TDengine

AI-Native Data Platform for the Industrial IoT

TDengine White Paper

Introduction

We now live in an era defined by the Internet of Everything and artificial intelligence, where data has become the core productive asset driving digital transformation. From devices and sensors to business systems, vast streams of real-time data are continuously being generated, yet often scattered across different platforms, formats, and semantics. Traditional databases, data historians, data warehouses, and data lakes struggle to meet the demands of high-frequency collection, multi-source aggregation, and low-latency analytics required in industrial and IoT environments.

At the same time, enterprises face high development, maintenance, and talent costs when trying to unlock the value of their data, leading to consistently low returns on investment.

TDengine was created to address these challenges. As an AI-native industrial and IoT data platform, TDengine rebuilds the data architecture from the ground up, integrating data collection, aggregation, storage, analysis, real-time computation, visualization, event management, and intelligent insight into a unified solution. It enables enterprises to fully unleash the value of their data with exceptional performance, minimal cost, and an intuitive user experience.

The core advantages of TDengine include:

1. AI-Ready Data Platform: Efficient Aggregation and Management of Multi-Source Data

   TDengine supports seamless integration with multiple industrial data sources, including OPC, MQTT, and Kafka, with built-in ETL capabilities for efficient data cleaning and transformation. A hierarchical tree structure organizes a clear data catalog, while templates for devices, attributes, analytics, and dashboards enable full data standardization.

   Additionally, TDengine allows configuration of metadata such as descriptions, units, thresholds, and tags to achieve contextualized data modeling. This comprehensive approach makes the platform truly AI-ready, laying a solid foundation for intelligent analytics and insights.

2. High-Efficiency Storage: 10x Performance at 10% Cost

   Designed specifically for time-series data, TDengine features a next-generation storage engine that leads the industry in write, query, and compression performance. Compared to general-purpose databases, TDengine delivers over 10× faster write and query speeds while reducing storage costs to as low as one-tenth.

   It supports tiered storage and integration with object storage systems such as S3, enabling efficient hot-cold data management. With native horizontal scalability, TDengine can reliably handle

hundreds of millions of data collection points and high-cardinality workloads, making it ideal for large-scale deployments in industries such as manufacturing, energy, and connected vehicles.

3. Advanced Analytics: From Raw Data to Real-Time Insight

   TDengine supports standard SQL and a wide range of time-series functions, powered by a high-performance stream processing engine capable of millisecond-level data processing and real-time analytics.

   Its built-in AI time-series agent, TDgpt, enables predictive analysis, anomaly detection, data imputation, and classification, all with a single SQL command. Under the hood, TDgpt leverages machine learning algorithms and large-scale time-series models to deliver intelligent, automated insights.

   TDengine also provides SDKs for easy integration of custom algorithms and models, allowing enterprises to accelerate the realization of data value and build smarter analytical applications.

4. Zero-Query Intelligence: Let Your Data Speak

   Powered by an LLM-driven AI Agent, TDengine introduces zero-query intelligence: an innovative paradigm where data analysis shifts from pull to push. The system automatically understands business contexts and generates real-time analyses, reports, and visual dashboards without requiring users to ask questions or write queries.

   This capability significantly reduces dependence on specialized expertise or IT support, enabling any business user to access instant, meaningful insights directly from their data.

5. Open Ecosystem, Unlimited Connectivity

   TDengine is built on an open ecosystem philosophy, with its core code fully open source and comprehensive support for standard SQL as well as mainstream interfaces like JDBC and ODBC. It integrates seamlessly with a wide range of visualization, BI, and AI tools, and supports major programming languages including Java, Python, Go, Rust, and C/C++.

   Applications can subscribe to data from TDengine via MQTT or Kafka, while TDengine itself can publish data to Kafka, MQTT brokers, and Flink, enabling flexible, bidirectional data exchange. This openness allows enterprises to build their own industrial data applications freely, without vendor lock-in.

As AI technologies continue to be integrated with the Industrial Internet and the Internet of Things, enterprises are placing greater emphasis on real-time data accessibility, openness, and intelligence.

TDengine is not just a high-performance data platform — it is the core data infrastructure for the AI-driven decision-making era. It empowers enterprises to move beyond merely “storing data” to truly using data and enabling AI, allowing data to speak for itself and fully unlocking its value.

Chapter 1

Data Management Challenges in the AI Era

Production and operational data collected by enterprises is often massive in scale but low in value density. Extracting meaningful insights from this vast volume of low-value data presents several major challenges:

1. Unprecedented Data Scale, Limited Processing Power

   The widespread adoption of IoT technologies has significantly lowered the cost and technical barriers of data collection and transmission, enabling enterprises to gather data from more devices at higher frequencies. As a result, data volumes are growing exponentially.

   However, traditional industrial data platforms, data historians, and even modern data lakes and data warehouses struggle to efficiently process and analyze such enormous datasets in real time. The common approach remains "store first, analyze later," which delays value creation and limits the potential of real-time insight.

2. Integration Difficulties Across Heterogeneous Data Sources

   SCADA and DCS systems, PLCs, and IoT platforms are often supplied by different vendors and use diverse data protocols (e.g., Modbus, OPC UA, MQTT) and storage formats. This fragmentation results in data being isolated across multiple systems, making it difficult to establish unified standards and hindering data flow between departments and sites.

   For example, equipment operating data, process parameters, and quality inspection records are often stored on separate platforms, creating data silos that prevent enterprises from gaining a comprehensive, real-time view of operations.

3. Loss of Semantics and Context

   Collected raw data, such as temperature or voltage readings, often lacks essential contextual information, such as which asset it belongs to (e.g., "Reactor A's real-time temperature"), its physical unit (e.g., V, kV), or its valid range. When this data is transmitted to IT systems such as ERP or MES, critical metadata is frequently lost, making it difficult to perform accurate analysis or early warning.

For instance, if a tank's temperature data cannot distinguish between the shell temperature and the internal liquid temperature, and lacks a clearly defined range, its analytical value is significantly reduced.

4. Inconsistent Data Quality

   Industrial data often suffers from uneven sampling frequencies, high signal noise, and missing fields. Issues such as sensor drift, which distorts measurements, and communication interruptions, which create data gaps, directly impact the accuracy of models used for predictive maintenance and other analytics.

   Moreover, due to limitations in storage and computing costs, enterprises are often forced to reduce data collection frequency, resulting in the loss of key variation patterns — and consequently, a decline in analytical precision.

5. Real-Time Decision-Making Bottlenecks for Business Users

   Even after data is centralized in databases, the lack of semantic context, metadata, and business logic relationships between tables makes it difficult for data analysts to quickly respond to requests or craft effective SQL queries. For business users without SQL expertise, the challenge is even greater.

   As a result, developing new dashboards, reports, or analytical views can take days or even depend on vendor updates, preventing organizations from achieving the agility required for real-time decision-making.

6. High Barriers to Industry Knowledge

   Every industry has its own unique terminology, key performance indicators, and analytical methodologies, and these continue to evolve over time. This demands deep domain expertise, often requiring five to ten years of experience to accurately interpret data and identify meaningful insights.

   As a result, even enterprises that have amassed vast amounts of data may struggle to extract value from it — unable to ask the right questions, define proper evaluation standards, or apply effective analytical methods — leaving their "data goldmine" untapped.

In response to these challenges, TDengine has applied AI-driven thinking and design principles to completely reimagine the underlying architecture of industrial and IoT data systems—creating a next-generation, AI-powered data platform.

Chapter 2

TDengine Architecture

TDengine consists of two integrated products: the high-performance, distributed time-series database **TDengine TSDB** and the AI-native industrial data management platform **TDengine IDMP**. The two components are designed to work seamlessly together, forming a unified data infrastructure.

The overall architecture is illustrated as follows:

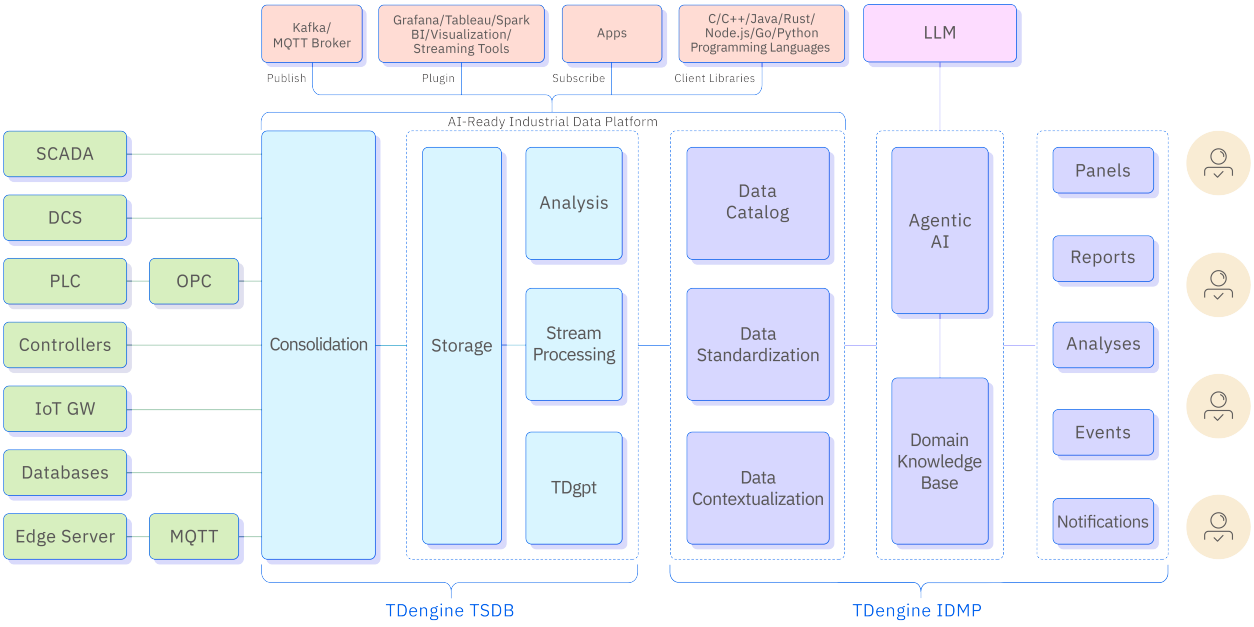


Figure 2.1: TDengine Architecture

Data flows from the data sources on the left, into TDengine TSDB for aggregation, storage, and computation. It is then standardized and contextualized by TDengine IDMP, after which the AI Agent delivers intelligent insights and analysis.

2.1 Integration of TDengine TSDB and TDengine IDMP

Together, TDengine TSDB and TDengine IDMP deliver a comprehensive end-to-end solution for industrial and IoT data management — covering every stage from data acquisition, storage, analysis, and real-time computation to data cataloging, standardization, contextualization, visualization, event management, and root cause analysis.

