

WHITE PAPER

Use IoT To Advance Railway Predictive Maintenance

By Hitachi Vantara

Antonio Lugara, Industrial IoT and Transportation SME (Subject Matter Expert)

June 2018

Contents

- Executive Summary..... 3**
- A Predictive Maintenance Framework..... 4**
- Predictive Maintenance in the Railway Sector 4**
 - Maintenance Strategies 5
 - Approches to Railway Predictive Maintenance 6
 - An Example..... 7
- Effective Railway Predictive Maintenance 8**
 - Identify Prediction Viability and Effectiveness 8
 - Extract the Right Data..... 10
 - Let the Domain Expert Influence the Data Analytics 11
 - Identify the Achievable Value-Added Outcomes 11
- From Data to Insights..... 12**
 - Data Acquisition 12
 - Data Transformation 12
 - Data Evaluation 12
 - Data Visualization 14
- A Robust IoT Platform..... 15**
 - An Integrated Ecosystem 15
 - Hitachi’s Lumada and Railway Predictive Maintenance 16
 - Concepts 18
 - The Lumada Platform 18
- Customization Capabilities and Railway Predictive Maintenance 19**
 - Predictive Maintenance and Smart Manufacturing 20
 - Mathematical Methods..... 21
 - In-Memory Database 22
- The Business Side of Railway Predictive Maintenance..... 23**
 - A Financial Perspective..... 23

Business Partnership Approach	25
Conclusion	25
Acknowledgements	26

Executive Summary

The internet of things (IoT) is made up of billions of smart devices, such as cameras, sensors and mobile devices, all capable of wirelessly communicating with each other and with us. According to various estimates, there are already about 20 billion internet-connected devices or “things” in the world, and by 2020, one estimate says that this figure will exceed 50 billion. Of these devices, connected transportation vehicles are the third fastest growth category, behind smartphones and tablet PCs. These vehicle connections are generating data that is used for everything from diagnosing engine problems to monitoring cargo loads.

The availability of new technologies and the huge amounts of data they deliver are the key factors able to revolutionize maintenance for transportation vehicles in the 21st century. Vertically integrated development of IP smart sensors, computational performances, and big data analytics frameworks are making transportation such as rail more punctual, cost-efficient and safer.

By combining operational technology (OT) with information technology (IT), the conditions are right for a new framework. In this new approach, all data output from operational devices is collected, stored, normalized and analyzed through effective algorithms based on inferential statistics, machine learning and artificial intelligence.

This white paper introduces predictive maintenance in the smart rail sector, emphasizing railway engineering elements, IT and data-mining aspects, and the business benefits you can reap through this innovative framework. Within the fourth industrial revolution, the information extracted by data is the new currency.

A Predictive Maintenance Framework

It's no surprise that organizations are constantly looking at their operations for ways to reduce costs. All of these businesses are challenged by global supply chains, aging assets, and price volatility of raw materials, increasing compliance requirements, and dealing with an aging workforce. Due to strong competition in an increasingly globalized marketplace, organizations need to maximize asset productivity and ensure that associated processes are as efficient as possible, resulting in strong financial returns.

The development and application of a predictive maintenance framework¹ can help organizations achieve these results. This new paradigm is pushed by the availability of large amounts of data from instrumented and connected assets; requirements to do more with less (for example, stretching the useful life of an asset); reduced costs of computing, network and storage; and the convergence of information technology (IT) with operational technology (OT). Predictive maintenance, intersecting with IT and OT, helps organizations gain key insights into asset failure and product quality, enabling them to optimize their assets, processes and employees' activities. Predictive maintenance is the "killer app" that helps businesses compete against a globalized, high-pressure marketplace.

Advantages

Major advantages of predictive maintenance include:

- Optimize maintenance intervals.
- Minimize unplanned downtime.
- Uncover in-depth root cause analysis of failures.
- Enhance equipment and process diagnostics capabilities.
- Determine optimal corrective action procedures.
- Reduce and optimize inventory costs.

Predictive Maintenance in the Railway Sector

The internet of things (IoT), pushed by technological progress and cost reductions, is starting to impact public transport in a big way. With millions of data points captured and transmitted from sensors on critical train components, analytics can monitor the degradation of parts and detect impending parts' failures. In this way organizations can ensure that maintenance is performed when required. The benefit of the ongoing analysis of predictive maintenance is that the maintenance is "right-time," occurring well before a fault but not unnecessarily early, so the lifespan of the part is optimized. A good analytics package includes the ability to distinguish between maintenance data that is critical and requires immediate action, and data that is informative but does not indicate the need for action.

When you can forecast which parts are likely to fail in the near future, you can also achieve close to 100% uptime. You can fix impending faults when units are out of service according to an efficient planning routine, avoiding downtime and untimely breakages. An outage on a rail line during peak commute hours can mean disruption on the local network for several hours, with thousands of productive hours lost. Minimizing unplanned rolling stock outages through railway predictive maintenance (RPM) is fundamental to ensure stability and reliability throughout transport networks.

Sensors that are available today can be retrofit on existing fleets, where they can collect trillions of data per year, and use that data to develop deep engineering knowledge. Data analytics capabilities allow humans (and, soon, automated maintenance systems) to predict component failures and carry out root cause analysis, enabling a

¹ Levitt, J., *Complete Guide to Preventive and Predictive Maintenance, Second Edition* (New York: Industrial Press, 2011).

continuously improving process. This allows tailored maintenance planning, improved availability and reduced overall maintenance costs.

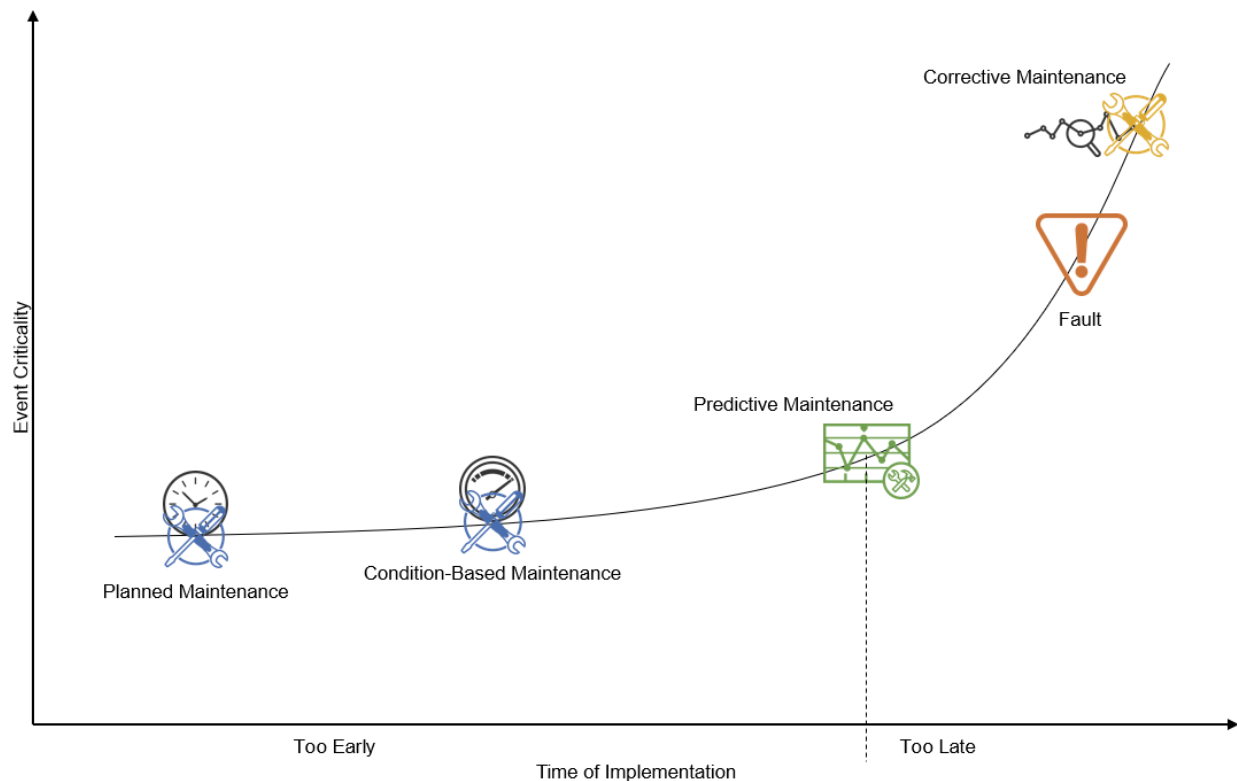
Railway operators and leasing firms expect the virtually fault-free operation of rolling stocks during their service life of 30-40 years. From a railway service provider's perspective, reliability and maximum availability are critically important to ensure the cost-efficient operation of rail vehicles and the infrastructure they use. Predictive maintenance allows them to create the conditions to improve punctuality, reliability and satisfaction. Due to budget cuts and spare parts and train reductions, operators demand availability that's higher than 99% from the rail industry to avoid downtimes that lead to both direct costs (for example, corrective maintenance) and indirect costs (for example, compensation claims).

To achieve these results and prepare for the future, it is fundamental to switch from diagnostic tasks performed under human supervision with long-term experience to a new paradigm. This new approach can guarantee repeatability of results, record the processes and experiences, and transfer knowledge efficiently, avoiding the inevitable human errors.

Maintenance Strategies

A key factor to consider is time. That is, consider the time interval between individual maintenance operations and the time available before a fault occurs. You can significantly increase the availability of systems while reducing maintenance costs if you can make timely predictions of the maintenance needs of systems. Figure 1 shows the most common maintenance strategies.

