WHITE PAPER

The Data Science Value Chain for the Intelligent Enterprise

Building an IoT roadmap with Hitachi Digital Manufacturing Suite for SAP® solutions

By Hitachi Vantara

January 2020

Summary

Technology advances, especially artificial intelligence and machine learning, are changing how manufacturers operate. Harnessing technology effectively and adopting Industry 4.0 are key factors for organizations to remain competitive. Hitachi worked closely with SAP to create accelerators on SAP Leonardo® that help manufacturers transform their businesses into intelligent enterprises, ready to compete in today's digital economy.

This paper explores why organizations should integrate advanced technologies into their manufacturing operations, introduces the Hitachi Digital Manufacturing Suite for SAP® solutions and details how Hitachi accelerators on SAP Leonardo help optimize the Industry 4.0 experience across an organization.

Introduction

It's true that our brains are absurdly marvelous and powerful and capable of astonishing leaps of intellect. For millennia, we have done quite well with our biological and metaphysical senses. It has been possible to devote time and resources to come to the conclusions and decisions we needed.

Over the last 30 years, we've seen technology crunch data and numbers with increasing speed and accuracy. The speed at which information is now available has turned us into data hostages. We must understand the data and know how to use it to stay ahead of the competition, to streamline operations, reduce costs and increase speed to market.

So, we've done what has worked in the past. We've planned and planned again; we've tightened every process possible to shorten the time to knowledge and resolution. But that reality is our past reality. It is not our present or future reality. We cannot process and analyze information fast enough for the rapid pace of data gathering and innovation that is around us today. We can't do it by ourselves. But we can do it with the help of one of our creations: artificial intelligence (AI).

A pragmatic approach with AI, machine learning and SAP Leonardo

Over the last decade, innovations in technologies have reshaped many aspects of how companies work and create their products–even leading to the birth of entirely new industries. Manufacturing organizations are going through major transformations as these technologies are integrated into the manufacturing experience, driving these enterprises to innovate to stay ahead of the competition. This shift brings AI and machine learning (ML) into the heart of the business to collect and analyze data and provide insights for informed decision-making and greatly improved outcomes.

We're already seeing the power of data science in a variety of industries. Forward-looking manufacturers are embracing Industry 4.0, which focuses on interconnectivity and AI with real-time data. An important advantage of Industry 4.0 is that it lets manufacturers continue to use their existing technology investments, connecting sensors and other equipment from their operations, and bringing modern instrumentation and smart sensors only when and where they are needed.

However, it takes time to understand these new technologies and digitally transform a business. To start developing your digital manufacturing landscape, we suggest you evaluate your available information, identify data that's missing, and gain better insight into how your factory is running today.

The power of experience

As one of the largest manufacturers in the world, and with 108+ years of experience building products across a wide range of industries, Hitachi has the know-how to help you reshape your business. We have developed a set of intellectual property and insights designed to empower organizations to create smarter manufacturing solutions.

Hitachi is working closely with SAP to help manufacturers like you create a comprehensive, effective ML and automation landscape–and a pragmatic internet of things (IoT) roadmap. In this paper, we'll explore how our complete suite of Hitachi accelerators on SAP Leonardo enables you to extend your SAP Digital Core, transforming your business into an intelligent enterprise for today's digital economy.

Advanced technology backed by deep experience

Hitachi's Digital Manufacturing Suite for SAP solutions is backed by intellectual property from our global solutions encompassing optimized factory, predictive maintenance and enhanced logistics. Hitachi brings more than 108 years of experience, working across a variety of industries, to develop specialized solutions for chemicals, aerospace and defense, discrete manufacturing, and other industrial segments.

The Hitachi accelerators on SAP Leonardo enable you to further extend these capabilities and shorten your time to value. The solutions address specific requirements for quality control and predictability; dynamic order scheduling; work cell monitoring and automation; automated visual inspections; global track and tracing of goods and operating assets; and more.

The accelerators are independent, yet they complement one another. These flexible offerings enable you to enhance your Industry 4.0 experience by choosing from a menu of options to address your needs for a modern, digitized, connected manufacturing operation.

Data science in the value chain

To drive operational excellence and improve decision-making, the Hitachi Digital Manufacturing Suite for SAP solutions helps you:

- Connect and visualize workers, material flow and manufacturing processes
- Digitize institutional knowledge
- Integrate with customers and suppliers across the value chain

The accelerators in the suite use specific predictive algorithms that provide insightful analysis of data to anticipate and proactively mitigate issues. The rapid throughput of data in motion and the large volume of data at rest require both powerful distributed computing and applied statistical analysis. Distributed computing architectures provide the framework for distributed storage and processing of big data using the MapReduce programming model, and advances in system software facilitate computation across any number of distributed nodes. Although this dexterity comes at a financial cost for both storage and compute, you can obtain many quantities of interest within a satisfactory uncertainty by sampling the population.