Through data subscription and publication, the combined platform provides third-party applications with an AI-ready data foundation, allowing users to stay independent from vendor lock-in while fully leveraging emerging technologies, algorithms, and applications to maximize data value.

This combination serves as a complete replacement for traditional industrial data historians, including the PI System, while offering superior scalability, analytical power, and openness.

Even more distinctively, its built-in AI capabilities — including Zero-Query Intelligence and Chat BI — dramatically reduce reliance on data analysts and IT engineers, enabling business users to gain real-time insights instantly and intuitively.

Compared with general-purpose big data platforms, the combination of TDengine TSDB and TDengine IDMP delivers highly optimized storage and analytical performance, a unified end-to-end solution, and powerful AI capabilities—all made possible by a system architecture purpose-built for time-series data.

TDengine IDMP is designed with openness and compatibility in mind, ensuring that the database layer remains independent. Future versions will support integration with multiple third-party time-series and relational databases. However, in the current release, IDMP operates exclusively with TDengine TSDB to provide the best integrated experience.

2.2 TDengine TSDB: High-Performance, Distributed Time-Series Database

TDengine TSDB is a high-performance, distributed, and cloud-native time-series database, with its core code fully open source. Unlike general-purpose time-series databases, TDengine goes far beyond basic data ingestion and querying — offering a powerful suite of additional capabilities that make it ideal for industrial and IoT workloads:

1. Data Ingestion and Aggregation: Without writing a single line of code, users can easily configure TDengine to ingest data from multiple interfaces such as OPC, MQTT, and Kafka directly into the database. It also supports importing from CSV files, relational databases like MySQL, and data historians such as PI, enabling seamless integration across heterogeneous systems.
2. Stream Processing: TDengine features a built-in stream processing engine that supports multiple trigger types, including sliding windows, event windows, state windows, and session windows. The computation logic is fully decoupled from the triggering mechanism, allowing for flexible and efficient real-time analytics.

3. Data Subscription: Acting as a message queue, TDengine allows users to subscribe to data via Kafka or MQTT. With SQL-based topic definitions, users can precisely control the granularity of data distribution and enforce fine-grained security policies, ensuring the right data reaches the right consumers.
4. Read Cache: Whether it's an industrial device's operational state, a vehicle's location in a connected fleet, or a smart meter's live reading, real-time values are critical to business operations. TDengine automatically caches the most recent data, allowing applications to access current values directly — no need for external caching tools like Redis.
5. Data Publication: TDengine can also act as a real-time data source, publishing data directly to Kafka, MQTT brokers, Flink, Spark, and other systems—greatly simplifying integration across enterprise data platforms and enabling smooth downstream processing and analytics.
6. TDgpt: As an external AI agent for time-series data analysis, TDgpt provides powerful capabilities such as forecasting, anomaly detection, data imputation, and classification, all accessible through simple SQL functions. It comes with a rich set of built-in algorithms and offers an open SDK that supports various machine learning algorithms and time-series foundation models, including TDengine's own TDtsfm, designed specifically for time-series understanding and prediction.

2.3 TDengine IDMP: AI-Native Industrial Data Management Platform

TDengine IDMP is a data management platform built on top of the TDengine time-series database, purpose-designed for industrial applications. It extends TDengine's high-performance data foundation with powerful modeling, management, and intelligent analysis capabilities. The platform provides the following core functions:

1. Data Modeling: TDengine IDMP digitizes physical and logical entities through a hierarchical tree structure, creating a unified and scalable digital representation of industrial assets. By defining entity types, attributes, and relationships, it transforms fragmented sensor, device, and system data into a business-semantic "digital twin" model.

   The model supports dynamic node expansion, automatically synchronizes changes in entity topology, and maintains a complete record of data relationships and references—providing a traceable, extensible structure for unlocking deeper business value.
2. Data Contextualization: IDMP brings business meaning to industrial data by dynamically linking raw measurements to specific operational entities. Through KPI definitions, equipment states, and multi-dimensional tags (such as location, organization, thresholds, and categories), low-level operational data is transformed into intuitive business objects.

   Context dimensions can be extended on demand, ensuring that real-world changes are reflected in real time and eliminating the disconnect between data and business understanding—offering an immediately usable analytical perspective.
3. Data Standardization: IDMP establishes a unified data governance framework for industrial information. Using entity templates, it standardizes device and asset attributes, analytics models, and

dashboard specifications. Event templates unify alarm and work-order logic and formatting, while a unit conversion engine automatically harmonizes measurement systems across heterogeneous sources.

This eliminates naming conflicts and unit inconsistencies, ensuring comparability and consistency of data across systems and time periods—and providing a zero-ambiguity foundation for advanced analytics.

4. Intelligent Data Visualization: Powered by an AI-driven context awareness engine, IDMP automatically identifies application scenarios and business priorities to dynamically generate visualization dashboards that support decision-making. By analyzing real-time data characteristics and business context, the system proactively recommends the most relevant chart combinations, eliminating the need for manual configuration.

   Even without deep domain knowledge, users can instantly access key metrics, alerts, performance benchmarks, and anomaly tracking views. Of course, users can also ask questions or manually create dashboards when needed.

5. Real-Time Data Analysis: Leveraging both collected data and business context, IDMP uses AI to recommend real-time analytical insights with millisecond-level responsiveness. A multi-modal trigger engine enables real-time computation based on scheduled polling, data input, events, or state transitions. Built on a stream processing framework, it performs expression calculations, time-window aggregations, and cross-device analytics automatically.

   This empowers users — even without technical or business expertise — to gain instant analytical results for faster, more informed decision-making.

6. Event Management and Root Cause Analysis: IDMP automatically converts analytical results into actionable events and assigns severity levels. Through an intelligent routing engine, notifications are pushed in real time to the responsible personnel, who can acknowledge or respond directly.

   Users can drill down into event context with a single click, accessing the associated device snapshot, reviewing time-series trends before and after the event, and tracing anomaly evolution within seconds. This dramatically reduces downtime and decision latency while improving operational efficiency.

Chapter 3

AI-Ready: Efficient Aggregation and Management of Multi-Source Data

In industries such as manufacturing, oil and gas, and energy, data collection is no longer the main challenge. The real difficulty lies in the fact that data remains scattered across multiple systems, protocols, and standards, each operating in isolation and forming fragmented data silos.

In this reality, enabling AI to truly empower business operations doesn't begin with model building or training—it starts with aggregating, cleaning, transforming, and restructuring these diverse data sources into high-quality data assets with consistent structure and business semantics. Only then can enterprises establish the solid foundation required for intelligent analytics and decision-making.

3.1 Integrating Diverse Data Sources to Eliminate System Silos

To achieve efficient and unified data aggregation, TDengine supports connectivity with a wide range of mainstream industrial protocols and data sources, including but not limited to:

- Modern industrial protocols such as MQTT, OPC-UA, and OPC-DA
- Data collection agents like Telegraf and collectd
- Traditional data historians such as PI System and AVEVA Historian
- Relational databases including MySQL, Oracle, SQL Server, and PostgreSQL
- Message queues such as Apache Kafka and FlashMQ
- CSV file ingestion for batch data import

Through a flexible connection architecture and unified data ingestion workflow, enterprises can consolidate data from geographically dispersed and structurally diverse systems without writing any code, seamlessly integrating all sources into a single, unified platform.

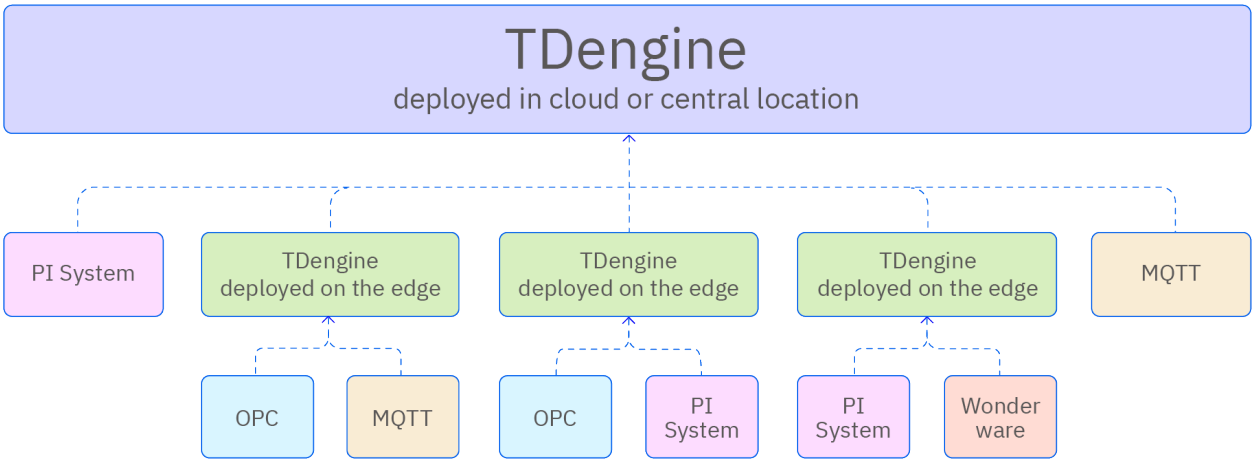


Figure 3.1: Example of data ingestion and consolidation

Unlike traditional data aggregation tools, TDengine incorporates data governance requirements directly at the ingestion layer. It comes with built-in ETL capabilities that support field mapping, unit conversion, expression transformation, and data type unification — ensuring that aggregation is not just about "piling data together," but about making the data truly consistent and comparable.

3.2 Edge–Cloud Synchronization

TDengine allows one instance to subscribe to data from another TDengine instance, enabling a seamless cascading architecture. At the edge, collected data can be stored locally in a TDengine instance, while a cloud-based TDengine instance can, through the subscription mechanism, continuously aggregate data in real time from one or multiple edge instances.

In addition, TDengine addresses critical technical challenges such as resumable transmission after network interruptions, firewall traversal, and data backfill. With simple configuration, the entire system gains full edge–cloud synchronization capabilities, significantly reducing both system complexity and maintenance costs.

3.3 Building a Data Catalog Through a Structured Perspective

TDengine organizes and manages industrial data using a hierarchical tree structure, clearly representing the relationships from enterprise → factory → production line → equipment → sensor. Each node corresponds to a physical device or logical entity that can not only store data but also have its own attribute configurations, visualization panels, analytical logic, and event management functions, serving as a complete business object.

