kx | it's about time

Kx Data Refinery
Product Overview

CONTENTS

1 INTRODUCTION

Making decisions has, paradoxically, become both easier and harder. Easier because there is more data to inform and validate options, but harder because it is sometimes difficult to get the very data you want. It might exist within terabytes or even petabytes of other data, it may be streaming in at micro or even nanosecond speeds or it might be stored in one format in one location and a different format in another - the so-called Volume, Velocity and Variety of Big Data. On top of that it might contain duplicates, omissions or inconsistencies. So add Veracity to the mix. It's a difficult problem to manage – but a potentially lucrative one to solve.

For the last 20 years, financial markets have arguably been at the forefront of the Big Data arena (only it didn't have a name then!). Over that time, Kx has become the benchmark technology for time series data and analytics within Finance. The techniques learned and technologies developed over that period continue to lead the market today and as other business verticals advance into the arena of Big Data and complex analytics Kx has become increasingly relevant to the complex and interesting data problems outside of the financial markets.

Kx Data Refinery is designed to wrap all of the hard won experience and battle hardened technology developed for financial markets into a platform that can be used for real-time and historical data capture and analytics across any business vertical.

Built on the strong foundations of Stream and Control for Kx, Kx Data Refinery takes our software solutions one step further – from being a set of core technology (kdb+ and q) and foundation libraries (Stream & Control) to an implementation ready, data-centric analytics platform ready to deliver to your business.

The "Four V's" of Big Data (Volume, Velocity, Variety and Veracity) continue to task the technology industry to deliver solutions that can address the challenges therein. Often this results in companies implementing custom solutions using a variety of technologies that address one of these domains well but not all. Kx Data Refinery offers an integrated solution that covers all of these data challenges using a single core technology – kdb.

- Volume – Our architecture scales both horizontally and vertically. We can run on anything from a Raspberry Pi to the largest servers available to the market. We can deploy on one machine in one location or hundreds deployed globally meaning we have limitless scalability.
- Velocity - Kx Data Refinery can process millions of inbound records per second while maintaining sub-millisecond outbound query response times. Data is made available for query and calculation immediately upon receipt and our client gateways provide a single entry point for both real-time and historical data in the same query.
- Variety – Kx Data Refinery Feedhandlers provide connectivity to multiple sources and formats with the flexibility to extend to others with accompanying data management functionality.
- Veracity - Kx provides managed services for the validation, cleansing and workflow management of data exceptions, a service for data queries and analysis and an analytics and data cleansing framework that is easy to configure and adaptable to data from any source.

In addition to sourcing and distributing data Kx Data Refinery provides highly resilient mission critical architecture and sandboxes for ad-hoc analysis, dashboards for streaming visualizations and IDE environments to support quantitative research and data science.

> Kx Data Refinery helps data owners to unlock and monetize the value of their information assets by enabling their clients to retrieve data when they want it and how they want it – fast, accurate and enriched.

2 KX DATA REFINERY OVERVIEW

As technologies to support Big Data evolve, the business around it evolves too. Born from financial markets where exchanges, regulators, data vendors and financial institutions seek to gather, store, process, distribute and ultimately monetize their data, Kx Data Refinery is a data capture and analytics platform for the modern era.

Performing analytics on real-time and historical data, processing queries, delivering results and feeding downstream applications requires an immediacy and agility that is often difficult to achieve. Doing it for multiple customers and markets is even harder. Kx Data Refinery can do both.

2.1 BUSINESS – KEY POINTS

Kx Data Refinery is deployable across multiple industries including utilities, pharmaceutical and retail as well as the financial markets. It supports both mission-critical operations and data science and quantitative research functions. Clear separation of these use cases ensures we can offer the operational performance and stability required to support your mission critical applications and processes as well as providing a powerful, flexible platform for research and idea generation that data scientists and researchers require.

Provider Benefits:

- Centralization and normalization of data and analytics for your entire organization.
- Support real-time downstream application use cases in addition to ad-hoc research and analysis
- Single coherent technology solution end to end.
- A company that not only understands technology but also data and data science disciplines.