Figure 1. Common Maintenance Strategies



The most expensive form of maintenance is corrective maintenance. It is performed after a fault has occurred, resulting in the need for a backup transport service to be organized and dispatched as soon as possible. A maintenance team must be dispatched to tow the train to the closest depot. Furthermore, the length of the track

blocked by the broken-down train causes delays for other trains on the same line. All these factors lead to very high costs and fees from the incurred service delay.

The costly inefficiencies associated with corrective maintenance have given rise to a call for change aimed at preventing the fault from occurring through better preemptive technologies. The most common preventive maintenance approach is to perform planned maintenance in which maintenance schedules are established according to the part manufacturers' schedules, recommended mileage-based maintenance, and other operational observations. While this method is an improvement over corrective maintenance, it often leads to unnecessary maintenance and premature parts' replacement. These scenarios are usually unavoidable because replacements are based on a one-size-fits-all schedule and are unrelated to the actual condition of the parts.

Condition-based maintenance is an improvement over, and logical extension of, planned maintenance. Performing direct measurements and estimations about the real conditions of parts based on their effective rate of usage can increase performance and reduce costs. This form of preventative maintenance is more proactive, looking to extend the life of parts by replacing them only when certain conditions are met.

Building on this knowledge and moving toward a more effective outcome leads a predictive framework to estimate the time when a fault is likely to occur and to adopt maintenance interventions accordingly. Recent advancements in smart sensors and IT have led to continuous data collection from various systems and subsystems in trains, enabling monitoring of mechanical and electrical conditions, operational efficiency and many other performance indicators. These new capabilities enable planning of maintenance activities with the maximum interval between repairs, while minimizing the number and the costs of unscheduled outages created by system failures.

Over time, the railway sector has been progressing through the maintenance strategies shown in Figure 1. The change from planned maintenance to conditions-based maintenance has been ongoing. The addition of predictive maintenance will minimize not just the maintenance costs of the train, but also the loss of revenues due to unplanned downtime.

Approaches to Railway Predictive Maintenance

Railway predictive maintenance or RPM can be performed following two different approaches:

- **Knowledge-based:** this approach considers competencies and know-how acquired by designers and maintainers. It also uses Failure Mode, Effects and Criticality Analysis (FMECA) and Reliability, Availability, Maintainability and Safety (RAMS) analysis. By combining those analyses with the knowledge acquired through experience, it is possible to identify deductively the train's abnormal behaviors based on known thresholds of relevant variables. The data is sampled at planned intervals and compared with the thresholds; when the detected values overcome the expected ones, alarms and triggers are sent to the appropriate stakeholders.
- **Data-driven:** driven by asset digitalization, maintenance engineering has an increasing volume of multisource and heterogeneous data. Often, however, those data are collected in different databases, creating silos that do not allow easy comparative analyses. To solve this problem, it is possible to use distributed file systems and big data platforms that allow the creation of heterogeneous and statistically representative data lakes containing structured, semistructured and unstructured data. These data are analyzable in their entirety through a holistic approach, applying artificial intelligence techniques, machine learning and predictive analytics. This approach has created the conditions to identify the relationships between apparently independent data, and extract insights. It can observe devices' previously unknown behavioral patterns and dependencies, which are now usable to predict abnormal behaviors and reduce performance issues.

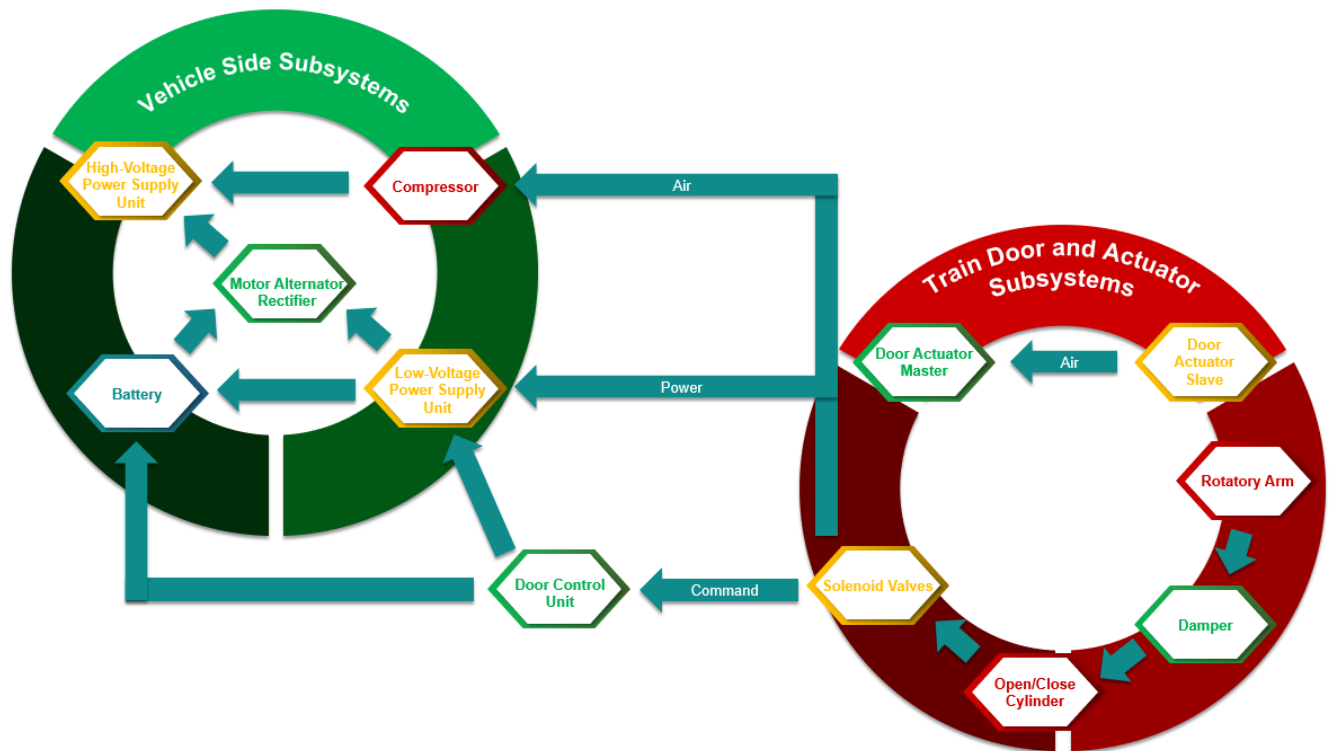
An Example

One example showing how RPM might prevent failures that can negatively affect onboard safety is the prediction of the door controller status². Door failure is not just a safety issue, but it is also at the top of the list of all urban services: it can lead to increased waiting time to ingress and egress from trains, creating delays in line service.

To maintain an equilibrium between quality of service and costs, a preventive maintenance approach based on manufacturers' schedules needs to be avoided due to the high expenditures incurred. The predictive maintenance approach is the right compromise to reduce recurrent costs related to the preventive maintenance. At the same time, it allows you to schedule the interventions at a point in time to maximize the asset life, but with replacement before the failure.

Door operations are managed by actuators that, through air pressure power, move a mechanical system of jacks and levers. The train management system, through electrical signals, activates the actuators and receives feedback on movement and status of the doors (opened or closed). A very simplified model of a pneumatic door subsystem is shown in Figure 2.

Figure 2. Simplified Model of the Door Subsystem



Based on Figure 2, a predictive diagnostic system should be able to assess different border conditions, such as air pressure, currents, velocity, voltages and so forth. For this reason, a system of smart sensors directly and digitally connected with the Train Control and Management System (TCMS) is required. Each component and functionality within the subsystem needs to be analyzed to identify degradation of performance that can lead to failure.

For instance, if the current of the door motor has not increased 10 seconds after open/close control, it could lead to the motor circuit failure. Moreover, if the door close switch has already been activated and the door is not locked, it demonstrates a failure within the door subsystem. With a computational-based comparison among real-time values

² Umiliacchi, P. *et al.*, "Predictive maintenance of railway subsystems using an Ontology based modeling approach", 9th World Conference on Railway Research, Lille, France, May 22-26, 2011.

and expected mean values, it is possible to analyze the deviations and to predict with reasonable effectiveness what part of the subsystem is going to fail and when. In this way, it is possible to identify anomalous conditions that will lead to recurrent faults, providing additional external information, such as correlation, between fault and position. When the alert is generated because the number of anomalous events in the same location exceeds the scheduled threshold, in addition to the technical data, the log file will add information about the train number, the railway line, the position along the line, the train mileage and so forth.

Rail as a Service

Digitalization ensures reliable prognoses for predictive maintenance that minimize failures and disruptions. With the vertically integrated development of IP-enabled smart sensors, computational performances, and big data and analytics frameworks, rail transport is becoming more punctual, cost-efficient and safer. The combination of OT with IT has created the conditions to give rise to a new framework where all data obtained from operational devices are collected, stored, normalized and analyzed. With end-to-end solutions becoming more customary, rail as a service will be an efficient and logical outcome of RPM.

Effective Railway Predictive Maintenance

To build an effective RPM program, consider four essential best practices:

- Identifying prediction viability and effectiveness.
- Extracting the right data.
- Letting the domain expert influence the data analytics.
- Identifying the achievable value-added outcomes.

Identify Prediction Viability and Effectiveness

When building an RPM solution, keep the scope narrow. Include only critical events that leave digital footprints large enough to build consistent predictive models. Trying to predict everything could result in misleading results and a waste of resources.

First, you want to identify what is possible to predict (which subsystem) and with how much accuracy. Start by mapping the available systems on a graph to identify a prediction possibility zone (see Figure 3) and a prediction effectiveness zone (see Figure 4). The goal is to identify the prediction feasibility of the most critical subsystems of the train, noting that often such systems leave little data for building a consistent model. So, the prediction possibility and viability zone are determined by the frequency of occurrence of the damage and its criticality level.