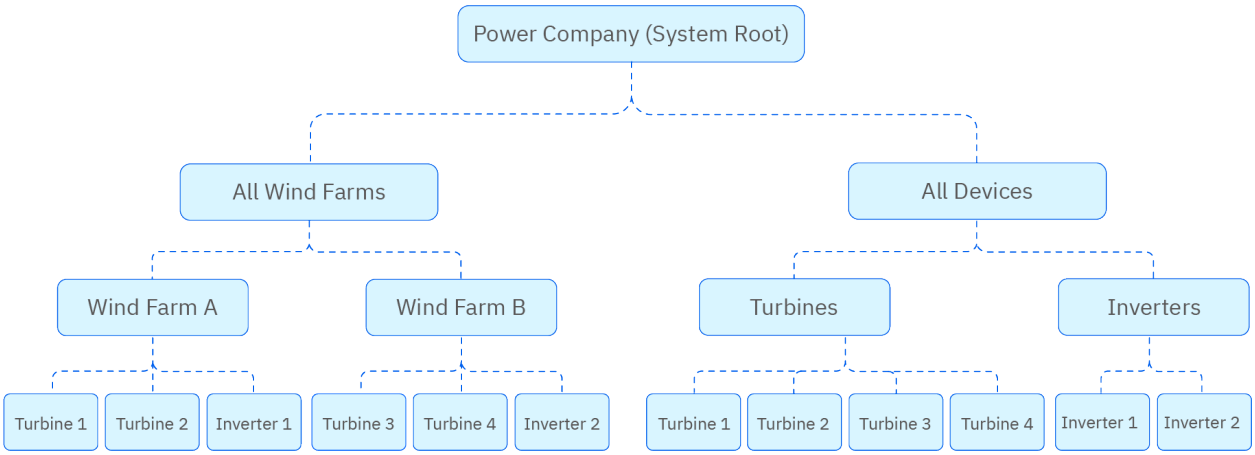


Figure 3.2: Elements in a tree hierarchy

The system supports the creation of multiple hierarchical structures based on different business perspectives. Data can be organized by organizational hierarchy (e.g., Group → Factory → Equipment) or by equipment type (e.g., Turbine → Inverter → Sensor), enabling unified data presentation and multi-dimensional analysis from various viewpoints.

Through this structured organization, enterprises can build a clear and manageable data asset catalog, allowing fragmented data to be semantically aligned, laying a solid foundation for subsequent data standardization and contextualization.

3.4 Aligning Data Structures and Definitions for Standardization

In real-world scenarios, even data of the same type often suffers from inconsistent naming, unit discrepancies, and irregular structures across different systems. For example, one system might record temperature as “tmp”, while another uses “temp”; some devices collect data in Fahrenheit, others in Celsius. For business analytics and AI algorithms, such inconsistent data cannot be directly utilized.

TDengine IDMP allows users to configure standard field names, target units, and conversion formulas for each data attribute, automatically performing data transformation and standardization. Through its data reference mechanism, it can also map data from different databases and table structures into a unified business schema—without manual intervention or data migration—achieving seamless multi-source, heterogeneous data modeling.

3.5 Enriching Business Semantics for Data Contextualization

Building on a clearly defined data structure, TDengine IDMP enables users to enrich every element and attribute with comprehensive business semantics, creating a context-aware data system.

Each element and attribute can include descriptive information to clarify its business meaning, as

well as custom tags for flexible classification and filtering. Users can define static properties such as equipment model, installation location, and specification parameters to enhance asset identification. At the attribute level, key indicators—including physical units, upper and lower limits, and target values—can be configured to provide essential baselines for analytics and alerting.

The system also supports additional attribute properties, such as whether a field is constant, visible, or participates in calculations, further enhancing the expressiveness and usability of data.

This mechanism transforms raw numerical data into rich, contextualized information—data that carries background, meaning, and business relevance—laying a solid foundation for intelligent analytics and automated decision-making.

3.6 AI-Ready: Built Differently from the Ground Up

Based on its unified data structure and contextual framework, TDengine IDMP establishes a complete set of AI-native capabilities — including automatic scenario awareness, proactive analytical recommendations, auto-generated visualization dashboards, and intelligent alert rules. These features enable AI to deliver real value without relying heavily on specialized IT or data analysis teams.

The reason AI in TDengine can proactively deliver insights is that it operates on top of a standardized, cataloged, and contextualized data foundation. What TDengine provides is not just a tool or an isolated model, but a robust data infrastructure that makes AI truly operational.

Built on this foundation, traditional reporting systems, BI tools, external AI services, and large language models can all run efficiently and respond instantly—achieving the vision of data that speaks for itself.

Chapter 4

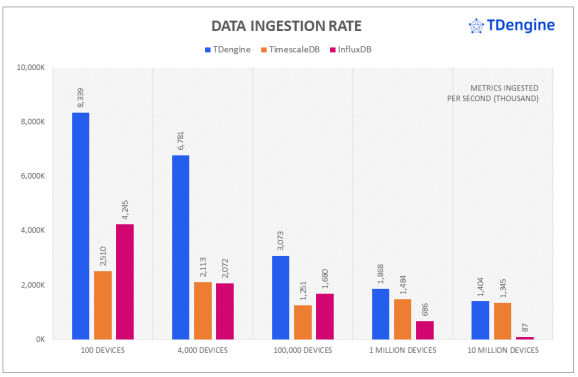
High-Efficiency Storage: Over 10× Performance Improvement

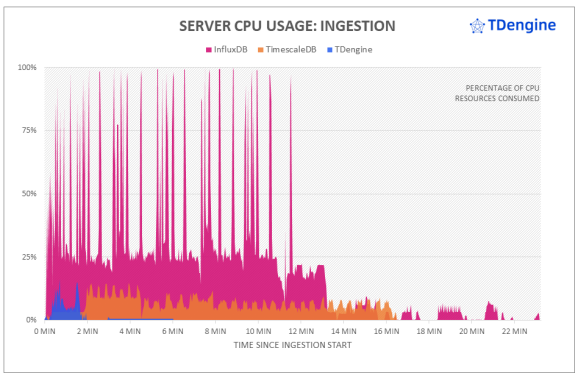
By fully leveraging the characteristics of time-series data, the TDengine team has developed an innovative storage engine that dramatically enhances both data ingestion and query performance, while significantly improving compression efficiency. Compared with general-purpose databases, TDengine delivers at least 10× higher performance while consuming less than one-fifth of the storage space. Even when compared to other time-series databases, TDengine's performance remains consistently superior.

4.1 Comparison with InfluxDB and TimescaleDB

Based on the internationally recognized Time Series Benchmark Suite (TSBS), TDengine TSDB outperforms both TimescaleDB and InfluxDB across all five IoT benchmark scenarios. TDengine's write performance reaches up to 3.3× faster than TimescaleDB and an impressive 16.2× faster than InfluxDB, while also consuming the least CPU resources and disk I/O during ingestion.

These results demonstrate TDengine's superior efficiency and scalability for handling massive time-series workloads in real-world industrial and IoT environments.

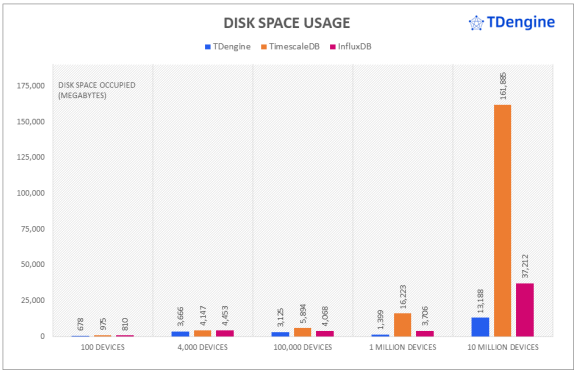


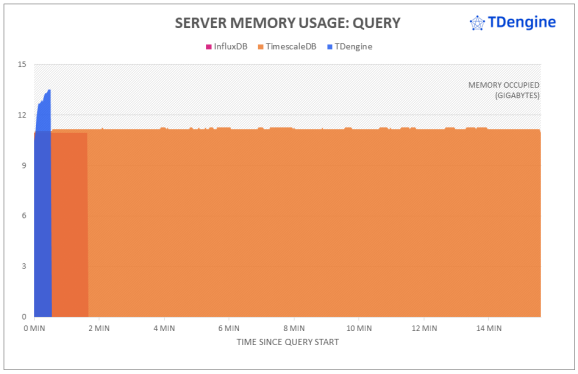


Across all benchmark scenarios, the data footprint of TimescaleDB was significantly larger than that of InfluxDB and TDengine TSDB, and this gap widened rapidly as dataset size increased. In the largest test cases, TimescaleDB's on-disk data volume reached up to 12.2× that of TDengine TSDB.

While InfluxDB's on-disk file size was comparable to TDengine TSDB in the first three scenarios, it expanded sharply in scenarios four and five, occupying up to 2.8× more disk space. These results clearly demonstrate that TDengine TSDB is better suited for large-scale time-series data storage.

In terms of overall CPU overhead, TDengine TSDB not only completed all queries faster than TimescaleDB and InfluxDB but also consumed far fewer CPU resources. Throughout the entire query process, TDengine TSDB's memory usage remained stable, further underscoring its efficiency and reliability under heavy workloads.

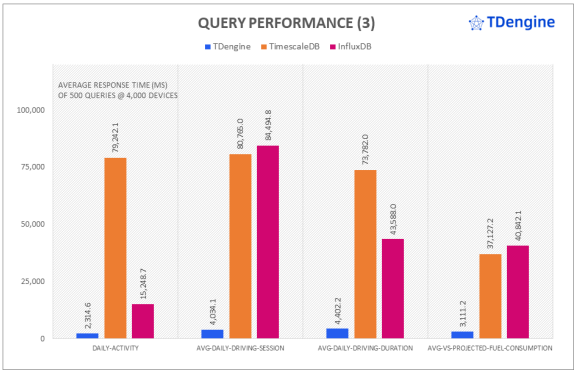


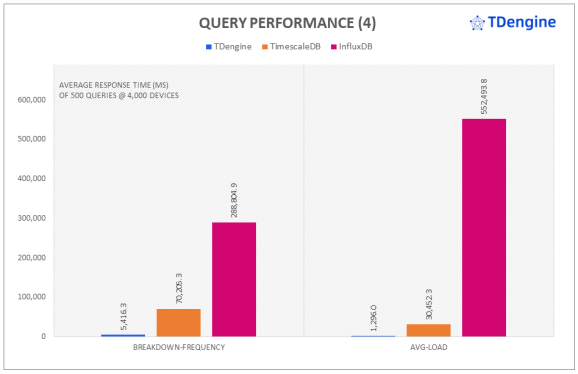


For most query types, TDengine TSDB outperforms both InfluxDB and TimescaleDB, with a particularly strong advantage in complex mixed queries.

In the avg-load and breakdown-frequency tests, TDengine TSDB achieved query speeds 426× and 53× faster than InfluxDB, respectively.

In the daily-activity and avg-load scenarios, it was 34× and 23× faster than TimescaleDB. These results highlight TDengine's exceptional ability to handle complex analytical workloads efficiently—delivering both high performance and low latency even at massive data scales.





For detailed test results and reproduction steps, see TSBS IoT Performance Report.

In DevOps scenarios, the TSBS benchmark results show that TDengine TSDB achieved write performance up to 6.7× faster than TimescaleDB and 10.6× faster than InfluxDB. During data ingestion, TDengine TSDB also consumed the least CPU resources and disk I/O, while maintaining superior storage efficiency—using only 25% of InfluxDB's storage space and just 4% of TimescaleDB's for the same amount of persisted data.