Flexible deployment options of Kx Data Refinery include:

- Can be run as an entirely hosted and managed service, nearshored with your business.
- Can be deployed locally within your own data centers and infrastructure.
- Can be a hybrid of both approaches enabling you to host non-sensitive data in the cloud whilst retaining local control of any sensitive data whilst still providing seamless access to both.
- Runs in the cloud or on 'bare metal'
- Can be historical / batch only, real time or full streaming, you only pay for what you need.

Sample Applications of Kx Data Refinery

- TCA Analysis
- Best Execution Analysis (RTS27 & 28)
- Market Impact Analysis
- EBBO/NBBO
- Systematic Internalizer Determination

2.2 TECHNOLOGY – KEY POINTS

Often vendor solutions are designed to fill a highly specific use case. Kx Data Refinery takes a different approach. It is designed to act as your centralized platform for data capture, management, analytics and distribution. It is our fundamental belief as a firm that once a solid data foundation is in place, the building of solutions to fill a particular need is greatly simplified. Your application no longer needs to do the heavy lifting of data capture, storage, cleansing and calculation. This is done by Kx Data Refinery instead, enabling you to focus on building the solution you need today, quickly and easily.

The following image shows an example from our financial markets implementation and highlights our modular approach to solution delivery.
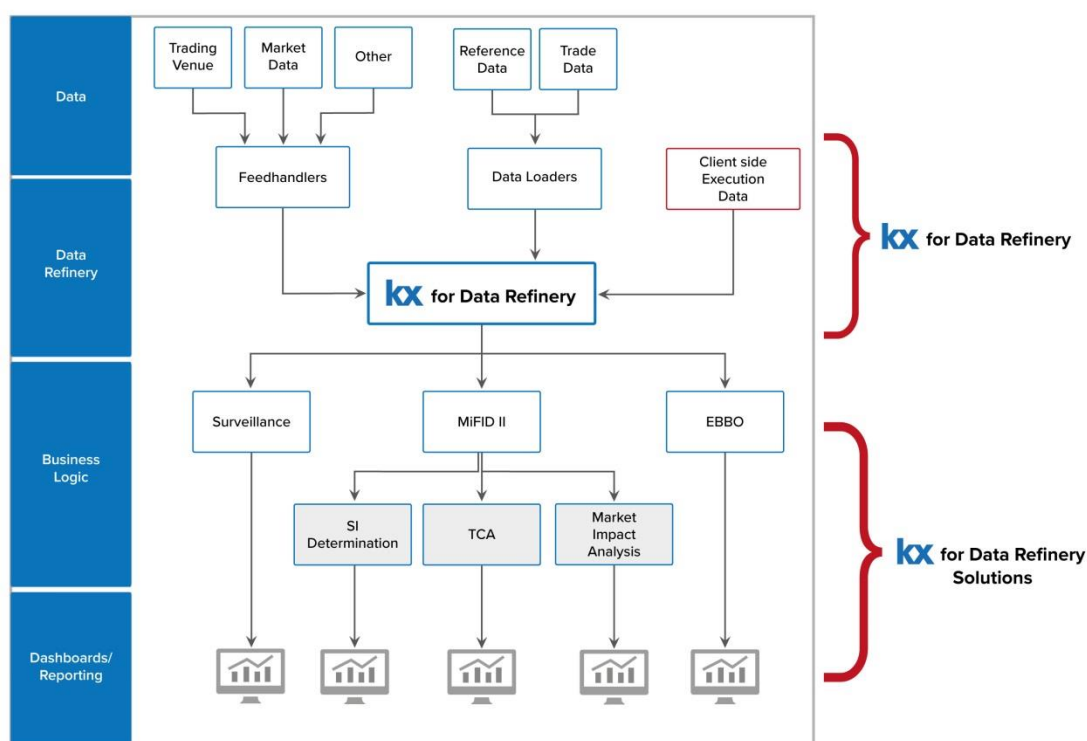
As the diagram shows, Kx Data Refinery provides both the market data and the order and execution data capture backbone. This data and associated analytics can then be served out directly to specific business solutions (e.g. Surveillance) without those solutions needing to recapture and persist their own copies of the data.
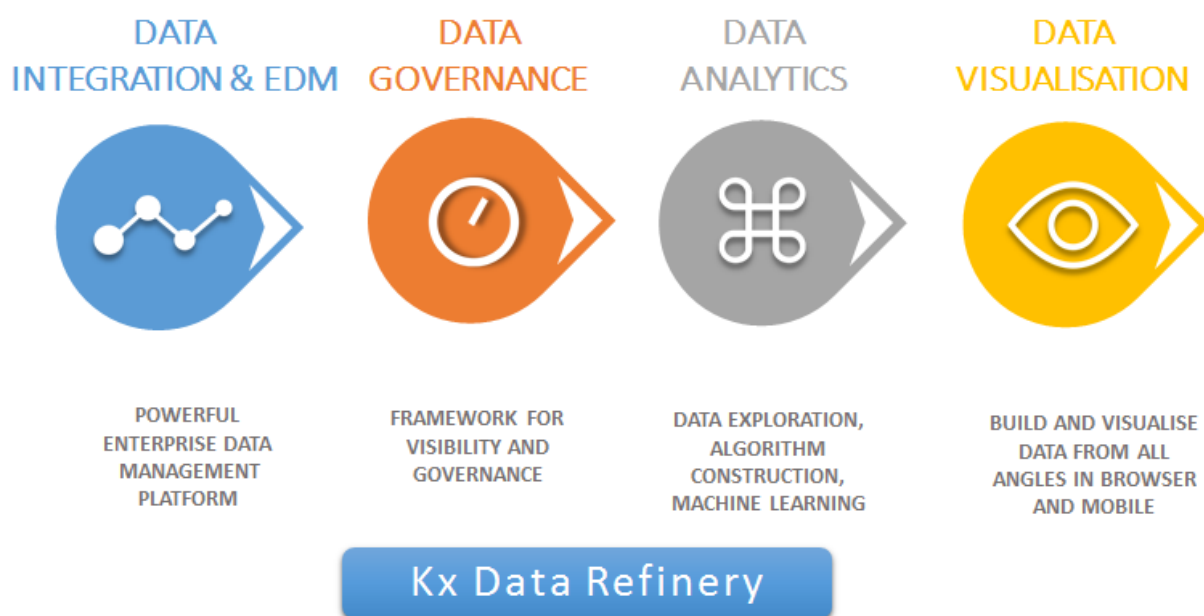
Additional points for consideration:

- Handles both structured and unstructured data.

- Can be globally or locally distributed and made 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.

- 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.

This combination of design principles and supporting components in the Kx platform provide the end-to-end capabilities of Kx Data Refinery.

DATA INTEGRATION & EDM	DATA GOVERNANCE	DATA ANALYTICS	DATA VISUALISATION
POWERFUL ENTERPRISE DATA MANAGEMENT PLATFORM	FRAMEWORK FOR VISIBILITY AND GOVERNANCE	DATA EXPLORATION, ALGORITHM CONSTRUCTION, MACHINE LEARNING	BUILD AND VISUALISE DATA FROM ALL ANGLES IN BROWSER AND MOBILE

Kx Data Refinery

2.3   DESIGN PHILOSOPHY

Outside in, Not inside out.
As far as is possible, the platform has been designed and developed from the perspective of the end user. We have attempted to 'shift the dial' in terms of ease of use for data and analytics platforms.

The technology is necessarily complex, however at all times we have attempted to develop the system in such way as to support the business focus of the end user and abstract them from implementation detail that commonly haunts the user experience with these types of platform.

For example, it is not necessary to understand the underlying platform architecture, process layout and database schemas to query the system, it is not even necessary to understand SQL or Q programming in any detail. Users access the system through a single entry point (the gateway) and interact with it through a series of easily understandable, human readable function calls.

Data is King
Kx Data Refinery provides a full suite of tools for managing your data and provides functionality to abstract the end user from the vagaries of data differences between markets and vendors.

For example, when dealing with financial market data from different exchanges, the same trade event on two exchanges can be represented with a different trade condition. (e.g. a regular trade on the LSE may come with a condition `A but on NYSE with an `R.) When multiplied across the number of asset classes, event types and exchanges this rapidly becomes an extremely complex and costly detail to manage.

Kx Data Refinery implements a system of named filters to abstract the end user from this complexity. When a request for data is made a predefined filter name that makes sense to the user can be sent, and the data returned cleansed according to these rules without the need to maintain complex SQL code.

Everything Changes.
The world is changing constantly and by its very definition, our platform is all about history and time series. Our most immediate data is what is happening right now, live, but our oldest data is measured in years and decades. Kx Data Refinery manages the changes in data over time so that our users can focus on what they want and not deal with the minutiae of change in the underlying data set.

