The Industrial Internet of Things Volume T3: Analytics Framework

IIC:PUB:T3:V1.00:PB:20171023

ACKNOWLEDGEMENTS

This document is a work product of the Industrial Internet Consortium Industrial Analytics Task Group, co-chaired by K. Eric Harper (ABB), Wael William Diab (Huawei), and William Sobel (VIMANA). Portions of this technical report are contributed from an IIC white paper: "Industrial Analytics – The Engine Driving the IIoT Revolution" (2017) [1].

EDITORS

The following persons (in alphabetical order) have edited the content in this document:

Diab, Wael William (Huawei)

Harper, K. Eric (ABB)

Lin, Shi-Wan (Thingswise, LLC)

Mellor, Stephen (IIC)

Waddell, Russel (AMT)

AUTHORS

The following persons (in alphabetical order) have authored the content in this document:

Anderson, Nina (AMT)

Diab, Wael William (Huawei)*

French, Terry (Denso Corporation)

Harper, K. Eric (ABB)*

Lin, Shi-Wan (Thingswise, LLC)*

Nair, Dinesh (National Instruments)*

Sobel, William (VIMANA)*

* Authored the original white paper and/or contributed substantive material to this framework

[1] Diab, Harper, Lin, Sobel, "Industrial Analytics: The Engine Driving the IIoT Revolution", March 2017, IIC, *http://www.iiconsortium.org/pdf/Industrial_Analytics-the_engine_driving_IIoT_revolution_20170321_FINAL.pdf*

CONTRIBUTORS

The following persons have contributed valuable ideas and feedback that have significantly improved the content and quality of this document:

Chatterjee, Chanchal (Dell EMC)

Guang, Liang (Huawei)

IIC ISSUE REPORTING

All IIC documents are subject to continuous review and improvement. As part of this process, we encourage readers to report any ambiguities, inconsistencies or inaccuracies they may find in this Document or other IIC materials by sending an email to *admin@iiconsortium.org*.

CONTENTS

FIGURES

TABLES

1 OVERVIEW

The Industrial Internet of Things (IIoT) seeks to connect industrial assets and machines—the things—to enterprise information systems, business processes and people who operate and use them. Advanced analytics is at the core of this next-generation level of integration and, when applied to machine and process data, provides new insights and intelligence to optimize decision-making significantly and enable intelligent operations leading to transformational business outcomes and social value. These new insights and intelligence can be applied across any level of any industry if the appropriate data can be collected and analytics applied correctly. If data is the new oil, data analytics is the new engine that propels the IIoT transformation.

1.1 INTRODUCTION

As a fledgling discipline combining advances in mathematics, computer science and engineering in the context of Information Technologies (IT) and Operational Technologies (OT) convergence, industrial analytics plays a crucial rule in the success of any IIoT system. Industrial analytics has unique requirements, characteristics and challenges compared to business analytics and so requires special considerations in its implementation. Since industrial analytics is in its early stage of development, much needs to be explored. This technical report describing an Industrial Internet Analytics Framework is intended to spur discussions and research, and speed up the development and maturity of this indispensable technology.

1.2 PURPOSE

This Industrial IoT Analytics Framework document provides guidance and assistance in the development, documentation, communication and deployment of Industrial Internet of Things Analytics Systems.

1.3 SCOPE

This document supplements the IIC Industrial Internet of Things Reference Architecture (IIRA)[1] by detailing Industrial Analytics crosscutting concerns. It provides the concepts and components required to create a viable analytical system and the characteristics of same so the technologies employed provide the necessary services to perform successfully and correctly in an industrial setting. This document presents an architectural framework of Industrial Internet of Things Analytics systems using the same approach as the IIRA, specifically the architecture viewpoints (business, usage, functional and implementation), which in turn was based on ISO/IEC/IEEE 42010:2011 architecture concepts.

1.4 STRUCTURE

This document is organized as follows:

[1] http://www.iiconsortium.org/IIRA.htm

- Chapter 2 – Business Viewpoint
- Chapter 3 – Usage Viewpoint
- Chapter 4 – Functional Viewpoint
- Chapter 5 – Implementation Viewpoint
- Chapter 6 – Artificial Intelligence and Big Data
- Chapter 7 – Analytics Methods And Modeling
- Chapter 8 – System Characteristics and Crosscutting Functions Related to Analytics
- Annex A – Example Analytics Standards

1.5 AUDIENCE

This document is intended for IIoT system architects and business leaders looking to successfully deploy industrial analytic systems.

1.6 USE

The Industrial Internet of Things Analytics Framework is intended as an architectural template for system architects to create a concrete architecture tailored for the requirements of a specific IIoT system; to assist in understanding and communicating the overall system among its stakeholders; and to implement the architecture to meet the unique system requirements.

1.7 TERMS AND DEFINITIONS

Since this document is intended to address both IT and OT, the following terms and definitions that are key to understanding this document are:

AI: Artificial Intelligence

Analytics: Analytics may be broadly defined as a discipline transforming data into information through systematic analysis

Architecture Framework: conventions, principles and practices for the description of architectures established within a specific domain of application and/or community of stakeholders1

ICT: Information and Communications Technology

Industrial Analytics: Industrial Analytics is the use of analytics in IIoT systems

IT: Information Technology

OT: Operational Technology

1.8 CONVENTIONS

Given that the document is non-normative, all 'must', 'may' and 'should' statements are to be interpreted as English language and not as in RFC 2119 [IETF-RFC2119].

[1] ISO/IEC/IEEE 42010:2011

1.8.1 Typographical and Linguistic Conventions and Style

Terms that require definition are rendered in *italics*. (As the usage immediately preceding demonstrates, italics may also be used as example, or for emphasis.)

Generally, only the first use of the term is italicized. However, when a term can be read in its usual English language mode, the first use of the term may be italicized as the discussion becomes technical. In the first example below, "safety" and "security" are used informally. In the second, it introduces a definition.

Example

> "Among the key system characteristics that must be considered, safety is perhaps the most important, followed by security."
>
> "*Safety* is the condition of the system operating without causing unacceptable risk of physical injury or damage to the health of people, either directly or indirectly, as a result of damage to property or to the environment."

1.9 Relationship with Other IIC Documents

This document fits into the IIC Technical Publication Organization as shown in Figure 1-1.

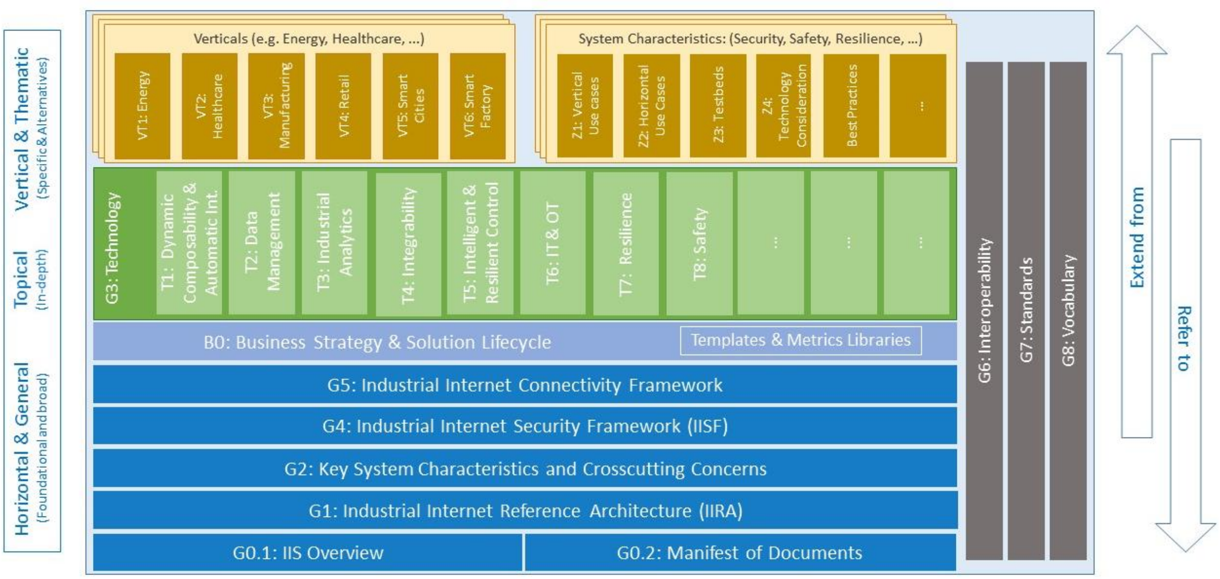


Figure 1-1 IIC Technical Publication Organization

2 BUSINESS VIEWPOINT

The business viewpoint attends to the concerns of the identification of stakeholders and their business vision, values and objectives in establishing an industrial analytics system in its business and regulatory context. It further identifies how the IIoT analytics system achieves the stated objectives through its mapping to fundamental system capabilities.

These concerns are business oriented and so of particular interest to business decision-makers, product managers and system engineers.

2.1 THE INDUSTRIAL INTERNET OF THINGS

The Industrial Internet of Things is a natural extension of the industrial and internet revolutions. IIoT will be a major force driving economic growth for the coming decades, at a greater pace than prior revolutions. As outlined by the World Economic Forum,[1] "The first Industrial Revolution used water and steam power to mechanize production. The Second used electric power to create mass production. The Third used electronics and information technology to automate production. Now, a Fourth Industrial Revolution is building on the Third, the digital revolution that is blurring the lines between the physical, digital and biological spheres."

To accelerate this digital revolution the Industrial Internet Consortium (IIC) is advancing the technology of IIoT across a diverse set of application domains. The industrial internet integrates the industrial assets and machines—the things—to enterprise information systems, business processes and people who operate or use them. With these connections to the industrial assets and machines, new technologies enable the application of advanced analytics to machine and operational process data to gain insights into the operations, optimize them intelligently to boost productivity, increase quality, reduce energy and material consumption, increase flexibility, and ultimately create new business values, while maintaining commitments to safety, reliability, resilience, security and data privacy as the trustworthiness of the systems, and conservation of the environment as social values.

Industrial analytics, applied to machine data for operational insights, is as an engine driving the convergence of OT and IT, and ultimately value creation for the Fourth Industrial Revolution.

2.2 CREATING BUSINESS VALUE

An enterprise needs to increase throughput, reduce expenses and inventory to generate higher margins and thus *create business value*. Sales drives manufacturing throughput, but it cannot exceed the enterprise's capacity to produce, lest it compromise on-time delivery and disappoint customers. One approach is to identify performance bottlenecks in overall operations continuously and remove them one-by-one to meet the demand from sales and profit targets.

[1] Schwab, K., "The Fourth Industrial Revolution: what it means, how to respond", World Economic Forum (2016).

This process needs to be grounded in reliable information about the demands, production, inventory and operational processes. IIoT and specifically industrial analytics enables stakeholders to gather data from machines and optimize processes more efficiently. The opportunity is substantial since over 86% of companies do not have a corporate analytics program that uses manufacturing data (LNS Research, 2016[1]).

According to the 2016 Global Manufacturing Competitiveness Index report by Deloitte Touche Tohmatsu Limited and the US Council on Competitiveness, Predictive Analytics and Smart Connected Products are #1 and #2, respectively (see Table 2-1). This shows that the desire to leverage analytics is a high priority for most CEOs. However, the awareness of how that technology will be applied and the areas of greatest impact are not well understood.

Advanced Manufacturing Technologies	US	China	Europe
Predictive analytics	1	1	4
Smart, connected products (IoT)	2	7	2
Advanced materials	3	4	5
Smart factories (IoT)	4	2	1
Digital design, simulation, and integration	5	5	3
High performance computing	6	3	7
Advanced robotics	7	8	6
Additive manufacturing (3D printing)	8	11	9
Open-source design/Direct customer input	9	10	10
Augmented reality (to improve quality, training, expert knowledge)	10	6	8
Augmented reality (to increase customer service & experience)	11	9	11

Table 2-1: Ranking of future importance of advanced manufacturing technologies by executives[1]

As our business systems and analytical capabilities increase, some enterprises will adopt data and analytical strategies; some will not. Enterprises that avail themselves of these technologies will far outpace the others in their ability to produce products and services quickly and safely, and maintain quality with full accountability for their processes. The cost of manufacturing will also drop, allowing developed countries to compete effectively in global markets because the value of skills and experience will more than compensate for the wage differential.

2.3 The Value of Industrial Analytics

Analytics may be broadly defined as a discipline transforming data into information through systematic analysis. Industrial analytics is the use of analytics in IIoT systems. It enables a better

[1] http://blog.lnsresearch.com/5-real-surprises-from-the-2016-metrics-that-matter-research-study

understanding of a system's operational states, performance and environment. It identifies and analyzes emerging information patterns to enable industrial system assessments under varied conditions. These assessments improve functionality and reduce inefficiency and operational cost. For example, utility companies may optimize electricity output based on dynamic usage patterns that factor in weather, season, events, pricing, resource availability, cost and electricity generation asset availability; support vehicle and equipment fleet management; optimize smart facility energy management and other unimagined capabilities. This is called *dynamic operations optimization*.

Industrial analytics can also optimize system missions. For example, metropolitan-area real-time traffic-pattern analysis combined with roadway conditions, roadway construction and maintenance, weather condition, time and day, seasons, accidents and other events can lead to vehicle control systems determining optimal routes to reduce travel time, congestion, pollution and energy consumption.

Industrial analytics can be applied to machine-streaming data received from disparate sources to detect, abstract, filter and aggregate event-patterns, and then to correlate and model them to detect event relationships, such as causality, membership, and timing characteristics. Identifying meaningful events and inferring patterns can suggest large and more complex correlations so that proper responses can be made to these events.

Industrial analytics can also be used to discover and communicate meaningful patterns in data and to predict outcomes. Traditional business analytics are typically applied only to business data to describe, predict and improve business performance.

