A white paper

Databases - A simple way of achieving Big Data in a production environment

Author: Harry Mulder

Executive summary

With the utilisation of digitisation in production environments, it seems more producers than ever are looking to amass Big Data for analysis. However, details of solutions and actual implementations are somewhat vague.

This paper shows how the latest industrial control systems come with built-in SQL database connections that allow users to make a start at collecting Big Data for themselves. Robust systems that are flexible and highly scalable, but still very cost effective can be deployed quickly and easily; enabling users to tap into the many benefits of Big Data.

Data is everywhere!

Data is everywhere! The information age is well and truly upon us, with the ushering in of *Industrie 4.0* – the fourth monumental phase of industrialisation over the last 200+ years. With the arrival of the IIoT (Industrial Internet of Things), connectivity to the internet is now a given, leading to the possibility of Big Data collection, and even cloud-based computing for both the storage and analysis of this data.

And it seems the sources of data are ever increasing. In the past an entire machine may generate limited production data. But now individual sensors, which traditionally only output digital on/off signals, can spawn continuous streams of data[1]. And as data sources increase, the amount of data generated increases exponentially.

But why?

Why have all this data and what to do with it? After all, it's not the amount data that's collected that counts; it's what's done with it. The real value of Big Data is not the data itself, its gleaning useful information that's hidden within it. It's the analytic processes that extract this information.

Competition between producers is demanding ever greater output with better quality, while costs must be reduced at the same time. So the drive for efficiency is on with unprecedented zeal. In the past, manufacturers relied on *gut feels* or *hunches* of plant floor personnel for suggestions on how to improve. But as many plants have found, processes of continuous improvement that are based on human judgements have stalled due to the complexity of the systems used in modern day manufacturing.

[1] One such example is the IO-Link international standard: IEC 61131-9.

The way forward, to gain the next round of quantum improvements, is the digitisation of a wide variety of parameters and the measurement of them over a long time. This is necessarily results in the accumulation of Big Data, and is often beyond the capacity of simple logging systems. However, raw numbers are almost meaningless to humans - data needs to be turned into information in order to be useful. This is where computer-based analytic tools become mandatory. Humans only come into play when results need to be analysed and changes actioned.

The application of Big Data in the manufacturing and process industries, and the advantages it brings, is slowly being understood by plant management. With the ability to now monitor not just every piece of equipment but every sensor as well, an accurate picture of what's actually happening in the plant floor can be painted. Many production managers have learned much from this type of scrutiny and many myths have been shattered; all thanks to Big Data.

It's not the amount data that's collected that counts;
it's what's done with it.

But not only has performance of individual machinery improved, by analysing flow of production throughout the plant, bottlenecks have been detected and overall productivity lifted. This has resulted in reduced manufacturing costs and improved competitiveness in the market.

Traceability and Preventative Maintenance

Apart from productivity improvements, Big Data has also opened up the whole area of traceability – the tracking of products throughout the sales channel and into the market place. Traceability helps overcome product counterfeiting, improve customer feedback and better manage product recalls.

Big Data also aids predictive maintenance, where machinery (or components) undergoes maintenance or replacement after a pre-set time, regardless of whether it's still operating correctly or not. The time period is determined by the original manufacturer, who knows the expected life cycle of the article. The work is done during scheduled down times and should result in less age-related breakdowns; which tend to happen at the most inconvenient times! Cost savings are enjoyed in the long term due to reduced downtime.

None of this would be possible without Big Data.

But how can it really be done?

So the benefits of Big Data are clear; the question moves to one of how to acquire it, and what's involved in realising it. It's at this point that discussions often become veiled and obscure. Perceived difficulty of implementation is one reason many plant managers have yet to embrace Big Data with actual deployment, despite seeing its benefits.

Some of the solutions available come from industrial automation vendors and are necessarily proprietary. They not only lock away the data making it inaccessible to the owner, but the methods of generating the data are often deliberately abstracted away, forcing owners to use services provided by the vendor. This is all to maintain that particular vendor's interest.