Figure 3. Prediction Viability Zone

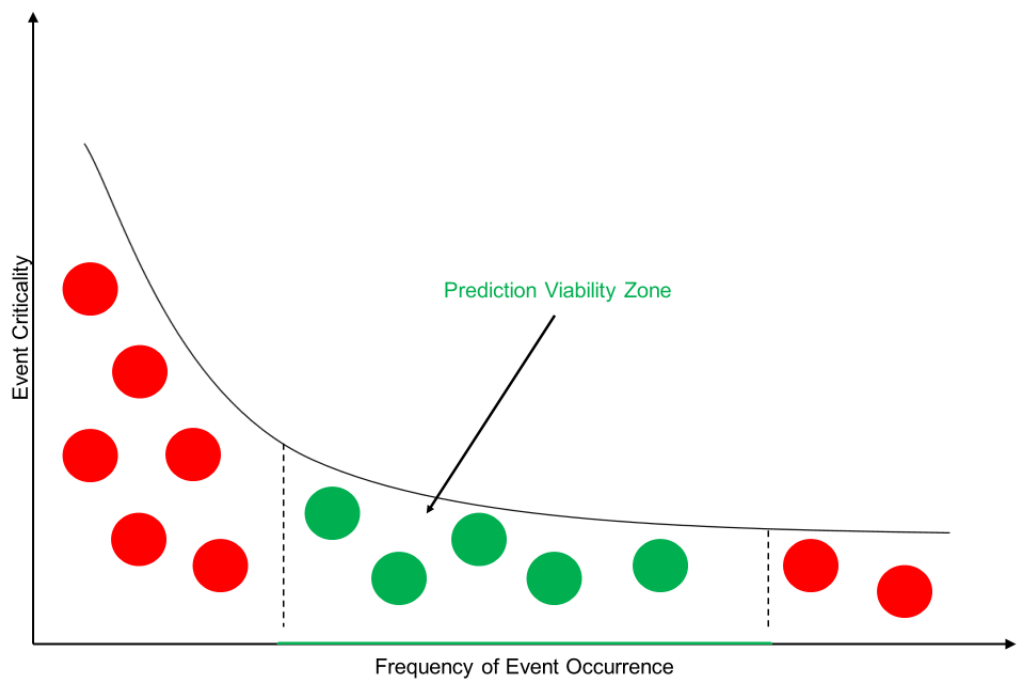
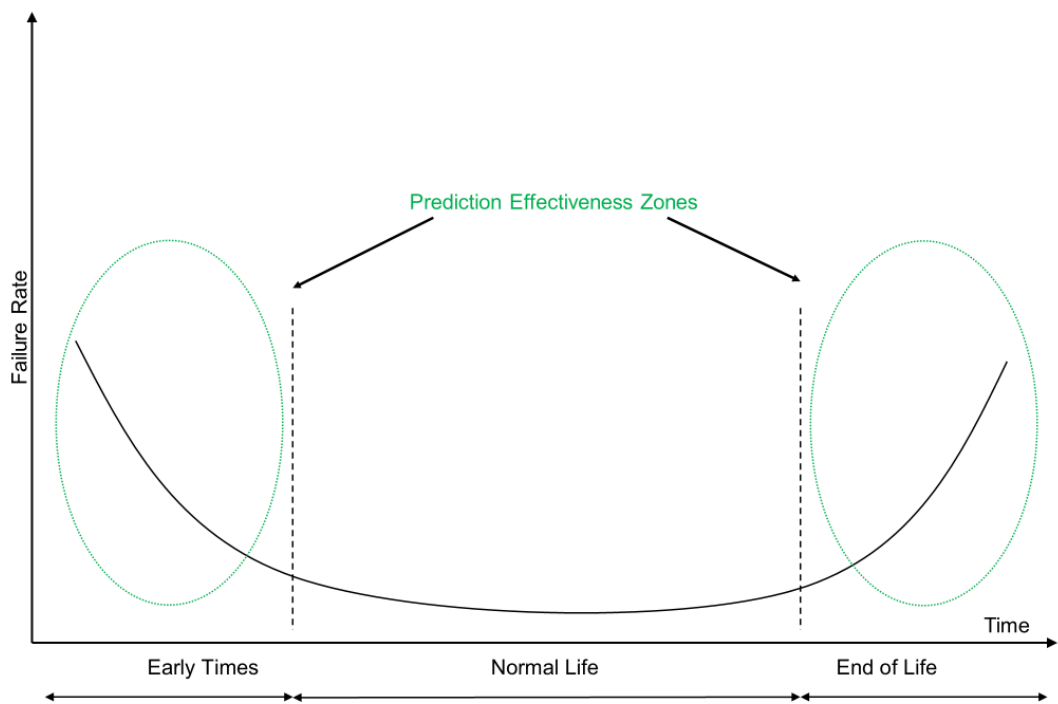


Figure 4. Prediction Effectiveness Zone



Next, identify in which time interval the prediction is more effective, from a maintenance point of view. Figure 4 reports the distribution of the failure rate of mechanical and electrical systems: it typically follows a bathtub curve. To achieve a considerable return on investment (ROI), the figure shows that both infant and end-of-life systems are most appropriate for deploying RPM solutions.

This approach helps to ensure that the outcomes are realistic before deploying resources to the development activity (identifying the required data sets, building algorithms and so forth).

Extract the Right Data

To build effective databases with meaningful information for predictions, consider all the variables potentially assessable and their measurement techniques. The following list is a sample of potential functions and components that might be monitored:

- Axles.
- Bogies.³
- Brakes.
- Door systems.
- Filters.
- Flat wheel (degradation of the steel wheel).
- Harmful currents or voltages.
- Pantographs.
- Rotating parts.
- Water and air pressure.
- Wheel bearings.

Normal mechanical failure modes degrade at a speed directly proportional to their severity. For this reason, if the problem is detected early, major repairs can usually be prevented. Different measurement techniques can be applied to collect data for predicting failures. The most common types of measurement are:

- **Sound:** vibrations generate acoustics. Measuring the acoustics level through an electromagnetic microphone can be an effective means of detecting vibrations.
- **Rotational speed:** a stroboscope or electrical counters could be used to measure rotational speed. Mechanical sensors fixed to the machine shaft could also meet this objective.
- **Temperature:** increased friction leads to an increase of temperature of the monitored asset. Thermistors or other temperature sensors can detect these variations. An inexpensive technique for measuring temperature is to coat the asset with heat-sensitive paint: The color of the paint changes when the temperature exceeds the normal level. Several authors have already provided details on the measurement of temperatures: Grudén *et al.*⁴ assessed bogie temperatures through three sensors. Plus, an additional sensor assesses the air temperature to consider even the border conditions to avoid both false positive and negative values. Kim *et al.*⁵, on the other hand, mounted a few surface acoustic wave sensors on the train's bogies to identify overheated wheel bearings.
- **Vibrations:** vibration is one of the most effective parameters to monitor. Shock pulse measurement, envelope technique and acoustic emissions are a few different techniques used to measure vibrations. Moreover, several

³ A bogie is a chassis or framework carrying wheel sets, attached to a vehicle, thus serving as a modular subassembly of wheels and axles.

⁴ M. Grudén, A. Westman, J. Platbardi, P. Hallbjörner, and A. Rydberg, "Reliability experiments for wireless sensor networks in train environment," in Proc. Eur. Wireless Technol. Conf., 2009, pp. 37–40.

⁵ J. Kim, K. S Lee, and J. Oh, "A study on the wireless onboard monitoring system for railroad vehicle axle bearings using the SAW sensor," Sens. Syst. Softw., vol. 57, pp. 52–58, 2011.

properties of the wagon can be analyzed using accelerometers installed along the train. Nejikovsky and Keller⁶ monitored the rail wagon body motion by mounting the accelerometers on the body of carriages of tilting trains. Wolf *et al.*⁷ installed them at the edge of rail carriages, and Gao *et al.*⁸ implemented accelerators on both the floor of locomotives and the chassis of carriages. Even the bogies can be the object of vibrations' measurement; Elia *et al.*⁹, for instance, mounted accelerometers on the bogies and on axle boxes to measure lateral acceleration.

- **Axle stress:** to measure the stress that affects axles, it is important to measure axle load, curvature of route, high-frequency dynamic forces, braking loads, changes in wheel profile and discrete irregularities (for example, wheel flats). The required attributes are collected using ultrasonic strain gauges, optical and electromagnetic sensors.
- **Oil compounds:** measuring the compounds in the lubrication of sensitive parts (for example, bearings or gears) is an effective way to detect whether there is too much wear or contamination.

All the sensors involved in these measurement processes can be federated through wireless communications over networks of spatial distributed smart nodes. A node represents a single sensor as well as the power supply, the microcontroller and the IP (internet protocol) data transmitter that allows the device to use the TCP/IP protocol.

To avoid replication of measurements from different sensors that will lead to misleading information, the nodes need to be suitably located along the train's subsystems. Furthermore, due to the radio communications range of the sensors, it is essential to use identifiers to connect each sensor with its train. Otherwise, due to the radio communications range of the sensors, data from one train could be collected and associated with another train if the two trains are close on a line or at a station. The use of IP allows the possibility to scale out the infrastructure, adding nodes and extending the coverage range thanks to different wireless technologies derived even by other domains¹⁰.

Let the Domain Expert Influence the Data Analytics

Achieving a successful RPM solution is a team effort where the railway domain expert plays the main role. It is always the domain expert who guides the data scientist in building the right algorithm that will be deployed through the right IT infrastructure. In fact, the success of an RPM solution lies in selection of the right systems, creating and preparing the right data, and getting the right combination of rail industry experts and data scientists on board, preferring hybrid profiles.