To understand the value of industrial analytics, consider this example. In an industrial setting, a major cause for unplanned downtime and expense is machine outage. This translates to billions of dollars[1] lost in unplanned failures of equipment and unnecessary maintenance. Currently, most companies use scheduled maintenance plans for preventive maintenance. This means that maintenance is performed on machines even if they are sitting idle, costing time and resources, as well as requiring unnecessary disruptive procedures that can reduce equipment reliability. On the other hand, critical problems are often missed due to insufficient diagnostics resulting in unplanned downtime and often-costly repairs. Both the over- and under-maintenance of assets contribute to higher operating expense.

To address these issues, maintenance needs to shift to prognostics that will schedule maintenance based on component lifetime characteristics and its usage, rather than to a predetermined schedule. Next, the practice needs to become predictive where sensor and machine operational data are analyzed to forecast the likelihood of certain failures in a given period. With this information, machine maintenance can be scheduled optimally to avoid interruption to the operations and at lower cost. Analytics can also be applied into the control

[1] J. Manyika, et al, "The Internet of Things: Mapping the Value Beyond the Hype", McKinsey & Company (June 2015)

loop to adapt the machines to avoid operational conditions that can have detrimental effects on the machines and processes.

Industrial analytics plays a vital role providing crucial insights needed for decision-making and for optimal deployment of resources in industrial applications. Those capabilities in turn effectively translate to an increase in the efficiency of labor and capital. Long-term GDP growth is fundamentally dependent on capital, labor and efficiency. Increasing the efficiency of capital and labor is paramount for GDP growth and competitive advantages.

Business leaders have increasingly recognized the importance of industrial analytics. A recent survey (see Figure 2-1) carried out by IoT Analytics GmbH[1] found that 69% of the surveyed business leaders or industry analysts consider industrial analytics crucial for their businesses within 5 years and want to apply analytics to strengthen their businesses. The same survey found that analytics on physical objects and machines (highlighted in Figure 2-1) rank high in importance. Predictive and prescriptive maintenance of machines (at 79% of surveyed considering it extreme or very important) ranks at the top, slightly surpassing consumer and market related analytics (at 77%).

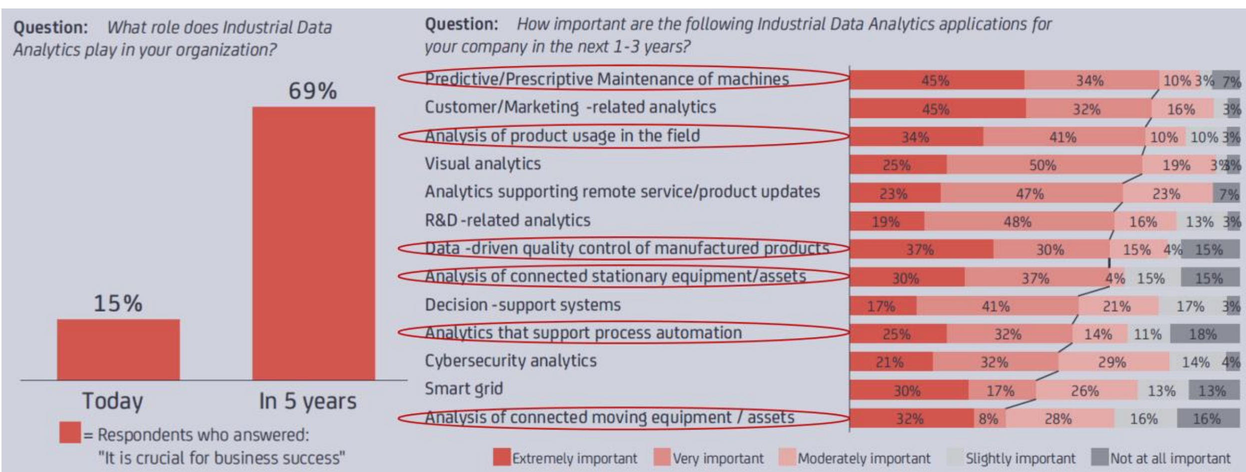


Figure 2-1 Importance of industrial analytics considered by business leaders and industry analysts

3 USAGE VIEWPOINT

The usage viewpoint addresses the concerns of expected system usage. It is typically represented as sequences of activities involving human or logical (i.e. system or system components) users that deliver the system's intended functionality.

System engineers, product managers and those involved in the specification of the analytics system representing the ultimate users are the typical stakeholders for these concerns.

[1] Knud Lasse Lueth, Christina Patsioura, Zana Diaz Williams and Zahra Zahedi Kermani, "INDUSTRIAL ANALYTICS 2016/2017", IoT Analytics GmbH, December, 2016

3.1 Analytics in Industry

Analytics as a discipline in applied mathematics has been widely adopted by organizations for decades, most prominently as business analysis in finance, banking and ecommerce. In industrial settings, analytics is currently used to identify and address potential faults in assets, improve uptime and reduce repair costs. This is called *condition-based monitoring*, or CBM. With the rapid development standards and innovations in sensor and computer technology, it is now possible to extend advanced analytics to large numbers of machines across the globe. Advanced analytics algorithms and techniques, including machine learning, are now used to analyze large amounts of data gathered from industrial control systems. The insights gained from the analytics can be applied automatically to increase operational efficiency of machines by, for example, anticipating peak usage, streamlining the supply chains for parts needed for preventative maintenance, and for business planning and decision-making. Using insights drawn from the machine data to drive intelligent operational and business processes, industrial analytics enables the convergence of analytics in the OT and IT worlds.

3.2 Getting Started with Industrial Analytics

Industrial analytics are used to identify and recognize machine operational and behavioral patterns, make fast and accurate predictions and aid in optimal decision making with greater confidence.

Analytics generally fall into three major categories:

Descriptive analytics gain insight from historical or current data streams including for status and usage monitoring, reporting, anomaly detection and diagnosis, model building or training.

Predictive analytics identify expected behaviors or outcomes based on predictive modeling using statistical and machine-learning techniques, e.g. capacity demand and usage prediction, material and energy consumption prediction, and component and system wear and fault predictions.

Prescriptive analytics uses the results from predictive analytics as guidance to recommend operating changes to optimize processes and to avoid failures and the associated downtime. An example of prescriptive analytics is on-demand production from a solid geometric assembly model to find the optimal set of manufacturing processes to achieve the final product.

The analytics results can be applied automatically to the machines and systems, or used to support human decisions through visualization of the analytics results to enhance human understanding and generate confidence in a decision.

Industrial analytics has unique challenges because the results can alter the operation and safety of things in the physical world. These effects may be undesirable or harmful, inadvertently affecting the safety of people or damaging property and the environment. Moreover, because industrial analytics often interpret data from different sensors and machines that may conflict with one another, we need to understand and synthesize the diverse information streams to reach a correct conclusion. Table 3-1 below, identifies the requirements to consider when planning for industrial analytics.

Topic	Description
Correctness	Industrial analytics must satisfy a higher level of accuracy in its analytic results. Any system that interprets and acts on the results must have safeguards against undesirable and unintended physical consequence.
Timing	Industrial analytics must satisfy certain hard deadline and synchronization requirements. Near instantaneous analytic results delivered within a deterministic time window are required for reliable and high quality actions in industrial operations.
Safety	When applying industrial analytics, and interpreting and acting on the result, strong safety requirements must be in place safeguarding the wellbeing of the workers, users and the environment.
Contextualized	The analysis of data within an industrial system is never done without the context in which the activity and observations occur. One cannot construct meaning unless a full understanding of the process that is being executed and the states of all the equipment and its peripherals are considered to derive the true meaning of the data and create actionable information.
Causal-oriented	Industrial operations deal with the physical world and industrial analytics needs to be validated with domain-specific subject matter expertise to model the complex and causal relationships in the data. The combination of first principles, e.g. physical modeling, along with other data science statistical and machine learning capabilities, is required in many industrial use cases in order to provide accurate analytics results.
Distributed	Many complex industrial systems have hierarchical tiers distributed across geographic areas. Each of these subsystems may have unique analytic requirements to support their operations. Therefore, industrial analytics must be tailored to meet the local requirements of the subsystems it supports. The requirements on timing (avoiding long latency) and resilience (avoiding widespread outage of service because of faults in the network or in a centralized system) require a distributed pattern of industrial analytics in that the analytic will be implemented close to the source of data it analyzes and to the target where its analytic outcome is needed.
Streaming	Industrial analytics can be continuous or batch processes. Because of continuous execution in industrial systems, a large proportion of industrial analytics will be streaming in nature, performing analysis of live data and providing continuous flow of analytics results in support of the operations. Traditional batch-oriented analytics will still be performed either for building or improving analytic models, or for human decision-making.
Automatic	For the industrial analytics to support continuous operations, the analysis of streaming data and the application of analytic outcomes must be automatic, dynamic and continuous. As the technologies in industrial analytics advance, improvements in analytic modeling e.g. through learning may also be automatic.
Semantics	Analytical systems require data that has meaning and context. Unstructured data, when reported without attribution to the source and the component or system it represents, makes deriving value complex since it requires the analytics to guess or infer the meaning. Inference unnecessary adds significant uncertainty into the system. Most data can be properly attributed at the source, and if this information is communicated, it can significantly increase the success and accuracy of the analytical systems.

Table 3-1: Industrial Analytics Requirements

4 Functional Viewpoint

The functional viewpoint focuses on the functional components in an industrial analytics system, their structure and interrelations and the relation and interactions of the system with external elements, to support the usages and activities of the overall system.

These concerns are of particular interest to system and component architects, developers and integrators.

4.1 ANALYTICS ARCHITECTURE OBJECTIVES AND CONSTRAINTS

Industrial analytics can be applied in different domains distributed across an IIoT system and tuned to varying time-scale horizons.

The required architecture can be analyzed in the context of the functional domains of the Industrial Internet Reference Architecture (IIRA) developed by the IIC, as shown in Figure 4-1 above. An end-to-end IIoT system in the IIRA is functionally decomposed into five functional domains:

- *Control*: sensing, communication, execution, motion, and actuation;
- *Operations*: provisioning, management, monitoring, diagnostics and optimization;
- *Information*: data fusion, transforming, persisting, modeling and analyzing;
- *Application*: logic, rules, integration, human interface; and
- *Business*: enterprise and human resources, customer relationships, assets, service lifecycle, billing and payment, work planning and scheduling.

The control domain is a collection of functions performed by the industrial assets or control systems, in exercising the closed control loops within themselves. The operations domain is a collection of functions for assets and control systems management and maintenance to ensure their continuing operations. The information domain is a collection of functions for collecting, transforming, analyzing data to acquire high-level intelligence of the entire system. The application domain is a collection of functions for applying use-case-specific logic, rules and models based on the information obtained from the information domain to achieve system-wide optimization of operations or other business objectives. The business domain is a collection of functions for integrating information across business systems and applications to achieve business objectives.

Industrial analytics results can be applied to the control domain at the edge providing real-time operational insights to the control loops in a *machine-time horizon* that typically requires analytic response in milliseconds or less. Examples include autonomous vehicles and robotics. Analytics in this time horizon tend to be streaming in nature and applied automatically. Industrial analytics results can be applied to the application and operations domains to provide machine insights that enable advanced maintenance such as automatic fault-detection and diagnosis, and preventive maintenance, or to drive optimal operations across fleets of machines or assets. The analytics result is applied in an *operation-time horizon* that typically requires an analytic response in the range of seconds or more. Analytics in this time horizon also tends to be streaming in nature and applied automatically. Industrial analytics results can also be applied to the business domain as well, providing insights to enable in intelligent business processes, including aiding business planning and engineering processes. The analytics result is applied in a *planning-time horizon* that typically requires an analytic response in the range of days or more. It consists of both streaming analytics results that are applied automatically (e.g. to work and machine part scheduling for on-site repair) and batch analytics results based on on-demand queries.

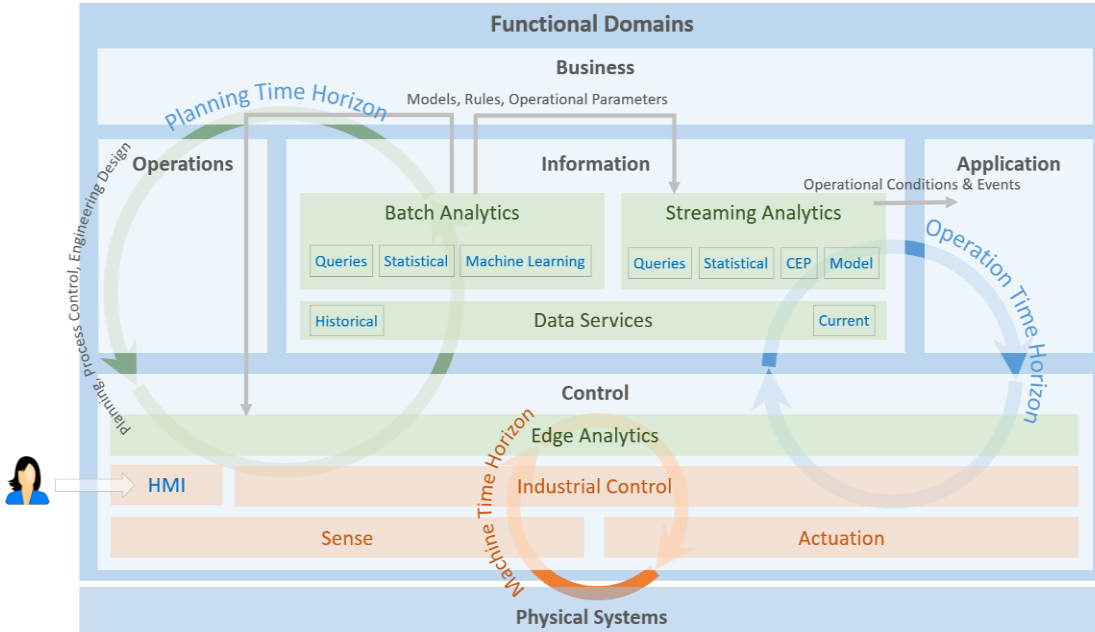


Figure 4-1. Analytics Mapping to the Industrial Internet Reference Architecture

4.2 Analytics Functionality

Industrial analytics functionality is deployable throughout the IIoT architecture. The capabilities needed for successful industrial analytics solutions are shown in Table 4-1, below. Each capability is realized by a set of functions defined by use cases that meet the stakeholders' expectations, especially with regard to non-functional requirements.

