WHITE PAPER

INTELLIGENT BUSINESS STRATEGIES

Edge Advanced Analytics

By Mike Ferguson
Intelligent Business Strategies
December 2018

Prepared for:

tellmeplus Predictive Objects

Table of Contents

AN INTRODUCTION TO IOT

Many organisations want to improve operational efficiency but lack of data has often made this difficult

In many businesses today, there is a huge demand to make use of data to improve efficiency of business operations in order to reduce cost and maximise throughput. There is also a need to minimise operational risk. Up until recent years, monitoring business operations has been challenging, primarily because of lack of data. Data warehouses have offered little help because they have been more focussed on business data which usually encompasses marketing, customer transaction activity and finance.

Telemetry embedded in consumer products and industrial equipment has made it possible to collect data from an Internet of Things

However today, the ability to deploy telemetry (sensors) in consumer products and in business operations has enabled companies to collect the data they need to monitor and analyse usage behaviour and operational activity to provide the insights needed to better support operational decisions.

This new capability has spawned a new era in computing. The era of the Internet of Things (IoT). So far, IoT has resulted in billions of devices now being online with billions more still to come. Much of the focus around IoT is consumer devices. However, the focus of this paper is Industrial IoT - the use of telemetry in the manufacturing industry. For example, to monitor and manage supply chains, production lines, asset utilisation and distribution to maximise business benefit.

Analysis of IoT data in industry can be used to optimise supply chains, maximise asset performance, improve manufacturing and distribution as well as reduce risk

Industrial IoT is becoming strategically important in many industries. Broadly speaking the objective is to collect, analyse and use data to optimise efficiency and to minimise risk by preventing things from happening.

For example, Industrial IoT can be used for:

- Asset management and field service optimisation
- Manufacturing production line optimisation
- Health and safety risk management
- Minimising unplanned operational cost
- Distribution optimisation

IoT is enabled by edge devices, edge gateways, IoT platforms and analytics

IoT platforms link edge devices and edge gateways to the enterprise

Many different things can be measured and analysed using IoT data to help optimise business operations including vibration, pressure, temperature, and movement

IoT is enabled by sensors in edge devices, gateway devices, IoT platforms and analytics. Deployed sensors in industry operations enable companies to measure everything from vibration, movement, pressure, temperature, location (e.g. GPS in vehicles), light, chemicals, airflow, liquid flow, and more. Edge gateways are often used to collect sensor readings from edge devices. They typically batch them up before sending them on in single timestamped messages via IoT platforms into the data centre or cloud for further analysis. The IoT platforms therefore link edge gateways and devices to the enterprise and provide services such as data collection, communication, enterprise security, automatic device discovery, device data schema registration, and configuration management. They allow data experts to access services to leverage their data and provide access to edge infrastructure. It is this infrastructure that permits edge computing to occur close to where the data originates at much less cost than going to the cloud. For example, lightweight data preparation and analytics can run on IoT platforms, edge gateways and in some edge devices. The last component is analytics – the ability to use machine learning to build predictive and prescriptive models to automatically analyse IoT data looking for patterns that can drive actions. These models can be developed centrally in the cloud or in the data centre but are typically deployed in edge gateways to run close to the data they are analysing.

In manufacturing, companies can create models to analyse sensor data to predict how equipment will perform

Manufacturing, transportation, logistics, retail, utilities and oil and gas are already heavy investors in Industrial IoT. In manufacturing, analysing industrial sensor data can help companies create models and run simulations on assets to prevent equipment from failing, predict problems and optimise performance on a continuous basis. This means companies can make use of machine learning to develop predictive and prescriptive models to optimise operations and reduce risk. In addition, it becomes possible to measure the time to perform manufacturing tasks and on that basis, to measure and predict throughput. This enables companies to automatically identify and predict bottlenecks in a production line and predict high cost activities. By pairing this with rules for decision management, it becomes possible to automate decisions to help improve overall business performance on a continuous basis.