A simple example from financial markets is dealing with symbology changes over time. Companies get acquired, merged, spun off etc. all the time yet the client may need all the history for that entity presented as a single view e.g. Blackberry decided to change their exchange ticker from RIMM to BBRY on 2013.02.04. This means we have to design the system such that the client can use the BBRY ticker and yet still get the tick history stored under the RIMM ticker.

Abstraction
Kx Data Refinery is intentionally designed to be end use case agnostic. The platform is designed to operate as the core time series and analytics platform for the entire organization as opposed to serving one use case only. By sticking to that principle, we ensure that our clients extract the maximum positive impact across their organization by centralizing their entire application and research practice on a single source of data and analytics.

That way the VWAP number that you see in your post trade report is the same one you see in your algo monitoring system and the same one you provide to compliance and your clients in your Best Execution reports.

2.4   SAMPLE APPLICATIONS

All screen have been created using Dashboards for Kx, an integral component of Kx Data Refinery which enables power used and business users alike to create, and share, customize views of data and analytics results
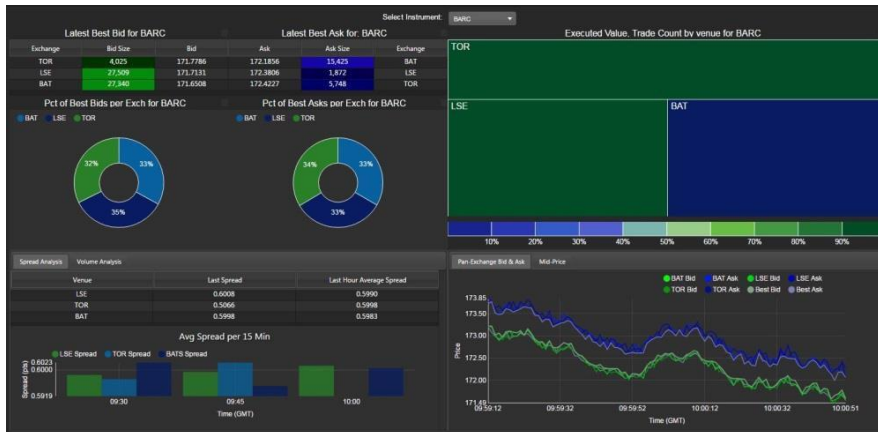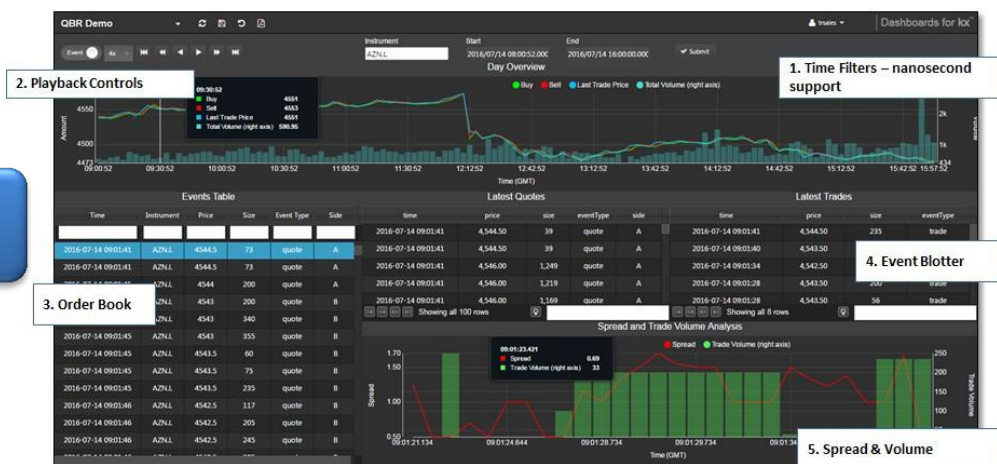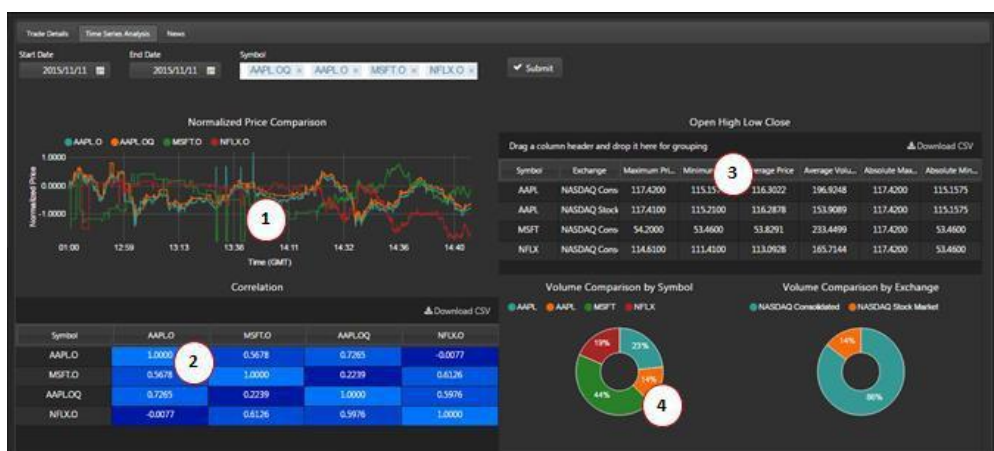
2.5   ARCHITECTURE