In a manufacturing use case, the plant operators visualize the conditions of the process line using graphical displays. When an alert appears, the operator drills down to the time-series records of the key parameters. Planning for the next run, the operator might use a similar drill-down to determine if there are any concerns that need to be addressed based on anticipated environmental conditions.

The fundamental prerequisite for industrial analytics is availability and access to data from the industrial process and related assets. Data is collected close to the process through connections and stored, at least temporarily, where the readings can be scanned and evaluated depending on the type of analytics. The stored values may be discarded or archived for further calculations. Data scientists can explore the archived data using statistics to compute correlations, and apply algorithms to classify and cluster the evidence over time. Industry subject matter experts have a good understanding of the context and condition of the process and assets, and can interpret and validate the readings and recommend cleansing filters. It is this combination of data science and subject matter expertise that produces the best results.

Capability	Description
Visualize	Display and manage data readings and analytics results using a common framework
Explore	*Ad hoc* experiments with historical data
Design	Automation of the data analytics stages; data quality, data mining, and business intelligence algorithm composition
Orchestrate	Delegate work requests over a cluster of computing resources, and collect and aggregate intermediate and final results
Connect	Exchange data and work requests between components using a common framework
Cleanse	Merge data sets from different data sources based on suitable criteria; remove irrelevant data and clean noise from data
Compute	Perform computation of statistical, first-principle and machine-learning model analytical calculations, including live analysis on streaming data, batch or ad hoc data mining and operation and business intelligence analysis
Validate	Ensure analytics results when applied in the context of the application and environment will not harm people or property. This function should be independent from the core analytics processing and act as a governor
Apply	Apply analytics results to various subsystems, including the automation systems (e.g. adjusting control parameters or models), operations and business processes, increasingly automatically or as information aiding human decision-making
Store	Archive and reproduce measured and calculated data streams, especially time-series sequences
Manage	Manage the information model, including data sources, computing resources and data analytics metadata
Supervise	Manage system reliability by ensuring processes are started and maintained, and that computer resources are not exhausted

Table 4-1: Industrial Analytics Capabilities

The industrial analytics workflow can be automated as the data and relationships are better understood. Given an appropriate framework, the automation is designed, configured and orchestrated to transform raw data into actionable results. The workflow and algorithmic content is versioned and deployed both on premise and in the cloud as needed to meet stakeholder expectations. The entire process is supervised to ensure all steps are completed and validated. Ideally, the industrial analytics solution evolves over time to produce better results using a minimum of resources, and to improve accuracy as more experience and historical data are acquired.

The final step is to communicate and present the industrial analytics results in a compelling and easily understandable format, including charts, graphs, and recommended actions. Most important is to provide means for humans to interact with the results, starting with a summary and allowing drill down into the evidence that supports the recommendations.

As analytics advances, more meaningful operational patterns, especially anomalies, will be detected, identified and reported as alerts, along with relevant supporting data, automatically. Root causes of faults can be automatically diagnosed and remedies or repairing actions can be prescribed; faults and failures can be prevented by ruling out improper operating parameters that are outside of the normal ranges and be predicted based on the historical experience of making similar parts. Machine operational efficiency can be monitored and optimized based on analytics results as well as orchestration of manufacturing resources and coordination and interaction of the prices of equipment and the people who operate them.

Clearly, all these will improve the operational efficiency of manufacturing and operations and at the same time reduce the stress on human operators in keeping the machines operating at their best capacity. On the other hand, analytics is no magic by itself—it requires a combination of obtaining the proper data at the proper time, applying the proper analytics algorithms and models that are guided by the necessary engineering domain knowledge from both the machine manufacturers, system integrators and the plant operators themselves. This is a process of continuing learning and improvement. For example, the types of data collected can be extended and their quality can be enhanced, the algorithms and models can be subjected to refinement and deeper domain knowledge can be injected into the models.

5 IMPLEMENTATION VIEWPOINT

The implementation viewpoint deals with the technologies needed to implement functional components (functional viewpoint), their communication schemes and their lifecycle procedures. These elements are coordinated by activities (usage viewpoint) and supportive of the system capabilities (business viewpoint).

These concerns are of particular interest to system and component architects, developers and integrators, and system operators.

5.1 DESIGN CONSIDERATIONS

To determine where analytics should be performed, the following considerations should be taken into account:

Scope: Ultimately, it is the derived information (not the raw data) and how it can be acted on that determines what kinds of analytics are deployed, and where. For instance, if the goal is to optimize machine uptime at one site, then analytics performed on data gathered there may be sufficient. In this case, the analytics can be performed anywhere, provided that, if done remotely, the normal local operation is not critically dependent on network latency and the availability of the analytics results. On the other hand, if the value proposition is to optimize production across sites requiring comparison of factory efficiencies then analytics needs to be performed on data gathered from these sites and thus be available in a higher tier of the system architecture.

Response time and reliability: In an industrial setting, some problems require deterministic analysis, computation and response; others can be done after the fact. The former almost always requires analytics to be local for reliability and performance.

Bandwidth: The total amount of data generated by these sensors, together with data gathered from the control systems can be huge. The increased network and infrastructure required to fuse data from one domain with others will be compensated by the creation of valuable insights.

Capacity: In some cases, it may be optimal to perform the analytics at a particular tier in a system architecture, but the existing infrastructure may not be able to support it, so a different tier is selected. Key properties of the ICT infrastructure include latency, bandwidth and computational capacity.

Security: The value from moving data must be balanced with concerns for transferring raw data outside of controlled areas, and the associated costs. It may be more efficient to perform some analytics locally and share necessary summary, redacted or anonymized information with other domains.

Volume: All data needs to be stored, at least temporarily. The storage required depends on the application.

Velocity: Industrial measurements are typically captured cyclically. High-frequency data, such as vibration or acoustics data from aircraft engines and wind turbines, can significantly increase the speed of data processing needed. Another consideration is transient events where readings must be captured with accurate time recording to determine order of occurrence, causality and root cause. With these low latency requirements, analytics using high velocity data is better performed close to where the data is measured.

Variety: When many pieces of equipment have been acquired over the years with dissimilar controls, interfaces and available data, effective analytics depends on shared information models to understand the data highly varying in both format (syntax) and content (semantics) to deliver the expected insights.

Analytics maturity: Analytics involves the processing of raw data (measurements) into descriptions (information) and then contextualizing the results (knowledge) and benefiting from historical experience (wisdom). The maturity of this process is not limited to where the analytics can be performed and as a design consideration is equally valid regardless of where the analytics are performed.

Temporal correlation: One of the common issues in IIoT systems is correlating data between multiple sensors and process control states, including the temporal order at which the data is generated. Applying analytics closer to where the data is generated reduces the burden of correlation when the analytics is applied.

Provenance: Performing analytics at a lower architecture tier maintains the genuine sources of the data as they progress through the IIoT system.

Compliance: To illustrate how compliance may impact the analytics as a design consideration, national security is used as an example. National security concerns may place restriction on the architectural decision about data management and sharing with government regulations in industries like aerospace and defense. This will influence where the analytics must be placed to meet the regulatory requirements, for example, possibly preventing performing large-scale computation in a public cloud facility to lower the cost.

Table 6-1 below is an example.

The combination of all these factors determines what capabilities are required and where the analytics will be deployed. Generally speaking, which analytics and where the processing is located depends on the maximum acceptable network latency and jitter in response to events, criticality of the analytics in the normal operations (e.g. how bad it would be if the external network is cut or an upper-tier system becomes unavailable) and the cost of uploading large amounts of data. From the deterministic response, reliability and resilience perspectives, it is optimal to perform analytics close to the sources of the data, and the analytics results need to be accessible for decision-making. In most systems, some form of hybrid approach with local and centralized analytics will be required.

Evaluation Criterion	Industrial Analytics Location		
	Plant	Enterprise	Cloud
Analysis Scope			
Single site optimization	✓	✓	✓
Multi-site comparison		✓	✓
Multi-customer benchmarking			✓
Results Response Time			
Control loop	✓		
Human decision	✓	✓	
Planning horizon	✓	✓	✓
Connectivity Reliability			
Site	✓		
Organization	✓	✓	
Global	✓	✓	✓
Connectivity Bandwidth			
Raw data	✓		
Processed results	✓	✓	
Summarized results	✓	✓	✓
Storage and Compute Capacity			
Server	✓	✓	✓
Multiple servers		✓	✓
Data center			✓
Data Security			
Secret	✓		
Proprietary	✓	✓	
Shared	✓	✓	✓
Data Characteristics			
Volume			✓
Velocity	✓		
Variety	✓	✓	✓
Analytics Maturity			
Descriptive	✓	✓	✓
Predictive	✓	✓	✓
Prescriptive	✓	✓	✓
Event Correlation			
Sub-seconds	✓		
Seconds	✓	✓	
Tens of seconds	✓	✓	✓
Data Provenance			
Sensor	✓		
Asset	✓	✓	
Site	✓	✓	✓
Regulatory Compliance			
Asset	✓	✓	✓
Process		✓	✓
Industry			✓

Table 6-1: Industrial Analytics Design Considerations

5.2 ANALYTICS CAPACITY CONSIDERATIONS

The functionality for industrial analytics is informed by experience and wisdom from both IT and OT. Both expect reliable operation and repeatable response times. However, the way these goals are achieved is different. Information technology relies on elasticity to provide the required capacity and operational technology ensures determinism with engineered capacity.

Elasticity is a cloud-technology measure of “the degree to which a system is able to adapt to workload changes by provisioning and de-provisioning resources in an autonomic manner, such that at each point in time the available resources match the current demand as closely as possible.”[1] For example, most retail businesses generate significant revenue during holiday seasons. All the IT systems need to be prepared and have enough capacity to avoid affecting profitability. The remainder of the year those resources can be scaled back or repurposed to reduce costs.

Determinism is the ability to support computation and transmission of data within a predetermined time between connected devices and applications. Deadlines must be met with expectations that a work request will complete within the same response time for every request. The analytics and results must be communicated in a defined time-period and confirmation must be provided. Operational systems are designed with a capacity for continual processing, regardless of the state or condition of the plant. So, when a production plant starts up or shuts down, which can generate rapidly changing values and multiple alarms, the requirement for deterministic response in the IIoT system is no different from when the plant is at steady state.

These two philosophies are in many ways complementary. Close to industrial processes, it is appropriate to provide reliability and predictability using dedicated resources. In the cloud with multiple tenants sharing the same set of resources dynamically, available capacity can be shifted to support different service level agreements. There are early signs that the manufacturing industry is moving in the direction of on-demand services to deliver just-in-time parts using shared assets. The World Economic Forum, in collaboration with Accenture, have predicted that this will become the long-term trend of manufacturing and there will be a convergence between the concerns between shared IT capacity on demand and manufacturing services.[2] In other industries, such as transportation and energy, there is movement towards services and shared capacity.

5.3 ANALYTICS DEPLOYMENT MODELS

The design and capacity considerations described above determine where the analytics will be deployed. Most IIoT systems use a hybrid approach to analytics deployment with analytics that

[1] Herbst, Nikolas Roman; Samuel Kounev; Ralf Reussner (2012). "Elasticity in Cloud Computing: What It Is, and What It Is Not" (PDF). Proceedings of the 10th International Conference on Autonomic Computing (ICAC 2013), San Jose, CA, June 24–28.

[2] O’Halloran, et. al., (2015). “Industrial Internet of Things: Unleashing the Potential of Connected Products and Services” (PDF). World Economic Forum in collaboration with Accenture.

need to be performed with very low latency and in a deterministic manner deployed closer to the I/O at the edge; predictions that are not time sensitive or require data from sources that are distributed or historical in nature are deployed in the cloud.

5.3.1 Types of Analytics

Figure 5-1 illustrates the different types of analytics that can be performed and the types of data required to perform the analytics.

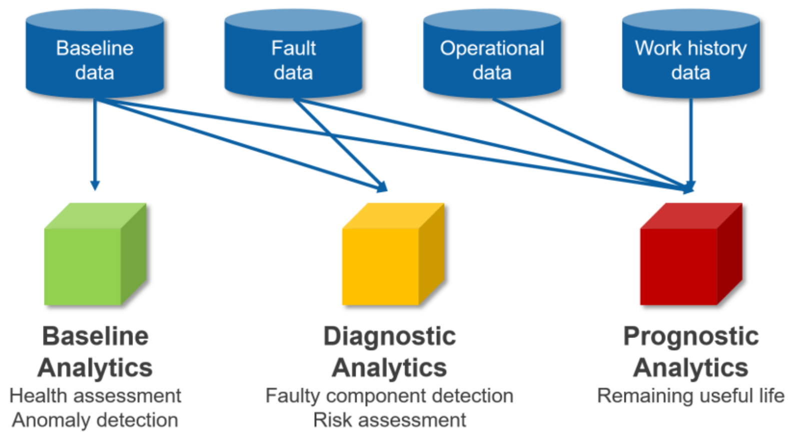


Figure 5-1 Types of analytics based on its applications

Baseline analytics detect irregular behavior of the asset within milliseconds of the actual event. The data used to perform these analytics is usually local to the asset under consideration and relies on data acquired from the asset when it was working normally.

Diagnostic analytics that identify the root cause of the anomaly such as a failing bearing in a motor requires previous knowledge of fault states. Diagnostics results can be returned in the order of minutes.