Cost structures are probably the biggest hindrance to adopting vendor-specific Big Data systems. Yet it is well known that cloud storage is relatively inexpensive, as is the use of analytic tools. And even as the amount of data concerned increases substantially over time, costs do not rise appreciably.

It need not be hard

However, dealing with Big Data and analytics need not be difficult. Nor do costs need to be prohibitively high. There are open solutions on offer that are easy, robust and well tested, but yet still very economical.

Databases have been available for many years and have proven to be capable of handling enormous amounts of data. They can now reside on the cloud, where they can manage proportionately larger amounts of data. They are an open technology approach to Big Data, as they are well known by IT specialists. Traditionally, they have not been the domain of the OT (Operational Technology) professionals in manufacturing, but this is starting to change with the convergence of IT and OT professions.

Data needs to be turned into information in order to be useful.

Internally, relational databases (where all data is organised in a series of tables) are almost universally accepted and SQL (Structured Query Language, pronounced "sequel") has established itself as *the* language for accessing the data. Furthermore, databases can export data to other software products ranging from simple spreadsheets (e.g. MS-Excel) to very sophisticated report generating programs[2] These programs provide both data analysis and its visualisation – the open display to those who can interpret it. Graphical means are utilised in visualisation.

Using commercial databases greatly reduces costs barriers. They come standard with features like data robustness and security but yet are still relatively inexpensive[3]. They often support a bewildering array of features, including analytics – far more than most applications will ever need! Also, as their configuration and the SQL dialect are well understood, the need for dedicated professionals can be somewhat mitigated.

---

[2] Many Report Generating software packages exist, both free and commercial: https://en.wikipedia.org/wiki/List_of_reporting_software. A discussion of them is beyond the scope of this paper.

[3] Microsoft SQL Server Express is freely downloadable and handles databases up to 10Gb, albeit with numerous performance and other constraints. MySQL by Oracle, is another database package that has a free edition. While neither may prove suitable for a true industrial application, they do at least demonstrate that databases can be sourced for reasonable cost.

The current situation

Manufacturing has been dominated by industrial controllers (such as PLCs) that perform discrete control. These controllers have proved themselves to be highly reliable, offering many years of trouble-free service. Being at the heart of manufacturing, these controllers hold the key as they contain the data that needs to be *mined* (i.e. extracted) for storage and analysis.

Over the years, these controllers have become more and more powerful, and taken on more and more functionality. These days, Machine Automation Controllers (hereafter MACs, shown in Figure 1) have gone beyond normal PLC functionality, to include motion control, safety and even robotic functionality! They also support high-speed field networks (e.g. EtherCAT) and high capacity data networks, such as EtherNet/IP. However, databases do not use these well-established industrial communication protocols, and are therefore unable to communicate to industrial controllers. This has thus far been the missing link.

Figure 1 – Omron's MAC: Machine Automation Controller

However, in the case of Omron's NJ series MACs, their functionality has been further enhanced by support for SQL client functionality. The network ports of these controllers are able to directly connect to up to three SQL database servers[4], without any intervening hardware or software. This means the controller can execute its normal program and simultaneously exchange data straight with the SQL server. This allows the logging or recording of a huge, almost limitless amount of production data. And once data has been captured, in-depth analysis can begin. Figure 3 shows a simple topology with a single MAC (client) connected to one SQL server[5].

Database architecture

SQL databases commonly use a client/server model, where client and server nodes are distributed across a network, typically using an Ethernet connection. One or more server node(s) are data repositories, while numerous client nodes access and manipulate this data. Servers can be either local or reside on the cloud, in which they connect to client nodes via the internet, through a firewall.

MACs connect directly and seamlessly to SQL servers, as client nodes. Having a connection straight to a server, as opposed to middleware or any other hardware or software, offers several immediate advantages, like a simpler and far more reliable system. The connection is also faster and more cost effective, as there are fewer components.

[4] The database vendors supported by Omron are: Microsoft SQL Server; Oracle Database and MySQL Community Edition; IBM DB2 for Linux, UNIX and Windows; Firebird and PostgreSQL.
[5] It's also possible to incorporate an Omron visual inspection system, which writes image files directly to the database. Data from the controller can be collated in the database with images received from the camera.