Identify the Achievable Value-Added Outcomes

The data you obtain through an effective RPM solution does not deal with just predicting failures; in fact, it can inform root-cause analysis related to the design of the parts, the construction processes, the life cycle and much more. RPM can be used to identify various business scenarios and appropriate prescriptive actions. The value added from RPM solutions includes:

- Predicting when, subject to specific border conditions, a part will fail, and which maintenance actions are required.
- Planning the maintenance actions in advance, allowing a just-in-time sourcing for replacement of parts, and optimizing procurement and inventory.
- Identifying systems that might be affected by potential design problems based on their history of poor performance.

⁶ B. Nejikovsky and E. Keller, "Wireless communications based system to monitor performance of rail vehicles," in Proc. IEEE/ASME Joint Railroad Conf., Newark, NJ, USA, 2000, pp. 111–124.

⁷ P. Wolfs, S. Bleakley, S. T. Senini, and P. Thomas, "An autonomous, low cost, distributed method for observing vehicle track interactions," in Proc. IEEE/ASME Joint Rail Conf., Atlanta, GA, USA, 2006, pp. 279–286.

⁸ C. Gao *et al.*, "Design of train ride quality testing system based on wireless sensor network," in Proc. Int. Conf. Electron. Mech. Eng. Inf. Technol., Harbin, China, 2011, pp. 2636–2639.

⁹ M. Elia *et al.*, "Condition monitoring of the railway line and overhead equipment through onboard train measurement-an Italian experience," in Proc. IET Int. Conf. Railway Condition Monitor., Birmingham, U.K., 2006, pp. 102–107.

¹⁰ Hodge, V., *et al.* "Wireless Sensors Networks for Condition Monitoring in the Railway Industry: A Survey". IEEE Transactions on Intelligent Transportation Systems, vol. 16, n° 3, June 2015.

- Identifying a track's problem when a train goes through a specific point in line, treating the vehicle like a sensor on wheels.

By understanding the reasons behind various failure patterns and categorizing them into various action buckets, it is possible to address both short-term and long-term objectives.

From Data to Insights

Sensors will create both exogeneous data that measures external factors, such as the weather or line conditions, and endogenous data synthesized from within the train's subsystems. Once the data is created, the flow required to convert raw data into useful information is from data acquisition to transformation, evaluation and visualization.

Data Acquisition

Data acquisition is the process of gathering and measuring information from heterogeneous sources (such as the different trains' subsystems, railway line and weather) and related targeted variables in an established, systematic trend. This makes it possible to capture quality evidence that is then translated into rich data analysis, to build an effective and credible data set. The acquisition process requires a converged IT infrastructure per each train (including software, networking, server and storage) to:

1. Collect and store the data produced by IP sensors and other external sources.
2. Perform a first analysis of the data in real time, providing useful information to the driver about the route and the health of the systems.
3. Share through wireless connectivity all the data acquired during the trip. This data is then consolidated and processed through Extract, Transform and Load (ETL) data transformation tools.

To perform an effective data acquisition, you must develop specific connectors able to interface with batch sources and real-time flows. And you must collect all the data into a data set that will be transferred to the normalized master data lake within the main data center of the railway service provider.

Data Transformation

Through visually interfaced data integration tools, it is possible to move data from many different sources, to aggregate and transform those to allow domain experts to analyze a heterogeneous set of data of any format, schema and type (data lake). In a good data integration tool, this mapping is depicted visually so that it is easy to follow the path of the data, and to understand precisely where each piece of data originates, how the data is processed or transformed as it passes through the system, and exactly where the transformed data is going. For example, Hitachi can perform these actions through Lumada, its comprehensive big data integration and analytics solution.

The data transformation and integration process provide standardized data in a format and a place where it is consumable from a maintenance life-cycle point of view. In this way, it is possible to build data lakes where, by coding the right algorithms, information is extracted from raw data. Indeed, the process addresses the problem of incoming and stored data from many different fragmented places and in many formats.

Data Evaluation

Through data evaluation, data scientists analyze data and search for patterns that predict potential faults through advanced algorithms, expertise, domain know-how and best practices. For example, patterns might predict the circumstances in which a traction drive, electronic door motor or a wheel set will fail.

The data evaluation phase deals with both short-term and long-term analysis. The short-term analysis is performed on board and provides real-time information to the driver about the running trip. The long-term analysis provides an end-to-end view of the maintenance framework to make it more efficient, identify new patterns, and improve decision-making and future planning.

It is possible to use several capabilities and technologies to achieve these results by gaining insights from data. The following list is a sample of some potential techniques:

- **Descriptive analytics** techniques provide simple summaries and observations about the data.
- **Data mining** analyzes large quantities of data to extract previously unknown interesting patterns and dependencies. (See the “Data Mining Techniques” sidebar for more information.)
- **Machine learning** enables the software to learn from the data and predict accordingly. For example, when a train’s subsystem fails, several factors come into play. The next time those factors are evident, the software will predict the failure.
- **Simulation** enables what-if scenarios for specific assets and/or processes; for example, how running specific components for a certain period of time impacts the likelihood of failure.
- **Text mining** is a subset of data mining, where data is composed by natural language texts. It enables the understanding of and alignment between computer and human languages. For example, by analyzing maintenance logs it becomes possible to determine that a specific operator performed specific operations, which led to extended asset life.
- **Predictive analytics** uses machine learning and data mining techniques to predict future outcomes. The holistic approach of sophisticated analytics tools is applied to develop models and estimations about the behavior and the useful life of assets.
- **Prescriptive analytics** adds a decision-management framework to the predictive analytics outcomes to align and optimize decisions according to analytics and organizational domain knowledge. The goal to achieve is not just to identify when an asset fails, but also to suggest actions, and to show the implications of each decision.

The data analysis can provide a precise forecast about how long a component or a drive unit will continue to function under specific conditions. The analysis also determines, with good level of accuracy, which actions must be taken when a behavioral pattern registered by the data, and based on past experience, indicates that an acute failure can be expected in short time. To achieve these goals, it is fundamental to apply a holistic approach made by the implementation of advanced algorithms, expert domain know-how and best practices.

Data Mining Techniques

The following are the most common key data mining techniques:

- **Anomaly detection** deals with the discovery of records and patterns that are outside the norm. For example, if the door motor current has not increased 10 seconds after open or close control, it could lead to the motor circuit failure of the door system.
- **Association rules** search and identify dependencies, relationships, links or sequences among variables in the data. For example, wheel bearings tend to fail under specific different conditions [external temperature, forces, wind speed and direction, hours of operations, mileage, singular points (GPS positions) within the track, and so forth].
- **Clustering** creates groups of objects that satisfy the same properties.
- **Classification** correlates the new data points collected with the most appropriate set by identifying the level of affiliation; for example, a vehicle can be classified as “old” or “new” according to the mileage.
- **Regression** assesses the relationships among variables and calculates how much a variable changes when another variable is modified. For example, the brakes reduce their useful life faster in function of the route and of the driver (the way in which the driver leads the train).

Data Visualization

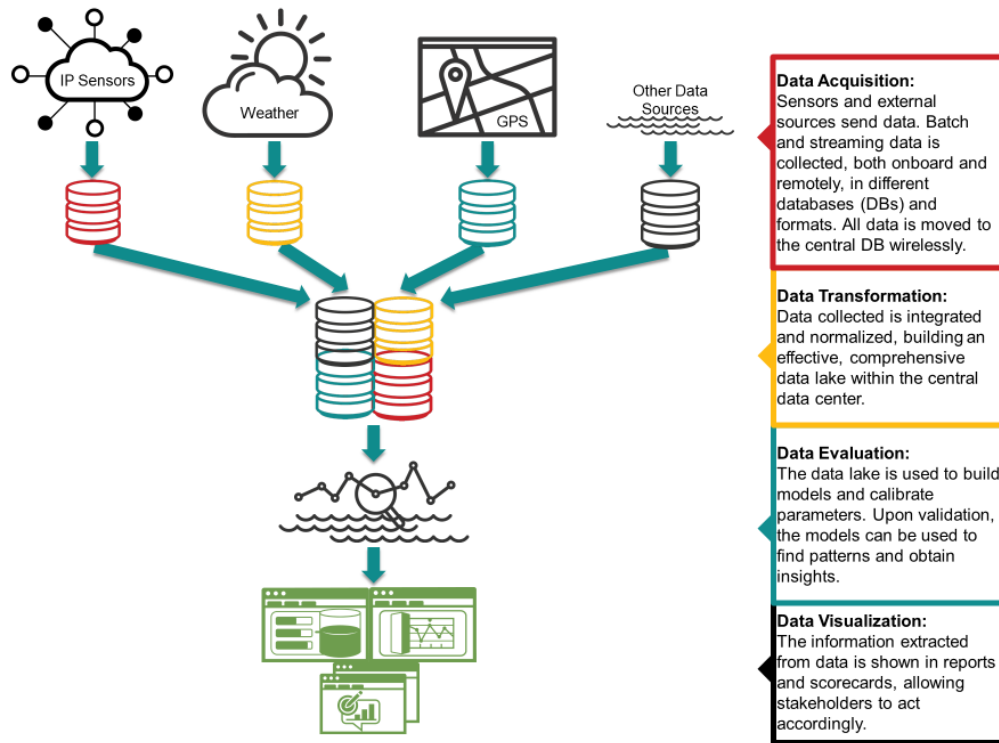
After the data has been correlated and analyzed and new patterns have been discovered and validated, the visualization phase allows the stakeholders to take actions accordingly. Within the data visualization types, the most common are the dashboard, infographics and balanced scorecards.

