The term “connected plant” may evoke mental images of advanced automation, robotics and artificial intelligence. One key component that shouldn’t be overlooked, however, is human collaboration. In fact, one of the most important characteristics of a connected plant is the ability to increase and enhance collaboration to help solve previously unsolvable problems. The emergence of the cloud-based historian is a key component, therefore, to bringing the connected plant to life because it greatly enables this type of collaboration.

Consider the following – using these historians, a new type of technology platform that supports broad data analysis greatly expands what’s possible in today’s manufacturing facilities. Here are a couple of examples:

Example 1: The Corporate Expert Accessing Enterprise Data in the Cloud

When a typical event happens in a plant, the operators can usually deal with the immediate issue but are often not able to fix it or understand how to prevent it from reoccurring. A process engineer can find data correlations that could indicate a root cause, but that often requires looking at much larger data sets.

With all the data stored in the cloud from the entire enterprise, however, it’s possible for subject matter experts to look at all the data for similar assets or processes. From there, they can use more advanced analysis tools to find and identify the causes. The asset or process templates in the cloud can then be updated with enhanced issues detection and/or with process changes, or notifications to change the process.

Those templates can be pushed out to all plants in the enterprise. In the future, when a similar event is about to happen, the operators are notified of the impending issue and corrective actions.

Example 2: The Cloud Community Environment

The cloud also can make data available to other vendors if allowed. These could be:

- Equipment OEMs who can troubleshoot their equipment on site
- Knowledge vendors with key expertise in the specific process or industry
- Applications connected to the data, for various types of solutions: advanced process control, alarm management, performance monitoring, scheduling
- Joint customer industry projects, used to generally optimize the industry practices
- Utilities and suppliers that can collaborate with the manufacturer on (for example) power usage and optimization
Put simply, the connected plant should connect people just as effectively as it does previously disparate subsystems. Now comes the (perceived) hard part: how does one enable these connections in an efficient and secure architecture?

**Using Edge Gateways to Build Secure Cloud Connections**

Bringing connected plants to life shouldn’t endanger operations with a data breach. How secure is that cloud connection, and is the collaboration worth the risk?

Connecting any sort of critical system to the cloud understandably brings out concerns about protecting data that are essential for successful operation. There are steps that can and should be taken, though, to help insulate important data housed by historians from potential outside threats.

It's important to remember that moving and storing data in the cloud for analysis and visualization will impact traditional process and enterprise historians; specifically, the cloud historian replaces the enterprise historian as many manufacturers look to simplify their hardware infrastructures. But while the cloud historian enables problem solving at a larger scale, the process historian still has an important role to play to ensure data at the control level is captured reliably and serves local users.

The cloud historian architecture therefore must respect the fundamental requirement of isolating control system data from the outside world. This means simply connecting the process historian to the cloud is not practical or wise. The solution here is an edge gateway (either as software or a physical appliance) as an intermediary responsible for secure communication between site and cloud, as well as guaranteed data delivery – even in outages.

From a security perspective, the gateway must be explicitly provisioned and authorized to communicate with the cloud, and all communication must be initiated from edge to cloud. Bi-directional communication is possible, but it must originate at the edge, meaning “southbound” messages are effectively pulled down from a cloud connection.

Another benefit of the edge gateway is that it can facilitate integration of independent devices in an Industrial Internet of Things (IIoT) architecture. This strengthens security by providing a single, secure connection to the cloud rather than opening a myriad of different channels.

Employing an edge gateway can simplify the historian architecture and communication between process and enterprise systems – and simple, of course, usually means more secure. Next is the question, though, of whether the connection itself is secure. A strong provisioning process is required to establish connection and associated keys. In some scenarios, the keys may be baked into a gateway appliance in the historian vendor’s factory and pre-provisioned in the cloud.

Once the connection is established, secure encrypted communication protocols can send data to the cloud. All communication is initiated by the edge gateway, even for permitted “downloads” from the cloud.

Data Partitioning for Streamlined Access

As mentioned before, collaboration for problem solving is the top benefit of cloud-based historians that ultimately enable the connected plant. Naturally, the historian uses a multi-tenant architecture, allowing various end-users to share the same cloud resources to spread system and support costs across them. Data privacy, however, is a key concern.
Each user is considered to be a tenant, and system resources and functions are segregated by tenant, as needed. Static and persisted data must be strictly partitioned by tenant. There are different implementations depending on the data storage technology used; some storage approaches like Microsoft Azure SQL natively support tenant-partitioning, others require this to be designed into the solution. Likewise, data security may be enforced by the database itself but more often by data access services.

Data partitioning is also important for scalability and availability. Partitions can be assigned to different resources in a data center or even across different data centers, as needed. Larger customers may desire separate, dedicated storage.