OMRON

Programming wise, MACs comply with the well-established IEC 61131-3 and PLCopen standards. These define, among other things, data types and Function Blocks (hereafter FBs). Here, additional FBs have been created to handle database operations, like opening and closing connections, formatting data structures and mapping transfer variables. Users do not need to learn SQL commands, as common functions like Insert, Update, Select and Delete are embedded into the FBs provided.

By utilising the existing programming standards, the database implementation in the MAC becomes open and easy to understand. It is not an esoteric system, set out by one particular vendor.

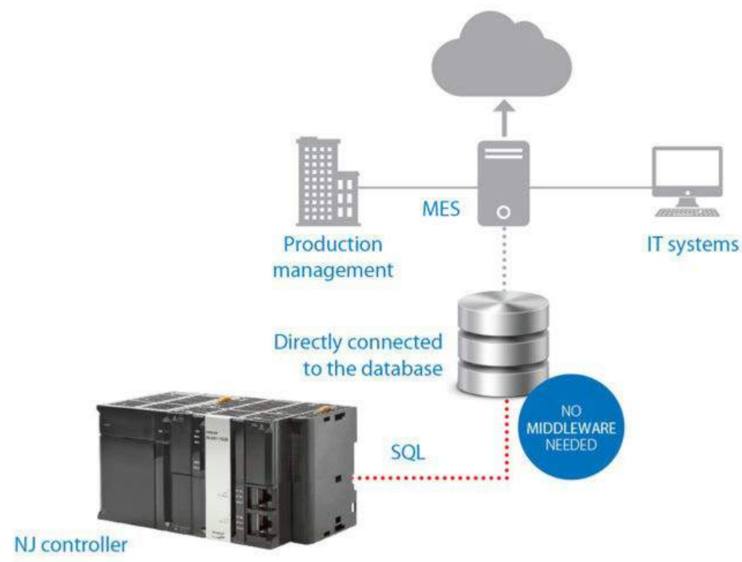

Figure 2 – Basic system diagram, showing a single SQL client node, the NJ Controller.

Depending on requirements, such as amount of data and analysis that needs to be undertaken, it may be advantageous for the database server to be hosted onsite, rather than in the cloud. This may be suitable for simpler or smaller applications. The scalability and flexibility provided by the database's client/server architecture is such that additional servers can easily be added, and at any time. Regardless where the additional servers reside, the rest of the hardware does not need to be retrofitted. Significantly, client nodes such as MACs do not need to be disturbed from plant control when changing server configurations; they can continue running.

Most importantly, databases put the end-user in control of their own data. Users are able to maintain the configuration, analyse the data produced, and make changes as they see fit.

Further advantages

Industrially hardened controllers like MACs are fault-tolerant and are more dependable SQL clients than PCs. Also, as data comes straight from the source – the controller - all programming and configuration is contained in one IDE (Integrated Development Environment). This is far easier to manage than having multiple, dissimilar devices distributed across the network, each with their own configuration methodology.

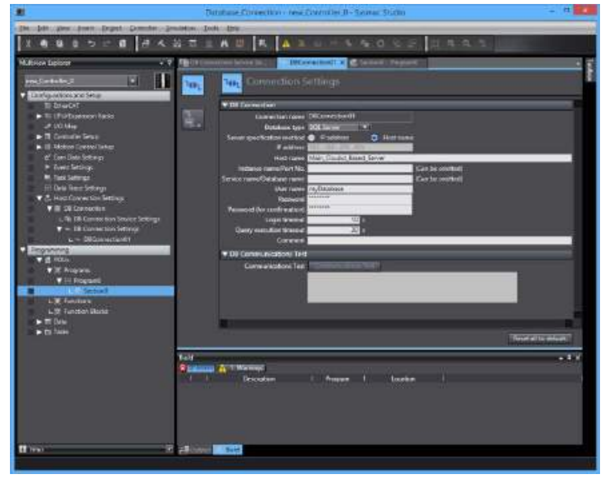

Figure 3 – MAC Database Setup screen in SYSMAC Studio

In the event of the server being unavailable or the connection to it being down, the MAC is able to spool a large amount of data and commands. These are transferred and executed when the server or its connection is restored. This means that no data is lost, even without an operational server.