Through the application of applied statistical techniques and ML, you can enhance your business investment in your enterprise resource planning (ERP) platform and increase your return on investment by scaling quantitative analysis from the desktop to the operational data store, data lake and data warehouse. You can also use disciplined design of experiments (DOX) to select a refined simple random sample for desktop analysis to prove value and concept before you employ expensive, time-consuming distributed storage and compute calculations.

A proper predictive-analytical theory that uses ERP for applied statistical analysis will support the following capability maturity roadmap:

- **Descriptive analytics** of past system state, perhaps in near-real-time backward-looking resolution
- **Diagnostic analytics** for near-real-time backward-looking fault isolation, restoration and resolution
- **Predictive analytics** to apply information theory, statistical mechanics, deduction, simulation and statistical inference to predict near-term future events in real-time forward-looking resolution
- **Prescriptive analytics** for executive decision support to proactively optimize industrial processes with continual forward-looking resolution

We’re our own customer

Our suite is based on digital practices from our own top-performing manufacturing facilities. It is our framework for operational excellence across Hitachi Group companies. This approach helped to completely reshape the business of Hitachi Rail–a global leader in high-speed train manufacturing for passenger transportation—by making “trains to order” and providing a complete passenger experience as a service, including the assets involved in high-speed urban transportation systems.

The seven workflows of the Hitachi Digital Manufacturing Suite for SAP solutions

You can apply quantitative techniques to workflows throughout your enterprise value chain to monitor, tune and remediate quality, yield, efficiency and effectiveness. This enables you to enhance margin, profit, client satisfaction and brand loyalty and reduce costs, time to delivery, defects and idle time.

The Hitachi Digital Manufacturing Suite for SAP solutions offers analytics capabilities that span seven workflows:

1. Kepware simulation analytics
2. SAP procedural SQL stored procedures
3. SAP predictive analytics libraries to apply ML libraries
4. Extraction, transformation and load to the analytic data set
5. Extraction, transformation and load to dashboard staging
6. Maintenance to refresh ML models
7. Ongoing research

1. Kepware simulation analytics

The first workflow focuses on analytical theory to investigate and bound the predictive power of Kepware simulation data. Kepware sensor telemetry simulates the condition of a hydraulic test rig, based on multisensor data made up of streams of double-precision numbers. ZeMA gGmbH provides these raw data streams as a publicly available data set. They contain 2,205 instances of 17 sensors that measure physical attributes such as pressure, motor power, volume flow, temperature, vibration, cooling efficiency, cooling power and an efficiency factor.

This workflow also uses a profile file that classifies each of the production processes. Processes are expressed as a Boolean vector, with 0 for a normal successful production run and 1 for a defect. Using ML, the process evaluates the sensor telemetry to classify production runs as “in the class” for a defective production run or “without the class” for a normal production run.

Each sensor periodically samples its environment, using its own fixed sampling rate, as demonstrated in the chart:

Sensor	Physical quantity	Unit	Sampling rate
PS1	Pressure	bar	100 Hz
PS2	Pressure	bar	100 Hz
PS3	Pressure	bar	100 Hz
PS4	Pressure	bar	100 Hz
PS5	Pressure	bar	100 Hz
PS6	Pressure	bar	100 Hz
EPS1	Motor power	W	100 Hz
FS1	Volume flow	l/min	10 Hz
FS2	Volume flow	l/min	10 Hz
TS1	Temperature	°C	1 Hz
TS2	Temperature	°C	1 Hz
TS3	Temperature	°C	1 Hz
TS4	Temperature	°C	1 Hz
VS1	Vibration	mm/s	1 Hz
CE	Cooling efficiency (virtual)	%	1 Hz
CP	Cooling power (virtual)	kW	1 Hz
SE	Efficiency factor	%	1 Hz

In addition:

- Individual sensors broadcast to their own landing area, whether that is a data table in a data lake or a delimited file on a disk. For example, 17 sensors would provide 17 total data files on a disk, or 17 result sets from the data lake.
- Each table or file has one record per production run, resulting in 2,205 records per table or file.
- Each record for the telemetry data set represents one minute of history; for sensors that broadcast at 1 Hz, the dimensions of the file or table will be 2,205 x 1. For sensors that broadcast at 10 Hz, the data will land with dimensions of 2,205 x 10; and for 100 Hz sensors the dimensions will be 2,205 x 100.

