TDengine

Supercharging PI System for Data Sharing & Advanced Analytics

Extending PI System's Time Series Data Capabilities with TDengine

Contents

In the age of IIoT, effective data management, including accessibility, is critical.

The proliferation of connected sensors generates a massive volume of timestamped data that needs to be shared, centralized, and analyzed. This whitepaper discusses the value of collaborating on time series data and the challenges associated with it. Specifically, it explores the capabilities and limitations of PI System, a popular data historian used throughout manufacturing, and illustrates how TDengine can be used to extend this system to enable data centralizing and sharing.

01. Introduction
Managing & Sharing Data in the IoT Era

1.1. The Proliferation of IoT Data

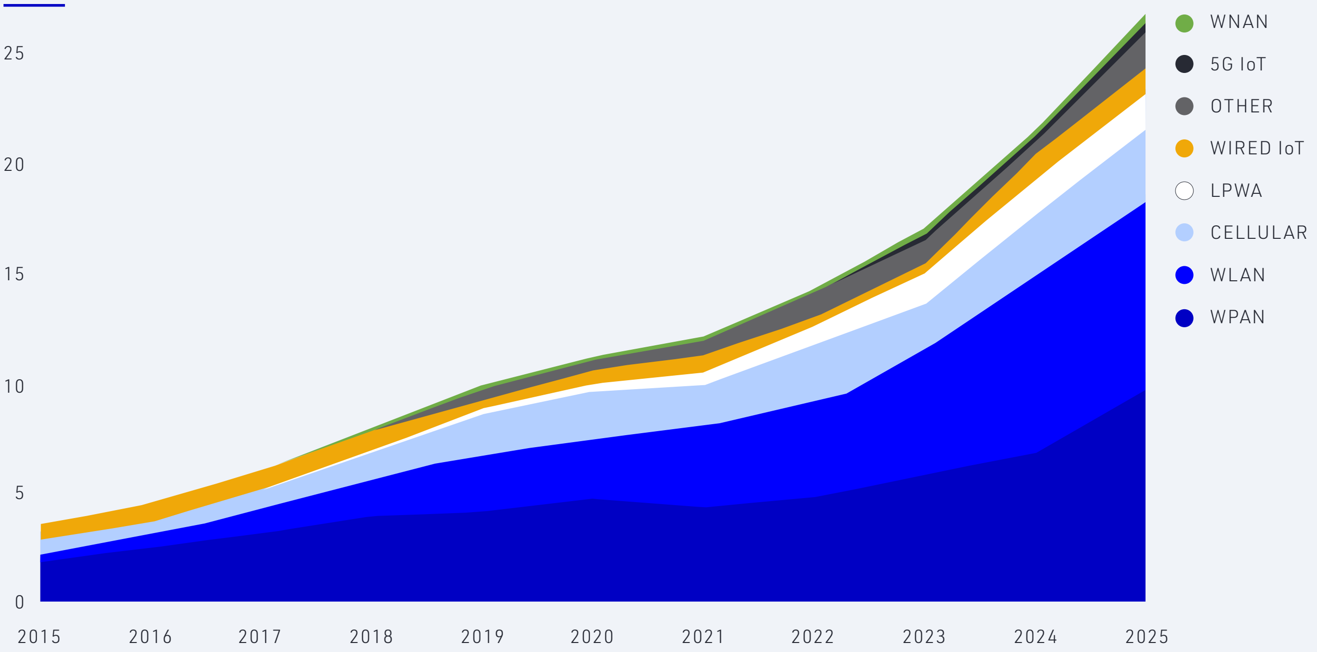
Enabled by the internet of things (IoT) and the digitization of manufacturing, the industrial internet of things (IIoT) is the greatest advancement in manufacturing since the electrification of factories in the early 20th century.

In practice, the IIoT applies a connected network of devices and sensors that automatically monitor equipment and processes in real-time, vastly improving efficiency, reducing downtime, and supporting advanced business intelligence. Naturally, this approach generates a massive volume of data, virtually all of which is timestamped, that needs to be collected and analyzed.

With the number of connected sensors increasing exponentially, it's critical for businesses to implement effective strategies, processes, and technology to ingest and manage this data at scale.

GLOBAL ACTIVE IoT CONNECTIONS (IN BILLIONS)

Although growth slowed during the height of the pandemic, the number of IoT connections is expected to double every three years.



Source: IoT Analytics Research, State of IoT 2022. May 2022
IoT connections do not include computers, laptops, fixed phones, cellphones or tablets. Counted are active nodes/devices or gateways that concentrate the end-sensors, not every sensor/actuator. Simple one-directional communications technology is not considered (e.g., RFID, NFC.) Wired includes ethernet and fieldbuses (e.g., connected industrial PLCs or I/O modules); Cellular includes 2G, 3G, 4G; LPWA includes unlicensed and licensed low-power networks; WPAN includes Bluetooth, Zigbee, Z-Wave or similar; WLAN includes wifi and related protocols; WNAN includes non-short range mesh, such as WI-SUN; Other includes satellite and unclassified proprietary networks with any range.

1.2. Adapting to IIoT: The Value of Centralization & Sharing

Adapting to this industrial revolution requires introducing new technologies, systems, and approaches.

For example, with IIoT bridging the gap between physical devices and digital applications, legacy embedded systems are being replaced in smart factories by cyber-physical systems (CPS) which combine computing, communication, and control capabilities with physical procedures and elements.