Our design philosophy and some of the key points on our technology have already been covered at a high level. The following section aims to provide some additional depth and detail.

Whilst reading this document, we encourage you to explore our website which contains further detail on many of the technologies referenced here.

In terms of the Kx software stack, the simplest way to think about the Kx Data Refinery platform architecture is as follows:

- Kdb+ is at the core of everything we do.
- Stream & Control for Kx provide the base software frameworks to support our architecture.
- Kx Data Refinery is the ultimate expression and implementation of these core technologies.
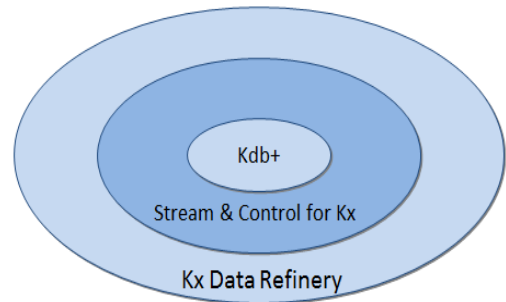
The following diagram provides a more complete picture and shows how Kx Data Refinery services combine to provide a comprehensive data management solution.
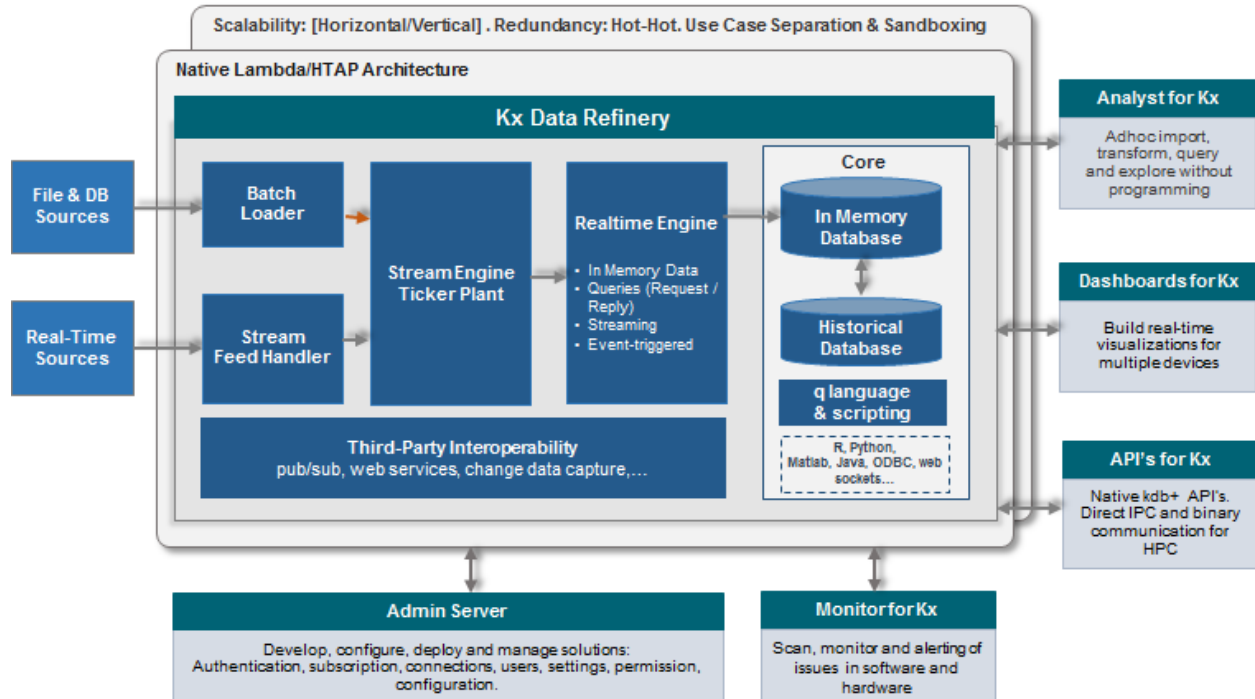
2.6   SCALABILITY