For most query types, TDengine TSDB again outperformed both competitors, with particularly strong results in complex query workloads—achieving up to 37× the query performance of InfluxDB and 28.6× that of TimescaleDB.

In both benchmark comparisons above, TDengine provides complete benchmark scripts that can be executed with a single command, allowing any third party to independently verify the test results.

4.2 A Uniquely Innovative Storage Model

Designed specifically for time-series data, TDengine TSDB introduces an innovative architecture based on the concepts of "one table per data collection point," "supertables," and "virtual tables."

This design is the core foundation of TDengine's exceptional performance, enabling efficient data organization, fast querying, and optimal compression tailored to massive time-series workloads.

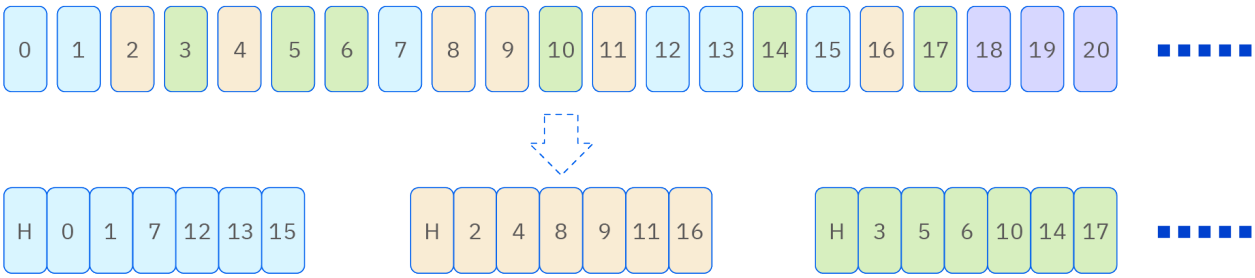
One Table per Data Collection Point

To fully leverage the characteristics of time-series data, TDengine TSDB creates a dedicated table for each data collection point (DCP). Data within each table is stored in blocks, with each block containing multiple time-ordered records and precomputed metadata. The system maintains a block index based on time ranges, which serves as the foundation for its superior performance.

This architecture provides several key advantages:

- When reading data within a specific time range, it dramatically reduces random I/O, improving query speed by orders of magnitude.
- Because each data collection device operates independently and its data source is unique, each table has only one writer—allowing lock-free writes and significantly faster ingestion.
- For an individual data collection point, data is inherently sequential, so append-only writes can be used to further boost write performance.

By adopting the one-table-per-DCP model, TDengine TSDB ensures that insertion and query performance for each data point is optimized to the greatest extent possible.



Supertables

Because each data collection point in TDengine TSDB has its own table, the total number of tables can be extremely large, making management and aggregation across multiple points challenging. To solve this, TDengine TSDB introduces the concept of a supertable.

A supertable represents a collection of data tables of the same type—that is, multiple data collection points that share the same structure but differ in their static attributes (tags). In TDengine's design, an individual table corresponds to a specific data collection point, while a Supertable corresponds to a group of similar data points.

When creating a new table for a particular collection point, users define it using the supertable schema as a template and assign unique tag values to identify that table. TDengine TSDB stores time-series data and tag data separately.

When performing aggregation queries across multiple data points, users can simply query the supertable with filtering conditions. TDengine then filters the relevant data points based on tag values before performing aggregation, thereby reducing the amount of data scanned and greatly improving aggregation efficiency.

Virtual Tables

The concepts of "one table per data collection point" and "supertables" solve most time-series data management and analytics challenges in industrial and IoT scenarios. However, in real-world applications, a single device often contains multiple sensors that collect data at different frequencies, making it difficult to represent the device with a single table. Typically, this requires multiple tables, and when cross-sensor analysis is needed, multi-level join queries must be performed—leading to usability and performance issues.

To address this, TDengine TSDB introduces the concept of a virtual table.

A virtual table does not store actual data itself but is a logical view for analytics and computation. It dynamically combines data from multiple physical tables (subtables or regular tables) by aligning and merging columns based on timestamps. Each virtual table can flexibly reference different columns as needed, enabling customized data models for different analytical perspectives—essentially achieving a "one view per need" effect.

In queries, virtual tables behave just like real tables. The only difference is that their data is generated dynamically at query time—only the columns referenced in the query are merged. This means the data presented by the same virtual table may differ between queries depending on context.

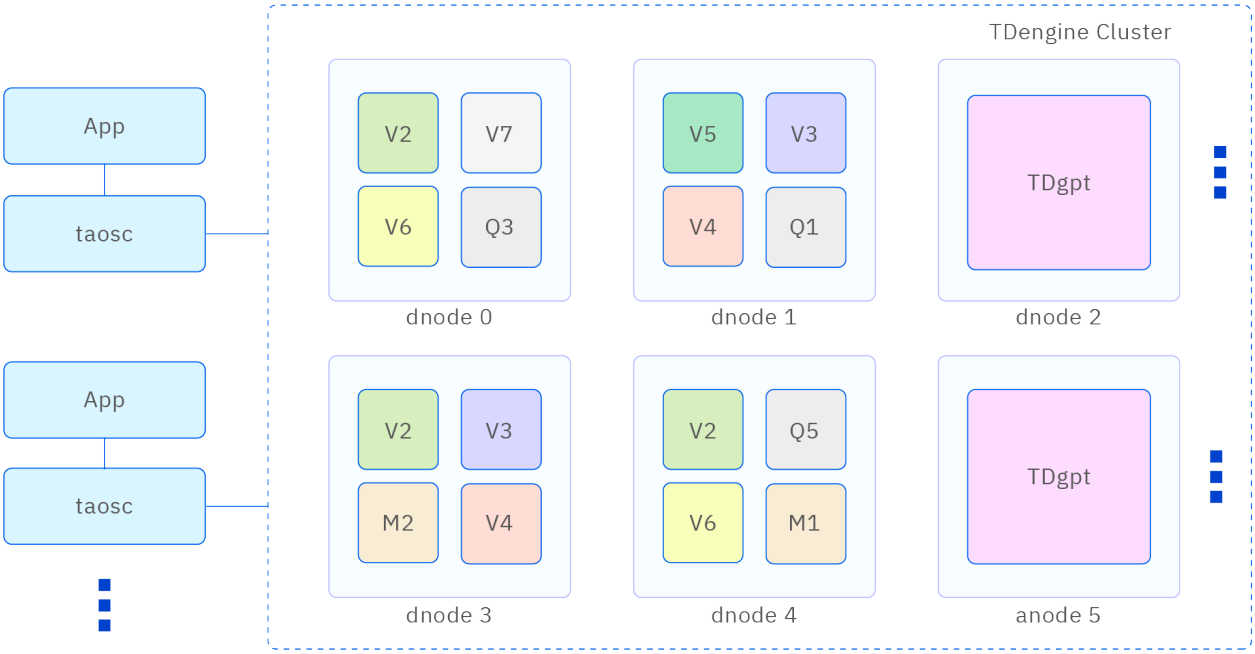
The virtual table mechanism makes “write first, model later” and “one device, multiple sensors” architectures practical. During data collection and ingestion, there’s no need to predefine complex schemas; data can be written directly into TDengine in a structure closest to the device protocol (e.g., single-column models). Later, users can dynamically create business-oriented data models through virtual tables—simplifying data ingestion, reducing modeling overhead, and making the entire data workflow more flexible and efficient.

4.3 Horizontal Scalability for Over One Billion Data Collection Points

TDengine TSDB is built on a core assumption: no single hardware or software system is fully reliable, and no single machine can provide sufficient compute and storage capacity to handle massive-scale time-series data. Therefore, from day one, TDengine TSDB was designed as a high-availability, distributed system.

By leveraging virtual node technology, TDengine partitions and shards data across multiple nodes. Its storage-compute separation architecture ensures that query and computation capabilities can scale horizontally. When additional processing power or capacity is needed, users can simply add more nodes to the cluster—no downtime, no complex reconfiguration.

The logical architecture of a TDengine TSDB cluster is illustrated below:



A complete TDengine TSDB system runs on one or more physical nodes. Logically, it consists of data

nodes (dnodes), the application driver (taosc), and applications (apps). One or more data nodes form a cluster, and applications interact with the TDengine TSDB cluster through the taosc API. Below is a brief overview of each logical component:

- Data Node (Dnode): A dnode is an instance of the TDengine TSDB server process (taosd) running on a physical node. At least one dnode is required for a TDengine TSDB system to operate properly.

  Each dnode contains zero or more logical virtual nodes (vnodes). Additionally, each dnode may include up to one instance of a management node, an elastic compute node, and a stream computing node, which together support the system's storage, computation, and orchestration capabilities.

- Virtual Node (Vnode): To enable data sharding, load balancing, and to prevent data hotspots or skew, TDengine TSDB introduces the concept of virtual nodes (vnodes).

  A single data node is virtualized into multiple independent vnode instances (e.g., V2, V3, V4 in the diagram). Each vnode operates as an independent work unit, storing time-series data and corresponding tag data for multiple data collection points.

  To enhance high availability and data reliability, each vnode can maintain three replicas, forming a vnode group. TDengine employs the Raft consensus protocol to ensure strong data consistency among these replicas.

- Management Node (Mnode): An mnode is the logical management component within a TDengine cluster. It monitors the operational status of all dnodes and performs load balancing across them. In addition, the mnode is responsible for managing system-level metadata, including information about users, databases, and supertables.

  To ensure high availability and reliability, a TDengine TSDB cluster can have up to three mnodes (as shown with M1, M2, and M3 in the diagram). These mnodes automatically form a virtual mnode group and use the Raft consensus protocol to guarantee data consistency and operational reliability across the cluster.

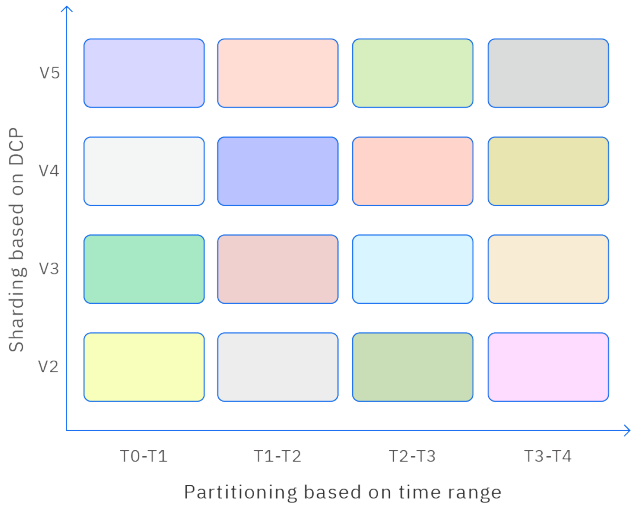
- Compute Node (Qnode): A qnode is the logical unit in the cluster responsible for query execution and computation tasks. To improve query performance and parallel processing capabilities, multiple qnodes can be deployed in a cluster, and they are shared across the entire system (as shown with Q1, Q2, and Q3 in the diagram).