To maximize the effectiveness of these systems, and the business benefits they can provide, data from individual sites must be centralized and shared. This enables an organization to benefit from a global view of an entire network of connected sensors and devices, and develop insights that lead to company-wide efficiencies. Operating in silos doesn't cut it when it comes to IIoT.

BENEFITS OF CENTRALIZING IIoT DATA

UNIFIED WORLDVIEW

- Instead of analyzing fragmented, siloed data, centralizing allows users to view a single dashboard.
- Users can review global operations while retaining the ability to zero in on individual sites.

SEAMLESS BENCHMARKING

- By pulling together data from multiple sites, users have greater context of the performance and outliers in operational data.
- It's easier for users to see which sites need support, and allocate staffing and other resources accordingly.

CONTROLLED SHARING

- Entire datasets or limited segments can be shared throughout the organization with fine-tuned privileges.
- Data segments can be shared with external partners, vendors or regulators.

CENTRALIZATION & SHARING CHALLENGES

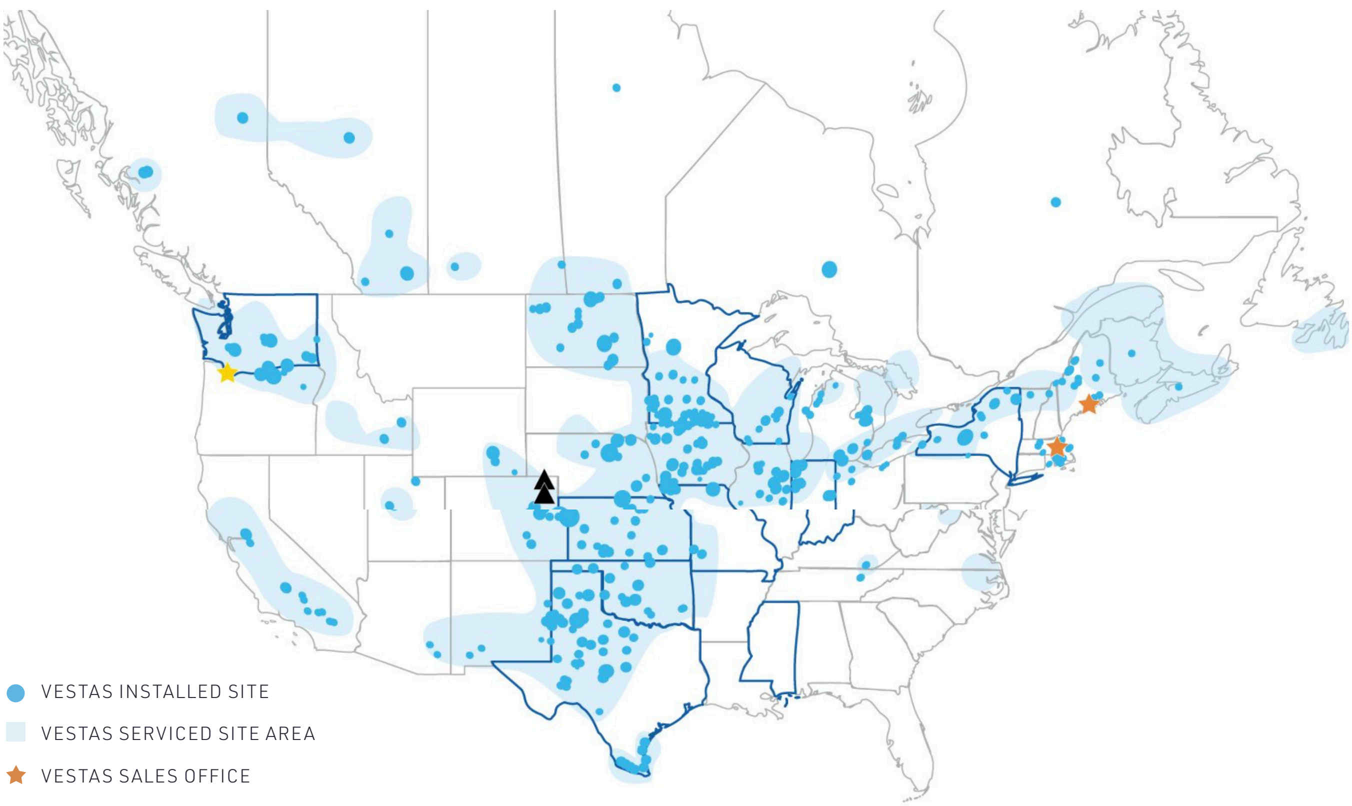
Many large manufacturing companies, especially in mature industries such as utilities and automotive, have developed or acquired unique data collection and operation systems within many of their sites.

Aveva PI System, a data historian that collects and stores time-series data, has been widely adopted in manufacturing to monitor and report on individual sites. To ensure adequate reliability and low latency, a PI instance is deployed in each site, often physically close to the data it's collecting. This architecture often requires a significant investment in software, hardware, and time for deployment and administration.

Although PI System excels at collecting data for individual sites, it has considerable limitations when it comes to centralizing data from multiple locations. Having made such a significant investment in this architecture however, ripping and replacing the solution may not be feasible, despite the benefits of a centralized approach. This issue is compounded when businesses use varying data historians across sites or protocols such as MQTT.

VESTAS NORTH AMERICA: COMPANY SITES & TURBINE INSTALLATIONS

Like many IIoT companies, Vestas has to centralize data across tens of thousands of devices and multiple factories – in North America alone.



