What is kdb Insights Enterprise?

Reliable data capture and persistence is the bedrock of our business and our greatest strength. Use kdb Insights Enterprise to capture and analyze your high-volume, high-velocity, machine-generated, time-series data. Enterprise is developed for use on any of the three major cloud providers: AWS (Amazon Web Services), Microsoft Azure, or GCP (Google Cloud Platform) and has our super fast columnar, time-series database, kdb+, at its heart.

Here are the key features:

Data discovery and exploration

- Use kdb Insights Enterprise to take data from multiple sources (historical and real-time), process, and use the generated insights to inform decision-making processes in your business.
- Interact with kdb Insights Enterprise through our user interface or command line interface.
- We provide integrated tools to import, transform, and query data (using q, Python or SQL).
- Create graphs and charts to spot patterns and trends, and to present your results.

Scalability

Due to its deployment in the cloud, kdb Insights Enterprise is almost infinitely scalable. Scalability is dynamic, meaning that you can increase or decrease database capacity according to your requirements.

Familiar query and scripting languages

Explore your data using familiar languages; q, Python, SQL.

Access

Access to kdb Insights Enterprise is restricted by robust authentication (user identity checking) and authorization (permission to use system features is defined in user accounts).

Available and fault tolerant data

- kdb Insights Enterprise retains multiple copies of data, ensuring that no data point is duplicated or lost.

- Multiple process redundancy for both streaming analytics and query processing, allowing process and node failure.
- Deployed across multiple availability zones, allowing a single AZ failure.

Diagnostics

We provide a self-help diagnostics facility in the user interface, where you can view logs of both historical and recent events.

Support

We provide support for kdb Insights Enterprise through one of the following plans:

- Customer managed plan, where you manage your own Enterprise deployment and we supply upgrades and platform support when you need it
- KX managed plan, where we can request access to the relevant resources to manage the application on your behalf

What's included?

kdb Insights Enterprise comprises kdb+, our column-based, relational, time-series database, a graphical user interface and command-line interface, and a full complement of our pre-configured microservices.

More about kdb+

More about the UI

More about our microservices

We provide a reference architecture to get you started; you can then tailor the system to meet your needs.

How do I access kdb Insights Enterprise?

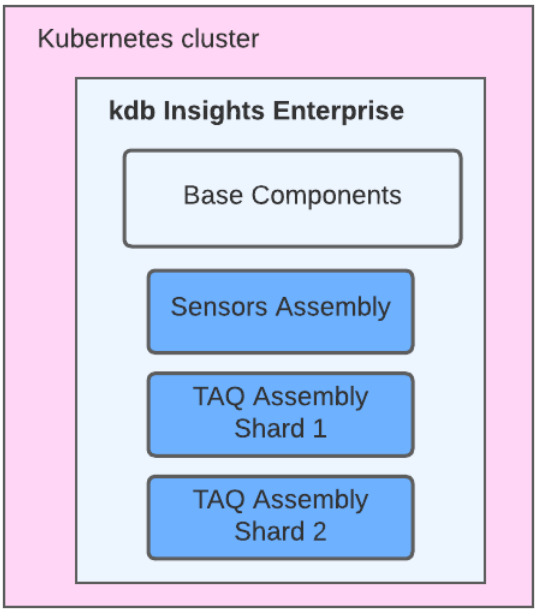
You can deploy kdb Insights Enterprise:

- In Azure Marketplace
- To AWS (Amazon Web Services), Microsoft Azure, or GCP (Google Cloud Platform)

kdb Insights Enterprise architecture

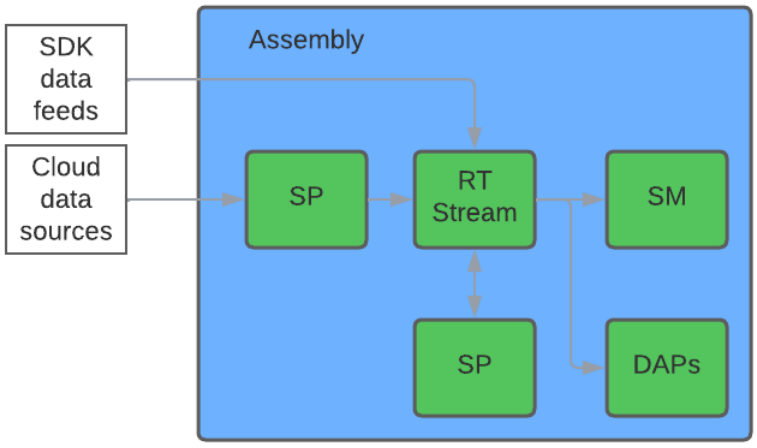
kdb Insights Enterprise provides the following core components that are used to create data-capture and analytics workflows:

- Databases (Storage Manager and Data Access)
- Streams (Reliable Transport)
- Pipelines (Stream Processor)
  These are grouped into an *assembly* (similar to a database shard). A single kdb Insights Enterprise installation can deploy any number of these assemblies, for either related or disjointed data-sets.



Dynamic workload creation

The life cycle of each assembly deployed into kdb Insights Enterprise is managed by an internal Kubernetes operator. The operator takes care of creation and placement of all required configuration files, deployments, and services, to bring an end-to-end data workflow online. The configuration of these workloads can be changed dynamically, and kdb Insights Enterprise automatically updates the underlying workflow components.