- Unlike dnodes, qnodes are not bound to any specific database, meaning a single qnode can handle queries from multiple databases simultaneously. By introducing independent qnodes, TDengine TSDB achieves true storage-compute separation, allowing compute resources to scale elastically based on workload.

- TDgpt Node (Anode): An anode is the node that runs TDengine's time-series data intelligence agent. It provides capabilities such as time-series forecasting, anomaly detection, data completion, and classification. In addition to built-in algorithms, the anode can integrate with time-series foundation models and various machine learning frameworks, enabling intelligent analysis natively within the TDengine cluster.

To handle massive volumes of data, TDengine TSDB partitions the data from collection points into multiple shards, with each shard corresponding to a vnode.

Each vnode stores data from a specific set of collection points, and the data from any single collection point is stored exclusively within one vnode. A vnode contains not only time-series data but also the metadata for each collection point. Within each vnode, time-series data is further partitioned by time range and stored in separate files. Through this dual mechanism of sharding and partitioning, TDengine TSDB achieves true horizontal scalability in storage.



When an application inserts or queries data from a table, the system uses the table name's hash value to route the request directly to the corresponding vnode. Because there is no central coordinator, the architecture avoids bottlenecks entirely. For aggregation queries across multiple points, the query is distributed to all relevant vnodes in parallel — each vnode performs its own aggregation and returns intermediate results to taosc or qnode, which then perform the final aggregation. During queries, the user only needs to specify the time range and table name; TDengine TSDB automatically locates the corresponding data files, eliminating performance bottlenecks.

Thanks to its native distributed architecture, benchmark tests have shown that even with 1 billion time-series, spread across 100 data nodes, TDengine TSDB maintains excellent performance. The notorious "high-cardinality" problem in time-series data processing is thus fully resolved.

4.4 Efficient Storage: Reducing Costs to One-Tenth

In industrial IoT and enterprise IoT scenarios, data volumes are enormous — terabytes of new data can be generated every single day. As a result, efficient storage becomes a critical requirement.

TDengine TSDB employs a series of advanced techniques to maximize storage efficiency while maintaining high performance. Through innovations in time-series compression, tiered storage, and data lifecycle management, TDengine TSDB is able to reduce storage costs by up to 90%, enabling enterprises to store massive datasets at a fraction of the cost of traditional databases — all without sacrificing speed, scalability, or reliability.

- High Compression Ratio: TDengine TSDB adopts a columnar storage architecture and employs

a two-stage compression mechanism to minimize storage space while maintaining fast access performance.

In the first stage, TDengine uses encoding-based compression. For example, it applies Simple8b encoding for short integer data, and delta-of-delta encoding for long integers and timestamps. When the variation between consecutive data points is relatively small — which is common in time-series data — these encoding methods achieve excellent compression efficiency.

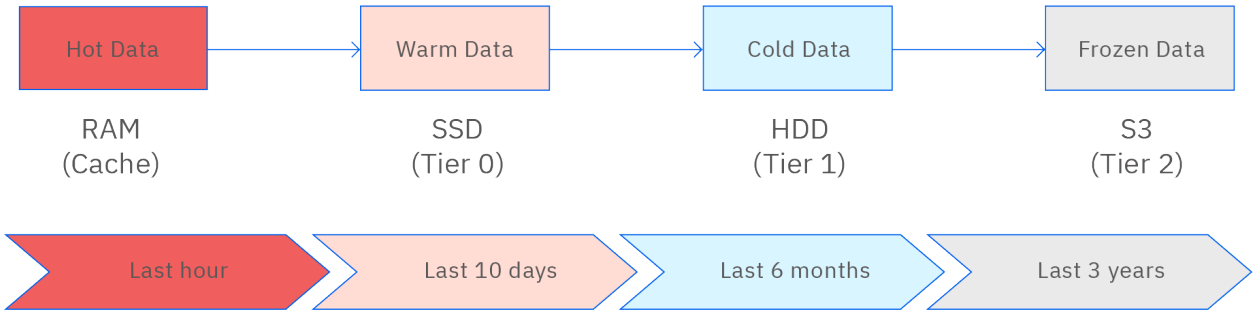
In the second stage, TDengine applies data-type–specific compression algorithms to further enhance compression ratios after encoding.

Because TDengine TSDB is designed on the principle of "one table per data collection point," the variation within a single dataset is much smaller than the variation across multiple combined sources. As a result, compared with other time-series databases that also use columnar storage, TDengine TSDB achieves a significantly higher compression ratio, effectively reducing storage costs without compromising performance.

- Tiered Storage: TDengine TSDB partitions data by time intervals, enabling tiered storage management based on data "temperature" — how frequently the data is accessed.

  The newest data resides in memory for ultra-fast reads and writes. As it ages, it is gradually moved to SSD, then to local disks, and finally, the coldest data can be offloaded to object storage systems like Amazon S3. The retention duration for each storage tier is fully configurable, allowing enterprises to balance performance and cost according to their needs.

  Importantly, TDengine TSDB abstracts away the complexity of tiered storage and keeps it completely transparent to applications. When an application queries data, TDengine TSDB automatically retrieves and merges results from the appropriate storage tiers based on the time range — eliminating the need for manual database sharding, table partitioning, or archival management at the application layer. This dramatically simplifies system architecture while maintaining optimal query performance.



By combining high compression ratios with tiered storage, TDengine TSDB achieves a dramatic reduction in storage costs compared to traditional databases.

In most real-world deployments, storage costs are reduced to one-tenth of those of general-purpose databases — and in certain specialized scenarios, the savings can be even greater, reaching as low as one-hundredth of the original cost.

Chapter 5

From Raw Data to Real-Time Insights

Storing data efficiently is only the first step — true value emerges through analysis. Beyond its robust analytical capabilities, TDengine TSDB supports real-time stream processing and integrates an AI-powered time-series analytics agent, enabling users to derive actionable insights instantly.

It also integrates seamlessly with a wide range of visualization, BI, AI, and machine learning tools, empowering enterprises to fully leverage the technological advantages of the AI era and turn raw industrial data into intelligent, real-time decision support.

5.1 Built for Advanced Time-Series Analytics

TDengine TSDB offers a suite of powerful features that make time-series data analysis efficient and intuitive:

1. High-Performance Aggregation Across Multiple Data Points: Leveraging the unique characteristics of time-series data, TDengine TSDB introduces the innovative supertable concept, which separates time-series data from tag data for optimized performance. This architecture eliminates the need for costly JOIN operations — users can perform aggregations across similar data collection points simply by specifying tag-based filters, significantly simplifying data organization and retrieval.

   Each data collection point can be assigned up to 128 tags, with full support for dynamic updates and deletions. This design allows data to be structured into multidimensional cubes, enabling flexible, multidimensional analysis that scales seamlessly across industrial and IoT environments.

2. Unified Historical and Real-Time Analysis: TDengine TSDB automatically partitions data by time, eliminating the need for manual sharding or archiving — even when managing over a decade of historical data.

   To optimize cost, it employs a tiered storage strategy based on data age, yet this process remains completely transparent to users. Whether querying data from the past 10 seconds or 10 years, the only difference lies in the selected time range — not in data accessibility or performance.

3. Time-Series–Specific Analytical Functions: Built on standard SQL, TDengine TSDB extends query

capabilities with functions tailored for time-series data, including cumulative sum, time-weighted average, moving average, rate of change, and interpolation. It also supports multiple window types — sliding, count-based, state, session, and event windows — enabling flexible time-based analysis.

Using time-window segmentation and interpolation, TDengine TSDB can align timestamps across multiple data sources, simplifying comparative and correlation analyses.

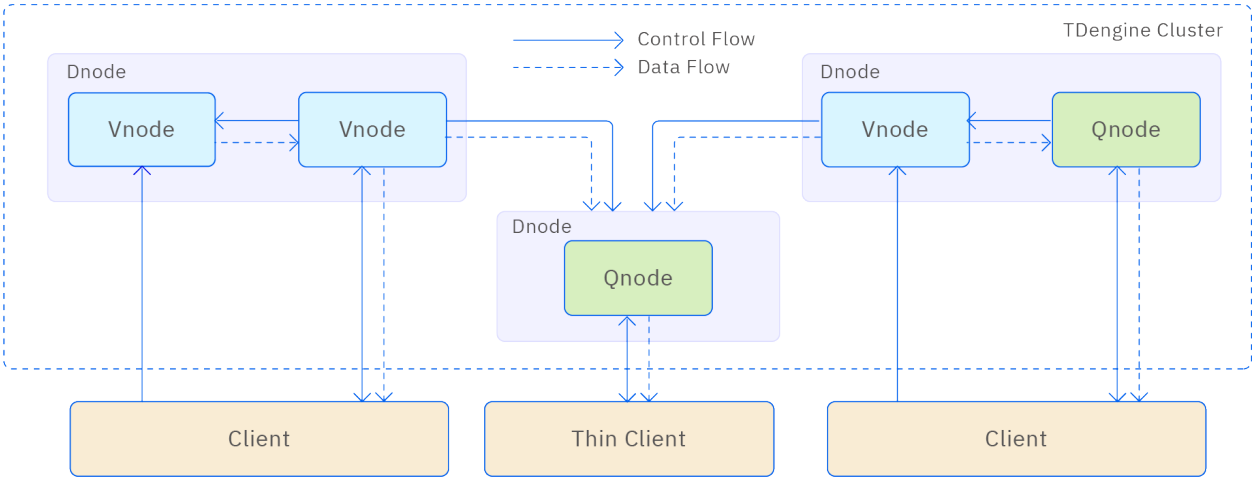
4. Separation of Compute and Storage: Starting from version 3.0, TDengine TSDB adopts a compute-storage separation architecture, allowing users to scale compute resources independently. One or more query nodes (qnodes) can be dynamically launched to accelerate complex queries and minimize latency.

   On cloud deployments, these nodes run in containers that can be rapidly started or stopped, enabling full use of elastic cloud compute resources for efficient and cost-effective analytics.

5. Open and Interoperable Architecture: TDengine TSDB fully supports standard JDBC, ODBC, and REST interfaces, enabling seamless integration with a wide range of modern AI, ML, and BI tools such as Power BI, Grafana, Seeq, and more.

   Moreover, its data subscription mechanism allows third-party applications to subscribe to real-time data streams from TDengine TSDB, enabling continuous analytics and instant insight generation.

   This open ecosystem ensures that systems built on TDengine TSDB can freely leverage the latest AI algorithms and analytics frameworks to unlock maximum data value — without vendor lock-in.