VESTAS INSTALLED SITE

VESTAS SERVICED SITE AREA

VESTAS SALES OFFICE

NORTH AMERICAN HEADQUARTERS

VESTAS FACTORY

STATE WITH STEELHEAD AMERICAS

TDengine

02. Sharing & Centralizing Data with PI System

2.1. PI System for Time Series Analytics

The PI System, launched in 1980 by OSIsoft, was developed in response to the increased automation of industrial processes, where devices gather large amounts of data that can be valuable when analyzed and processed.

In addition to timestamped data, legacy databases struggled to collect and store vast amounts of metadata including time, location, device type, etc. Instead, the PI System is specifically designed to handle time series data by facilitating quick data ingestion and querying, while maintaining the full context of each data point, which is essential for developing meaningful insights.

2.2. Centralization and Sharing Limitations

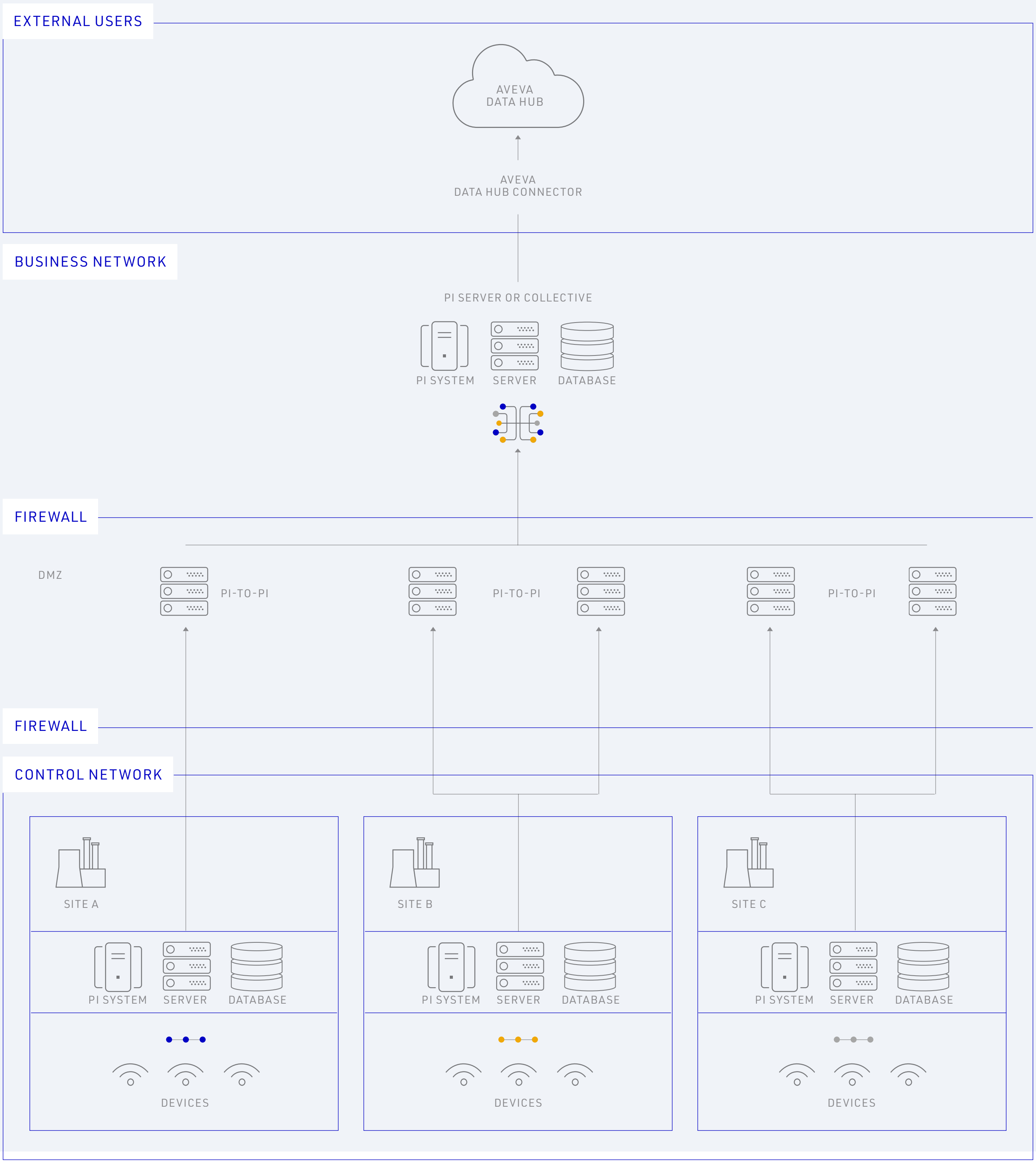
Although PI System is well-suited for time series data, it has significant limitations when it comes to centralizing data from multiple sources. It usually requires either complicated IT implementations or exporting and categorizing vast amounts of information manually, which prohibits any form of real-time analytics.

Unfortunately, PI System also lacks easy-to-use tools for sharing data, such as creating specific privileges for access to a segment or filtered data. This creates difficulties when businesses need to share data, including sensitive information, with customers, partners, and regulators.

In some cases, the limitations of PI System have resulted in further fragmentation of a business's data stack, as rather than finding an agile solution that supplements PI System's capabilities, additional solutions are brought in to manage each scenario that PI cannot handle.