THE CHALLENGE OF IOT

Along with the benefits that industrial IoT brings, there are many challenges associated with devices / equipment where telemetry is deployed, the data being collected and where it is best to process and analyse this data.

Automated device discovery is critical in organisations with a large number of assets in business operations

With respect to devices and equipment, a key challenge is keeping pace with what is deployed. This is especially true in organisations with multiple manufacturing plants, or a large number of assets (equipment, machines, processes etc.) deployed in field operations. Keeping pace with the number of devices and associated schema of data they emit, makes connectivity a challenge. IoT platform automatic device discovery is the only practical way to handle this.

IoT data poses a number of challenges

The data never stops

Newly deployed assets can introduce new schema and formats without notice

Asset upgrades can introduce schema changes without notice

With respect to IoT data, itself, there are a number of characteristics that really challenge the way we have done things in the past. Key characteristics include the fact that, IoT data is continuous, streaming, real-time, high velocity data. It never stops. Also, it is potentially huge in volume and with so many types of sensors and devices, there are no standard formats. Therefore, newly deployed assets can introduce new formats and new schema without notice. Equipment upgrades can also result in schema changes without notice as new telemetry replaces old. Also, data can be missing, arrive out of sequence and in bursts. In addition, timestamps associated with devices readings are often out of sync. To get around this problem, device gateways are often used to collect data from multiple devices and create larger messages containing multiple device readings and a single timestamp. Even then, problems can still occur. For example, data emission delays can result in the number of readings in each gateway message varying from message to message.

Model development may take place centrally but data can be analysed more rapidly at the edge close to where it is created

In terms of where industrial IoT data should be processed, collecting data and bringing it into a data centre or into a cloud for central processing is possible but not always practical, secure or fast. For example, if readings indicate a high risk equipment failure then why should you have to wait for data to be collected and brought into a central data store or platform for analysis when it can be done more rapidly at the edge close to where data is created and potentially in real-time? Also bringing very large data volumes of data to a central data store isn't necessary when in many cases it could be analysed at the edge. Edge analytics are inevitable and are a crucial part of an IoT solution but given this situation, there are still challenges.

Need to take the analytics to the data rather than the data to the analytics

Deploying analytics at the edge means that thousands of models may need to be managed and monitored which is impossible to do manually

Edge analytics means taking the analytics to the data rather than the data to the analytics. That implies that models will need to be developed centrally and deployed at the edge. Given the number of assets that may need to be monitored at the edge, it means that an industrial machine learning process is needed to rapidly develop, test and evaluate analytical models, deploy them into the network, manage all the deployed models and continuously monitor their accuracy. Furthermore, given that analysis will happen locally, actions will also need to be taken locally either automatically or by operations personnel based on analytical model predictions and scores. This means there is a challenge to ensure that computational resources are available at the edge to support this, in the equipment itself (if practical), on device gateways and on IoT platforms. Also, all of this needs to happen in a secure environment.

EDGE ANALYTICS PROCESSING REQUIREMENTS

A number of requirements need to be met if edge analytics are to deliver business value

Given the challenges that industrial IoT brings, what then are the requirements for edge analytical processing that need to be met to enable efficient, safe, optimised operations in vertical industries like manufacturing, logistics, utilities, oil and gas etc.? The following list provides some of the core requirements for analytical processing at the edge.

It should be possible to:

Automatic discovery of devices and schema

Data needs to be transformed at the edge

Trained analytical models need to be deployed at the edge

- Use an IoT platform to automatically discover, classify and register industrial IoT edge devices or gateways that collect data for analysis.
- Define and/or automatically discover, catalog and manage schema for data emitted by each type of registered device.
- Automatically detect new devices or upgraded devices in the network, devices that are now offline, device data schema changes (also known as schema drift).
- Pre-process and transform raw data at the edge.
- Integrate raw data from multiple edge devices or assets in the same class in preparation for analysis e.g. to understand how an asset in a manufacturing process is performing.
- Integrate raw data with enterprise data (e.g. on assets) at the edge.
- Manage master data at the edge to apply context during analysis.
- Develop and/or automate development of analytical models either on-premises or in the cloud for deployment at the edge.