5.2 Real-Time Stream Processing

In time-series data scenarios, real-time stream processing is essential for tasks such as tiered storage and intelligent downsampling, scheduled reporting, pre-computation acceleration, anomaly detection, and low-latency alerting.

Traditionally, these workloads require deploying complex stream-processing frameworks such as Kafka or Flink, which significantly increase development and maintenance costs.

TDengine TSDB eliminates this complexity with its built-in stream processing engine, which enables real-time computation directly on incoming data streams — using only standard SQL. When data is written to a source table, computations are automatically triggered based on the defined mode, and results are pushed to a target table. This provides a lightweight alternative to traditional streaming systems while maintaining millisecond-level latency even under high-ingestion workloads.

Unlike conventional streaming architectures, TDengine's approach separates triggering from computation, maintaining the ability to process continuous, unbounded data streams, but with key architectural enhancements that simplify configuration and improve performance.

- Expanded Trigger Mechanisms: Beyond traditional write-based triggers, TDengine TSDB's stream processing engine supports a wide range of window-based triggers, including event, state, session, sliding, count, and time windows.

  Users can flexibly define trigger behavior — at window open, window close, or both — to meet different analytical or operational needs. In addition to event-time–driven triggers, TDengine also supports time-based (scheduled) triggers, enabling computations independent of event streams. Before triggers are activated, pre-filtering can be applied so that only data meeting specified conditions participate in the trigger evaluation, improving efficiency and precision.

- Enhanced Computation Model: TDengine TSDB decouples triggers from computations, meaning that the triggering table and the computation source table can be entirely different.

  This flexibility allows calculations to be performed either on the triggering table itself or across other databases and tables, with no restriction on query type — any SQL statement can be executed. Computations can be triggered at window open, close, or both.

  The results can then be handled according to user preference — sent as notifications, written into a destination table, or both simultaneously.

  TDengine's stream engine also provides fine-grained control over latency and resource utilization, allowing users to balance real-time responsiveness against system load. In cases of out-of-order data writes, it offers configurable handling strategies to ensure data integrity and consistency even in high-throughput industrial environments.

5.3 TDgpt: AI Agent for Time-Series Analysis

TDgpt is TDengine's built-in AI-powered time-series analysis agent, designed to handle tasks such as forecasting, anomaly detection, data imputation, and classification with minimal setup.

It seamlessly integrates with a variety of time-series models, large language models (LLMs), machine learning frameworks, and traditional statistical algorithms, allowing dynamic algorithm switching. With just a single SQL command, users can perform advanced time-series analysis — no external tools or pipelines required.

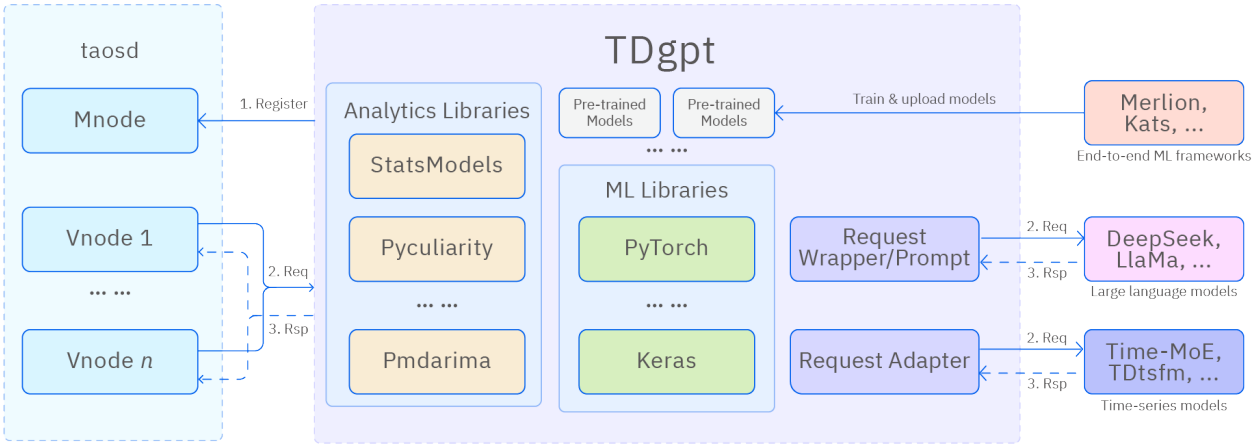
Through its open SDK, developers can easily integrate their own AI models or algorithms into TDgpt, instantly making them available to all TDengine users.

TDgpt is composed of multiple stateless analysis nodes (Anodes) that can be deployed flexibly across the system cluster or assigned to specialized hardware — for instance, GPU-accelerated nodes for model-heavy workloads.

TDgpt provides a unified interface and execution model for different algorithms. Depending on the parameters of each user request, it automatically selects and invokes the appropriate high-level analytical package or engine, then returns the computed results to the TDengine TSDB main process (taosd).

The architecture of TDgpt consists of four key modules:

- The first module is the built-in analytics library, which includes frameworks such as statsmodels, pyculiarity, and pmdarima. These provide ready-to-use algorithms for forecasting and anomaly detection, allowing users to perform predictive analytics directly within TDengine without external dependencies.
- The second module is the integrated machine learning library, featuring PyTorch, Keras, and scikit-learn. These frameworks enable pretrained machine learning and deep learning models to run directly within TDgpt's process space. The model training phase can be managed with open-source end-to-end ML frameworks like Merlion or Kats — once training is complete, models can simply be uploaded to TDgpt's designated directory for immediate use.
- The third module is the LLM adapter layer, which transforms time-series forecasting requests into prompt-based queries for general-purpose large language models such as DeepSeek and LLaMA (this capability is not yet open-sourced).
- The fourth module provides adapters for specialized time-series models, such as Time-MoE and TDtsfm. Unlike general-purpose LLMs, these models do not require prompt engineering, offering a lighter, more efficient, and hardware-friendly approach for local deployments. Additionally, the adapter layer can interface directly with cloud-based time-series analysis MaaS platforms like TimeGPT, enabling TDengine to invoke external model services for localized, high-precision time-series analytics.



During query execution, any portion of a query in TDengine TSDB that involves advanced time-series analysis is automatically delegated to the Anode. Once the analysis task is completed, the results are seamlessly assembled back into the main query flow, ensuring an uninterrupted and fully integrated analytical process.

TDgpt is designed as an extensible AI agent for advanced time-series analytics. By following the simple steps outlined in the algorithm developer guide, users can easily add their own analytical algorithms or models to the system. Once integrated, these algorithms can be invoked directly through SQL statements, lowering the barrier to advanced analytics to virtually zero.

Best of all, newly added algorithms or models can be immediately utilized by existing applications — no code changes or system modifications required — allowing enterprises to continuously expand their analytical capabilities with minimal effort.

5.4 Additional Analytical Capabilities

TDengine TSDB also includes (or plans to include) a growing set of advanced analytical functions to support deeper time-series insights.

1. Correlation Analysis, including:
   (a) Autocorrelation (ACF): Detects the correlation of a single time series with its own past values, useful for identifying patterns and periodicity.
   (b) Pearson Correlation Coefficient: Measures the strength of the linear relationship between two time series.
   (c) Time-Lagged Cross-Correlation (TLCC): Evaluates dynamic relationships between two time series — for example, identifying leader–follower behaviors.
   (d) Dynamic Time Warping (DTW): Computes the optimal alignment between two time series, even when they differ in length or sampling rate.

   These correlation and similarity analysis capabilities are currently under development and are planned for release by the end of 2025, further extending TDengine's analytical depth for complex industrial and IoT time-series scenarios.
2. Regression Analysis: TDengine TSDB supports multiple regression techniques — including linear, polynomial, and exponential regression — to model and predict trends in time-series data. These methods help users uncover relationships between variables, estimate key parameters, and forecast future outcomes with greater accuracy.
3. Batch Analysis: TDengine TSDB enables comparative analysis across different production batches, including benchmark or "golden" batches. This functionality helps identify deviations, optimize manufacturing processes, and fine-tune production formulas. Batch analysis can be visualized along a timeline or aligned by start time for direct comparison, with the flexibility to add and correlate multiple time-series datasets for deeper insight into process performance and variability.

Chapter 6

Zero-Query Intelligence: Let Data Speak for Itself

TDengine enables true hands-free analytics by automatically generating real-time dashboards and reports based on collected data — no queries, no setup, no coding required. Even without deep business expertise or SQL knowledge, users can instantly understand whether operations are running smoothly, identify efficiency improvement opportunities, and detect potential risks.

This "intelligent push" approach transforms traditional data analysis from pull-based to AI-driven, dramatically lowering the barrier to extracting value from data and empowering every user to make informed, data-driven decisions.

6.1 Comparison with Traditional Analytics and Chat BI

When using conventional BI or visualization tools, users must understand data sources, schema, field meanings, and relationships between tables. They need to know how to clean, transform, and model data (e.g., star or snowflake schemas), define business metrics, and apply analytical methods and algorithms. On top of that, users must be familiar with different chart types, configuration options, and scripting languages like SQL or Python — a process that demands both technical expertise and business acumen.

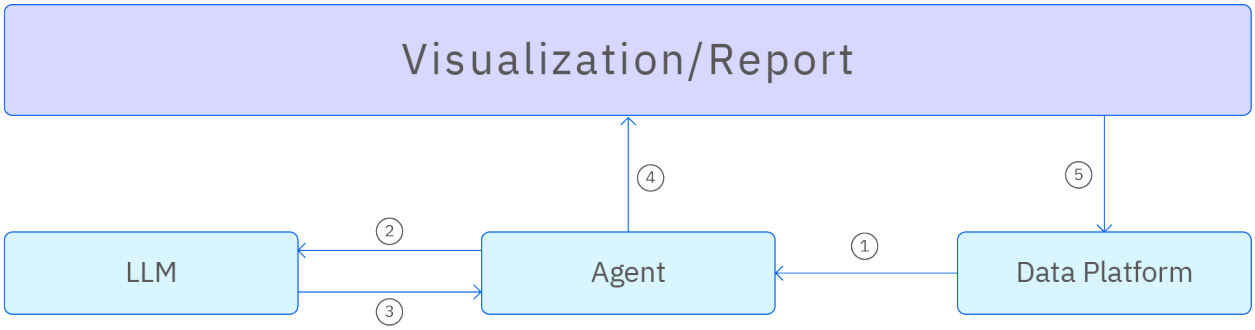
The rise of Chat BI tools powered by large language models (LLMs) has lowered some of these barriers by allowing users to generate dashboards and reports using natural language prompts. However, this still depends heavily on the user's ability to ask the right questions — and even experienced professionals may overlook important insights due to limited perspective or focus. As a result, data-driven intelligence remains partially accessible, not truly democratized.