PI <> PI DATA CENTRALIZATION & SHARING

Moving data from individual control networks into a business network is often achieved with the PI to PI Interface, with sharing to external users via Aveva Data Hub.



2.3. Aveva Solutions to Facilitate Data Sharing

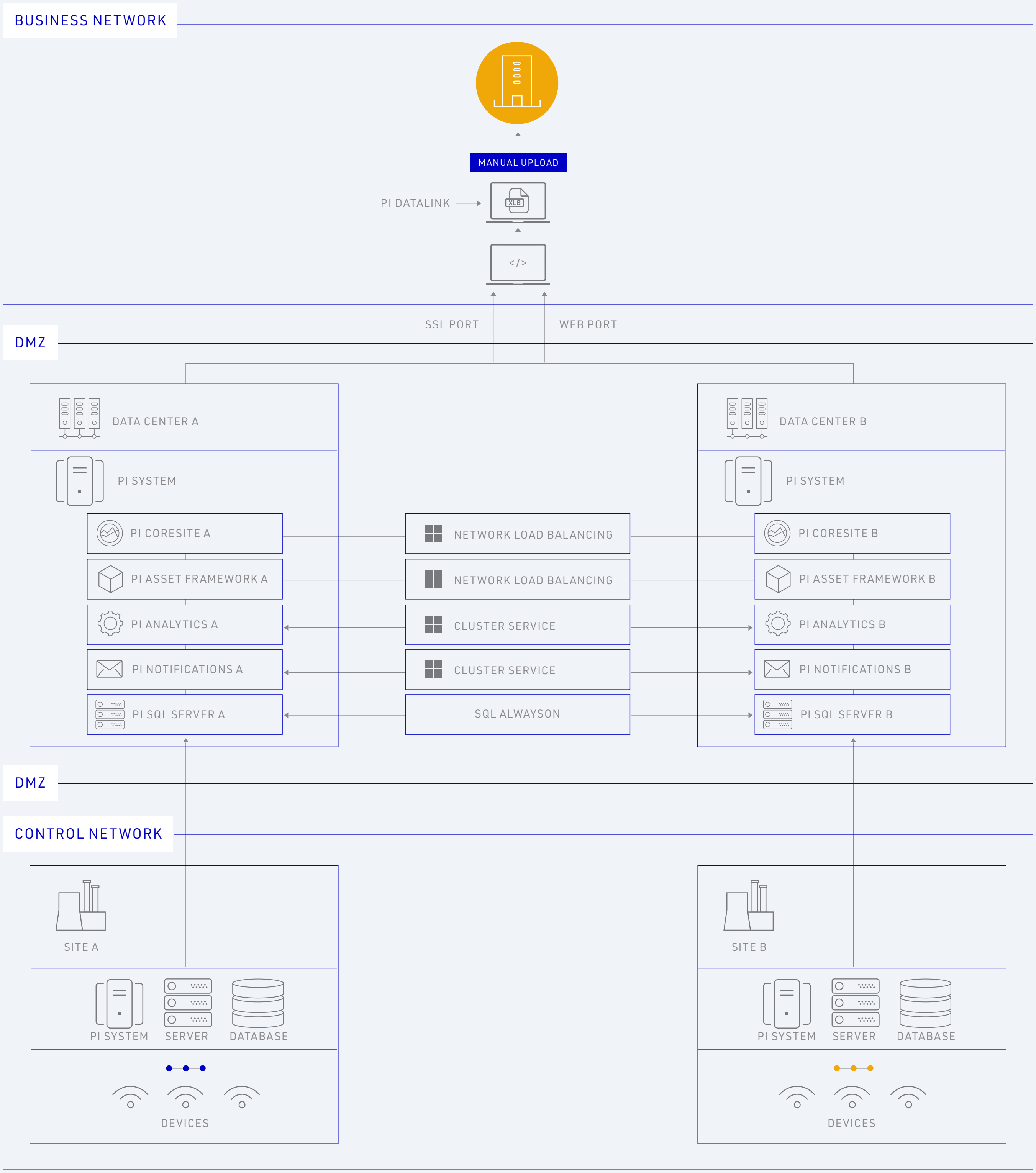
AVEVA PI INTEGRATOR FOR BUSINESS ANALYTICS

In a bid to address PI System's limitations for centralizing data, Aveva launched the PI Integrator for Business Analytics in 2016. For businesses hoping for a seamless solution to centralize PI System data however, PI Integrator has several weaknesses.

- While backfilling of data is supported, it cannot automatically backfill after interruptions in data collection.
- Customers with an enterprise agreement can only export a small subset of their PI points.
- Significant expertise is required to configure data for PI Integrator and run the solution in general.
- The cost of the license, along with the cost of the employees required to run and maintain it, is expensive.
- Data that is shared using PI Integrator is transformed into flat CSV files, removing the contextual metadata that is valuable for analysis.

EXAMPLE: HIGH AVAILABILITY PI SYSTEM ARCHITECTURE ACROSS TWO DATA CENTERS

As this diagram shows, even a carefully planned system for PI integration requires complex architecture, which is expensive to run and maintain.



AVEVA DATA HUB

Shortly after acquiring OSIsoft in 2020, Aveva launched the Data Hub to facilitate sharing across PI systems. This cloud-based system removes many of the barriers to sharing data from disparate PI servers, providing secure access to real-time data for users inside or outside of the company's network. Bi-directional sharing is also available for a number of services within Aveva's ecosystem.