Prognostic analytics that tell you the remaining useful life of a bearing can take in the order of hours to return a result and requires access to multiple types of data and from multiple sources to make the prediction.

5.3.2 Deployment Location

Figure 5-2 is an example of where these analytics could be deployed.

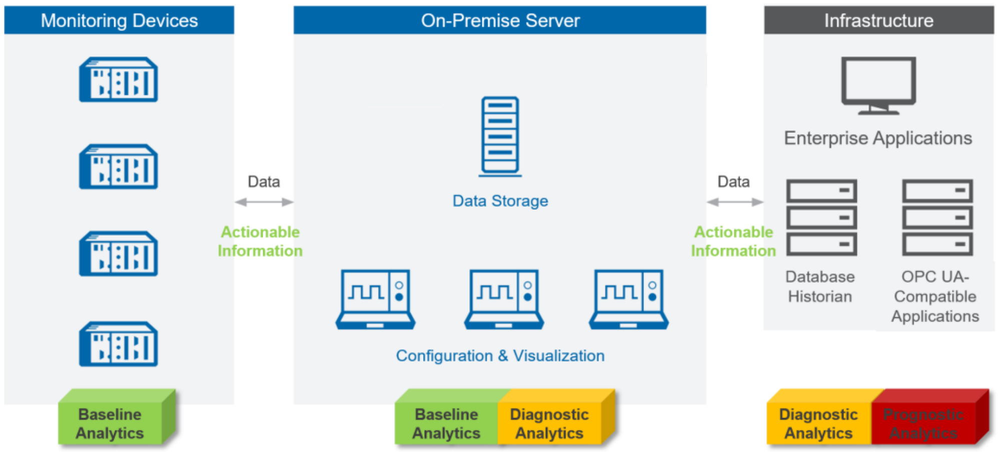


Figure 5-2 A deployment pattern of various type of analytics

5.3.3 Deployment Phase

Deployment of analytics typically consists of three steps:

- *train* a (predictive) analytics model,
- *test and validate* the model on previously unseen data and
- *deploy* the model to make predictions on real (streaming) data.

Training and deployment of analytics models can be broadly categorized into the following three workflows:

1. training of the model and the deployment (or inference) of the model is done in the cloud,
2. training of the model is done in the cloud, while the model is deployed at the edge and
3. training and deployment of the model is done at the edge.

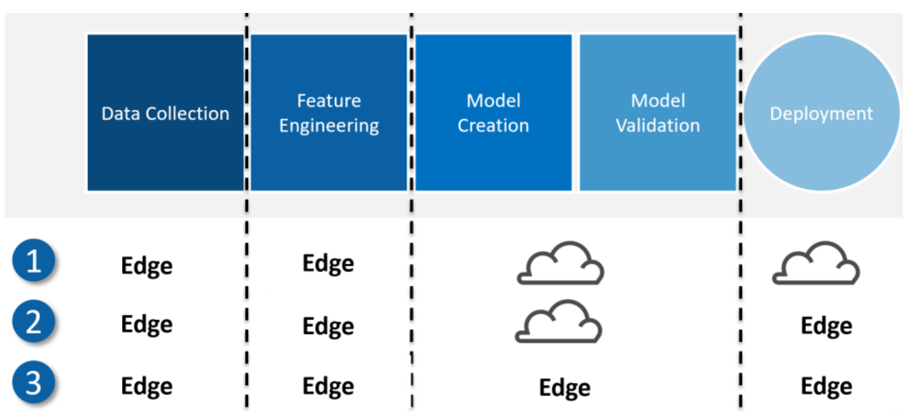


Figure 5-3 A deployment pattern of analytics during its different lifecycle stages

Figure 5-3 illustrates the three workflows for the training and deployment of analytic models.

For the first workflow, once the model is trained in the cloud and the deployment of the model is usually deployed as a web service. Data acquired from sensors at the edge is sent to the model via the web service. Predictions from the model can be returned to the edge through the web service if required.

In the second workflow, once the model is trained in the cloud the model is deployed to the edge. For deployment, the model is usually exchanged between the cloud and analytic algorithms running at the edge using a standard interchange formal such as PMML[1] or JSON[2].

Analytic models can be deployed at the edge on heterogeneous computing elements, such as a CPU, GPU or FPGA. For example, FPGAs are ideal deployment targets for very low latency applications.

5.4 Data Pre-Processing, Transformation and Curation

Data preparation accounts for about eighty percent of the work of data scientists.[3] "Messy data is by far the most time-consuming aspect of the typical data scientist's workflow". Exploratory analysis of raw data will often reveal distributions that are skewed, affecting how relationships in the data can be effectively characterized with machine learning classifiers. The relationships in the data may be inconclusive, however, transforming the records may clarify these relationships.

The same survey shows that data scientists spend 60% cleaning and organizing the data, and 19% of their time getting access to and collecting the records. Identifying the significant features of a data set is the first step to transform raw data into information. It is also beneficial to reason about relationships that may exist in the data. For example, if a feature is correlated with another variable in the dataset, converting it into a ratio that cancels the effect of that bias may improve the accuracy of the analytics results.

Several researchers go into more detail regarding their work[4] with big data. They recognize that data is imperfect due to missing values, inaccurate measurements, and too many independent variables. To address these challenges, techniques are used to simplify and clarify the dimensions. For example, the number of independent variables used in the analysis can be reduced using statistical techniques to identify correlations. Consulting with subject matter experts helps to understand why variables depend on one another, providing guidance which combinations of variables should be used as features to train machine learning models. Another technique is to reduce the number of records that need to be processed by classifying the states and maintaining the distributions of the readings. On the other hand, curated data can be used in artificial ways to reinforce the original measurements to fill regions of the domain of the

[1] http://dmg.org/pmml/v4-1/GeneralStructure.html

[2] *http://www.json.org/*

[3] "Data Science Report", CrowdFlower (2016).

[4] Garcia, S., Ramirez-Gallego, S., Luengo, J., Benitez, J.M., Herrera, F., "Big data processing: methods and prospects", Big Data Analytics, 1:9 (2016).

problem which have no representative examples in the original data. Finally, raw data can be discretized into buckets to reduce the noise of the measurements and reduce the complexity of the algorithms.

When reviewing the time dimension in historical data, a combination of transformations may need to be considered based on either: seasonality and trend patterns, or to stabilize the variance in the data. This reduces the impact of time in order for the data to be better statistically assessed. These transformations aid in generating forecasted values with greater accuracy. Transformations can be mathematical or based on adjustments such as using indexes to represent a current or scaled value of the series. Additionally, differencing is a type of transformation that adjusts for the seasonality and trend patterns to stabilize the mean of the series prior to using the data in certain time series algorithms. Differencing seasonality produces a representation of the current data with its corresponding data from the previous year.

These transformation approaches optimize how to measure information by selectively altering the shape of the distribution in a way that is still aligned with the broader business and strategic perspectives of each use case. The goal of the data transformation phase in machine learning is to simplify complexities which may exist in the data such that the information more appropriately falls within the parameters of the algorithms.

6 ARTIFICIAL INTELLIGENCE AND BIG DATA

Innovation in artificial intelligence and big data is expected to play an increasingly important role in industrial analytics.

This chapter looks at taxonomies of artificial intelligence and emerging computational techniques in big data in relation to industrial analytics.

6.1 BIG DATA ANALYTICS

Big data analytics cuts across IT (information technology) and OT (operational technology), data and roles. Big data requires computational systems and networks to be designed around the data. It will transform how businesses operate and the digital/physical divide.

There are numerous emerging definitions for big data. One definition is "data set(s) with characteristics (e.g. volume, velocity, variety, variability, veracity, etc.) that for a particular problem domain at a given point in time cannot be efficiently processed using current/existing/established/traditional technologies and techniques to extract value".[1]

Big data differs from traditional data storage and processing applications in five ways:

- *volume*: too big,
- *velocity*: arrives too fast,
- *variability*: changes too fast,
- *veracity*: contains too much noise and
- *variety*: too diverse.

Applications generating these data or requiring their analysis may have one or more of the above aspects present.

In addition to the "5Vs" above, the machine and operational data have their own features such as higher correlation, sensitivity to time order and historical context. Industrial big data are processed and analyzed for various application scenarios and purposes such as industrial automation, system health monitoring, predictive maintenance and remote operation.

To support these application scenarios, diverse big data analytics functions are performed, including but not limited to:

- *complex aggregation analysis*: to profile information of different time periods or locations,
- *multi-dimensional query and analysis*: to examine and deep-mine the machine data from different perspectives,
- *log data analysis*: to monitor system and operational health,
- *time-window based stream data analysis*: to identify temporal features and trends and

[1] ISO/IEC JTC 1 Big Data Preliminary Report 2014

- *complex event processing*: to detect patterns and anomalies.

As shown in Figure 6-1, the big data analytics platform need to address and support the processing of multi-typed input data from a large quantity of sensors or machines. To thoroughly analyze and mine the data (either real-time or historical) for value, diverse types of queries and analyses need to be applied. For in-time condition detection and decision making, these big data analyses need to be completed under throughput and latency requirements.

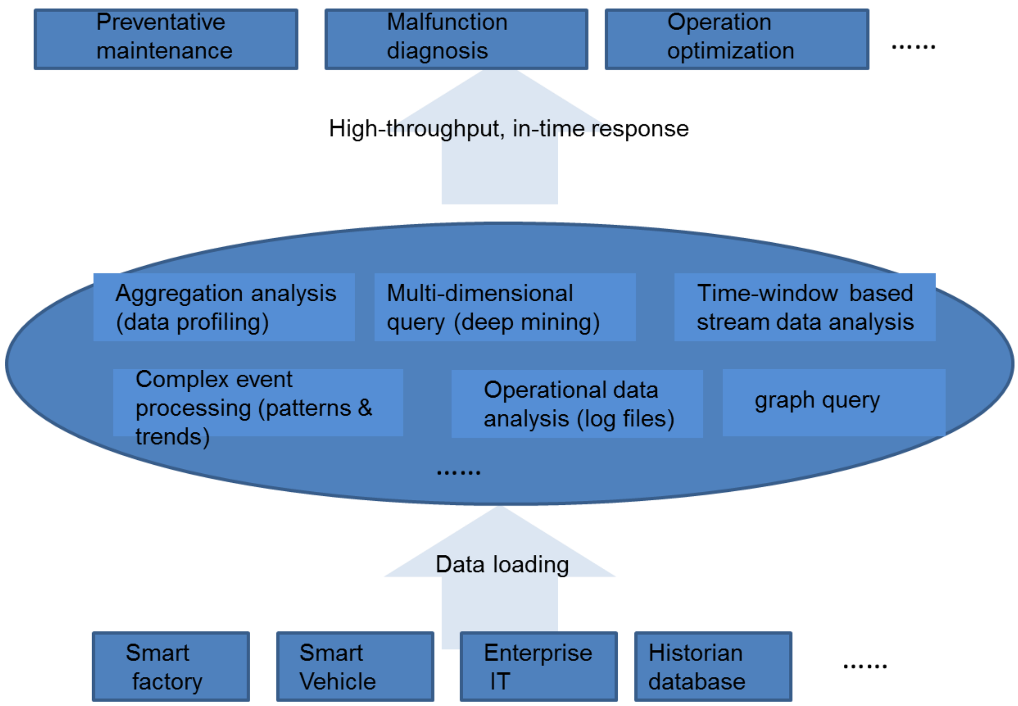


Figure 6-1 Example of Multi-Typed Data Processing in Big Data Analytic Systems

Analytics functions usually face stringent requirements in an industrial environment such as:

- high-performance in data loading,
- query and analysis,
- a single copy of input data for different types of analytics and
- fast response to concurrent queries and commands.

6.2 Artificial Intelligence

Artificial intelligence (AI) technology, with its rapid development, is increasingly used in industrial analytics to improve the analysis efficiency and accuracy.

The term *artificial intelligence* (AI) describes the analytics algorithms and frameworks used in IIoT. Two branches of AI are considered:

- machine learning (ML) and

- deep learning (DL).

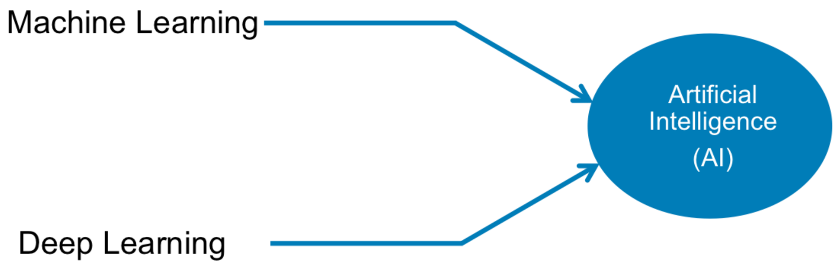


Figure 6-2 Artificial Intelligence (AI)

In IIoT applications, machine learning and deep learning provide new approaches to build complex models of a system or systems using a data-driven approach. Instead of using the physics-based models of the systems to describe the behavior of the system, ML and DL contain algorithms that can infer the function of the model from sample input data. These models are then used to make predictions on the state of the system. This is commonly referred to as predictive analytics.

6.2.1 Machine Learning

The workflow for developing and deploying models based on machine learning can be divided into a number of steps.

Figure 6-3 illustrates the steps involved in deploying a machine-learning model.