The principal approach for developing an analytic data set (ADS) from streaming data is to invoke a sliding time window of petite width and apply a piecewise aggregate approximation (PAA) to the window for dimensionality reduction. For a large stream of double-precision numbers, a sliding time window is applied, and a single aggregate value is computed for the window, which gives immediate dimensionality reduction.

For example, if we have 60 seconds of telemetry from several sensors of different cut rate, say both a 100 Hz telemetry stream of 6,000 double-precision numbers and a 1 Hz telemetry of 60 double-precision numbers, and we wish to analyze the two streams pairwise, we could apply a sliding time window of one-second duration to the 100-Hz telemetry stream and calculate aggregates on the 100 values in each one-second sliding time window for dimensionality reduction of 99 percent.

Several characteristics of the data are important. Both parametric and nonparametric measures of central tendency and dispersion are calculated, as well as raw measures of skewness and kurtosis. The latter features guide the application of logarithm transforms to reconstitute Gaussian behavior in severely skewed raw data.

The outcome of this analysis is that, for each sensor, several candidate features are assembled and tested for predictive power, including:

- Minimum value
- First quartile
- Arithmetic mean
- Second quartile = median
- Standard deviation
- Inter-quartile range
- Third quartile
- Maximum value
- Area under the curve per unit time
- Maximum gradient within the window
- Minimum gradient within the window
- Gradient across the window
- Thermodynamic entropy

These engineered features are calculated for each sensor, and error conditions are trapped and eliminated. For example, for a sensor that broadcasts at 1 Hz, no measures of dispersion, gradient or integration are computationally available. The resulting ADS comprises the surviving engineered features for each of the 17 sensor-telemetry streams. This ADS is submitted to ML analysis.

Zeroth-order ML analysis and data science discovery occurs on the data scientist's laptop. Large data populations are sampled, and statistical-inferential methods are applied to obtain indicia of predictive power among the engineered features. The result of the data science discovery is a recipe that is transmitted to the SAP engineering team for implementation in the SAP ecosystem.

2. SAP procedural SQL stored procedures

The second workflow focuses on SAP procedural SQL stored procedures to craft custom algorithms and test harnesses. The initial touch point in the SAP ecosystem is an SAP procedural SQL stored procedure. This procedure gives the algorithm the appropriate chain of custody to bridge the data science discovery on the scientist's laptop in sample space with the SAP ecosystem at scale. Vendor implementations of ML algorithms are black-box implementations, and documentation is often obscure, if it is available at all. It's simplest for the quantitative team to code the algorithms from scratch, using SAP procedural SQL.

For this SAP accelerator project, this workflow focuses on coding the algorithm for classification. Through a simple coding workflow, the SAP SQL engineer executes the recipe comprising coefficients, parameters and hyperparameters provided by the data scientist. The SAP engineer then analyzes a set of test vectors to replicate the results at scale in the SAP system, as found by the data scientist at sample on the laptop.

3. SAP predictive analytics libraries to apply ML libraries

The third workflow employs SAP predictive analytics libraries (PALs) to apply ML libraries within the SAP ecosystem. Once the SQL stored procedure is tested and integrated, the team may invoke and tune the black-box algorithms of the SAP PAL. Each algorithm in the PAL is seeded with several parameters, so this workflow focuses on tuning the parameters to replicate the test vectors, as calculated both on the data scientist's laptop at sample and in the SAP stored procedure at scale.

Each algorithm in the PAL has defined data-ingestion staging and egestion landing tableaus within the table definitions of the SAP system. Much of this workflow involves preparing data for ingestion and tailoring the PAL calculation to consume it. The statistical advice of the PAL algorithms is confederated among several egestion landing tables and further processed to give a consolidated report of results.

After careful data preparation, algorithm parameter tuning and results consolidation, this workflow replicates test vectors among the laptop at sample, SQL at scale and PAL at scale for common algorithms.

4. Extraction, transformation and load to the ADS

The fourth workflow consists of extraction, transformation and load (ETL) from the edge to the ADS.

A principal workflow of this accelerator is preparation of the ADS. Although the ADS is little more than a database table or delimited file, it is also the result of disciplined, meticulous transformations. A typical ETL process–even a production-grade one–may leave NULL or blank values among the data, which are then tabulated by an online analytical processing (OLAP) cube. An ML ADS requires manicured data that is free from such imperfections.

This means ETL for data staging for ingest is the most critical workflow of this accelerator. There is no successful data science without meticulous, disciplined, journeyman data engineering.