This service isn't an accessible extension of the PI database however, as it operates largely independently – Aveva Data Hub is built on a separate data structure and lacks PI's widely-used asset framework (AF). The cost and limitations on the number of data points that can be shared are also a drawback for some users. Sharing data from 3,000 streams costs mid-five figures annually, which adds up to an extremely hefty outlay for companies with tens of thousands, or millions of metrics to access.

03. A New Approach
TDengine for PI System

TDengine has built a powerful hybrid solution that extends PI System to enable seamless data sharing and centralization, making it easy for businesses to break down data silos and generate actionable insights.

BENEFITS

RAPID AND SIMPLE IMPLEMENTATION

It can be installed, configured, and start streaming data in under an hour, ensuring a quick return on investment.

NO RIP AND REPLACE

It's built to centralize PI System data without replacing the entire system. Data is streamed to TDengine and can be viewed and processed natively back within PI AF.

OPERATING AT SERIOUS SCALE

TDengine's time series database is proven to perform successfully on one billion IoT devices. TDengine for PI System can easily scale by using multiple connectors and has been tested on hundreds of thousands of PI points for a single connector.

SEAMLESS DATA SHARING

Once the data is centralized, users can leverage the powerful data-sharing capabilities of TDengine.

3.1. Replicating Data

TDengine for PI System allows users to quickly replicate data to the cloud, reducing deployment time from months to minutes. Users have the option of streaming data continuously from PI points or AF templates to TDengine Cloud. After a quick setup, the integration handles the messy business of extract, transform, and load and enables automatic backfill after any disruptions.

REPLICATING FROM AF

To replicate data from AF, users just select a template, and TDengine for PI System will start streaming all relevant data.

REPLICATING PI POINTS

For those who aren't using AF, a convenient point builder tool allows users to stream individual PI points to the cloud.

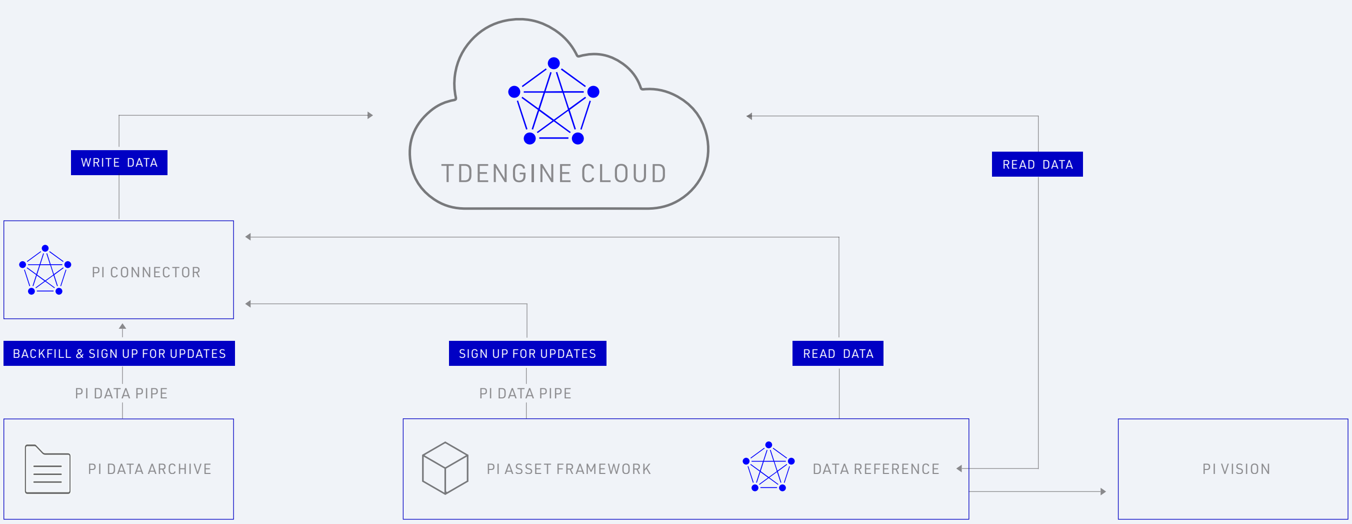
3.2. PI System Integration

TDengine for PI System uses the AF SDK to query historical data from the PI Data Archive, set up PI and AF data pipes for streaming data, and connect to PI AF to query the AF structure. It also creates the corresponding tables and writes this data over a secure RESTful API to TDengine Cloud.

The TDengine Data Reference is an AF Custom Data Reference that queries data within TDengine and allows users to interact with TDengine as though it were a PI Point within the PI Data Archive. In this way, TDengine data can be used alongside PI data in PI Vision.

