



it's about time

# Kx Data Refinery

## An Introduction

---



# What is Kx Data Refinery?

---



Kx Data Refinery is a real-time, interactive time series and analytics system that aims to provide a 'shrink wrapped' solution to data and analytics.

The technology is agnostic of business vertical and all analytics, data structures and functionality are configurable and portable.

It is an interactive platform that allows access via both request/reply and pub/sub mechanisms so it can be integrated with high performance downstream client applications but also supports sandboxes for research and analysis.

The software is location agnostic and can run either as a hosted or deployed system and on any hardware, including cloud platforms.

Kx Data Refinery is designed to be generic and data centric, rather than customized for a particular function (e.g. algo execution). This means it can be used anywhere and everywhere within a firm (buy or sell side) where market data content and analytics are required.

The user base can be broken into two conceptual groups - Applications and Quant Research.

Some example use cases in the Application space include: Pre-Trade analytics, Algo trading, Post-trade reporting, End of day risk and position keeping, compliance monitoring, TCA, best execution, surveillance, MiFID reporting i.e. it can be used across the entire trade lifecycle.

Quant Research (or direct human interaction with the platform) is both logically and physically separated on the server side as the usage pattern vs Applications is markedly different.

The Kx Data Refinery platform serves the quant user base by providing fast access to full order book data away from the compute stack in use by client applications. Rapid downloads of large datasets to client environments (Matlab / R etc) are supported and it is also possible to setup client side or hosted instances of the environment specifically for quantitative research.

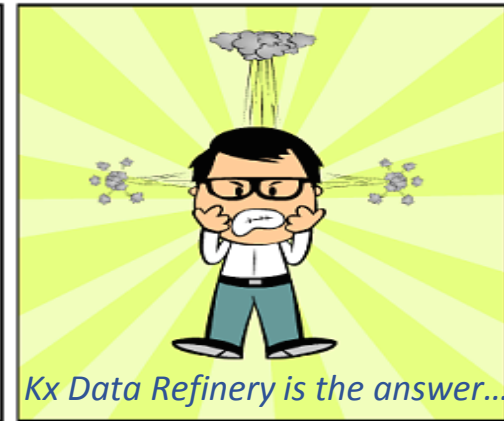
Pre-Trade Analytics

E-Trading Engines

Algo Models

Unification of Fragmented Data Infrastructure

Risk and P&L



Quant Research

Trade Data Capture

Compliance

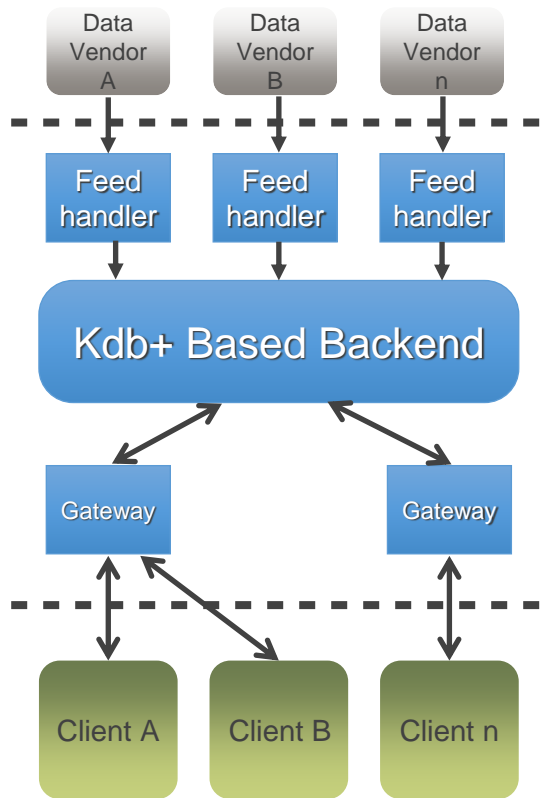
Execution Monitoring

Post-Trade Analytics

- All asset classes supported
- Level 1 and Level 2 data
- Dynamic orderbook snapshots and rebuild / replay
- Historical symbology management
- Symbology mapping (e.g. BBG ID->Ric)
- Corporate action adjustment to the tick level
- Condition code management and filtering
- Analytical (calculated) filtering on any column
- Dynamic bucketing, racking and filling at any level
- Cancellation & corrections
- Timezone management
- Data permissioning per user to the instrument level
- Nanosecond timestamps
- Seamless historical and real time data access
- Streaming and snapshot analytics

- Cost reduction - staff and technology costs for large scale tick data platforms are extremely high. Kx Data Refinery optimizes efficiency by spreading load and technology cost across a shared, hosted platform. By optimizing the technology and staffing costs, we can in turn, offer better performance with reduced cost vs in-house solutions.
- Increased functionality, scalability, performance and lower time to market. No need to build it all and take the associated risks and costs. Use a platform built and maintained by a team with proven industry success at designing, deploying and operating tick data systems.
- Opportunity to access data from multiple vendors. The Kx Data Refinery platform will contain data from many vendors, all of which can be accessed by our clients (with the right permissions and payments of course). This can all be done without the time and cost of running your own market data infrastructure.
- Standardize data and analytics across your firm. You no longer need to build analytics and data cleansing routines into individual applications - connect them all to the Data Refinery and remove the common (and expensive) issue of multiple data reconciliations.