In an industrial production scenario, sensor telemetry will be time-stamped, either by the sensor or the edge device. A typical edge device will serve as a sliding time windowing appliance and can calculate parametric and nonparametric statistics for data reduction before sending the reduced stream to the data lake for staging. This staged data is to be time-ordered and cleansed of imperfections before being staged for ingest. Failure of this ETL workflow can doom downstream ML and statistical inference.

5. ETL to dashboard staging

Our next workflow focuses on ETL of the ML output to the dashboard staging area. The data results of SQL-stored, procedure-based algorithms are defined by the data scientist and replicated by the SAP SQL developer. With a stored procedure-based algorithm, there is little reason to be confused by the resulting statistical advice. In the case of a PAL-based algorithm, however, the statistical advice is confederated among several egestion tables. Consult the documentation to identify the location of the statistical advice and to understand the format and form of the output.

Additional ETL is required to move the egested statistical advice to staging areas for consumption by the reporting engineer. The engineer will then bind the algorithmic classifications and scores to a dashboard widget for visualization and alerting.

This workflow requires close collaboration among the data scientist, data engineer and visualization engineer. In general, permissions between and among staging tables are nontrivial in an SAP environment. Organizations should identify a stakeholder for each constituent team of the blended team of SAP system administrators, data engineers and visualization engineers.

6. Maintenance to refresh ML models

The sixth workflow is built around maintenance to refresh the ML models for features with proven predictive value. On a timescale longer than a production run, but shorter than the time it takes to retool a production line, machinery behavior often drifts slightly within normal operating ranges. For example, for a production environment where a widget is produced once every minute, with production versions that last months or years, some drift could occur in as little as one week.

In this case, the same engineered features that are considered statistically significant for the initial release of this accelerator would also be significant in the slightly drifted, near-term operating environment. Mathematically, organizations should reanalyze the data *in situ*, recalculate the regression coefficients for the same statistically significant features, and publish the new regression coefficients for application interoperational ingestion.

The new coefficients are then regression-tested against near-term historical outcomes. Finally, the statistical quality of the prediction is marked to model for the engineering and production lab book.

7. Ongoing research

The last workflow focuses on ongoing research to sunset existing features that have lost predictive power–and discover new features with predictive power. Over a longer-term timescale, calculations should be performed from scratch to reestablish the statistically significant engineered features that characterize the drifted process going forward.

In this case, the first workflow explained in this paper, Kepware simulation analytics, is reperformed under the direction of a senior data scientist. The data scientist then records the statistical quality of the prediction, such as accuracy, true positive and true negative rates, and functions on the true positive and true negative rates, in the production lab book.

New regression coefficients, parameters and hyperparameters are then published to the SAP egestion tables for visualization interoperability.

Conclusion

Digital manufacturing offers rich possibilities for organizations seeking to build more efficient, agile, dependable manufacturing environments across the value chain. From the front office to the back office, the modern industrial enterprise requires finely tuned processes. Building on the experience and intellectual property we have acquired through our global solutions, Hitachi, in close collaboration with SAP, is poised to enable you to transform the manufacturing experience. Together, with the Hitachi Digital Manufacturing Suite for SAP solutions, we can position you to unleash the value of data science on your factory floor and beyond.

For more information, please contact info@hitachivantara.com.

About the Authors

Carlos Pinheiro Guimaraes Chief Technologist, SAP Practice, Hitachi Vantara

Carlos helps enterprise businesses drive value and transform with IoT, artificial intelligence, big data, blockchain and digitization. With more than 25 years of leadership experience managing sales and solutions teams, he is instrumental in creating Hitachi solutions for SAP products across industries including utilities, aerospace and defense, industrial manufacturing, luxury retail, and real estate.

Michael A. X. Izatt Chief Scientist, Insights & Analytics, Hitachi Vantara

Michael has more than 30 years' experience as a consulting physicist, statistician, actuary, financial mathematician and analytical programmer, working across a wide range of industries. He's a visionary analytics champion focused on delivering advanced quantitative methods to the operating business to help increase revenue and profit, decrease operating costs, and optimize organizational effectiveness.

About Hitachi, Ltd.

Hitachi, Ltd. (TSE: 6501), headquartered in Tokyo, Japan, is focusing on Social Innovation Business combining its operational technology, information technology and products. The company's consolidated revenues for fiscal 2018 (ended March 31, 2019) totaled 9,480.6 billion yen ($85.4 billion), and the company has approximately 296,000 employees worldwide. Hitachi delivers digital solutions utilizing Lumada in five sectors including Mobility, Smart Life, Industry, Energy and IT, to increase our customer's social, environmental and economic value. For more information on Hitachi, please visit the company's website at www.hitachi.com.

GDC Jan 2020