TDENGINE FOR PI SYSTEM

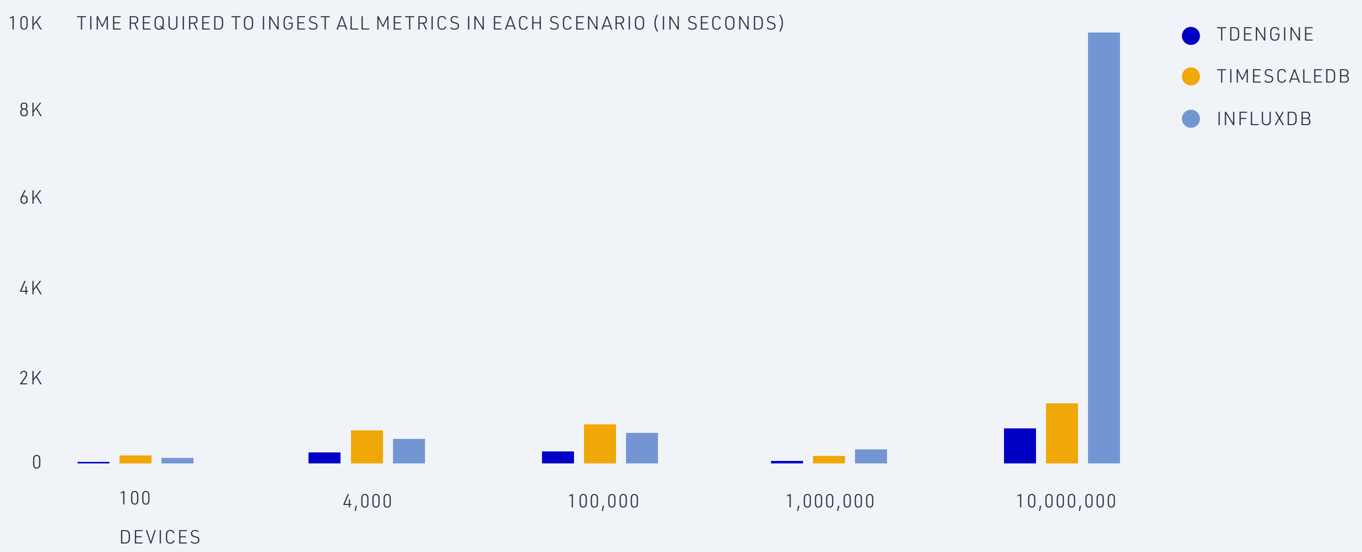
TDengine for PI System creates multiple data pipes to the PI Data Archive and AF to stream data. TDengine's Data Reference allows users to interact with TDengine in AF as though it were a PI Point within PI's Data Archive.



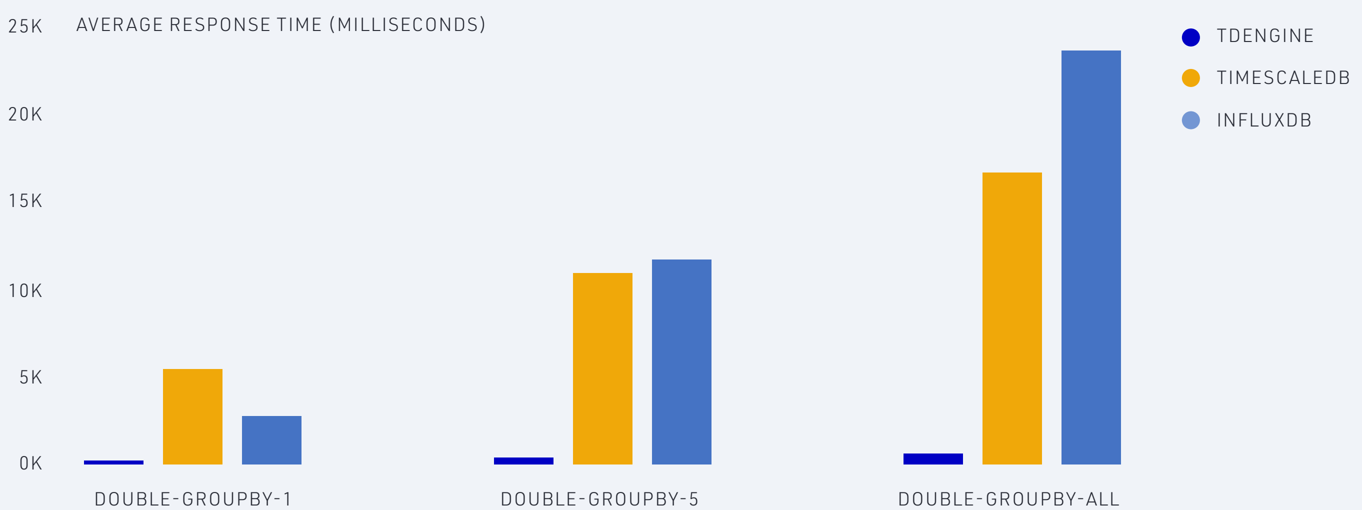
04. Modernizing the Time Series Database

TDengine is a cutting-edge time-series database purpose-built for IoT. Designed as a distributed database from day one, it can support billions of IoT devices effortlessly while consistently outperforming other time-series databases in data ingestion, querying, and data compression.