# How Does it Work?



- Market data vendor stack provides exchange connectivity & data.
- Standard client/vendor connectivity layer for data ingress.
- Kx dev team to establish connectivity to each provider.
- Real time data publication to core platform.
- Data consumption, persistence and client requests handled here.
- Asynchronous parallel processing of client requests.
- Secure single entry point for any and all client requests.
- Async TCP based connectivity for simplicity, security & performance.
- System designed to scale horizontally to support any number of clients and data vendors.

- **Isn't this just another historical tick data provider?**

There is a common misconception that tick data platforms simply enable users to download historical (T+1) time series data. There are some systems out there that operate in this fashion but Kx Data Refinery goes much further.

The platform differentiates itself in the following ways:

- It's a live, interactive platform. Users connect and open a session. Queries are executed in real time with responses in milliseconds, not hours. Traditional request/reply (query) functionality is available, as are real time publish/subscribe mechanisms.
- It makes all data available - from the oldest historical tick to the very latest tick from the exchange.



- It's a highly scalable compute and analytics platform. It can handle everything from CEP style, real-time analytics to wide ranging historical calculations. It is designed to take the data management & processing weight away from the client.
- The platform has data intelligence. We make filtering functionality available to clients so they can remove data they don't want included in their calculations (e.g. OTC trades are generally not included in Equity Algo calculations since they cannot participate in this flow).
- It's a customizable platform. Functionality is designed to be extensible. The pre-defined analytics are simply those we believe are the minimum required. New, custom calculations and content can be added on a per client basis (and can be made private / restricted to that client).

- **Ok, so it's a database then?**

Not in the traditional sense, no.

Yes, at the backend, database like technology is in use in some areas of the system. However, the client does not need to know a query language or anything about the underlying system design or schemas to access the content they desire. The system is specifically designed to minimize the 'technology weight' placed on the client so they can focus on what they really want - the data.

Kx Data Refinery offers an intuitive programming interface that allows clients to connect and form a 'query' by simply stating what data they want to access and providing any additional parameters necessary to calculate the response / return the data.

Data is returned in a self describing format that can be parsed programmatically by the client i.e. there is no need for pre-defined data dictionaries and lookups.

- **I don't want to share an environment with other users – I believe my needs are too large/complex/specific.**

This is fine. The Kx Data Refinery can be deployed as your own externally hosted instance or on your site directly.

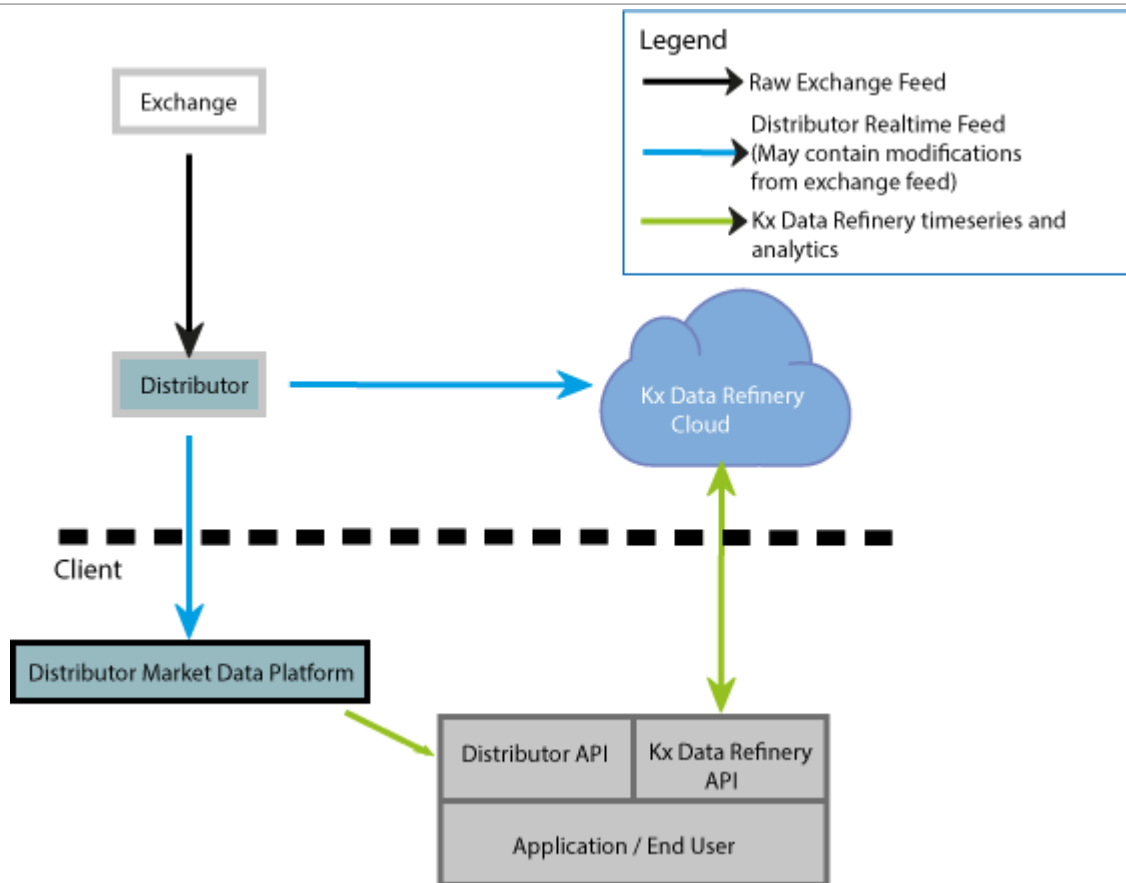
Our technology team will assess your requirements and design and deploy the system for you. We can also fully manage and support the system for you as required.