Speed of data access is very fast, down to 10-20 mS for data inserts. This allows real-time recording, which some applications require. Gateways and alike on the other hand introduce latency.

Security of data is built into the database design, which is vital for servers that are hosted over the internet. Individual user access to the data is also controlled by normal password mechanisms.

As MACs are only database clients, mostly sending data to the server, future-proofing can be assured. Servers are subject to both minor updates and major upgrades, both of the software and the operation system used. Server hardware will also become dated or superseded and will eventually need to be replaced. But changing a server will not affect any clients, which work independently of the server. Clients are abstracted away from the server and are often unaware of its structure – they merely see an interface to it.

More importantly, databases put the end-user in control of their own data.

Non direct connections

While it may be possible to use an intervening data concentrator, middleware or some other gateway to collate data from controllers, such architecture introduces numerous drawbacks.

Most importantly, any additional hardware adds directly to costs and occupies cabinet space. Reliability is also reduced by having additional hardware and connections.

Furthermore, latency, (i.e. the delay in getting data to the collector and forwarding it onto the server) is introduced. These delays are unavoidable but their inconsistency destroys determinism which can preclude usage in real-time applications.

As the data collector runs on separate hardware from the main controllers, they will need their own cyber security and virus protection systems. These data collectors are also subject to their own maintenance schedules. Upgrades of its hardware and/or software need to be done separately from the main controller, making maintenance more difficult.

The best and most efficient method remains for the controllers to support client operations natively, and to have a connection straight to the database server.

Putting it altogether - a real live application at Omron's Kusatsu Factory

We at Omron have also found process improvements had stalled in our own factory in Kusatsu, Japan. To achieve further productivity gains, a far deeper reaching, digital analysis was needed – in other words Big Data.

Figure 4 – Omron's Kusatsu factory, in Japan

A NJ series MACs was installed to collect data from the factory floor, for loading into a SQL Server database. A Microsoft Big Data solution was used, including PowerPivot (an add-on for Excel that enables millions of rows of data to be imported) and Sharepoint to openly share files and visualisation data. MS-Excel was chosen as the main BI (Business Information) tool as it's so widely understood. This enabled data and formatting to be manipulated most easily by various staff, who often had very different areas of interest. The idea was to display the results very simply so that as many people as possible could see what was happening and make suggestions for improvements.