INGESTION RATE

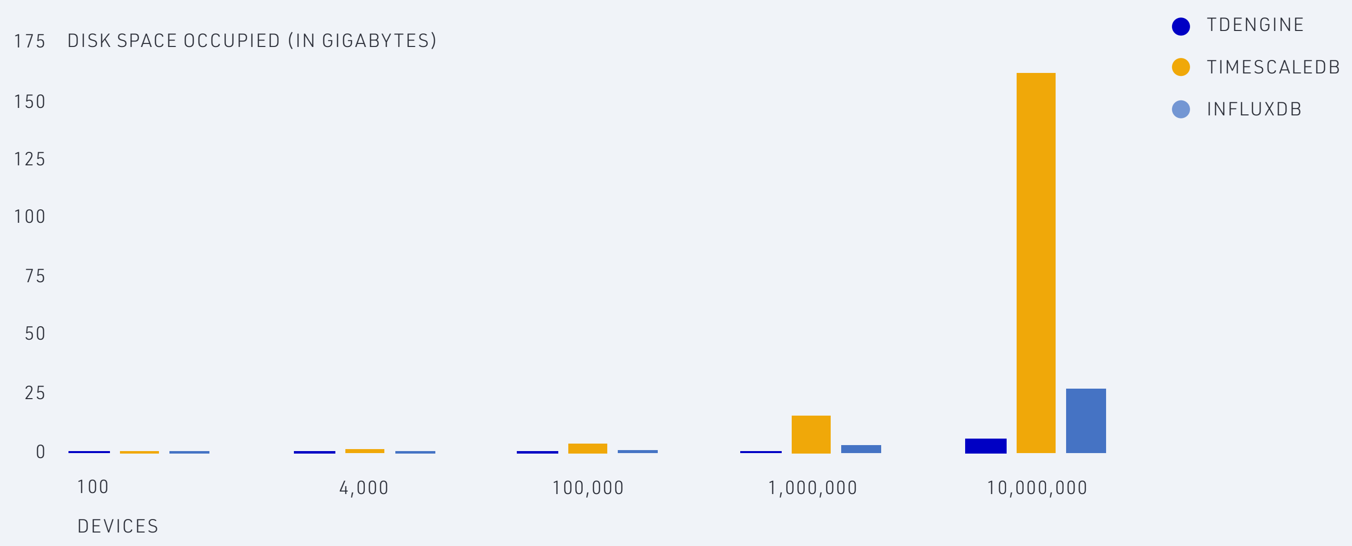


QUERY PERFORMANCE



TDengine, Q1'23 Time Series Data Performance Benchmarks. Feb 2023. All scenarios and datasets pulled from the Time Series Benchmark Suite

DISK STORAGE



TDengine, Q1'23 Time Series Data Performance Benchmarks. Feb 2023. All scenarios and datasets pulled from the Time Series Benchmark Suite

Despite its impressive capabilities, TDengine is user-friendly and intuitive to use. On the surface, it looks like a relational database that uses SQL, however, it's designed as a real-time data engine optimized for data in-flight. TDengine's unique architecture also makes it easy to share and replicate data across multiple instances, enabling rapid access to vital data.

4.1. User Management

TDengine's user management system is defined by two main concepts: privileges and resources.

PRIVILEGE

A level of permission, for example, "Read Database". Role-based access control allows admins to assign view and edit responsibilities to different roles or user groups.

RESOURCE

This is what the privileges apply to, which can be the entire organization, a specific instance, database, or even topic.

The user management system can be used to share data internally and, unlike many other systems, externally, allowing users to set fine-tuned data-sharing privileges for the entire organization, a specific instance, a specific database, or a specific SQL query through data subscriptions (e.g. a table filtered by a specific tag for a defined amount of time).

4.2. Data Subscriptions

SUBSCRIBING TO UPDATED DATA

TDengine data subscriptions and consumers act similarly to other widely-used messaging queue solutions. Applications can subscribe to real-time data updates, making it easier to process data in the order that events occur. TDengine doesn't require users to deploy Kafka or continuously query the database. Partners get notifications as soon as the data is received, simplifying time series data processing and eliminating the need for additional messaging queue solutions.

DEFINING TOPICS

TDengine's data subscription feature provides greater flexibility than similar products such as PI System, thanks to its use of topics defined by query conditions based on an existing table, subtable, or supertable – or in other words, a SELECT statement.

By leveraging SQL, data can be filtered by tag, table name, column, or expression, and then processed using scalar or user-defined functions.

This approach allows applications to control the granularity of data on demand, while filtering and preprocessing tasks are handled by TDengine rather than the application layer. As a result, the amount of transmitted data is reduced, and the complexity of applications is streamlined.

DATA SHARING

TDengine enables data sharing with third parties, such as customers or consultants, using subscriptions and topics. This approach provides fine-tuned access control to data while allowing third parties to receive real-time updates on specific topics. Consumers can be programmatically created to consume messages in the queue for each topic, and TDengine can also consume topics, allowing data to be replicated between instances. Additionally, both real-time and historical data can be shared via topics, with the option to set start and end dates or continuously stream data.

IMPLEMENTATION

TDengine is a real-time streaming database that uses a write-ahead-log (WAL) to not only prevent data loss, but also for stream processing and data subscriptions.

The system keeps track of where individual subscription consumers are in the WAL. Each consumer receives updated data, and multiple consumers can be grouped together as a consumer group to improve efficiency and availability. If the data in a supertable is sharded across multiple virtual nodes (vnodes), multiple consumers can consume it more efficiently than a single consumer through multi-threaded, distributed data consumption.

TDengine also includes an acknowledgment mechanism to ensure at-least-once delivery, even in complex environments where machines may crash or restart.

4.3. Integrating Advanced Analytics

It's hard to imagine an enterprise that isn't focused on improving their operations with the help of advanced analytics, and there's obviously been a large increase in the number of legacy companies and new entrants that have launched new AI-enabled tools.

As an open system, TDengine makes it easy for companies to integrate a range of third party solutions, including providers of advanced analytics. The data sharing capabilities noted above also ensure these third parties are only working with the data they've specifically been giving access to.

HYBRID SOLUTION

TDengine's PI System integration allows data in PI and TDengine to coexist. New points are stored in TDengine while legacy points remain in PI – and with the TDengine Data Reference, all data stored in TDengine can be viewed and manipulated like PI points within AF and PI Vision.