- **What about other datasets? Can your technology be used to capture my own data?**

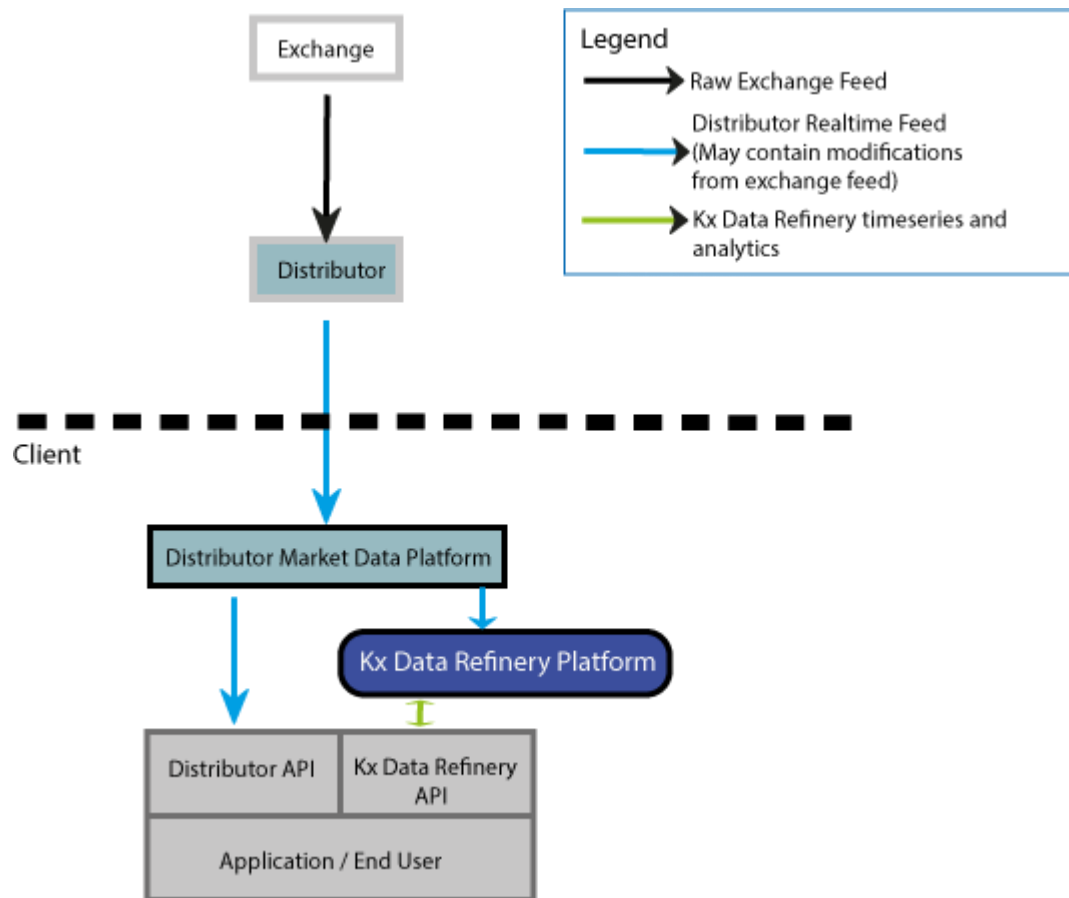
Yes, it is possible to deploy instances of our technology to capture datasets other than market data. A simple example is execution data capture. There are other concerns to consider here e.g. security & compliance but as previously mentioned, technology can be deployed at the client site where required to provide the required level of technology compliance.