Transforming data into meaningful and easy to understand information in reports or some other visual format can lead to the implementation of an effective business intelligence (BI) strategy. To achieve these results, the data visualization system must meet the following requirements:

- **Useful:** all the stakeholders (management, dispatchers, maintenance engineers and so forth), although with different aims, use the information on a regular basis and make relevant decisions by viewing all the insights they need in one place.
- **User friendly and visually appealing:** it is both easy to use and a pleasure to use.
- **Effective:** stakeholders who use it accomplish their goals quickly and easily.
- **Scalable:** it is accessible, and conducive to future maintenance and modifications.

The end-to-end flow to transform raw data into useful insights and maintenance patterns is shown in Figure 5.

Figure 5. The Transformation Flow From Raw Data Into Useful Insights



A Robust IoT Platform

To deploy an effective predictive maintenance framework, it is not enough to identify and collect the right data, calibrate the right models, and build the appropriate algorithms. You need a robust IoT platform to collect and store a large amount of raw data to convert into actionable insights and useful information.

The volume of data is rapidly growing and can be effectively administered through relational databases and non-SQL databases. Only intensive parallel processing systems and in-memory databases can handle and analyze such huge volumes of data with complex algorithms.

An Integrated Ecosystem

From an IT infrastructure point of view, there are different solutions that enable railway predictive maintenance. These solutions are designed to collect, store, manage and analyze a huge amount of heterogeneous data, and can interface with in-memory platforms to perform real-time analysis of structured data.

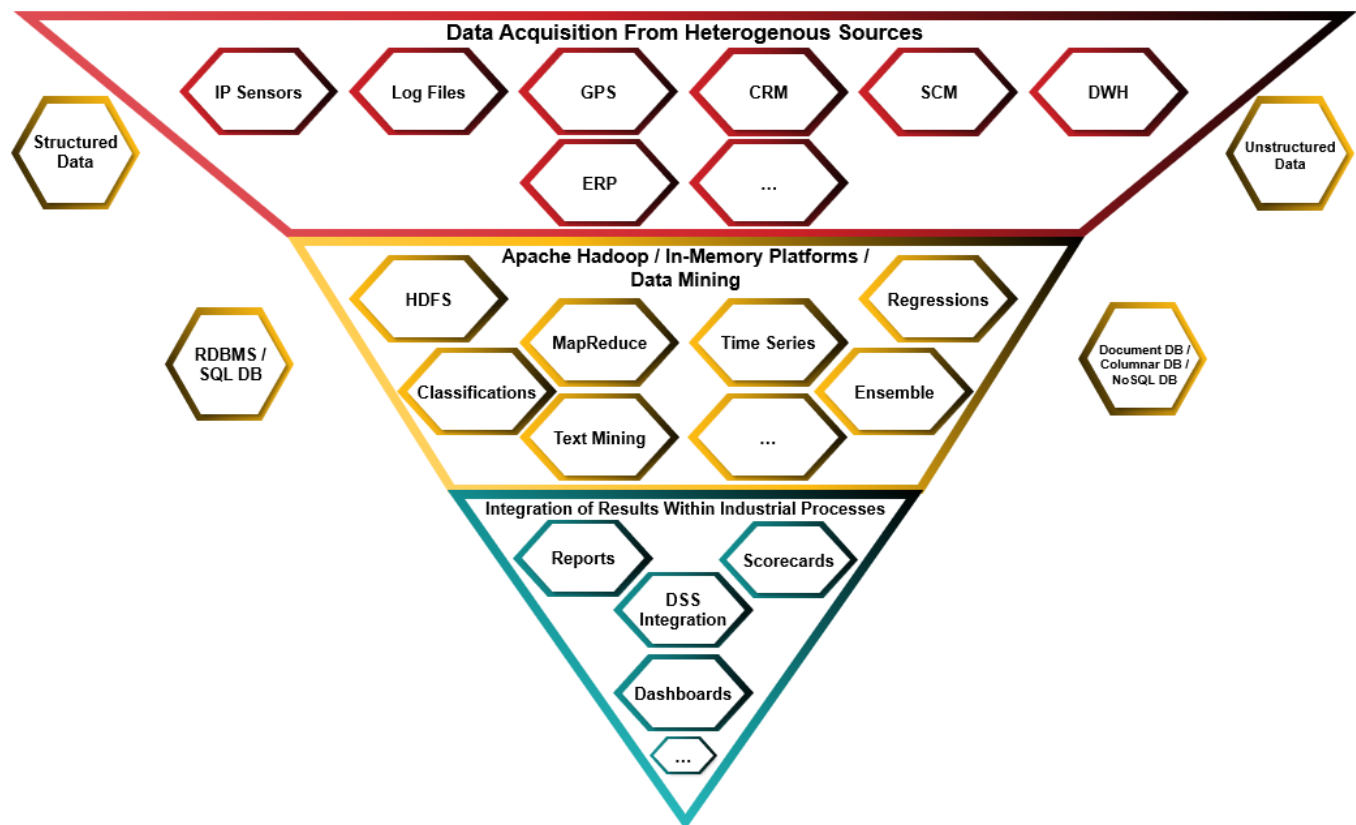
The central ecosystem (see Figure 6) might be constituted, for example, by Apache Hadoop: an open-source framework used to manage and process a huge amount of data through commodity hardware and both distributed computational (MapReduce or Spark) and storage (Hadoop Distributed File System) resources. Multiple data types from many sources (such as engine variables, bogie sensors, GPS position within the line, and atmospheric data) may be ingested into the data lake built over the infrastructure, satisfying the requirement to run Hadoop and other analytics suites across large, diverse data sets.

Hadoop, with its modular modules, can perform comprehensive analysis of structured, semistructured and unstructured data, identifying predictive models and dependencies among data seemingly not correlated, showing the results through highly customizable reports. In this way, it is possible to extract information from different

independent data, whose correlation could provide insights about the health of different trains' subsystems, in function of dynamic border conditions and variable exogeneity.

To maximize performance, data is automatically spread and balanced across the cluster's nodes, guaranteeing the required scalability. This ecosystem is designed to analyze both structured data derived by IP sensors and unstructured data (usually bigger) obtained by external sources not directly related to the train's diagnostics.

Figure 6. Integrated Ecosystem Hadoop, In-Memory Platform and Data Mining



CRM = customer relationship management, SCM = supply chain management, DWH = data warehouse, ERP = enterprise resource management, HDFS = Hadoop Distributed File System, DB = database, RDBMS = relational database management system, DSS = decision support system

Hitachi's Lumada and Railway Predictive Maintenance

To satisfy the hardware and software exigencies described above, Hitachi developed the Lumada internet of things platform. As a vertically integrated IoT platform, Lumada consolidates all the required components within a unique stack. Lumada, in fact, is designed to manage the entire life cycle of different types of assets and devices. It measures real-time performances, builds a statistically representative data set along the useful life of the asset, and performs technical and financial forecasts and optimizations.

Its modular and flexible architecture, along with public application programming interfaces (APIs), allows the implementation of third-party solutions (both proprietary and open source) even if already installed within the existing framework. This approach allows the preservation of previous investments. Lumada can establish both real-time and batch connections with single devices and with a fleet of assets, even if geographically dispersed or in movement.

Using its data ingestion tools, Lumada can visualize data coming from assets, store those on an SQL database and/or on big data platforms. It can analyze the entire data lake through artificial intelligence and machine learning tools,

implementing a specific workflow as a response to achieved results, and integrating with corporate IT systems. The modular approach does not deal just with integrable functions within the ecosystem. It also handles the scalability in terms of types and amount of monitorable devices, letting you add and remove assets without affecting the global availability of the solution.

Lumada Framework

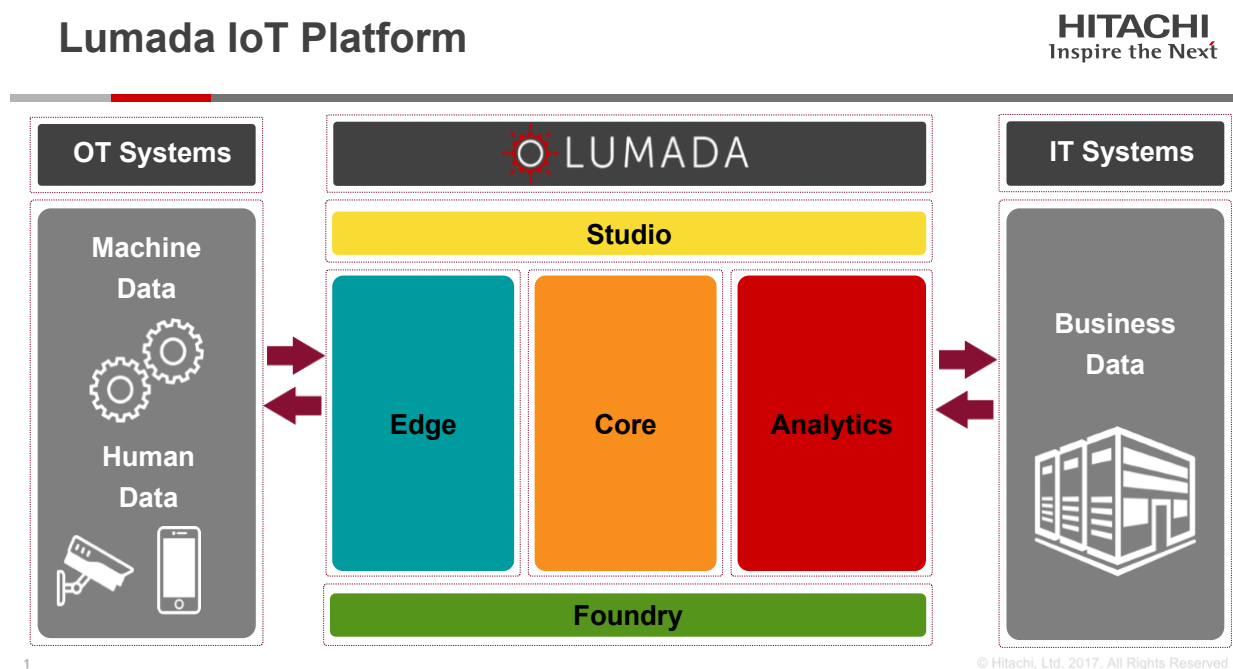
With respect to other solutions within the marketplace, Lumada is built as a semi-finished framework with a fixed schema (edge, core, analytics, studio, and foundry) and potentially unlimited number of integrable solutions (both Hitachi and third-party). In this way, Lumada can address not only the technical and operational exigencies identified during the design phase, but also any exigencies detected on an ongoing basis (see Figure 7).