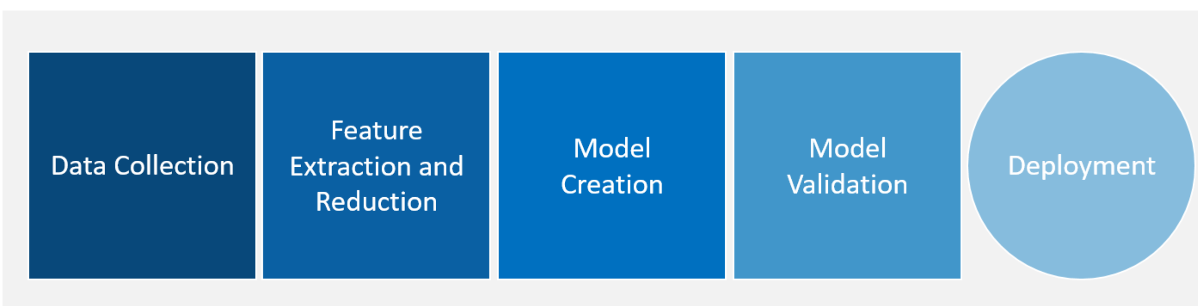


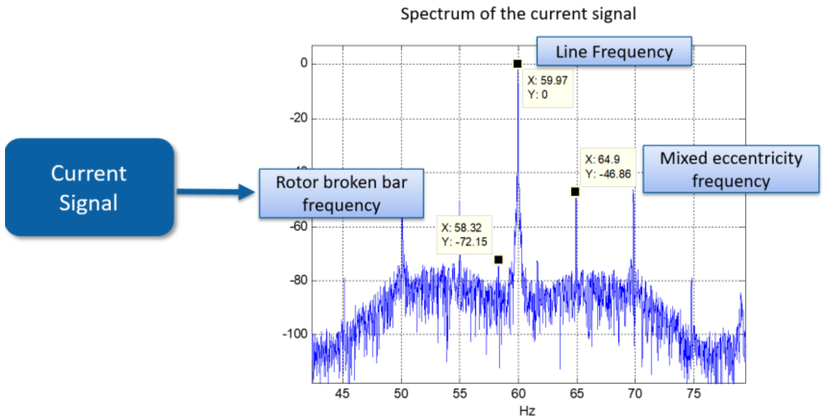
Figure 6-3 Workflow for developing analytic models

Data collection is the first step and involves data acquisition and data preparation.

Feature engineering extracts important information from the raw data that have been collected. This step usually comprises two parts: feature extraction and feature reduction.

Feature extraction is the process of converting raw data to information that relates to the physical state of source of the data (asset). Consider an analytics model that is created to monitor the health of a motor by measuring the motor's vibration using an accelerometer. Information in the frequency domain expresses more information about the state (normal and failure modes) of the motor than the time domain. Therefore, in the feature extraction step the waveform information is converted to the frequency domain and the amplitudes at certain harmonics are used as the features that are input to the model. In a typical application, data are collected from multiple sources and this may result in redundant information in the features that are extracted from the data.

The *feature reduction* step reduces the number of features by selecting a few features or transforming the features to a lower dimensional space without reducing the discriminatory properties of the features. The resulting reduced set of features is usually represented as a vector (or array) and input to the machine-learning algorithm in the next step.



Features	Significance
MCS ARMS	Max RMS of 3 phase current
Max Peak at (1+-2s)fs	Broken rotor bar
Max Peak at (1+-4s)fs	Broken rotor bar
Max Peak at fs +- fr	Shaft fault feature
Max Peak at fs +- 2fr	Shaft fault feature
RMS of the envelope signal	Energy after removing the line frequency
Amplitude of 2fs / DC value of Park's vector	Stator fault feature
RMS of AC level of Park's vector	Electrical failure feature

Figure 6-4 Mapping features to failure conditions for a motor

Figure 6-4 is an example of a feature-engineering step. It shows how the spectrum information of the current signal measured from a motor can be used to classify failure conditions in the motor. Figure 6-5 shows an example of a spectrum feature engineering process.

The analytics methods and algorithms used to create the models can be broadly classified in two categories: *supervised* and *unsupervised*. This classification is based on the nature of information contained in the input data that algorithms use to infer the function of the model. Both types need validation.

In supervised methods, the algorithm is presented with pairs of input data and desired output. This data is referred to as labeled data. The algorithm learns the function that maps the inputs to

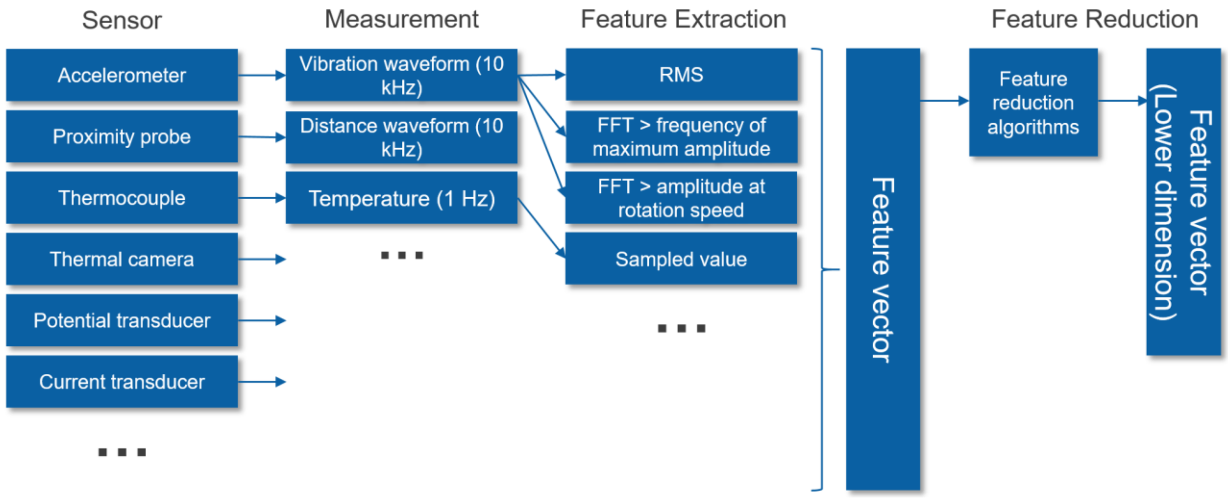


Figure 6-5 Spectrum feature engineering process

the outputs. Consider the example of a model that is created to monitor a water pump for a cavitation fault. The pump is instrumented with sensors that measure various parameters of the pump such as the horizontal and vertical vibration of the motor and the in-flow and out-flow pressures. The training data for a supervised algorithm will contain example features extracted from sensor measurements for both normal and fault condition of the pump, with each example clearly indicating the operating state.

Supervised algorithms are useful when it is feasible to acquire training data for the different states (or classes) that need to be modeled. The training data needs to be balanced, that is, there should be enough samples for each output state or class. In the pump example, supervised methods can be used when data for both the normal and the cavitation operating state is available. In the case, there are two classes. Supervised algorithms can be easily extended to learn multiple classes or states. For example, a supervised machine learning algorithm can be used to detect various fault conditions in a bearing such as outer race fault, inner race fault etc.

Unsupervised methods do not use labeled data. These algorithms find structure in the input data on its own. These methods are good for discovering hidden patterns in the data. Unsupervised methods are typically used when labeled data is not readily available; where the instances of normal behavior are more frequent than faults. A common use of unsupervised learning algorithms is for density estimation of a function. For example, an unsupervised learning algorithm such as the Gaussian Mixture Model (GMM) can be used to estimate the density function that represents the normal operating condition of an asset. In this case, the model is learned based on data acquired during the normal operation of the asset. Once the function has been learned, the GMM model can then be used to determine the health of the asset by computing the distance of the current estimate of the function based on the latest data from the center of the estimate of the normal (or base) function. This distance (or probability) is used to detect when the asset starts to deviate from normal operation.

Another commonly used category of unsupervised learning algorithms is clustering. Clustering group objects (input data) together in such a way that objects that belong to a group or cluster

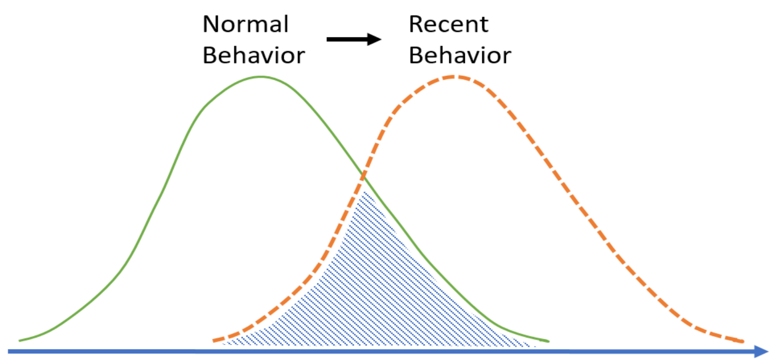


Figure 6-6 Health assessment using Gaussian Mixture Model

are more like each other than those in another cluster. We can use clustering where the asset may operate in different "normal" regimes (for example when the load on a motor may increase or decrease over time). In this case, clustering is used to detect different regimes and GMMs are used to estimate the normal behavior within each regime.

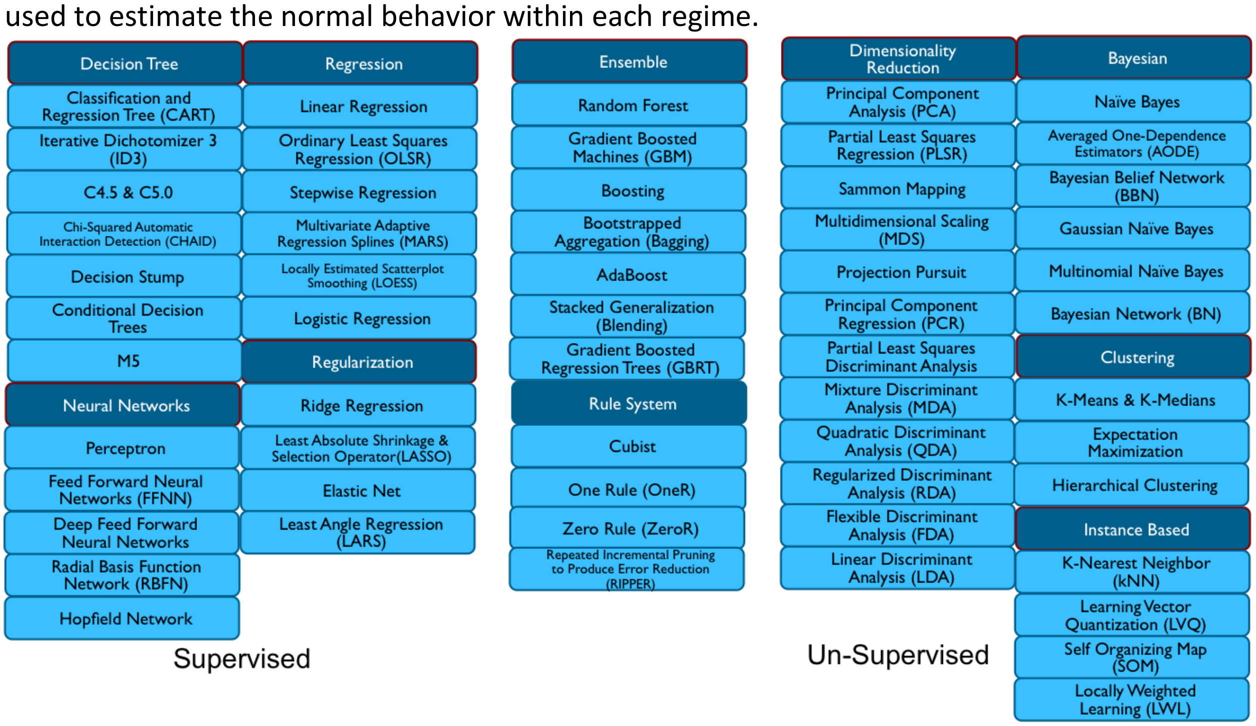


Figure 6-7 Taxonomy of machine-learning algorithms

The taxonomy of machine learning algorithms is given in Figure 6-7 for a brief overview.

One of the challenges with traditional machine-learning techniques is the need for features to be extracted from the raw data for creating good models. Feature extraction is usually a brittle process and requires the knowledge of a domain expert. It is usually the point of failure in the analytics workflow.

6.2.2 Deep Learning

Deep-learning algorithms have recently gained popularity because they remove the need for the feature-engineering step. Figure 6-8 illustrates the deep learning workflow.

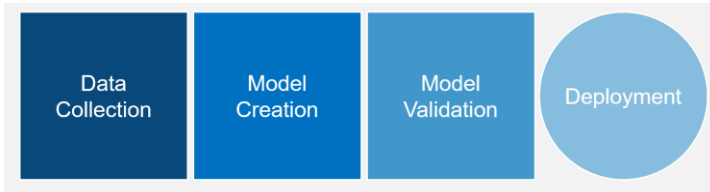


Figure 6-8 Deep learning workflow

Note that in contrast to the machine learning workflow, the feature-engineering step is not present. Data acquired from the sensors (raw measurements) can be directly input to the deep learning algorithms.

Deep learning algorithms are largely based on artificial neural networks (usually called *neural networks*). Artificial neural network learning algorithms are inspired by the structure and functional aspects of biological neural networks. These algorithms are structured in the form of an interconnected group of computational nodes (artificial neurons) organized in layers.

The first layer is the *input layer* that interfaces to the input signal or data. The last layer is the *output layer* and the neurons in this layer output the final prediction or decision. In between the input and the output layer there are one or more *hidden layers*. By using multiple hidden layers, deep learning algorithms learn the features that need to be extracted from the input data without the need to input the features to the learning algorithm explicitly as in the case of machine-learning algorithms. This is called *feature learning*.

Deep-learning algorithms require large amounts of training data and are computationally intensive. Deep learning has seen recent success in IIoT applications mainly due to the coming of age of technological components:

- advances in computation power with GPUs, FPGAs and CPUs,
- availability of large data repositories with labeled data to train these networks (e.g. ImageNet) and
- access to established open source deep learning software frameworks such as TensorFlow, Caffe, Theano, CNTK etc.