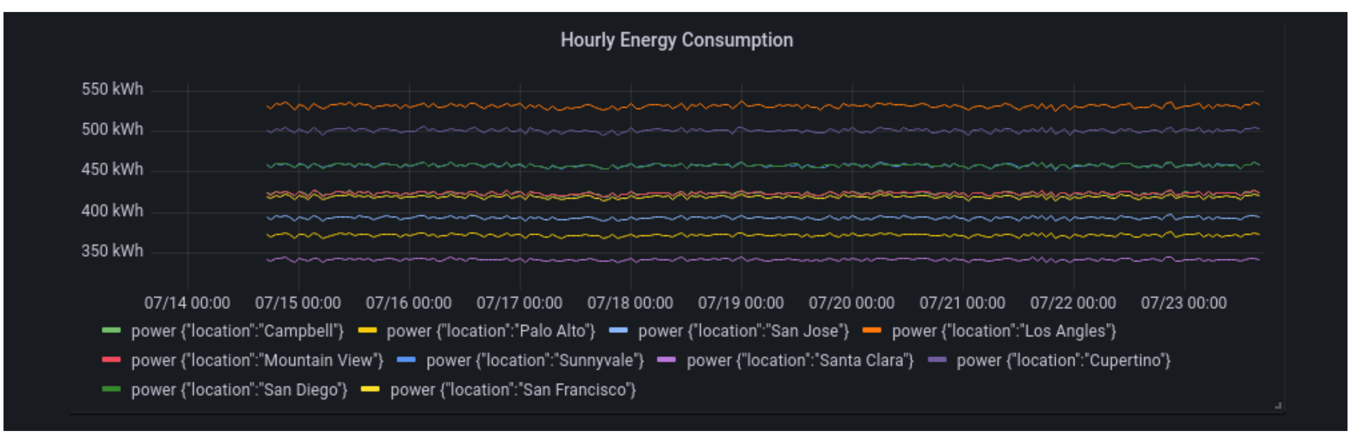
Users can replicate data to TDengine Cloud for analytics and then view the results in PI Vision dashboards alongside PI data, without creating any new PI points.

VISUALIZATION

Users can also view new data stored in TDengine together with existing data stored in PI in the TDengine Data Explorer or other modern visualization tools. A number of options exist for building dashboards based on these analytics:

Use the Grafana or Google Data Studio integrations for ad-hoc dashboards.

Build custom dashboards with TDengine's programming connectors.



4.4. The Open System Future

TDengine is aligned with the open-source future of IIoT - meaning that breaking down silos and centralizing data is just the start. Users can integrate TDengine seamlessly with modern analytics and visualization tools, such as Grafana or Google Data Studio, or build custom dashboards by leveraging TDengine's programming connectors. The open nature of TDengine means there are endless opportunities to generate useful insights that drive greater efficiencies and improve outcomes.

PI SYSTEM DATA CENTRALIZATION AND SHARING WITH TDENGINE

TDengine centralizes data from multiple sites to the cloud, where users can manage permissions to share all of the data or select portions.

TDengine

CASE STUDY

Data Centralization and Sharing for Electric Utilities

Improving a power generator's financial, operational or environmental outlook requires insights that can only be developed with a full view of the system's operations. TDengine partnered with a management consultant operating several electricity plants, each with information that needs to be centralized and aggregated. This process is complicated with a system that includes 32 PI servers on separate networks, with different site owners and multiple data administrators.

Each site had solid monitoring in place for their facility, with turbine sensors reporting fuel consumption, power output, operating temperature, vibration and other key metrics. Collecting information across sites however, involved a largely manual process of accessing each PI server via VPN, querying the database, exporting files to the cloud, and aggregating the data. By the time a company-wide analysis was complete, it often represented information that was largely outdated.

TDengine was introduced to centralize this data and automate the process. Instances of TDengine for PI System were installed for each PI server and set to stream real-time data to TDengine Cloud. The occasionally spotty internet service doesn't derail the process, as each TDengine PI Connector backfills any missing data once the service is re-established. This new system provides the team with a single dashboard to view data from all of the plants or drill down on a specific site or turbine.

In addition to centralization, TDengine Cloud acts as a powerful extension of their PI System with a range of data-sharing capabilities. This gives the team fine-grained control over the specific data to share with regulators and consultants, without having to provide full database access.

1. **PERMISSIONS:** Can be given organization-wide, for certain instances, for certain databases or for a specific data subscription topic.

2. **DATA SUBSCRIPTIONS:** Allow topics to be created from individual SQL queries to select data from a specific table, device or time period.

3. **ACCESS:** All the team needs to add someone to a topic is an email address – the subscriber can then access the data programmatically or replicate it to their own instance of TDengine.

05. Conclusion
Future-Proofing PI System

While many users love PI System, it's had little development in recent years, partly due to multiple acquisitions which led to a significant loss in product knowledge and a divergence in development priorities.

TDengine provides a powerful extension of PI System that enhances its capabilities, enabling efficient data centralization and sharing without having to rip and replace a PI deployment. Organizations benefit from the freedom to use multiple systems at individual sites, using the right system for each site's needs, and benefit from a unified view of their data.

TDengine

TDengine™ is an open-source, cloud-native time series database optimized for IIoT. With its built-in caching, stream processing, and data subscription capabilities, TDengine offers a simplified solution for time series data storage, analytics and sharing.

To learn more visit tdengine.com

15559 UNION AVE
SUITE 135
LOS GATOS, CA 95032
BUSINESS@TDENGINE.COM