Subsystems, devices and assets can interact with Lumada by using the HTTP protocol through the REST (REpresentation State Transfer) architecture and other binary protocols with a different pattern, known as publish/subscribe. Common among those are:

- MQTT (Message Queue Telemetry Transport): designed to satisfy telemetry exigencies, it is very light and affordable, with three different levels of quality of service (QoS). It performs over all networks with low connectivity performance and stability.
- AMQP (Advanced Message Queueing Protocol): developed predominantly for server-to-server connections for enterprise systems, AMQP is heavier than MQTT, but it supports many different additional patterns.

Hitachi provides not just the technology, but also the domain knowledge to advise the involved stakeholders in defining technical specs, in industrial processes optimization, and within the end-to-end cycle of value creation.

Figure 7. Lumada Architecture



The Lumada IoT Ecosystem

Lumada is a complete IoT ecosystem built on the following:

Lumada edge analyzes, filters and secures data from sensors and assets and integrates the results into business operations for a comprehensive, real-time view of asset status.

Lumada core connects assets, collects data, provides identity and access management, and creates and stores digital representations of physical assets called asset avatars.

Lumada analytics uncovers patterns in device data with machine learning (ML) and data-mining tools to avoid costly breakdowns and support fast, data-driven decisions.

Lumada studio brings together analytic information, alerts and notifications, and business enablement in dashboards to give you fast, meaningful insights into your data.

Lumada foundry delivers a foundation for the rapid development of service-based apps used to deploy, compose and package product solutions.

Concepts

An asset is a physical object that you manage. Assets can connect directly to Lumada or through a gateway. When connected to Lumada, the asset avatars monitor and manage the assets. This section provides a brief description of gateways, asset types and asset avatars.

A *gateway* is a software that federates devices that will be interfaced with Lumada. Gateways are commonly used when:

- There is a spatial distribution of devices. For example, consider a device that is on a private OT network and needs to connect to the public IT network.
- Lumada and the devices might use different protocols and might be unable to understand each other. For example, within the industrial sector, the gateway can connect to a device through systems such as:
 - Programmable Logic Controller (PLC): an industrial digital computer that has been ruggedized and adapted for the control of manufacturing processes.
 - Supervisory Control and Data Acquisition (SCADA) system: a control system architecture to interface to the process plant or machinery.
 - OPC Unified Architecture (OPC-UA): a machine-to-machine protocol for industrial automation.

An *asset type* contains the digital features that will be provided by each single asset. Asset type *train*, for instance, will provide the train's speed, the voltage of the line, the GPS position and more. This element will teach Lumada about the values it will receive as input from each type of asset.

An *asset avatar* is a digital representation of each asset: it is the digital twin that digitally reproduces the machine that is necessary to map. Within the railway example, the asset represents the single train that belongs to the specific fleet (asset type).

Hitachi's Lumada Platform

To perform all the activities related to data management, the Lumada stack offers Pentaho's Data Integration capabilities. This suite of analytics acts as an orchestrator for extract, transform, load (ETL) processing, big data ingestion, *MapReduce* management, data analytics and visualization. With the Pentaho Data Integration (PDI) module, for instance, you can integrate, refine and correlate different types of data, including multisource data. This is accomplished using the Pentaho native console, or integrating models built with R, Python, Weka and so on.

PDI has a drag-and-drop interface that removes the need to write code, reducing complexity and increasing data scientists' productivity. It improves productivity during the ETL phases within the activities of analysis and results

visualization, during the interfacing with Hadoop and Spark distributions, with SQL and NoSQL databases, and during the export of data toward in-memory platforms (for example, SAP HANA platform).

Based on open source framework with a large community of developers worldwide, Lumada represents a continuously evolving solution that is constantly aligned to the new and mutable marketplace's exigencies. It guarantees deep compatibility with sources of heterogeneous data and with the most diffused proprietary applications.

Using a cybersecurity module designed ad hoc, advanced encryption tools protect all the communications. This protection extends among the different frameworks' layers and the distributed assets to be monitored, as well as to all the connections with the end users.

Examples of Deployable Global Functionalities

The following is an illustrative list of deployable global functionalities:

- Integration with third party-applications.
- Implementation of different communication buses.
- Data lake management.
- Analytics.
- Extract, Transform, Load (ETL).
- Complex event processing.
- Real-time and batch analysis.
- Extraction of predictive models with the possibility to interface in-memory platforms.
- Workflow execution in function of thresholds and iterative cycles.
- Digital twin (asset avatar) construction to digitalize real assets.
- Management of additional metadata.
- Management of entire asset's lifecycle.
- Predictive quality.
- Implementation of what-if scenarios.
- Chargeback and cost splitting according to activity-based costing.

Customization Capabilities and Railway Predictive Maintenance

The customization capabilities of Lumada enable the construction of a framework for railway predictive maintenance. This framework integrates domain competencies with technologies enabling predictive analytics and automated decision support systems. It can interface in real time with the fleet of trains along the lines, collecting events, diagnostic signals and counters. This approach allows the construction of a multisource and statistically representative data lake that enables the extraction of predictive models to perform *just-in-time* or *in-case* maintenance operations, increasing the trains' availability and reducing potential inefficiencies.

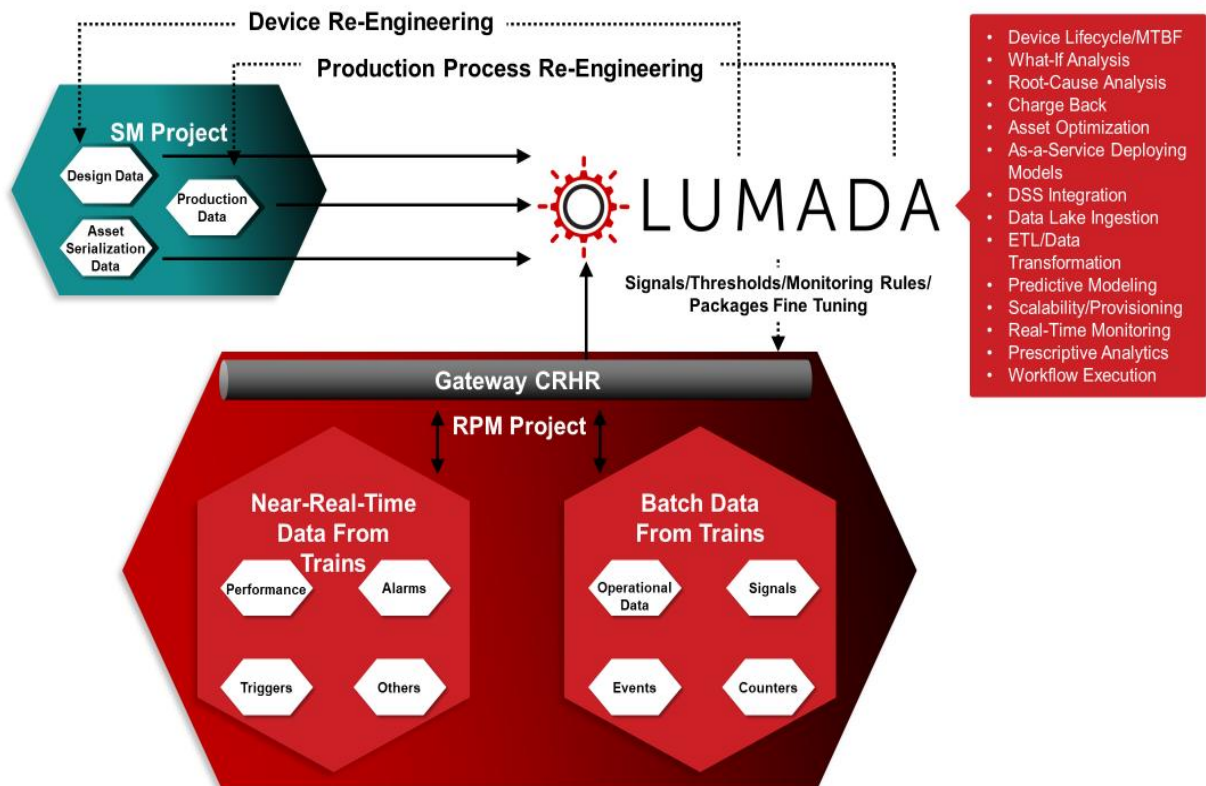
In parallel with this first phase, proceeding with an important digital transformation project within the production chains, Hitachi is working to map the production processes. This mapping integrates with the predictive maintenance data lake, additional data related to design of subsystems, their production, and the tests performed to certify the

compliance. With asset serialization, it is possible to build a digital representation of all the processes within the entire lead cycle. It enables root-cause analysis, process and product re-engineering, and chargeback policies, and extracts technical and economic information that can help increase the efficiency and effectiveness of key performance indicators (KPIs).

Predictive Maintenance and Smart Manufacturing

Lumada represents the common denominator for enabling both predictive maintenance and smart manufacturing (see Figure 8), blending different sources of data, also in real time, extracting information and generating value.