Edge rules also need to be defined to manage local decisions and actions

Rule management and rule version control in the network is key

Model management and version control of models deployed in the network

- Deploy analytical models to edge gateways and devices for execution.
- Manage analytical models and model versions running in the edge.
- Define and manage edge rules for execution at the edge including support for version control and parameters.
- Activate and de-activate edge rules.
- Define business conditions (e.g. in a manufacturing process) indicated by patterns in the data.
- Define the actions to take at the edge if a business condition is met.
- Define conditions (rule combination) upon which actions should be taken.
- Drive automated actions from analysing data at the edge e.g. threshold alerts.
- Define who should be using the results of edge analytics e.g. shop floor supervisors, asset operations managers, field service personnel etc.
- Track actions to provide an audit trail for governance and to provide lineage to understand what action(s) were taken and why.
- Filter, reduce and send data to a cloud or data centre for further analysis.
- Automatically monitor the accuracy of models deployed at the edge and trigger automatic re-training and re-deployment if it drops below thresholds.

Edge analytics need to be automatically monitored and retrained if they become stale

In addition, it is important to appoint an owner of an edge analytics programme responsible for defining and prioritising model development. They should also align model development with business strategy to maximise business benefit e.g. to enable preventive maintenance of assets, maximise throughput and avoid unplanned operational cost. They should also be responsible for managing the development, deployment and monitoring of edge analytical models in the enterprise and for integrating edge analytics with central analytical environments. This is needed to enable development of analytical models that can run at the edge, in-stream, in-Hadoop, in-memory and in database for end-to-end analysis coverage of core business problems.

An Architecture for Industrial IoT Analytics

Machine learning automation accelerates development of edge analytics

Model development can happen centrally with trained models being deployed at the edge

In terms of an architecture, Figure 1 shows how machine learning automation can be used to expedite the development of analytical models that can be deployed at the edge in supply chains, manufacturing production lines, field operations, and distribution chains. In addition, models can also be developed for deployment in the cloud or in on-premises data centres to look for patterns associated with recurrent business conditions over longer periods of time. In practice both are likely with edge analytical models continuously monitoring assets in operation for conditions like propensity to fail in order to prevent costly, unplanned outages and asset performance to maximise productivity. Edge analytical models analyse data in real-time taking action when a single instance of a pattern (e.g. predicted manufacturing delay) is detected while offline central analysis of this data allows things like recurrent more serious problems to be identified over longer time periods (e.g. predicting recurrent equipment failure and high risk, high unplanned cost business impact).

Machine learning automation can be used to develop models to be deployed anywhere in the enterprise