Both the backend data structures and schemas are extensible and the query interface is designed to support any definable dataset.

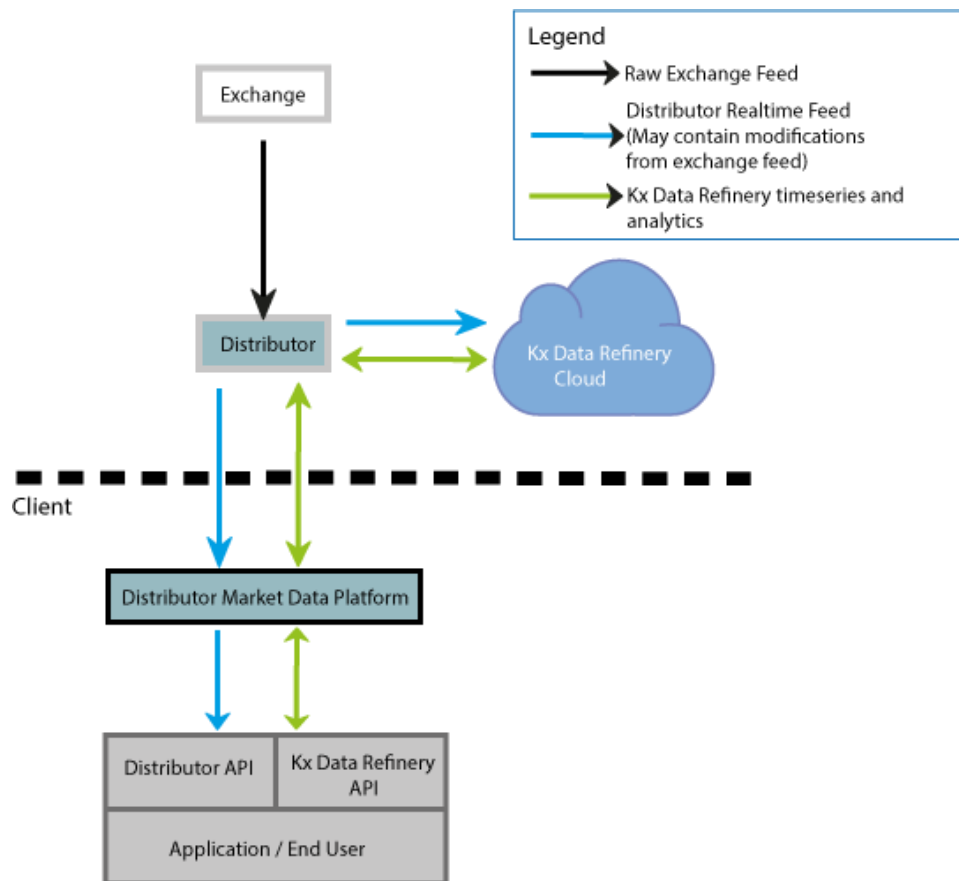
# Deployment Options – Cloud Deployment



# Deployment Options – Deployed



# Deployment Options – OEM/White Label



## Key Points

- Globally distributed and accessible from any location
- Clear separation of high performance, critical application usage from ad-hoc research & analysis environments.
- Co-Location of data and compute resources (Data on disk right next to CPU & Memory, no need to shift tick data around the network for processing, only results to clients)
- High redundancy and availability by design to ensure maximum uptime and minimum impact when issues are seen. This is inherent through the entire stack from the way data feeds are sourced to how data is distributed across the storage nodes.
- True horizontal scalability – theoretically there is no limit to the data volumes that can be captured and processed, just add hardware and distribute your datasets.
- Highly parallel execution. Need it faster? Sub-divide your data and multi-process.
- Designed to manage and expose both real time and historical data seamlessly.

- A single, common API and access layer that is extensible to accommodate new demands – the minimum possible number of functions with the maximum possible flexibility per function. Avoid the ‘n thousand function’ issue (don’t know what the system does anymore, replicated functionality, hard to navigate etc) and by-proxy produce a modular codebase that is as small as reasonably possible.
- Flexible data models and datasets – handles everything from Equity tick data to execution data to Rates Curve objects.
- Segregation of clients from backend complexity. Single entry point per client irrespective of the number of datasets they need access to.
- Common codebase globally and uniformity of hardware and software solutions to ensure the platform can be managed and supported with the minimum number of resources and that developers can work on different areas of the platform with confidence.
- The same technology that’s used to run the hosted tick platform can be used to deploy custom time series data capture solutions at clients sites (or hosted) with seamless integration between the two e.g. algo flow data & analytics, trade data capture.