gas company roughly $1.25 million.[3]

To that end, operational analytic processing on IoT data can be used for instant response to critical machine events and proactive, predictive maintenance. This information can be used to:

- **Maintain equipment uptime** by analyzing millions of equipment conditions and events with real-time monitoring
- **Reduce operating costs** by avoiding costly outages with optimized equipment operations and planned maintenance
- **Predict failures** to prevent high-risk operating conditions with real-time predictive scoring

[3] https://connectedworld.com/make-data-the-new-oil-iiot-enabled-predictive-maintenance-for-the-oil-and-gas-industry/

Use Case: Enabling Real-Time Visibility for Utilities

A utility operator used operational analytics processing to ingest and analyze smart grid data to enable real-time visibility for energy delivery. The solution also immediately identifies energy usage patterns to help customers lower power consumption.

Using the analyzed IoT data, the operator was able to:

- **Improve grid availability** by monitoring for and locating grid issues in real time to improve customer satisfaction and reduce service calls
- **Lower operating costs** by predicting grid operating issues and minimizing costly outages
- **Empower customers** by delivering real-time visibility of energy consumption to help lower customer energy usage

Use Case: Optimizing the Supply Chain

Another organization used operational analysis processing of IoT data to optimize its supply chain to improve demand forecast accuracy and reduce product delivery lead times. Supply chain analytics improves margins and revenue with minimal technology investment. The solution provided instant visibility into a variety of supply chain processes, including storage management, inventory management, transportation operations, and portfolio optimization.

Relying on the fast ingestion and analytic performance of the operational analytics processing solution, the supply chain optimization programs were able to:

- **Improve forecast accuracy** by analyzing a broad spectrum of historical and live events to predict inventory or production outcomes
- **Increase operational efficiency** by analyzing and responding to inventory, prices, and logistic deliveries
- **Improve business agility** by identifying new opportunities or by price hedging using ML-powered applications

Additional Use Cases

There are other ways to make use of analyzed IoT data, including:

- **End-to-end visibility:** IoT in manufacturing can enable the monitoring of production lines starting from the refining process down to the packaging of final products. This complete monitoring of the process in real time provides insights that can be used to make adjustments in operations for better management of operational cost.
- **Improved safety:** Analysis of IoT data might be used to improve worker safety and security in the plant. By monitoring the key performance indicators of health and safety (like the number of injuries and illness rates, near-misses, and short- and long-term absences), incidents and property damage or loss during daily operations might be avoided.
- **Enhanced collaboration with partners:** By analyzing IoT-enabled machinery data, a manufacturer can share that operational information with partners like original equipment manufacturers, and with field engineers.

CHAPTER 5

Recommendations: How to Move to an Operational Analytics Data Architecture

Given the benefits cited in the previous chapter, it is obvious that companies need to support a broad set of analytic requirements for IoT applications, including:

1. High-performance transactional processing
2. Streaming analytics
3. Real-time insights
4. Machine learning

By selecting the right architecture—one that plugs into an existing infrastructure with minimal disruption— such benefits can be realized.

An ideal solution would be able to perform all workloads within a single database. This would allow an organization to store and process data in a single, operational analytics processing platform, enabling synergies across operational areas. Additionally, the right solution can help deliver real-time insights with high availability to data 24x7, and low-latency access to the data.

Moreover, a suitable operational analytics processing solution must avoid errors and problems that occur in other analysis systems where data must be extracted, transformed, and loaded (ETL) before it is analyzed. The right solution overcomes these problems by delivering a real-time, trusted view of critical data. This ensures that information is accurate and helps guarantee that the same data is used across the organization.

An additional issue to consider when selecting an operational analytics processing solution is the ability to run transactional and analytic workloads at the scale required by today's industrial IoT efforts.

A Modern Analytics Platform

New processing requirements and analysis workflows need a platform that can easily ramp up capacity as data volumes and streaming speeds increase. Legacy relational databases can't be scaled easily. Limitations are so severe that transaction processing is run as a batch process, with the database unavailable or degraded for queries during processing. A complex ETL process must then be run to ready data for analytics; and analytics, even against this stale data, are often not available in real time. The need for 24x7 data collection and monitoring can generate very large tables, resulting in slow-performing queries.

A common trade-off is to break up the data into separate silos or batch the data into low-cost data lakes—which requires the use of caching, an additional layer that demands expert, ongoing tuning. The manual work involved leads to query delays, inaccurate results that can ultimately cause customer attrition, and increased costs for performance workarounds.

NoSQL databases overcome the scalability barrier but are designed for unstructured or semi-structured data. So, they can't natively maintain consistency and proper join patterns for a structured data store, as required for efficient analytics. This makes NoSQL less than ideal to fully address the problems of legacy operational data for real-time or near-real-time decision-making at scale.

A New Generation of Database for a New Class of Operational Workloads

The only way to deliver true real-time decision making at organizational scale is to combine ingest, transactions, and analysis in a single, continually updated, fast database. This is where a new generation of database can help.

Specifically, scalability is addressed using new operational analytics processing solutions that are fully distributed and can run as a single database across a pool of machines while maintaining consistency and stability. This allows an organization to scale out. This new generation of database can speed up transaction processing, avoiding batching on ingest. It can also speed up analytics processing, making real-time analytics at scale possible. And it can combine both processes and enable true, end-to-end, real-time decision-making.

These are all areas where MemSQL excels. MemSQL is a distributed, highly scalable, SQL database that can run anywhere. It delivers maximum performance for transactional and analytical workloads with familiar relational data structures.

MemSQL ingests data continuously to perform operational analytics for critical industrial operational systems. It can ingest millions of events per day while simultaneously analyzing billions of rows of data in relational SQL, JSON, geospatial, and full-text search formats.

Organizations that use MemSQL get high ingestion performance at scale, and eliminate the need for costly data integration tools with built-in batch and real-time data pipelines. MemSQL lets organizations achieve ultra-fast query responses across both live and historical data using familiar ANSI SQL.

This highly scalable, distributed system balances data and queries across a cluster of commodity servers or cloud instances for maximum performance, concurrency, and availability. Infrastructure elements that enable these characteristics include:

- **Distributed storage** that allows for the processing of data across clusters of machines for maximum resilience and performance. Distributed storage helps an application more efficiently leverage system memory and disk infrastructure.
- **Massively parallel architecture** that provides a robust parallel execution engine for both updates and queries, delivering ultra-fast performance.
- **Big data capacity** that allows for the storage of petabytes of data on low-cost disk and cloud storage (for archive requirements) while maintaining instant retrieval for fast, deep analysis.
- **ANSI SQL support** so the widest possible range of ad hoc queries, BI tools, application interfaces, and programming language interfaces can be supported.

Combined, these capabilities allow MemSQL to deliver:

- Extremely fast performance
- Massive scalability
- An easy-to-use SQL architecture
- Cloud-native, able to run on any cloud infrastructure or commodity hardware
- Reasonable price, with strong price-performance

Companies can accelerate their industrial IoT analytics efforts by partnering with a technology provider like MemSQL. The capabilities offered by MemSQL give organizations employing industrial IoT and operational analytics the processing power and flexibility to improve operations, reduce downtime, and enact ML-enabled predictive maintenance programs. Organizations that undertake such analytics efforts can use the derived information to revamp business processes, increase product yields, improve product quality, and drive revenue and growth.

For more information on MemSQL's operational analytics processing solutions, visit memsql.com.