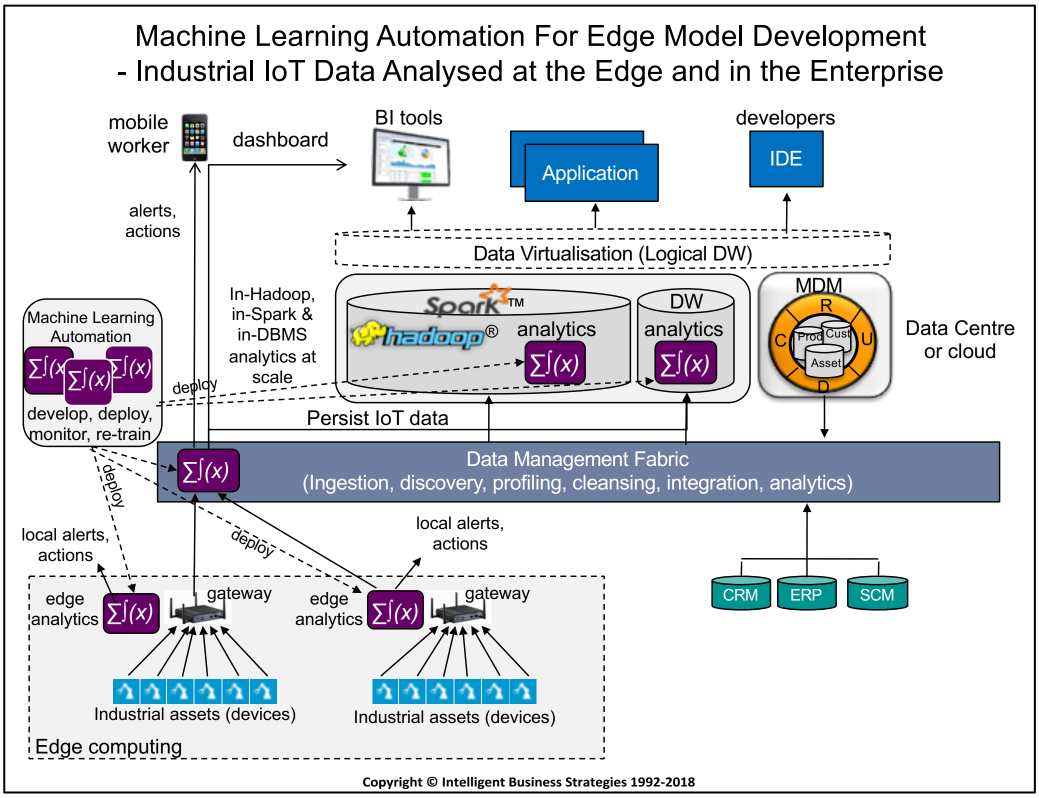

Figure 1

Developing and Deploying Edge Analytics Using Tellmeplus Predictive Objects

Having understood the challenges and requirements associated with analysing industrial IoT data at the edge, this section of the paper looks at how one vendor, Tellmeplus, is using its Predictive Objects software platform to develop and deploy edge analytics.

Tellmeplus Predictive Objects

Tellmeplus Predictive Objects is a machine learning automation platform designed to accelerate development and deployment of edge analytics use in industrial IoT environments

Tellmeplus's flagship solution is Predictive Objects which is aimed at business. Predictive Objects is machine learning automation software that uses artificial intelligence and allows organisations to speed up development of predictive and prescriptive analytical models. It can also manage the deployment of these models to industrial assets and edge device gateways to continuously analyse and act on IoT data in real-time. Model deployment to the edge is done in conjunction with industrial IoT platforms. Predictive and prescriptive analytical models can also be deployed to the cloud and on-premises data centres. In addition, it can manage the complete lifecycle of machine learning models by automatically monitoring deployed model accuracy and triggering retraining and re-deployment of any models where accuracy has dropped below user defined thresholds. This keeps models fresh and fit for purpose in everyday business operations.

Automated Model Development

Tellmeplus Predictive Objects organises model development into projects

Looking at Predictive Objects in more detail, development of a new model for example, to predict the propensity of an industrial asset to fail, starts by first creating a project.

Source data is imported into Hadoop HDFS

Automatic data discovery helps quickly understand source data

Time-series calculations can be performed on IoT data during data preparation to create a stronger data set for use in training models

Within a project, you can define one or more data sources from which data is taken and prepared for input into the model algorithm. Source data is imported into and persisted in Hadoop HDFS from where Predictive Objects automatically discovers, profiles, infers schema and displays statistics about the data to speed up the data expert's ability to understand the data available. The ability to support multiple sources in a project means that both univariate[1] and multi-variate[2] analysis is possible. In the context of industrial IoT, telemetry data sources typically hold time-series data and so Predictive Objects allows business experts to define windows on the data over a period of time and then define time series calculations to be done at runtime on the data within the window to add useful new data to a dataset prior to the data being fed into an algorithm. This helps improve accuracy and also allows Predictive Objects to detect and predict more complex business conditions indicated by patterns and derived calculations found in data. This is especially important when data coming

---

[1] Univariate analysis looks at data from a single data source (e.g. a temperature sensor) to identify patterns

[2] Multi-variate analysis is looks across data from multiple data sources (e.g. temperature sensors, pressure sensors, asset master data etc.) to identify patterns

Event sequences can be specified / outlined to define patterns in the data that denote a business condition

from multiple data sources. Also multiple time windows can be defined on data which allows data experts to look for one or more patterns in the data at the same time.