TDengine moves beyond Chat BI with its zero-query intelligence capability. Instead of waiting for users to ask questions, TDengine automatically analyzes the data and its context using LLMs to detect business scenarios, then proactively recommends real-time analyses, dashboards, and reports relevant to those scenarios. Users can simply like or dislike the recommendations to fine-tune future sugges-

tions. With a single click, TDengine generates and deploys the corresponding dashboard or report — delivering true zero-query intelligence that makes data insights instantly available to everyone.

6.2 Main Workflow of the AI Agent

At the heart of TDengine's automatic dashboard, report, and analysis generation is its built-in multi-tasking AI Agent. The main workflow operates as follows:



1. The AI Agent retrieves the table schema for each device or logical entity from the data platform, including the table name, description, column names, data types, descriptions, physical units, and other auxiliary metadata, as well as the corresponding information for each entity's subsystems.
2. Based on the metadata obtained from the data platform, the AI Agent constructs prompts and asks the LLM to propose the real-time dashboards, reports, and analyses needed for the described scenario, returning them in a specified JSON format.
3. After the LLM reasons and responds, the AI Agent performs necessary validity checks and filters out erroneous output.
4. Using the LLM's response, the AI Agent automatically generates the configuration files required by the visualization/reporting module and sends them to that module.
5. The visualization/reporting module uses the received configuration to fetch data from the data platform and presents the final results to the user.

6.3 Why TDengine Can Achieve Zero-Query Intelligence

The process described above may seem straightforward — and indeed, many might imagine building something similar — but in reality, it presents tremendous engineering and technical challenges. Industrial data platforms often contain numerous databases and tables; in large-scale industrial environments, there can be tens of millions of data points and thousands of device types. For an LLM to understand the relationships between these databases and tables, and to infer the business meaning behind each field, is extremely difficult. Likewise, performing text-to-SQL conversions for complex queries remains a major challenge.

So why can TDengine achieve this? The answer lies in several key design innovations:

1. Unique Storage Model: TDengine TSDB adopts a "one device, one table" modeling approach. If you have one million devices, you'll have one million tables. Even when devices contain multiple subsystems with varying sampling frequencies and dynamic measurement points, TDengine's innovative virtual table design allows each device to be logically represented as a single table. In addition, the supertable mechanism enables the aggregation of data from similar devices through a single query. By combining virtual tables and supertables, TDengine eliminates most JOIN operations and simplifies SQL structure — making automatic SQL generation not only possible but efficient and accurate.
2. TDengine TSDB is a high-performance, distributed time-series database capable of aggregating, cleaning, transforming, and storing data from diverse sources such as MQTT, Kafka, OPC-UA, and OPC-DA. It features a powerful stream processing engine with multiple triggering modes — including scheduled, sliding, event, state, session, and count-based windows — and supports expression evaluation, time window aggregation, and cross-stream aggregation. The engine can actively notify applications when a window is triggered or when calculations are completed. Because all stream processing is defined and managed through SQL statements, it's both developer-friendly and easy for LLMs to generate and manage automatically.
3. Built on top of TSDB, TDengine introduces the Industrial Data Management Platform (IDMP), enabling users to build a unified data catalog and perform data standardization and contextualization. IDMP allows configuration of templates for devices, attributes, dashboards, analyses, and notifications, supports automatic unit conversion, and includes expression calculations, naming patterns, string construction, and data references to standardize data structures. Each device and attribute can also be enriched with descriptions, thresholds, locations, physical units, and tags to embed business meaning into data, achieving true contextualization. Additionally, its hierarchical tree model helps users organize data logically and visually, establishing clear relationships among physical and logical entities.

Through these foundational capabilities, the massive amount of data stored in the TDengine data platform becomes an AI-ready dataset. Without the SQL simplification brought by supertables and virtual tables, the real-time analytics enabled by the built-in stream processing engine, and the business semantics achieved through data standardization and contextualization, it would be impossible to automatically generate real-time dashboards and reports.

6.4 From Pull to Push — A Paradigm Shift in Data Consumption

TDengine's technological innovations have fundamentally transformed the data consumption paradigm. Traditionally, data analysis has always followed a "pull" model — users manually request data (e.g., through SQL queries), and the system returns results. Now, powered by LLMs and the AI Agent, data can "speak for itself" — actively delivering business insights without being asked.

This evolution shifts analytics from Pull to Push, turning users into passive recipients of insight. Relevant, real-time insights are pushed automatically, and the barrier to data-driven intelligence effectively drops to zero.

Through extensive foundational work and the integration of LLMs, TDengine has built a self-driving, autonomous real-time analytics platform that no longer depends on users' technical knowledge or analytical expertise. While TDengine is pioneering this transformation, similar systems are expected to emerge and gain popularity in the near future.

Looking ahead, TDengine will further open its AI-ready data through public APIs, allowing third-party applications to access enriched, contextualized datasets. Instead of merely returning raw SQL query results, TDengine will deliver AI-ready responses — data enhanced with business semantics and context — empowering external AI applications and helping data owners unlock maximum value from their data assets.

6.5 Tenfold Improvement in Productivity

The shift in the data consumption paradigm has led to an exponential increase in productivity. Traditionally, data analysis relied heavily on communication between data analysts or IT engineers and business personnel. Business leaders, who best understand operational realities, often lack the technical expertise to perform data analysis, while engineers typically lack business context. This gap meant that analytical requests from business teams could not be fulfilled in real time. Compressing this process and enabling instant access to insights allows for faster, deeper, and more accurate decision-making.

At the same time, business users have traditionally needed years of industry experience — often five to ten years in fields such as steel, oil, or power — to pose meaningful analytical questions. With TDengine IDMP, most common analyses no longer require such extensive experience; what once took years can now be achieved in a matter of days. Of course, more advanced analyses still benefit from expert involvement in management and technical innovation.

In building industrial or IoT data platforms, enterprises need only adopt TDengine, establish proper data source management, and define governance standards. Using TDengine's built-in tools for data standardization and contextualization, organizations can efficiently construct an end-to-end intelligent data platform with minimal effort.

Chapter 7

Open System: Connect Everything, Avoid Lock-In

As enterprises increasingly demand real-time data interaction and cross-system collaborative analytics, an open ecosystem has become the key to successful digital transformation. From its inception, TDengine has been built around one core mission — breaking down data silos.

With its open-source core and standards-based architecture, TDengine provides an open technical foundation that ensures interoperability and flexibility. Through its data subscription and publishing capabilities, it creates a fully open data pipeline that spans data ingestion, real-time flow, and downstream applications. This approach prevents the traditional pitfalls of vendor lock-in and costly system upgrades, significantly reducing both technical and economic risks in the journey toward intelligent transformation.

7.1 Open-Source Foundation: No Lock-In, Collaborative Ecosystem

The core code of TDengine TSDB is fully open source. TDengine's open-source model attracts a wide community of developers and users, creating a virtuous cycle of feedback and innovation where user needs continuously drive community-driven optimization. This collaborative ecosystem ensures ongoing improvements in compatibility, reliability, and ecosystem maturity.

7.2 Standards Compatibility: Simplifying Integration, Expanding Tool Interoperability

To address the challenges of system integration, TDengine TSDB natively supports standard SQL syntax, allowing enterprises to query and manipulate data without learning new languages. It also provides a REST API, native drivers, and client libraries for more than ten mainstream programming languages — including Java, Python, C/C++, Go, and Rust — covering the full range of industrial application development needs.

This standards-based design eliminates the need for system reconstruction, enabling seamless integration with popular BI and visualization tools such as Power BI, Grafana, and Tableau. The result is a dramatically lower barrier to integration and significantly reduced time and cost for building cross-platform industrial data solutions.

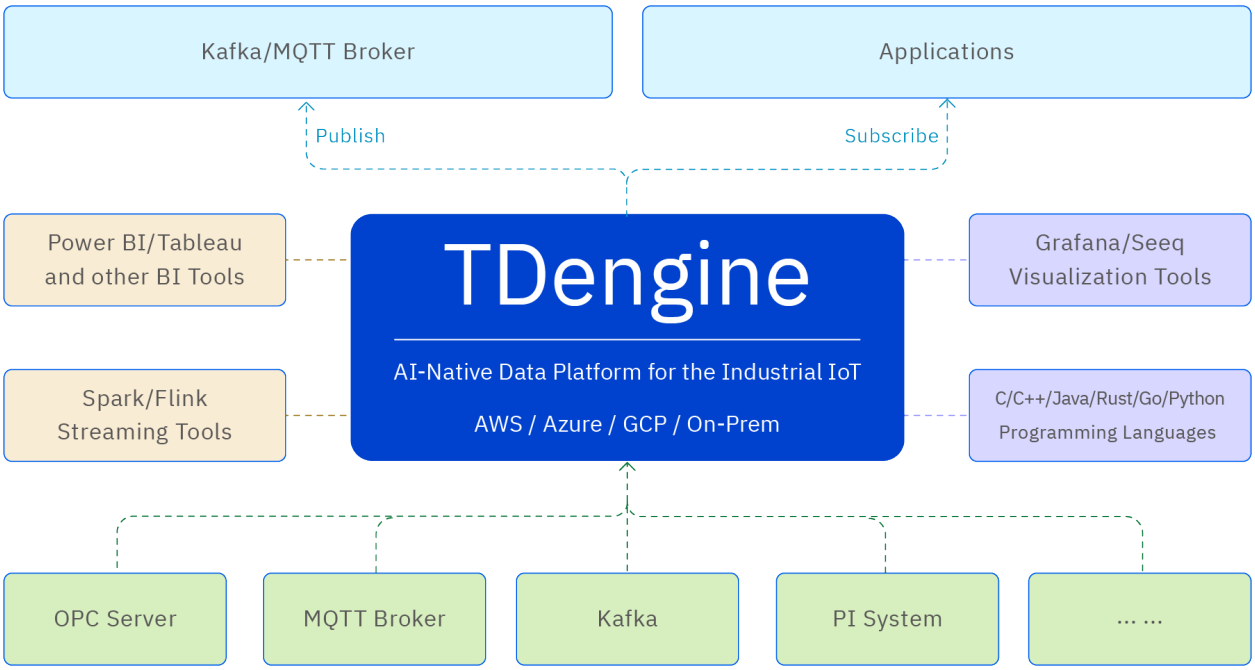
7.3 Data Subscription: Delivering Processed Data Directly to Applications

Since its first release, TDengine TSDB has supported a built-in data subscription mechanism, allowing applications to consume processed and aggregated data streams as if they were reading from a message queue. Data can be subscribed to using Kafka-like streams or the MQTT 5.0 protocol, enabling flexible and real-time data delivery across systems.

Similar to Kafka, users define topics within TDengine TSDB. A topic can represent an entire database, a supertable, or even a custom query based on existing supertables, subtables, or regular tables. Users can use SQL filters — by tag, table name, column, or expression — and even apply scalar functions and UDFs (excluding aggregation) directly within the topic definition. This gives TDengine's subscription model a major advantage over other message queues: it allows fine-grained, SQL-defined control over data granularity, filtering, and preprocessing, all performed natively within TDengine. The result is smaller data payloads, lower latency, and simpler application logic.