Figure 8. Lumada serves as common platform for both projects: predictive maintenance and smart manufacturing.



CRHR = control room, Hitachi Rail; RPM = railway predictive maintenance; MTBF = meant time between failures; DSS = decision support system; ETL = extract, transform, load

Those solutions can be delivered in the cloud and/or on-premises, and could also be proposed as a service to address the financial exigencies of small and medium enterprises that can prefer an operating expenditure (OPEX) investment in function of the real utilization of the resources.

The biggest part of data involved within the predictive maintenance processes is, nowadays, predominantly structured, and specifically *time series* – that is, data sampled at specific frequency. Often, the frequency is very high, and this leads to storing huge amount of data, increasing difficulties to manage and analyze those. And then there's

the additional concern that a big part of that data is not statistically representative and could be replaced with trends and recurrent cycles.

Mathematical Methods

The effectiveness of opportune mathematical methods to reduce the dimension of those time series has been demonstrated, while maintaining the intrinsic information. Among the most effective methods is the Discrete Fourier Transformation (DFT). DFT can convert a given collection of samples' functions with known frequency in a collection of coefficients of linear combination of complex sinusoids, sorted as frequency increases¹¹. In essence, it is possible to replace the original series with a linear combination of *sin* and *cos* keeping just a reduced number of initial coefficients.

An additional aspect to consider in analyzing time series is the recognition of recurrent schemas and predictive models. This identification can be performed within a single series to recognize anomalous behaviors with respect to the expected values. Or, as an alternative, to analyze more time series at the same time to identify the target time series that represent an effective approximation of the optimal operation values. Once the target time series is identified, by applying dynamic time warping (DTW) methods, it is possible to measure the distance $\delta(t)$ between the aligned sequences received as input by the system (for example, from n trains in line belonging to the same fleet) with respect to the target time series that shows the expected optimum behavior (see Figure 9).

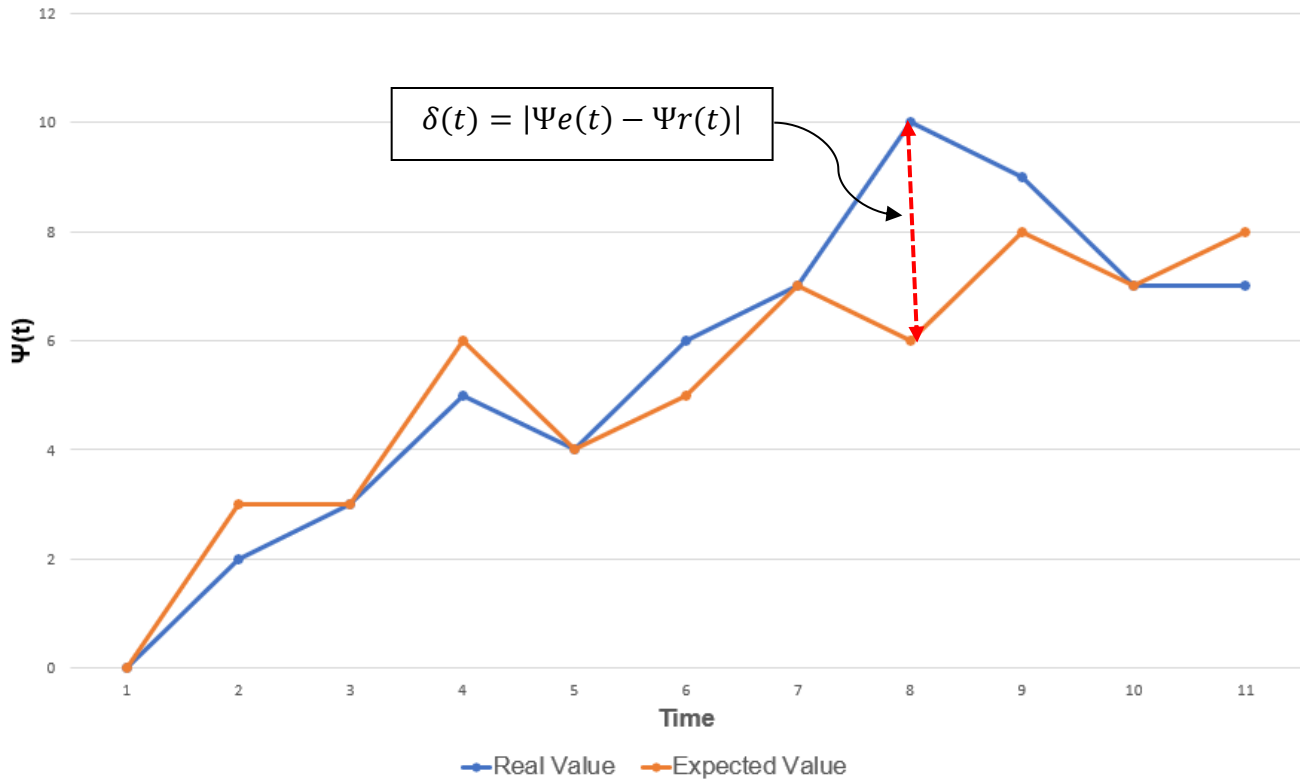
Let $\Psi_e(t)$ be the expected value extracted by the target series. It can be higher or lower with respect to the value measured onboard in real time $\Psi_r(t)$; let $\varphi(t)$ be the threshold, known *a priori*, that, once overtaken, certifies that the behavior of the asset can be considered as anomalous. Per each sampling time stamp, in order that the monitored subsystem runs with optimal operating conditions, it will be required to satisfy the following disequation:

$$\delta(t) = |\Psi_e(t) - \Psi_r(t)| < \varphi(t)$$

It is fundamental to dispose of appropriate computational capacity to deploy similarity analyses in real time. This includes, simultaneously, as many time series as possible to obtain triggers and alarms during the phases of incipient malfunction of fleets along the lines.

¹¹Smith, Steven W., "The Discrete Fourier Transform," in *The Scientist and Engineer's Guide to Digital Signal Processing, Second Edition*. (San Diego: California Technical Publishing, 1999).

Figure 9. Differential analyses between real values sampled and expected values are extracted by the predictive model.



In-Memory Database

Hitachi has its own solutions to perform streaming analyses, but because Lumada is an open platform, it is possible to integrate it, based on actual results, into environments where it is already installed and operative an in-memory platform. There are different platforms within the marketplace able to perform analysis over random-access memory (RAM); one powerful option is represented by the HANA solution provided by SAP. In fact, this platform lets you build an in-memory database able to extract real-time information from a huge amount of heterogeneous data.

From a hardware point of view, SAP has certified several vendors with different architectures that can run HANA, starting from 768GB of in-memory database up to 20TB for scale-up solutions for SAP suite on HANA (SoH) and suite for HANA (S4H) environments. Within the scale-out configurations, on the other hand, there are solutions able to reach 94 nodes of 4TB each, addressing 376TB of in-memory database per each single solution.¹²

To contextualize, one train, in function of the number of sensors installed, can produce up to 500MB of data within 24 hours of service – 25GB if the data coming from onboard video surveillance systems are also considered. The scale-out architecture includes:

- A set of blade servers (in function of the total amount of memory required).
- One or more storage subsystems that provide block storage for the nodes and network-attached storage (NAS) platform.
- A few NAS (in function of the size of the database) that will provide network file system (NFS) for SAP HANA binaries and cluster-wide configuration files.
- Up to two additional rack servers that run Network Time Protocol (NTP).

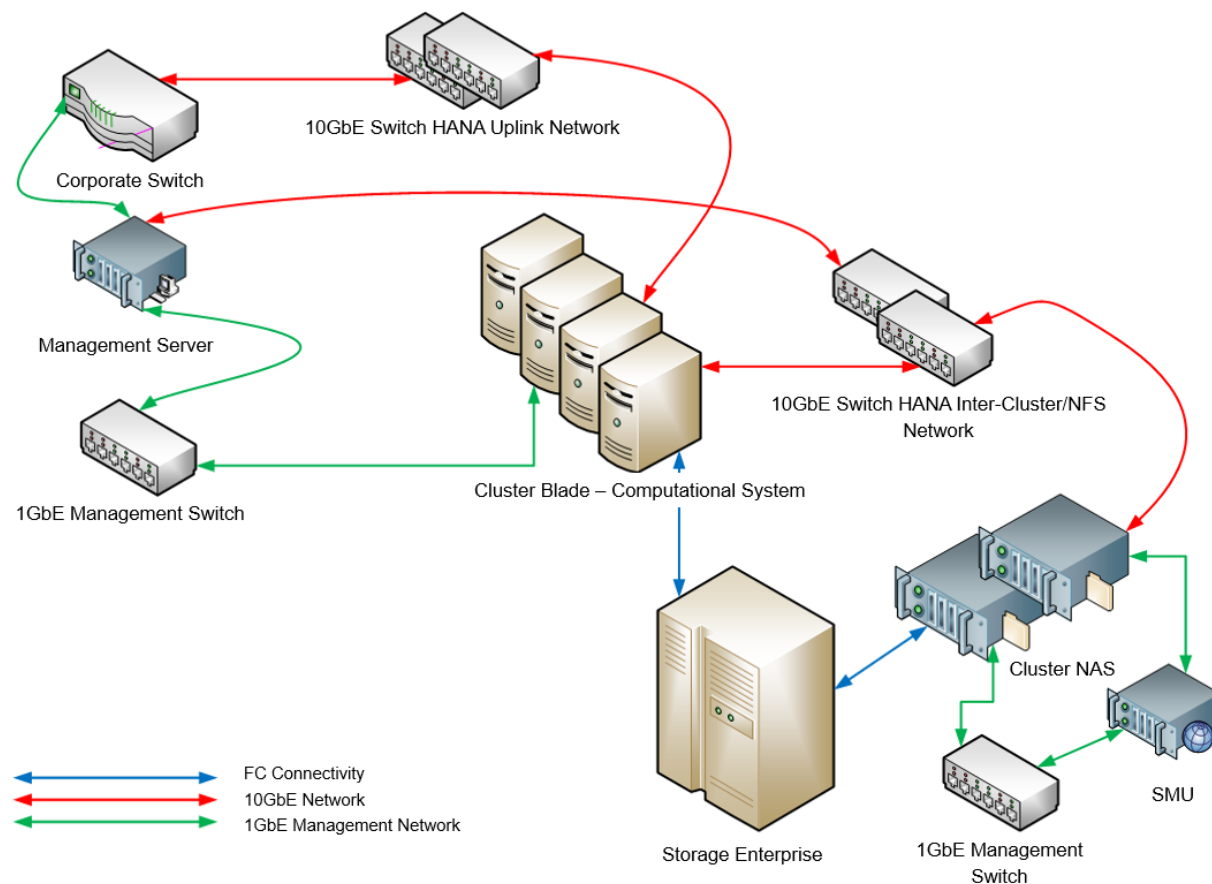
¹² SAP certified appliances (Apr.. 2017) <https://global.sap.com/community/ebook/2014-09-02-hana-hardware/enEN/appliances.html>.