In addition, it is also possible to define sequences of events that indicate a pattern in the data.

Predictive Objects automatically generates, trains, tests and evaluates predictive models, comparing accuracy before recommending the best model

Once this is done, Predictive Objects automatically runs a simulation process to generating Scala to train, test and evaluate multiple models passing the prepared data into different algorithms to determine the best model. Business analysts can manually configure how long the process runs for by selecting a small, medium or large simulation. Small simulations are good to get an overview of model performance. However, medium and large simulations will produce more accurate models but will run for longer. At the end of the simulation process, Predictive Objects shows the evaluations of all the models it generated, the most important data attributes contributing to the prediction, and visualisations to such as lift curves, ROC curves and model metrics to explain the accuracy of each model. A predictions chart also shows the times when business conditions were predicted which could relate to specific aspects in a business operation. Finally, it also recommends the best model. The result is not only rapid development but also a relatively easy way to compare multiple models to understand why Predictive Objects has recommended a model as the best one.

It can also explain model accuracy by indicating the relative importance of each attribute to a prediction and by providing visualisations to see accuracy at a glance

Model Deployment

Predictive Objects can also automate deployment of models on-premises, in the cloud, on IoT platforms or on edge devices

Once a best model has been selected, Tellmeplus Predictive Objects can be used to deploy the model through an IoT platform or directly to run in real-time at the edge of the network either within one or more specific industrial assets that need to be monitored or on an edge gateway device located near the asset(s) that need to be monitored. To deploy a model, Predictive Objects 'pushes' the model into production by creating a REST API to it. It can be deployed in a Docker container to run on-premises, in any cloud environment, on IoT Platforms, in an edge gateway device (e.g. a Cisco router by generating code that is natively integrated into a Cisco Kinetic Dataflow solution deployed at the edge) or an specific asset. Single model deployment can be done by Tellmeplus. However industrialised deployment of models to many different devices is done in conjunction with an IoT platform. IoT platforms that Tellmeplus works with include Atos Mindsphere, Cisco Kinetic, GE Predix, Google IoT Core and Microsoft Azure IoT.

Several IoT Platforms are supported

Automated Management and Monitoring of Edge Analytics

Predictive Objects can automatically monitor the accuracy of deployed edge analytics and trigger automatic retrain and redeployment when accuracy drops below user defined thresholds

Once models have been deployed, Tellmeplus Predictive Objects can continue to manage and monitor their accuracy. This is done by setting accuracy thresholds to monitor model accuracy against. Model accuracy is then continuously measured and when it drops below a set threshold, Predictive Objects can automatically trigger retraining and redeployment to keep models running at optimal accuracy levels.

TELLMEPLUS EXAMPLE USE CASE - AI AT THE EDGE

One example of Tellmeplus Predictive Objects in action is in the area of predictive maintenance and failure prevention for remote heavy equipment.

Predictive Objects is used to analyse sensor data and data on operating conditions to develop a model that can be used to prevent equipment breakdown

A global natural resources company is operating a fleet of heavy equipment (trucks, cranes, loaders, etc.) in several locations around the world. Most of these locations are remote, badly linked to maintenance infrastructures, and have sporadic/unreliable connectivity.

The challenge was to prevent equipment breakdown, a very costly occurrence when it happens away from repair facilities (cost of transporting the equipment for repair, loss of productivity on the site, etc.)

The model is deployed to run in each asset on an IIoT gateway embedded in the equipment

Data is analysed in real-time and when failures are predicted, equipment operators, site managers and central operations teams are notified and preventive maintenance performed

Predictive Objects collects sensor data from the equipment (temperature, vibrations, pressure, noise, etc.) as well as operating conditions (load weight, speed, elevation, wind speed, etc.). The model, calculated in the cloud to take advantage of all data from all assets in the same class, is then pushed down to each asset, running at the edge on an IIoT gateway embedded in the equipment - or in the case of smaller/less modern equipment, on a network router local to the facility.