Once a consumer subscribes to a topic, it can receive data in real time. Multiple consumers can form a consumer group, sharing consumption progress and enabling multi-threaded, distributed data consumption for higher throughput. Different consumer groups can consume the same topic independently, each maintaining its own progress. A single consumer can also subscribe to multiple topics simultaneously.

When topics correspond to large datasets — such as supertables or databases distributed across multiple nodes — TDengine TSDB efficiently handles data partitioning and balancing among consumers. Its message queue system also implements an ACK (acknowledgment) mechanism, ensuring at-least-once delivery even under node failures or restarts, providing reliable and consistent data delivery across industrial and IoT applications.



7.4 Data Publishing: Enabling Bidirectional Data Flow

TDengine TSDB not only allows data to be subscribed to but also supports active data publishing to third-party applications — addressing the long-standing limitation of traditional industrial software that could only receive data but not share it. This feature enables true bidirectional data flow across enterprise systems, bridging industrial and IT ecosystems.

Currently, TDengine TSDB supports the following publishing mechanisms:

- MQTT Broker Publishing: Push processed data directly to an MQTT Broker for downstream applications to consume.
- Kafka Cluster Publishing: Publish data streams to Kafka topics for integration with enterprise data pipelines.
- Flink Integration: Deliver real-time data to the Flink stream processing engine to build end-to-end, low-latency analytics workflows.
- Spark Integration: Push live data into the Spark streaming engine or Databricks Cloud, supporting large-scale, distributed analytics.

All of these data publishing features are lightweight and configuration-driven — no complex coding required. With simple configuration file updates or a few SQL commands, enterprises can easily connect TDengine TSDB to downstream systems, dramatically reducing technical integration costs and ensuring smooth, real-time data collaboration across platforms.

Chapter 8

Enterprise-Grade Applications

8.1 Multi-Layered Data Security

Data security is vital to enterprise operations, and TDengine provides comprehensive, multi-level protection mechanisms to ensure the integrity, confidentiality, and reliability of data throughout its lifecycle. These include:

- Access Control: Permissions can be configured for databases, tables, views, and data subscriptions in TDengine TSDB. TDengine IDMP further enhances security with role-based access control (RBAC) for fine-grained permission management.
- Reliable Storage: Through database log files and multi-replica mechanisms, TDengine TSDB ensures that once data is successfully written, it will never be lost.
- Data Backup: TDengine TSDB supports both full and incremental backups, while TDengine IDMP provides traditional file-based backup methods for flexible disaster recovery planning.
- Disaster Recovery: Using real-time data replication, TDengine TSDB can synchronize data from one database instance to another, ensuring business continuity even in case of hardware or site failures.
- Encrypted Transmission: All communications between client and server are encrypted in transit, safeguarding data from interception or tampering.
- Encrypted Storage: TDengine TSDB supports at-rest data encryption using industry-standard algorithms.
- Additional Security Features: TDengine also provides IP whitelisting and user behavior auditing, offering enterprises enhanced monitoring and control capabilities.
- Security Certifications: TDengine has passed internationally recognized data security certifications, including SOC 2 and ISO 27001/27017, demonstrating its commitment to global best practices in data protection.

8.2 Enterprise-Grade Features

Beyond data security, TDengine delivers a comprehensive suite of enterprise-grade capabilities and services designed to meet the demands of large-scale, mission-critical deployments:

- Version Control: TDengine IDMP introduces Git-based version control for data models — a first in industrial data management. It enables version merging, rollback, and traceability, ensuring data model evolution remains transparent and manageable.
- Single Sign-On (SSO): TDengine IDMP supports enterprise-level single sign-on, seamlessly integrating with corporate identity systems for secure and unified user authentication.
- Event Notifications: Events can be automatically dispatched via email or any Webhook-supported channel to designated personnel. If no acknowledgment is received within a set timeframe, the system can escalate notifications to higher-level contacts.
- Observability: Both TDengine TSDB and TDengine IDMP provide extensive system observability metrics, enabling users to monitor system health, performance, and stability in real time.
- Flexible Deployment Options: TDengine supports deployment on Windows and Linux environments, across bare metal, virtual machines, containers, and Kubernetes. One-click deployment scripts simplify setup and reduce maintenance effort.
- Private Cloud Support: In addition to public cloud, TDengine supports private cloud deployment, and can also connect to large language models (LLMs) hosted within private environments — ensuring data sovereignty and compliance with enterprise security policies.

8.3 Enterprise-Level Services

TDengine provides comprehensive enterprise-grade support through its professional solution engineering and customer success teams, ensuring customers receive continuous and expert assistance throughout the entire lifecycle of system deployment and operation.

- 24/7 Technical Support: A dedicated support team is available around the clock, offering immediate troubleshooting and rapid resolution to minimize downtime and ensure business continuity.
- Performance Optimization: TDengine experts assist customers in optimizing system configurations, ensuring the platform operates with maximum efficiency, stability, and security in production environments.
- Training & Consultation: For every new feature or capability, TDengine provides comprehensive training, best practice guidance, and technical consultation, helping enterprises fully leverage the platform's value.
- Customer-Driven Development: TDengine values customer feedback — real-world needs and improvement suggestions are continuously incorporated into the product roadmap, driving ongoing innovation and product evolution aligned with user expectations.

8.4 TDengine Specifications and Performance Metrics

- Supported Data Types: tinyint, smallint, int, bigint, float, double, binary, nchar, bool, vchar, geometry, decimal, blob
- Maximum Record Length: 64 KB per record
- Maximum Number of Tags: 128 per table
- Maximum Number of Tables: Limited only by the number of nodes in the cluster
- Single-Node Insertion Speed: 20,000 records per second (single core, 16 bytes per record, single replica, one record per batch)
- Single-Node Query Speed: 20 million records per second (single core, 16 bytes per record, full in-memory query

For a more comprehensive list of parameters and performance data, please refer to the TDengine documentation.

Chapter 9

TDengine Application Scenarios

TDengine is widely applicable across industrial and IoT domains, covering industries such as smart manufacturing, power generation, electric grids, oil and petrochemicals, automotive, mining, renewable energy, pharmaceuticals, and IT infrastructure. Its high-performance time-series capabilities enable real-time monitoring, predictive analytics, and optimization in mission-critical systems.

1. Industrial Process Monitoring and Optimization
   (a) Manufacturing: Real-time monitoring of production line equipment (e.g., temperature, pressure, vibration) to optimize efficiency and reduce downtime.
   (b) Chemical / Petrochemical: Tracking real-time data from reactors, pipelines, and storage tanks to ensure process safety and energy efficiency.
   (c) Power and Energy: Managing real-time performance of generators, grid loads, and renewable energy sources such as wind and solar.
2. Equipment Health and Predictive Maintenance
   (a) Analyzing sensor data (vibration, current, temperature, etc.) to predict failure risks and prevent unexpected downtime.
   (b) Lifecycle management of critical assets in aviation, power generation, and rail transportation.
3. Energy Management and Sustainability
   (a) Monitoring consumption of water, electricity, and gas in real time to identify energy-saving opportunities (e.g., building efficiency, factory carbon footprint).
   (b) Tracking carbon emissions data to help enterprises achieve ESG and carbon neutrality goals.
4. Infrastructure and Smart Cities
   (a) Public Utilities: Real-time monitoring of water, gas, and electricity networks with leakage detection.
   (b) Smart Buildings: Integrating real-time data from building automation systems (HVAC, lighting, etc.).

(c) Transportation Infrastructure: Monitoring and maintaining tunnels, bridges, railways, subways, and airports with real-time analytics.

5. Life Sciences and Pharmaceuticals

   (a) Biopharmaceutical Production Monitoring: Real-time tracking of fermentation, cell culture, purification, and downstream processing.

   (b) GMP-Compliant Batch Monitoring: Ensuring that all production parameters strictly meet regulatory and quality standards in highly controlled environments.

6. Data Centers and IT Operations

   (a) Infrastructure Monitoring: Real-time performance monitoring of servers and network devices (CPU, memory, bandwidth) to optimize resource allocation.

   (b) Environmental and Energy Monitoring: Tracking power usage, temperature, and cooling efficiency across data centers.

   (c) Service Monitoring: Observing the real-time performance of diverse internet and enterprise services to ensure reliability and uptime.

7. Smart Device Monitoring

   (a) Retail Chains: Real-time monitoring of smart devices across multiple stores to collect operational data and deliver actionable insights.

   (b) Robotics and Mobility: Continuous monitoring of robots, drones, and other moving assets to provide operational intelligence and performance analytics.

As of October 2025, TDengine TSDB has achieved 880,000 global installations and serves over 500 enterprise customers, including McDonald's, Siemens, Tesla, Sinopec, Nevados, General Electric, Nio, Mingyang, CATL, Lotus, marking its position as a trusted data infrastructure provider across industries.

Chapter 10

TDengine Future Outlook

Through its innovative storage architecture, native distributed design, and built-in features such as ETL, stream processing, caching, and data subscription, TDengine provides a high-performance, horizontally scalable big-data platform for the Industrial Internet and IoT. It simplifies system complexity while reducing R&D, operational, and maintenance costs.

By building a unified data catalog and supporting data standardization and contextualization, TDengine has evolved from a traditional big-data platform into an AI-Ready data platform. Based on this AI-ready foundation, the TDengine team became the first in the world to introduce zero-query intelligence — a capability that enables the system to automatically sense application scenarios from collected data, generate visual dashboards, reports, and real-time analytics without human intervention. This innovation significantly reduces dependence on IT engineers and data analysts, empowering business users to instantly uncover insights and fundamentally shifting the data-consumption paradigm.

With its Zero-Query Intelligence and Chat BI capabilities, TDengine is enabling millions of small and medium-sized enterprises worldwide—many without dedicated data-analysis teams—to gain actionable insights from their data.

Looking ahead, TDengine will continue to:

1. Support more data source types and expand its SQL function library.
2. Offer richer visualization panels and more intuitive analysis tools.
3. Leverage AI to enhance root-cause analysis and automated diagnostics.
4. Enable TDengine IDMP to connect with third-party time-series and relational databases.

All these efforts aim to help enterprises gain real-time, in-depth insight into their operations, strengthen decision-making, and maximize data value. TDengine's mission is to make time-series data processing Accessible, Affordable, and Valuable — allowing everyone to see and realize the true value of time-series data.

TDengine

TDengine® is an AI-powered data platform designed for industrial applications, combining the high-performance time-series database TDengine TSDB and the AI-native data management platform TDengine IDMP.

With TDengine TSDB handling data ingestion, storage, and processing, and TDengine IDMP providing contextualization, standardization, and AI-powered analytics, TDengine enables industrial enterprises to unlock the true value of their time-series data.

To learn more, visit https://tdengine.com

15732 Los Gatos Blvd
Suite 135
Los Gatos, CA 95032
business@tdengine.com