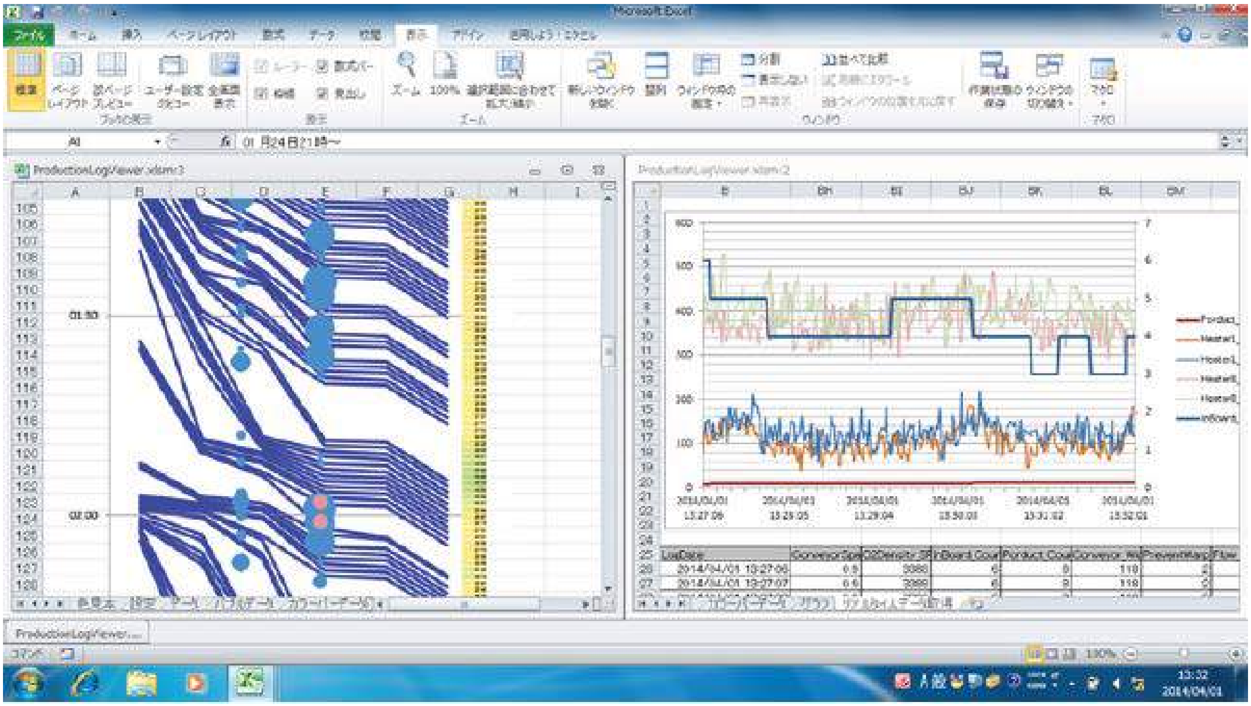

Figure 5 – A visualisation display, from MS-Excel

One of the resultant displays is shown in Figure 5. The diagram on the left hand side presents production. The tighter the lines are clustered together, the better the throughput. Larger gaps between the lines indicate hold ups and bottle necks. The diagrams on the right show the corresponding overall effect on production over time.

The end results have been outstanding, with the time taken to examine data and make improvements being reduced by a factor of six!

“Big Data has existed for a long time in production lines, but the environment to use it effectively has not been created,” observed Mr. Takeuchi, Automation Systems Division HQ Senior General Manager of Omron. He continued “Data does not make sense until people who support production sites read and use the data. The environment to connect between data and people in the best way was necessary.”

Rather wisely, he concluded: “The challenge to improve never ends.”

For more information: https://industrial.omron.eu/en/products/sysmac-platform/kusatsu

Much has been gained by Omron about their manufacturing processes, meaning both the functionality of MAC products and the manufacturing processes used to create them, have been improved. Both aspects have led to better end products for customers!

Summary

The creation, collection and analysis of Big Data is seen as highly desirable for modern manufacturers because it allows them to construct an accurate picture of what's really happening in their plant. This is mainly used to improve production processes, but also to provide traceability (in case of recalls) and preventative maintenance (to reduce down time).

While some propriety Big Data systems exist, many are expensive and need dedicated professionals to both implement and maintain them.

By embedding standard SQL database client functionality in its industrial controllers, Omron is able to offer a reliable alternative that's both low cost and highly effective. It offers many advantages, like open well-understood systems, simple implementation, real-time performance and future proofing due to the separation provided by the client and server architecture used in databases. This also provides scalability, enabling users to start with a small (often localised) system, and build up to expansive cloud-based systems, all without needing to retrofit controller hardware.

Best of all, it puts the end-users in charge of their own data, which they can manipulate and modify as required. Once a system has been successfully implemented, users gain invaluable knowledge about their operations, giving them opportunities to improve what they do, so as to gain competitive advantages.

References

NJ-series Machine Automation Controller Database Connection CPU Unit

NJ Database Connection CPU Units User's Manual

Big Data Drives Productivity Gains

Case Study of SQL Flyer

Understanding Microsoft's Big Data solutions patterns & practises

What is Big Data?

Planning a Big Data solution

About the author

Harry Mulder has been involved in the industrial control industry for almost 30 years, the last 26 of which have been with OMRON Electronics. With a degree in computer science, his experience includes sales, engineering and product management of industrial programmable controllers, HMIs, networking, software and even robots. He currently manages an engineering team across four states but still likes to get involved with day-to-day problem solving. He is happily married and has three girls who keep him on his toes.


OMRON
Industrial
Automation
eData DVD
Mar 2016
realizing