There are many deep-learning algorithms,[1] and the selection depends on the problem to be solved. The two most commonly used deep-network topologies for IIoT applications are convolutional and recurrent.

Convolutional Neural Networks (CNNs) are designed to take advantage of the structure of input signals such as an input image or a speech signal. A convolutional network comprises one or more

[1] http://www.asimovinstitute.org/neural-network-zoo/

convolutional layers (filtering layers) followed by a fully connected multilayer neural network. These networks are successful in problems such as image classification and object recognition.

Recurrent Neural Networks (RNNs) are based on algorithms that make use of sequential (or historical) information to make predictions. These networks are good for time-series analysis. A traditional neural network assumes that all inputs and outputs are independent of each other in time or order of arrival. RNNs on the other hand record state information that stores information about the past and uses the information calculated so far to make the next prediction. RNNs are good for learning historical behavior and predicting events in the future such as remaining useful life (RUL). The most commonly used type of RRN is the long short-term memory network.[1]

[1] https://en.wikipedia.org/wiki/Long_short-term_memory

7 Analytics Methods And Modeling

This chapter provides a survey of methods, models, algorithms and frameworks used for industrial analytics applications.

7.1 Analytic Methods and Algorithms

Streaming analytics is performed on data-in-motion to get high-throughput, low-latency actionable information so that industrial applications can respond to events in a timely manner. The speed layer, as shown in Figure 7-1, in the system performs fast, incremental algorithms on data as they are received and is usually performed close to the data sources to manage large data throughput, latency, reliability and security more effectively. For example, streaming analytics can be used to analyze wind turbine operations parameters (such as rotational speed, torque, blade pitch) and environmental data (such as wind speed and direction) to adjust the pitch values dynamically to maximize the electricity generation without causing damage to the turbines.

Batch analytics on the other hand is applied on all available data and aims to provide very accurate results, but with higher latency. This data can be historical in nature from a given source—months or years of data collected for a given wind turbine, and from sources that are spread out geographically—data from all wind turbines in a fleet. There are no restrictions on the type of computations that can be done in the batch layer, but they make take hours or days to complete. Batch analytics, usually performed in big data platforms, are widely used to identify and capture hidden patterns in the data set. Increasingly, it is used to build models, for example, to train and test deep learning models, before they are deployed back to streaming analytics environment to capture patterns in near-real time.

7.1.1 Streaming Real-time Analytics

The *lambda architecture* is a common pattern in industrial analytics that splits the data analysis into two separate concerns: *master/batch/serving* and *speed*. This section addresses the latter.

Stream analytics are not unique to the industrial internet. The financial industry used them for high-speed trading and arbitrage, and they have since spread to industry and media. Streaming analytics aggregate data sources comprising discrete events, digital, analog and waveform data. Each type of data requires special treatment and statistical processes to transform the data into a usable event stream that can be aggregated into higher semantic information models. Stream processing also takes repetitive data produced from sensors and control systems and reduces the repetition to make more efficient use of bandwidth. Statistical downsampling and outlier elimination are also used to increase the efficiency of connected devices where the raw data are not required for higher-level business systems.

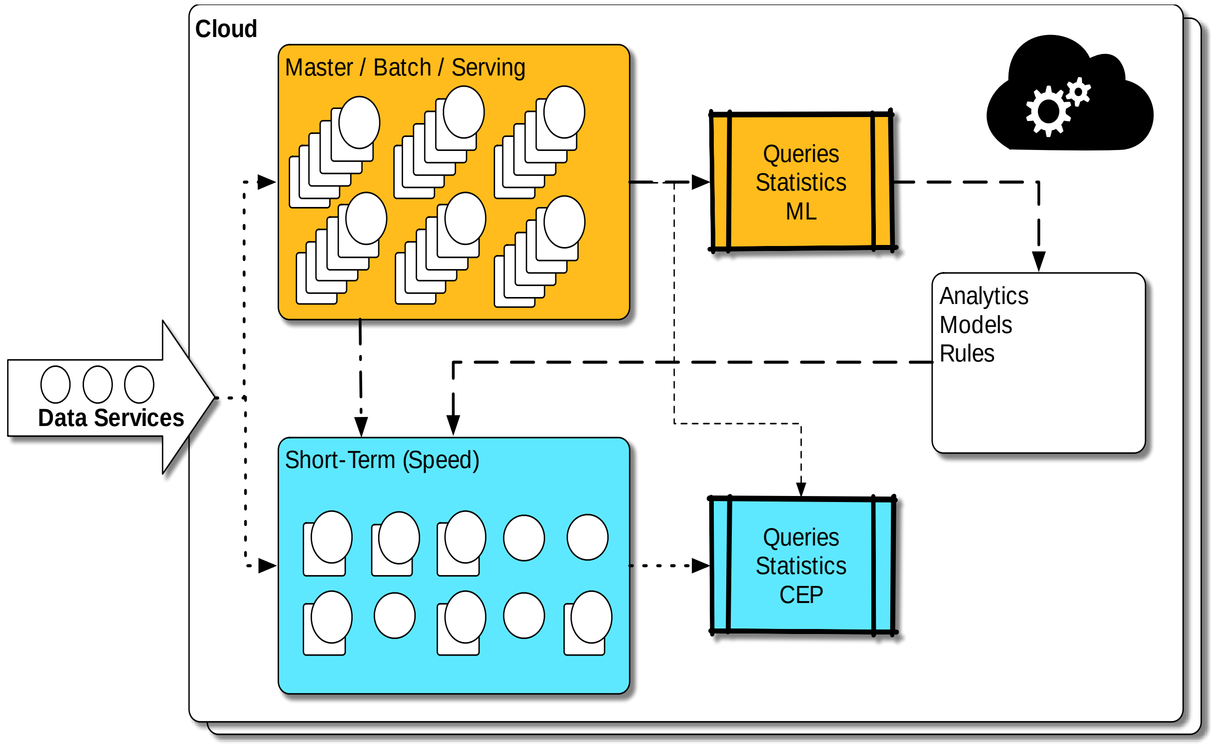


Figure 7-1 Lambda Architecture

One of the primary tools commonly used in stream analytics is *complex event processing (CEP).* CEP also started in the finance industry where high-speed analysis of market data and orders made deterministic statistical analytics a necessity and led to many of the tools that are now being applied to IIoT. CEP uses a combination of statistics and patterns expressed as rules to combine multiple event streams from multiple sources and create actionable higher-level events. The events can be re-introduced to the event stream and combined with additional data sources to provide context and infer higher-level understandings that are of greater value. CEP has been used in many production systems successfully and many open-source tools are available.

One of the main differences between stream analytics and batch analytics is the amount of data used for the analysis. Stream analytics assume that they will be operating on time series data-in-motion where the necessary reference data is known *a priori* and cached. Aside from the context of the data feeds coming into the analytical system, it is usually desirable to have minimal data access. The principal concern of stream analytics is fast computation of events where the intermediary computation and events are often kept in–memory and reference data is cached to reduce latency as much as possible.

In deterministic stream-based analytics, the end-to-end analysis and detection of actions must be completed within the window of time available to prevent or actuate the changes to ensure proper operation. For example, in CNC machining, CEP can be used to detect events like tool break in certain conditions by collecting contextualized data from the CNC controller and tool management system to determine when a certain set of events occur.[1] Rules of this type will need to be designed around the target process considering the material types and properties of the tooling and the machine. The action that occurs as a result of a rule can have direct effect on the equipment or may require a human to validate the inference. As the analytics become more reliable and the inferences are trusted, the human validation step can be removed and the analytics can directly change the operational state of the equipment.

7.1.1.1 CONCERNS ABOUT TIME

Time windows are a common construct of stream processing and CEP. They allow the application to apply a statistic (for example maxima, minima, average, counts) that spans an interval. All statistical operations in a stream analytics system have an implied time-period on which they act. Windowing also allows for the reduction of data that must be retained since only the necessary history is kept. For example, streaming analysis will capture data for a given period of time and then compute either at a given frequency (average over the last minute) or by using a rolling statistic (average every second using a one minute window).

Streaming data by its nature is temporal and must have a timestamp associated with each observation. The time stamp should be applied as close to the source as possible to ensure its accuracy. Causal relationships can be then be asserted across multiple data sources, which requires synchronization of clocks. In practice, assigning a timestamp at the time the data has been reported or observed is simplest for both statistical and time-series data sets.

The temporal dimension is a primary concern of stream processing. All the event-based patterns are built around time-dimensional, causal and contextual relationships between the data streams. The expression of the relationships between events and intervals allows the interpretation on whether an observation occurs before, after, within, or concurrently, thus constraining the analysis to “when” a certain state was observed or active.

7.1.1.2 STATISTICAL ANALYSIS

The first step in stream processing is to preprocess the data into a form that it can be analyzed. This usually requires various statistical- and signal-processing techniques such as Fast-Fourier-transformation (FFT) and high- and low-pass filters for analog data. Simple statistical methods, such as means and averages, are common as well as linear algebra kernel transformations for image and other complex data. The resulting data is a set of events and time series observations that can be used for pattern matching or higher-level analytics.

Deriving the initial transformation is often one of the harder tasks in stream processing. The development of the transformations is usually performed using batch systems to experiment with various methods on historical data to find the most effective method that will produce events that with highest certainty with respect to their use case. With many industrial use cases, there is substantial context that must be considered when performing the analysis. What may

[1] Such as a load less than 10 percent when the spindle is spinning and the machine is moving in a contouring type of motion for more than 30 seconds.

hold true when a physical system is in one state may not hold true when it is in another. This process is most often curated by domain experts who verify the behavior of the models.

7.1.1.3 EDGE, DATA AGGREGATION AND DOWNSAMPLING

In the cloud, analytics is usually a combination of both stream and batch, whereas at the edge, it is almost exclusively stream processing because of the constraints in computational capability, accessibility to data and the need for determinism. Common example use cases of edge streaming analytics include: downsampling of data, data aggregation, adaptive control, statistical analysis, process control, maintenance and error detection and correction.

By performing stream processing at the edge, the communications and computational overhead on upstream systems can be greatly reduced. Additional benefits may include the removal of unwanted detail for anonymity of the data and privacy of the supplier, balanced with the loss of detail in the data and the inability to reproduce a complete account of the events. It is difficult to balance the various concerns to collect just enough data to provide desired business value.

In many systems, there will be an initial dump of highly detailed raw data with no filtering and transformation. From this raw data, the analytical models will be developed and tested to ensure that the transformation are not too aggressive leading to missing critical events. When the models are sufficient for the purpose, they can be deployed in the streaming architecture. The models may not be complete, but they will be continually evolving and replaced as new discoveries arise. Most streaming platforms can deploy new models as needed in a continuous improvement cycle.

7.1.2 BATCH-ORIENTED ANALYTICS

Batch analytics can be used in multiple ways:

To improve the accuracy of the analytics (or models) deployed in the streaming layer: The analytics used in the streaming layer are sometime fast approximations of the exact analytics that need to be run on the data. This may be due to limited capacity available or the limited exposure to long-term trends in the data. The batch layer enables improvement of the model deployed in the streaming layer by periodically updating the model as more data are observed. This allows the approximate models running in the streaming layer to be corrected over time.

Perform analytics that require data from multiple sources that may be distributed geographically. In IIoT applications streaming analytics is usually performed at the edge, close to the physical assets. These analytics may run on an edge-computing node[1] that acquires data from sensors measuring information from one asset, such as a wind turbine or a single automated test system. The analytics running on an edge-computing node has limited exposure to data and state of other similar assets that are being monitored. This information on a class of assets can be useful in making more accurate predictions. For example, the anomaly detection analytics running on a wind turbine has limited knowledge on the current operating conditions of other turbines in a

[1] It in fact can be a controller, an IoT gateway, a server or even a cluster of servers, deployed in proximity to the assets.

fleet. By comparing the current state of a turbine to the others in a fleet, batch analytics can be used to detect true anomalous behavior of a turbine from operating conditions that may be affecting the entire fleet.

The master/batch/serving layer performs two functions. First, it manages an immutable master dataset that contains all the data that has been collected up to the current time. Second, it enables low latency analytics to perform on the entire master dataset on-demand. Instead of performing the analytics on the entire dataset when information is demanded through a query, the batch layer precomputes *batch views* of the data over time. The batch views can then be quickly queried to provide low-latency results. The master/batch/serving layer indexes the batch views to serve queries with low latency.

Here are some considerations when designing the master/batch/serving layer:

Data fidelity: In the batch layer, storing the data in the rawest form available is very valuable as keeping the raw data provides the flexibility to run new types of analytics and obtain new insights from the data. Storing only the preprocessed data (such as statistical features extracted from the data) limits the amount of information available from the data. The tradeoff is the cost of storage and data transportation.

Analytics: There are no restrictions on the type of machine learning or deep learning algorithms that can be performed in the batch layer. Deep learning frameworks rely on large amounts of data and are more effective to train on machines equipped with GPUs or FPGAs. There are a number of open source machine learning and deep learning tools that integrate well with open source big data platforms, the discussion of which is outside of the current scope of this document.

Reconciling speed and batch: Periodically, the results from the batch layer will override insights from the streaming layer. Given a query, the serving layer will merge real-time results with the batch views.

Drawbacks: The main drawback of the lambda architecture is the need to maintain two separate code paths and execution modules (speed and master/batch/serving). Emerging architectures unify both streaming and batch processing under a single execution model.