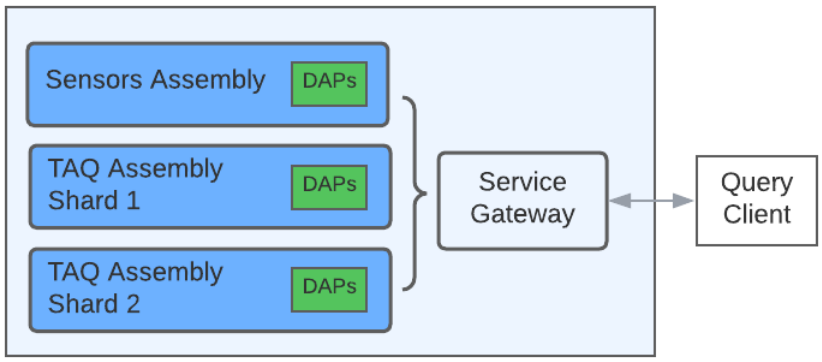


Data capture

The Stream Processor is used for getting data in, streaming data out, and creating streaming analytics such as derived data and alerts. Multiple input and output sources are built-in. For anything not built-in, data can be written directly to a Stream (Reliable Transport) via the C or Java RT SDK. The Stream Processor can then read the data from that stream if required, or it can be written directly into a database.

Data exploration

Each assembly is associated with a label set (a key/value set) of metadata which associates it with other assemblies that have an overlapping subset of labels. This allows queries to target a single assembly, or aggregate across multiple assemblies by specifying the common label set between assemblies.



Scalability

kdb Insights Enterprise is built on top of Kubernetes, allowing scaling and orchestration to be handled by the underlying cloud platform. kdb Insights

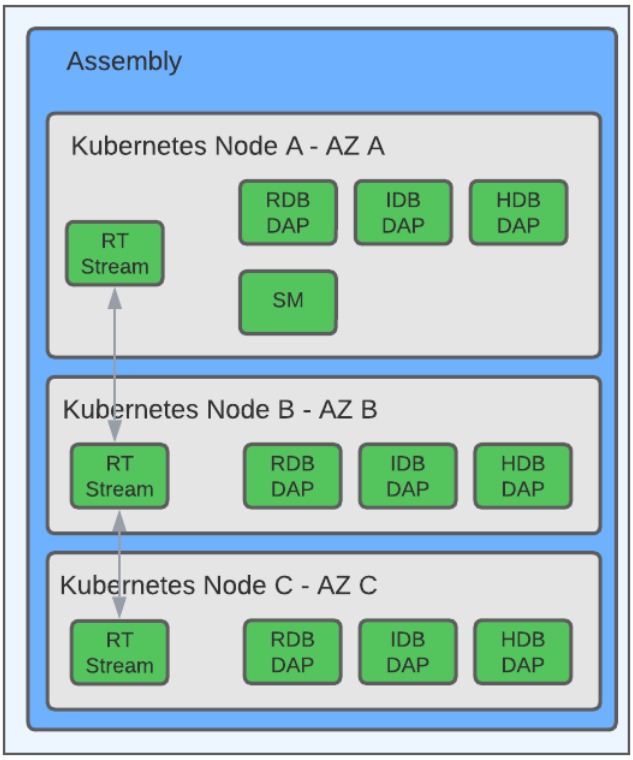
Enterprise leverages this capability by allowing scaling for a number of purposes.

Scaling for query load

Each assembly database takes the form of a single-writer, many-reader paradigm, with each of these being kdb+ processes.

Data is split into temporal tiers and automatically migrated between tiers (the most immediate data is held in memory, daily data is held on disk partitioned by arrival bucket, and historical data is held on disk and in object storage partitioned by arrival date). Each of these ranges is surfaced for queries by a separate class of data access process (RDB, IDB, and HDB respectively). Each of these allows multiple dynamic replicas to be set, and this number can change as required.

Data is written to fast shared storage, allowing query processes to span multiple nodes.



Scaling for increased ingestion

As data sizes grow for a given dataset, or if new datasets are added to an existing deployment, kdb Insights Enterprise can scale horizontally to accommodate the additional data.

Since each assembly database is a single writer, kdb Insights Enterprise scales for additional ingestion by creating more assemblies.

Fault tolerance

kdb Insights Enterprise is designed with process level redundancy and node assignment in mind; this allows for individual process, node, and availability zone failure without impacting primary data functionality.

All assembly components allow multiple redundant replicas:

- Multiple Stream (Reliable Transport) replicas making use of RAFT for coordination
- Optional multiple DAP database read replicas
- Optional multiple Stream Processor replicas with output deduplication handled by the receiving stream
  The exception is the database writer (Storage Manager), which is a single replica within each assembly. However, since the writer is not visible to queries, a failure or rescheduling of the writer does not impact ingestion and query workloads. This is because Kubernetes and the kdb Insights Operator bring the writer back online on a working node. During this time, additional data is held in memory in the real-time (RDB) data tier. When the writer comes back online, it continues to write data from the point where it left off.