- The management console of IT infrastructure.
- The remote monitoring system.
- SAP HANA Studio.

In addition, 10 gigabit Ethernet (GbE) IP switches will be required to address NFS and intercluster network, and additional 1GbE switches are required by the NAS platform private network.

Figure 10 shows a simplified example of a convergent solution for SAP HANA in scale-out configuration. To simplify the graphical representation, inter-switch link (ISL) connections are already considered present among the different couple of switches. Redundant links between each single switch and its correspondent device have also been omitted from the figure.

Figure 10. Simplified schema depicts convergent solution for SAP HANA in scale-out configuration.



The Business Side of Railway Predictive Maintenance

Affordable RPM solutions can positively impact the rail business, while completely transforming the maintenance landscape.

A Financial Perspective

Let the gross profit obtainable by the difference between revenues and costs ($GP = R - C$). An effective RPM framework can positively influence both revenues and costs, achieving efficiency and effectiveness improvements. For costs, it is possible to obtain the following results:

- **Reducing the exigencies for operational reserves and related costs:** train fleets typically need an operational reserve from 5 to 15% as backup in case of operational failure. Through an RPM framework, it is possible to optimize the rolling stock maintenance by predicting when a component will fail. Unplanned outages of rolling stock are minimized, so fewer trains need to be kept on standby. This leads to savings on both capital expenditure (CAPEX) and OPEX.
- **Extending the useful life of the assets:** RPM allows replacement of components when they are close to failure and not when the manual suggests. This means expensive components are used optimally, reducing the spending on parts, and minimizing labor costs related to maintenance.

RPM can also increase the revenues related to the railway service operators, achieving the following objectives:

- **Moving trains from operational reserve to the line:** by mitigating the risk of serious outages, it is possible to utilize trains that were earlier kept as backup to run new services and consequently increase the number of tickets sellable per day, without additional capex.
- **Attracting demand from other modes of transportation:** achieving a high degree of reliability allows the railway operators to be more attractive in customers' eyes, and to intercept new flows of passengers from other modes (for example, airplane for trips up to 700km).

Investing in RPM could be interesting for both railway manufacturers and service operators. A European national railway service operator can spend an average of 1.3 billion euros per year just for first- and second-level maintenance. First-level maintenance is referred to as ordinary maintenance, both planned and corrective. Second-level maintenance deals with interventions of greater impact.

It is possible to consider three scenarios: best, worst and most likely. In the first case, achieving a 5% potential savings can lead to 65 million euros saved in a fiscal year. In the worst case, with a savings of just 1% of the as-is scenario, it is possible to save 13 million euros per year. Finally, a most likely scenario could lead to savings of up to 3% of actual spending, achieving a savings of 39 million euros per year.

Of course, the initial investment can be amortized over a few years, generating positive cash flow in terms of cost savings, optimizations and new demand attracted from others means of transportation. The payback period is an important consideration in making the RPM project both feasible and attractive.

Essential Setup Activities

To reduce the latency between the initiation and execution of the project, it is essential to organize and manage the setup activities as follows:

- Cluster the families of rolling stocks that are to be analyzed by the vehicle type and its use (for example, high-speed trains, tilting high-speed trains, commuter trains, freight trains and so on).
- Create test scenarios with a few vehicles per cluster.
- Start the pilot projects for each family or cluster in parallel to reduce the deployment time and to build the models through a cross-fertilization approach among the different types of clusters.
- Perform validation and fine tuning for each model (function of the cluster).
- Install smart IP sensors and IT infrastructure for each train part in the project, and scale the actions taken within the test scenarios to all the vehicles.

The pilot phase could take 8-10 months including the data collection activities required to build a statistically representative data set to be analyzed. An additional 6-8 months might be needed to implement the pilot IT framework aboard the entire fleet of trains.

Business Partnership Approach

The industrial internet of things (IIoT) is creating the conditions for new businesses within the railway field that could be addressed if all the stakeholders (railway manufacturers, IoT companies and railway service operators) cooperate through a business partnership approach, co-creating value in a vertically integrated way.

There are already different partnerships around the world between IoT companies, railway manufacturers and railway service operators that led to the implementation of RPM scenarios. One of these is particularly interesting because it shows all the positive topics discussed previously. The scenario consists of two big cities with a distance between them of almost 700 km. Before the implementation of high-speed railway lines and the RPM framework, the time to complete a one-way trip was 5.5 hours, and the total number of passengers serviced was nearing 800,000 per year. At the same time, the airlines were able to cover the same route in 1.4 hours plus the time for check-in and security checks. The airlines were servicing 80% of the market between the two cities, even though they were the most expensive option. It is important to note that this particular air link between the two cities was one of the busiest air routes globally.

When the high-speed railway line opened, the rail operator was able to significantly reduce the journey time, covering the distance in 2.5 hours, making the plane and train trips comparable and giving passengers a real choice. To directly target the air route passengers, the rail operator offered full refunds for any journey that was delayed by more than 15 minutes. This policy was appreciated by the passengers, even though it exposed the rail operator to a considerable financial risk in the case of delayed trains. However, this serious risk has been mitigated through the implementation of RPM framework that allowed a high degree of reliability.

With unplanned outages minimized, there is little chance of mechanical failure en route or rolling stock availability delaying a train more than 15 minutes. With faster route time and reliable service, demand increased, and the rail operator increased its market share from 20% to 60%, reducing maintenance OPEX and increasing revenues.

From a financial point of view, the entire investment was repaid by savings, even generating positive cumulative cash flow. Without considering the extra revenue attracted by air routes, this project had a payback period of eight years and return on investment of 130% calculated over 10 years, just from the savings achieved through the implementation of the RPM framework. And by considering the additional cash flow derived from additional demand, the payback period was reduced to just three years.

From a transportation engineering point of view, there were several additional benefits:

- Rail services became more reliable, and customer satisfaction improved.
- Greater customer satisfaction causes market share to grow, thanks to service reliability of 99.98%.
- With fewer unnecessary component upgrades, maintenance costs were lowered.
- Reduced costs were passed on to passengers with reduced fares, improving ridership.
- Switching from plane to rail mode, there was a reduction of CO₂ emissions per passenger.

Conclusion

The availability of new technologies and the huge amount of data being created are the key factors able to revolutionize maintenance in the 21st century. When there is an ongoing transformation process, in front of disruptive events, new scenarios and opportunities to create businesses are leading to new equilibriums within the marketplace.

Sensors mounted on critical subsystem components gather thousands of data points within the unit of time, which allows engineers to understand the condition of the components. Leveraging deep engineering knowledge and data analytics capabilities, analysis of this data can be utilized to predict component failures and carry out root cause analysis when failures do occur, supporting continuous improving strategies and processes. This leads to a tailored maintenance strategy, extending the useful life of components, reducing labor costs and avoiding expensive corrective maintenance.

Recent advancements in smart sensors and IT have led to continuous data collection from various systems and subsystems in trains, enabling monitoring of mechanical and electrical conditions, operational efficiency and multiple other performance indicators. These new capabilities enable planning of maintenance activities with the maximum interval between repairs, while minimizing the number and the costs of unscheduled outages created by system failures. This minimizes not just the maintenance costs of the train, but also the loss of revenues due to the impossibility to utilize it to run passengers (or freight) services..

Minimizing unplanned rolling stock outages through predictive maintenance is fundamental to ensure stability and reliability throughout the transport networks. In this era of the industrial internet of things, a cross fertilization is occurring between railway engineering and IT. This change requires using vertically integrated knowledge, which will overturn the paradigms of classical railway engineering.

Predictive maintenance can, of course, be applied to more than the rail sector. Like a system of differential equations, it can also be scaled to different industries just by changing the border conditions. In fact, it can also be applied to other industries such as healthcare, aerospace and defense, automotive, energy and utilities, and telecommunications with significant impact.

Acknowledgements

The author would like to thank all the reviewers for their support and advice, which contributed to and improved the quality of this paper.

Hitachi Vantara



Hitachi Vantara

Corporate Headquarters
2845 Lafayette Street Santa Clara, CA 95050-2639 USA
www.HitachiVantara.com | community.HitachiVantara.com

Regional Contact Information

USA: 1-800-446-0744
GLOBAL: 1-858-547-4526 HitachiVantara.com/contact

HITACHI is a trademark or registered trademark of Hitachi, Ltd. Pentaho is a trademark or registered trademark of Hitachi Vantara Corporation. All other trademarks, service marks and company names are properties of their respective owners.