Each site should be able to manage its own users and permissions. Users can typically be managed directly in the IIoT solution (for example, Azure Active Directory). By default, multiple sites are managed in a single directory, with an overall security administrator (historian vendor) delegating limited administration rights. Optionally, a manufacturer might have its own dedicated directory that can be used, for example, when it already uses Azure Active Directory for other solutions. A more advanced company might choose to federate security in IIoT with its enterprise active directory so users and groups can be managed within the enterprise, but used in the cloud.

And that’s just for the internal users – outside users should be considered, as well, because some of the strong value propositions for cloud solutions require end-users to share data with third parties. This might include:

- Sharing quality or compliance data with their customers and regulators
- Sharing process and performance data with process suppliers and equipment vendors for analytics and optimization
- Sharing system health and diagnostics with equipment suppliers and out-sourced support service providers

This requires a multi-party approach to data security where a customer can grant selective permissions to third parties, as required. This necessitates a model where permissions can be assigned at a finer-grained level than just each customer.

A gas utility company, for example, may have gas meters at thousands of end customer sites – hospitals, colleges, light industry - and may want to share data with each end customer about their specific site(s).

The Underlying Infrastructure

Effective data partitioning is a key characteristic of new broad analysis platforms because the system supports not only a very secure access, but multi-party access to the data. This is done in a way so that access is only granted to specific areas of the entire enterprise data, and it can be easily and securely configured. This approach ensures OEMs are only accessing the relevant data for their equipment; utilities are only accessing the utility data from the plants and sites they are providing utilities for; software vendors only access the data from their applications, etc.

As for the foundational aspects, there are a variety of ways to implement a cloud historian using a mix of available technologies:

- OPC UA is a leading candidate for collecting data and sending it to the cloud. OPC is a well-established standard in the process industries, and OPC UA is adding cloud friendly options as pub-sub and protocols such as AMQP. Additionally, open protocols such as OData and OPC UA must be provided for the data to be universally available for visualization and analysis. This will allow for commercial visualization and analysis tools such as Tableau, Spotfire and other business intelligence tools to access the data natively, as well as traditional tools for the process industry using OPC today.

- Azure Event Hub and Kafka are good examples of event/message hubs that can process and distribute messages as they come into the cloud.

- Spark is a leading platform for running analytics on data streams as they flow through the system.

- Hadoop is the most common store for large volumes of unstructured data from multiple sources to support offline analytics. Cassandra and HBase are columnar databases that provide higher performance but less flexibility than Hadoop and are well-suited to storing specific types of data such as time-series. Time-series specific processing capabilities can be layered on top of these columnar stores using components such as OpenTSDB and KairosDB.
• SQL and NoSQL are examples of technologies that can be used to put time-series and other data into context. Graph and semantic database technologies, however, are proving to be popular approaches due to their extensibility and rich semantics.

• R, Python, HDInsight are examples of analytical tools that specialists and data scientists may use to visualize and analyze data once it’s in the cloud. Engineers may use traditional process-oriented applications like trend tools, plant graphics, and emerging tools that support visual analytics and pattern searches for process data. Business users may use dashboards created with a variety of Business Intelligence products.

Using these tools to create a broad analysis platform provides several advantages. First, the time-series data components allow traditional analysis to be performed on the data, which is now spanning the entire enterprise. This analysis would include real-time trending, comparing historical trends between batches and runs, correlating process trends with events, etc. The data would immediately be available for the same type of analysis typically performed today on smaller data sets by plants and sites.

The data stores can also be used for more advanced big data analysis in the cloud. This makes the environment not only useful for time series analysis, but also for data scientists to perform more advanced analytics.

Instead of connecting to multiple data sources via multiple protocols (typically seen with current technologies), the new data platform for the cloud allows all traditional process data types to be stored in the same broad analysis database, which ultimately becomes a platform for cloud applications for analysis, visualization and optimization.

The Cloud-based Historian as a Strategic Investment

Changes in manufacturing software are coming, driven by IIoT and cloud initiatives. The changes for historians will be very strategic as they develop into a platform for cloud applications. The data will not only be utilized as a larger-scale enterprise historian, but also must handle other needs, such as big data analytics. It also must be able to contain not only traditional process data as time series data, but other data types to be used for analytics and applications.

There is tremendous value that can be unlocked by the cloud. The change from traditional server rooms to cloud infrastructure will change the way applications are managed, deployed and configured. It will be much faster to deploy small applications to gain value.

Technologies are available and mature enough today, and used for very specific analysis applications. The major shift will be larger-scale application platforms in the cloud that will allow manufacturers to quickly deploy specific analysis on their data in the cloud, gain insight, and be able to use that insight to deploy remedies on their sites and plants.

Plant historians have always been very strategic for applications in plants, but the new enterprise historian in the cloud has potential to be even more significant.