Failures and dangerous, unsafe operating conditions are therefore predicted locally and in real time even when the network is down. They are communicated to the equipment operator if applicable through a local onboard computer integration, and used by site managers and central operation teams to remove equipment from service and perform maintenance before a failure happens.

CONCLUSIONS

OVERALL BUSINESS BENEFITS OF THE TELLMEPLUS PLATFORM

Industry use of IoT data gathered from telemetry is growing rapidly

As more and more organisations look to optimise their business operations, we are seeing a major increase demand for industrial IoT telemetry to capture the data needed to monitor and analyse efficiency and help minimise costs. Over the last several years this has been done by collecting data and bringing it all to a cloud or an on-premises data centre for big data analysis. In other words the data has been brought to a central location where analysis is done at scale. This increases data latency and could delay decisions that impact performance of every-day business operations.

As the number of devices grows, so does the demand to analyse data at the edge and not just at the centre

Management of deployed models is a critical success factor

As the number of devices grows, and IoT data volumes increase, bringing all this data over the network is less scalable than analysing that data at the edge near to where it is being generated. It makes much more sense from both a scalability and a responsiveness perspective to take the analytics to the data rather than the other way around. Deploying analytics at the edge is therefore the future in both consumer and industrial IoT. However to do this, means that potentially thousands of models will need to be developed, deployed, continuously monitored and managed. The only way to manage this complexity is via a machine learning automation platform.

Tellmeplus Predictive Objects automates the development, deployment, monitoring and retraining of analytics used in industrial operations

Tellmeplus Predictive Objects is an example of such a platform. It can be used to manage the entire model lifecycle, accelerating model development and edge deployment when working with IoT platforms. This speeds up projects delivery and reduces time to value. Once deployed, models become analytical assets of the company. In addition, it can also automatically monitor all analytical models that have been deployed at the edge retraining and redeploying them when their accuracy levels become suboptimal. It is this industrial approach to end-to-end model lifecycle management that makes Tellmeplus a serious contender in helping organisations reduce time to insight, conquer complexity and confidently manage a portfolio of edge analytical models deployed across a network of business operations.

About Intelligent Business Strategies

Intelligent Business Strategies is an independent research, education and consulting company whose goal is to help companies understand and exploit new developments in business intelligence, machine learning and advanced analytics, data management, big data and enterprise business integration. Together, these technologies help an organisation become an *intelligent business*.

About Tellmeplus

Every organization, efficiency depends on how well assets and processes are being used and leveraged. Predictive Objects helps businesses utilize AI to dramatically improve their efficiency and performance by embedding intelligence in every critical asset. The SaaS platform leverages machine learning, AI and big data to automate the creation and deployment of predictive models for faster, more accurate predictions, in order to achieve ubiquitous asset intelligence.

Contact : communication@tellmeplus.com

Website: www.tellmeplus.com

Phone: +33 1 39 62 43 64

Author

Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in business intelligence and enterprise business integration. With over 35 years of IT experience, Mike has consulted for dozens of companies on data strategy, big data, machine learning and advanced analytics, data governance, master data management and enterprise architecture. He has spoken at events all over the world and written numerous articles. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the Relational Model, a Chief Architect at Teradata on the Teradata DBMS and European Managing Director of Database Associates, an independent analyst organisation. He teaches popular master classes in Data Warehouse Modernisation, Designing, Managing and Operating a Multi-Purpose Data Lake, Machine Learning and Advanced Analytics, Big Data and Analytics Fundamentals, Enterprise Data Governance, and Master Data Management, Modern Data Architecture and Real-Time Analytics and Operational BI.

Telephone: (+44)1625 520700
Internet URL: www.intelligentbusiness.biz
E-mail: info@intelligentbusiness.biz

Edge Advanced Analytics