Figure 7-2 illustrates the application of the lambda architecture to a predictive maintenance application. The analytics (machine learning and deep learning) models are in the batch layer because of the time required to train these models, computational resources and the amount of historical data required for training. They are then run either in the speed or batch layer. The choice of the layer to deploy the analytics is based on criteria such as the maximum latency within which the predictions are expected, the sources of data (local or distributed) and security of the data. For example, event and anomaly detection analytics can run in the speed layer because of

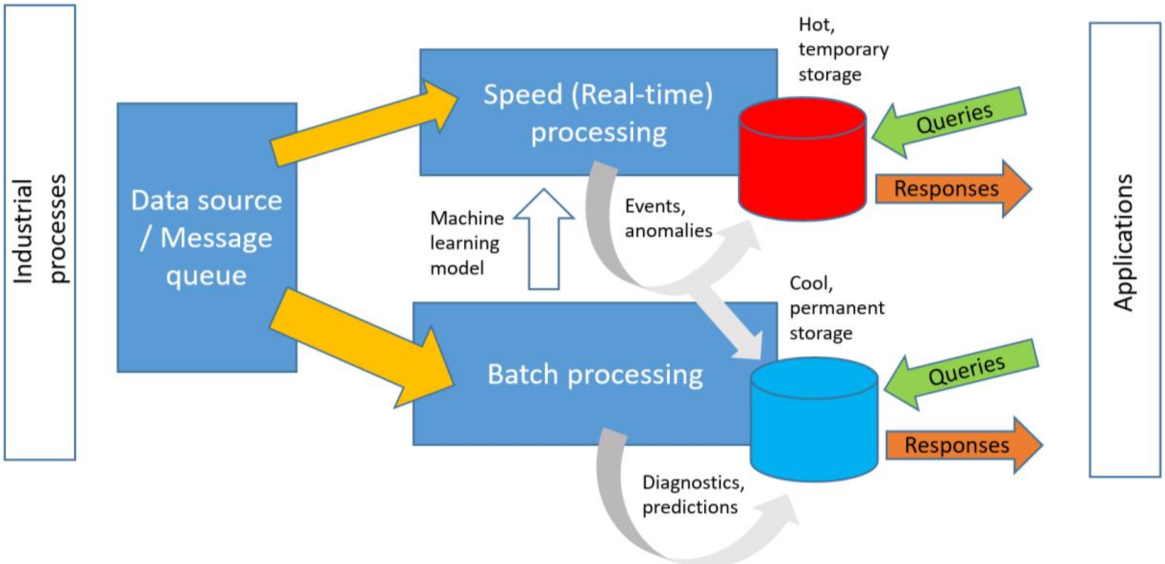


Figure 7-2 Lambda Architecture - Streaming and Batch for IIoT

the need to make predictions with low latency, while diagnostic and prognostic analytics can be run in the batch layer.

The *kappa architecture* is an emerging software architecture pattern that simplifies the lambda architecture. In this architecture, the batch layer is removed as shown in Figure 7-3.

The data is stored in an immutable, append-only log database using a distributed stream management system that start multiple stream processing jobs that can be used to combine speed processing as well as batch processing. For example, one job could be evaluating only current data and writing the output to the serving database. Another job could be looking at historical data in the stored log to get more accurate results.

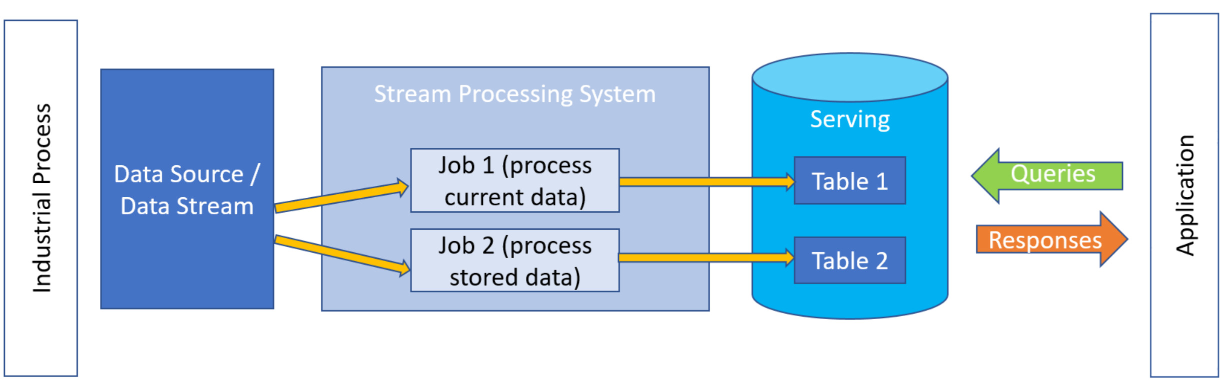


Figure 7-3 Kappa architecture

7.2 ANALYTICS MODEL BUILDING

The analytics process is shown in Figure 7-4. The first step is to collect and prepare the data by removing or filling in missing data, extracting features from the data that are relevant to the application and optionally reducing the number of features.

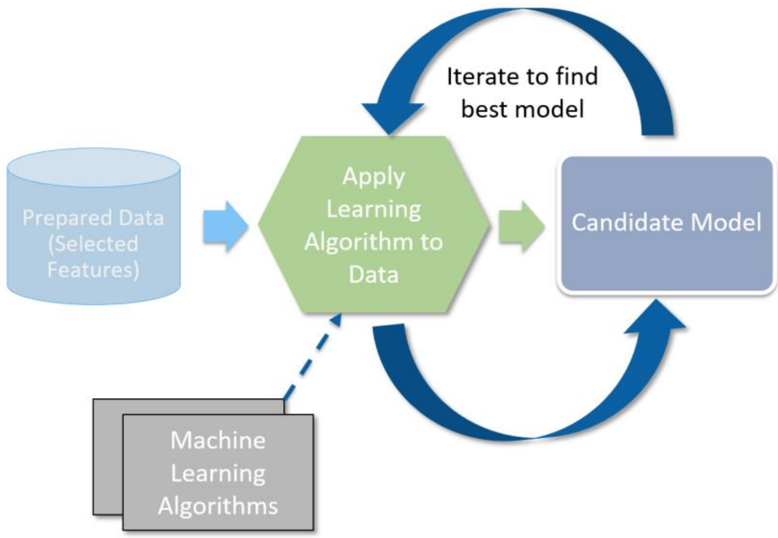


Figure 7-4 The model building process

7.2.1 ALGORITHM SELECTION

Analytics model building is next. Analytics models fall into three major categories: descriptive, predictive or prescriptive, as described in *Getting Started with Industrial Analytics* above. For each category, there a multiple machine learning or deep learning algorithms that can be used to build the model and the performance of these algorithms differ depending on the types of data being analyzed and the predictions to be made. So we should select a few algorithms in each analytics category and then iterate through them to identify the one that works the best for the current data and application. Table 7-1 shows some machine-learning algorithms that can be used to build analytics models, classified based on the major categories. For example, an anomaly detection model can be built using a one-class support vector machine (SVM) or a Gaussian Mixture Model (GMM).

Along with algorithm selection there are two other import aspects of the model building process. First is on how the data is used to train the model. The second is how to compare the performance of these algorithms and select the best one.

Algorithms		
Anomaly Detection (Baseline)	**Classification (Diagnostic)**	**Regression (Predictive)**
One-Class SVM	Neural Networks	ARMA
PCA-based	Support Vector Machine	Linear Regression
Gaussian Mixture Model (GMM)	Decision Forest	NN Regression
Logistic Regression	Bayes Classifier	Bayes Regression

Table 7-1 Machine-learning algorithm options for building analytic models.

7.2.2 Cross-Validation

To be able to test the performance of a model, we need two sets of data. One set is for training and the other is for validating the performance of the model on data it has not seen before.

Using the same data for training and testing the model can result in the model "memorizing" or *overfitting* the data. Overfitting manifests itself as a model that performs with high accuracy during training but poorly on new data. Cross-validation mitigates this issue.

With *cross-validation* the available dataset is split into training and testing data as shown in Figure 7-5. During cross-validataion, the master dataset is randomly divided into *k* folds (subsets) of equal size. *k-1* of these subsets are used to train the model and remaining fold is used to test the model. This process is then repeated *k* times. The model's performance is then the average of all *k* test sets. A split of 70% (training) to 30% (testing) is typical.

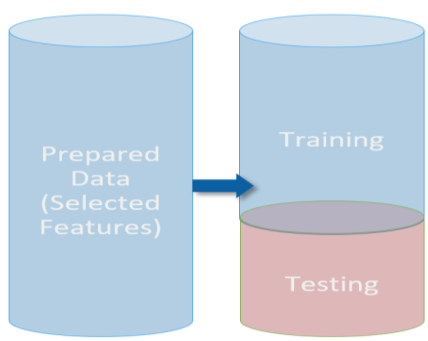


Figure 7-5 Splitting data for cross validation

7.2.3 Performance metrics

For supervised learning algorithms, the performance of the model is usually measured by looking at the classification errors. Because the training and testing datasets are labeled data, we know what the expected (or true) output is. This information can be used to bin the predictions results into different categories as shown in Figure 7-6, which shows the types of error for a binary classification. For an anomaly detection model condition positive is the detection of a failure or anomaly.

		Predicted Condition	
	Total Population	Predicted Condition Negative	Predicted Condition Positive
True Condition	Condition Negative	True Negative	False Positive (Type I Error)
	Condition Positive	False Negative (Type II Error)	True Positive

Figure 7-6 Confusion matrix showing types of classification errors for a binary classification problem

8 SYSTEM CHARACTERISTICS AND CROSSCUTTING FUNCTIONS RELATED TO ANALYTICS

The business, usage, functional and implementation viewpoints facilitate a systematic way to identify IIoT analytics system concerns and their stakeholders, to bring similar or related concerns together so they can be analyzed and addressed effectively. The deliberation of the concerns is often performed within each of the viewpoints to which they belong, but they should not be resolved in isolation to those in other viewpoints.

Industrial analytics requires many services from components of an industrial system, including connectivity and data management. The closest dependency will be the data management system that provides the raw material for the analytics. This *Industrial Analytics Framework* provides business and market drivers with guidelines in respect to the contexts and correctness, but stops short of addressing the ownership and chain of custody (data management) issues required. These concerns will be fully addressed in other documents. This framework provides a set of dependencies and requirements for both the collection, storage, and communication of data to other parts of the industrial processes.

8.1 SAFETY

The industrial internet has the potential to deliver value, but also to cause unintended loss. Unlike traditional information technology, the unintended harm from industrial analytics could result in human injury or loss of life. The mitigation of this risk is to design industrial analytics processes and computations to prevent unintended operation and independently validate that the resulting actions do not harm life or property.

The first line of defense comes with the security expectations for authentication and authorization. Role-based administration and control prevents an unauthorized actor from manipulating industrial analytics data or processes. The second line of defense is provided by the data management expectations for information models, reducing the unintended consequences of *ad hoc* configurations of system components. A common representation for types and their attributes enables simpler industrial analytics application design and realizations that business owners can verify and validate. The third line of defense comes with the connectivity expectations for reliable data synchronization between tiers. Missing or incomplete data sets could contribute to out-of-bound industrial analytics calculations. The fourth line of defense is to post-process industrial analytics results to confirm they are in the expected range, and engage a human in the loop to review anomalies.

8.2 SECURITY

The industrial internet is only as secure as its weakest link and should provide defense in depth so that if a malicious or unintended action compromises one security or accountability measure then another measure still guards the assets. On the other hand, there are architectural quality penalties such as reduced performance and scalability if security is the only consideration. The compromise is to secure only the sensitive (as defined by the business) data and govern who has access to critical resources. This approach increases the operational overhead to identify the sensitive content and manage it separately from everything else. These security mechanisms should be common across tiers and leverage a federated identity infrastructure for ownership, authentication and authorization.

First, the data management components should encrypt sensitive data-at-rest, and the connectivity protocols should perform the same functions for data-in-motion. Encryption is used both within a tier and for synchronization between tiers to prevent improper access. Only authorized clients, with appropriate credentials, should be able access the data based on established policies and protocols. Second, stakeholders can configure security domains to protect and manage access to industrial analytics processes and data based on defined attributes and scopes. The owner configures this role-based administrative control for authenticated clients, defining the policies by which sharing is allowed. This capability protects intellectual property and sensitive information. These policies synchronize across tiers so transfer of data does not implicitly cause transfer of ownership.

8.3 DATA MANAGEMENT

The industrial internet creates business value for a wide range of industrial processes and assets. Each asset has a set of static and dynamic characteristics tracked over time to gain insight, e.g. from historical records captured in the form of data, both raw and computed. The curated data (i.e. those stored in an historian) should be common across tiers and accessible using a federated information model that supports search, classification and markup to enable rapid industrial analytics application development.

First, a best practice for industrial analytics is to collect and store data in their rawest form, because the data-cleaning process to curate the data could remove the most important anomaly needed at a later date. But it is inefficient to clean data repeatedly for each computation if the results of that process are the same. The compromise is to store the data twice, once in its raw form and again in its curated form, updated periodically. The industrial analytics expectation for this curated form is common across the tiers. For example, the rules and controls must apply to the data uniformly. Second, asset types and instances are crucial aspects of the industrial ecosystem: discoverable, navigable and organized independent of naming conventions. It is unrealistic that all industrial internet applications will agree on a common taxonomy and attributes, but the mechanisms for organizing them can be common. Classification and markup of types in this information model apply to related instances and property values. The expectation is that federation of multiple information models within a tier provide a broad view of the available storage. Third, semantics must be maintained for the information to continue to be meaningful. Proper context of the data increases the success and accuracy of the analytics systems. Fourth, data lineage (provenance) provides considerable value to analytics as it supplies attribution of the data sources. For example, this lineage allows for quicker debugging and for tracking error propagation.

8.4 Connectivity

The industrial internet is a distributed architecture by design. This is necessary because, unlike traditional cloud computing (for example associated with online retail), the system produces raw data geographically separate from where the resulting information delivers business value. A distributed architecture requires connectivity between components, not only between collocated processes but also across wide-area and global networks.

First, connectivity within a tier is expected to be more *reliable* than communications across tiers. It is more likely that applications deployed within a tier, including local access to data and processing, will deliver valuable results compared to applications depending on resources across multiple tiers. The industrial analytics expectation is that there needs to be a capability for efficient and reliable synchronization of data between tiers when connectivity is available.

Second, IIoT systems typically incorporate multiple vendors in a deployment, potentially with *ad hoc* connectivity relationships and more than one communication protocol. There should be mechanisms for industrial analytics components within a tier to *synchronize* with each other without needing to know the deployment configuration or end-to-end protocols.

Third, *response time* is expected to be quicker within a tier compared to across tiers. If the analytics need to produce value close to real-time, then the expectation is that they need the connectivity at that tier.

Part I: Annexes

Annex A EXAMPLE ANALYTICS STANDARDS

This annex provides an example of an analytics standards ecosystem using the manufacturing vertical as an example.

ANNEX A SECTION 1 STANDARDS ECOSYSTEM

There are three primary categories of standards relating to industrial analytics that range from cross-domain to domain-specific: standards that address cross-domain systems engineering and simulation, those that provide domain specific performance metrics, and those that specify operational analytics for asset maintenance and function. There are other standards that will affect the analytics, such as design standards that are used to model the form and engineering function of the product and can be used throughout the lifecycle to provide a digital surrogate.

Only the first category, system-engineering standards, can be applied to almost any domain. The most prevalent standard in this category is SysML that is developed, owned and maintained by OMG. SysML provides a general-purpose language for systems engineering and a framework for expressing the specifications, analysis, design, verification and validation of engineering models (Figure 8-1). It was designed to address the concerns of model exchange between software

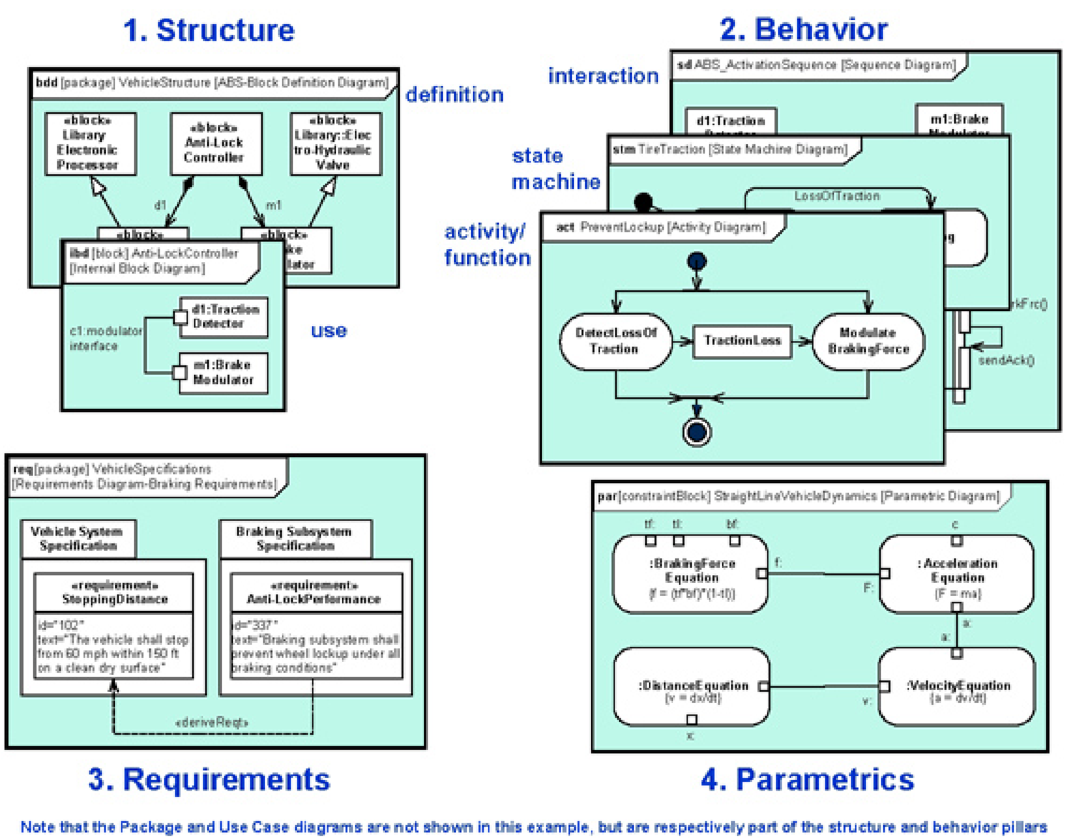


Figure 8-1 SysML for specifications, analysis, design, verification and validation of engineering models

systems and is the most widely adopted standard to date. There are other standards that are more domain-specific, such as STEP (ISO 10303 – AP233) in the manufacturing domain, but AP233 has been largely subsumed into SysML.

Beyond the engineering model exchange standards, the remaining standards tend to be focused on domain-specific concerns. In many domains these include specifications for operational metrics that are expressed in a set of objective and subjective formulae that can be applied to data from given systems to provide a business value metric. Since the business value metrics are all domain specific, the formulae are domain specific as well. These metrics can be applied at various degrees of specificity narrowly within a given domain.

For example, the OEE metric defined the Overall Equipment Effectiveness metric for manufacturing equipment comprises three components: the availability or utilization of the equipment, the performance of the equipment, and the quality of the parts. The components are then multiplied to provide a singular value that is a rough indicator of a piece of equipment's production capability. The standard is used across continuous process and discrete manufacturing, but in many cases, it fails to provide actionable information in discrete manufacturing. This is an example that even domain standards will have certain use cases within those domains where they perform better.

There are many standards in manufacturing that provide efficiency and effectiveness metrics that span from production performance to quality. These standards are often not known beyond practitioners in the domain and often have little applicability outside the confines of that domain. Therefore, the consideration of standards must evaluate the ecosystem in which the application lives and review the available standards within that ecosystem.

It is useful to evaluate other domains to see if there are developments within that domain that are applicable in other domains. For example, there may be standards or methodologies in maintenance that apply to manufacturing equipment that can be used to develop standards in transportation. There are similar ideas in each, both have actuators and mechanical components but the operational metrics and the failure scenarios are often different. An aircraft engine will have a different set of operational metrics than a machine tool since an engine failure is a more serious incident than a scrapped part—in most cases.

ANNEX A SECTION 1, SUBSECTION 1 MANUFACTURING STANDARDS ECOSYSTEM

Manufacturing has a numerous standards that relate to analytical methodologies, and four areas where standards apply: geometric design, system engineering and simulation, operational efficiency, measurement and verification, and equipment maintenance. Geometric design has been primarily represented by the STEP (ISO 10303) standards and the current revision, AP242ed2, takes the basic core geometric representation given in AP203 and adds product information and geometric tolerance.

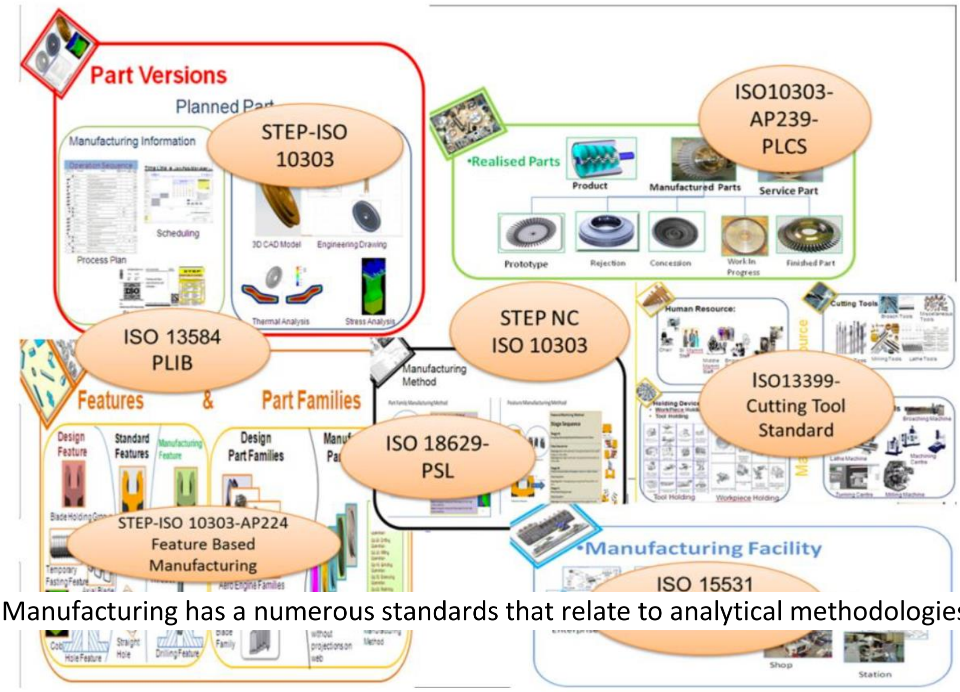


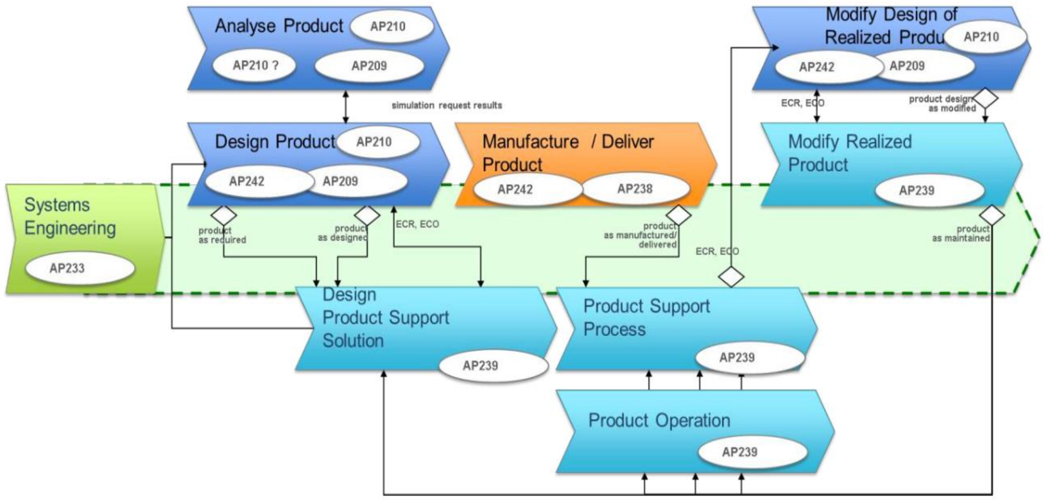
Figure 8-2 Manufacturing has a numerous standards that relate to analytical methodologies

Figure 8-2 is an illustrative example of some of the numerous standards in the manufacturing space.

Geometric design also needs to be coupled with engineering models to express the requirements of the product and simulate its function. That is where standards like SysML are applied to provide the foundation for communicating the functional information necessary to develop digital surrogates of the final product. The question may be raised about how this applies to the industrial internet that is primarily concerned with data from devices. The reason these models, often first-principles based, are necessary is they provide the initial intent from which the empirical data can be compared and the models refined.

Figure 8-3 illustrates geometric design related standards within the context of manufacturing IIoT.

Figure 8-3 Geometric design related standards



The standards ecosystem extends through the production of the product and standards that provide measurements of quality such as the Risk Priority Number (RPN) that evaluates the importance, stability, and the detectability of defects in a manufacturing process to standards that provide reports of the products lifecycle such as STEP AP239. There are multiple standards that also constrain manufacturing processes and material like ASME and ASTM that provide the information required to validate the correctness of a part and a product.

Analytics require the expression of the semantic intent of the product as well as the as produced and as built instances of the design. The models need to be communicated between systems using standards and open technologies.

The existing cross-domain and domain-specific standards as described above will help the semantic understanding for analytics, new standards with common abstract modeling of parts, products, equipment and processes may need to be developed to enable better communication of and greater degree of semantic understanding of them and their interactions.

Annex B REVISION HISTORY

Revision	Date	Editor	Changes Made
V1.00	2017-09-14	Diab, Wael William (Huawei) Harper, K. Eric (ABB) Lin, Shi-Wan (Thingswise, LLC) .	Initial Release

Annex C ACRONYMS

IETF	Internet Engineering Task Force
IIC	Industrial Internet Consortium
IIoT	Industrial Internet of Things
IIRA	Industrial Internet of Things Reference Architecture

Annex D GLOSSARY

Industrial Internet Consortium (IIC)

an open membership, international not-for-profit consortium that is setting the architectural framework and direction for the Industrial Internet. Founded by AT&T, Cisco, GE, IBM and Intel in March 2014, the consortium's mission is to coordinate vast ecosystem initiatives to connect and integrate objects with people, processes and data using common architectures, interoperability and open standards.

Industrial Internet of Things (IIoT)

describes systems that connects and integrates industrial control systems with enterprise systems, business processes, and analytics.
Note 1: Industrial control systems contain sensors and actuators.
Note 2: Typically, these are large and complicated system.

Internet Engineering Task Force (IETF)

The Internet Engineering Task Force (IETF) develops and promotes voluntary *Internet standards*, in particular the standards that comprise the *Internet protocol suite* (TCP/IP). It is an open *standards organization*, with no formal membership or membership requirements.

Annex E References

[IIC-IIRA2015] IIC: The Industrial Internet, Volume G1: Reference Architecture Technical Report, version 1.7, 2015-June-04
http://www.iiconsortium.org/IIRA.htm

[IIC-IIV2015] IIC: The Industrial Internet, Volume G8: Vocabulary Technical Report, version 1.0, 2015-May-07
http://www.iiconsortium.org/vocab/index.htm

[IETF-RFC2119] IETF, Bradner, S.: Key words for use in RFCs to Indicate Requirement Levels, 1997. Accessed 2016-06-29.
http://ietf.org/rfc/rfc2119.txt

Index

USE OF INFORMATION—TERMS, CONDITIONS AND NOTICES

TRADEMARKS