As noted previously, the platform is designed to be both horizontally and vertically scalable and supports cloud deployment as well as a 'bare metal'.

The following diagram shows a typical deployment within financial markets. There are three 'tiers' in this example with real-time feed processing, real-time capture and historic data separated onto dedicated hardware.
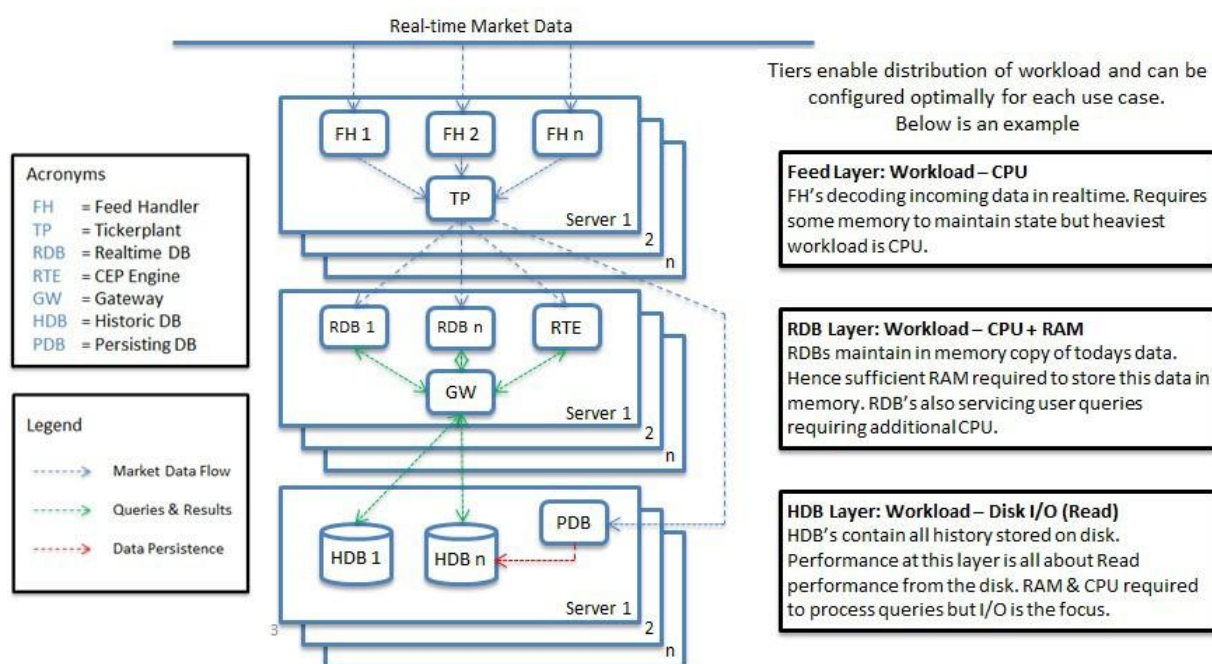

The diagram shows Tiers 1-3, servers 1..n and processes 1..n as an illustration of how the system can scale horizontally across many servers for performance. This is achieved by sharding the processes and the dataset that those processes are consuming.

- Processes can be run over 1 or many Tiers supporting scale up of performance as required.
- RDB's are all retained in memory while HDB's are stored on disk. Kx Data Refinery multi-processes at both levels but in different ways.
    o Data is sharded across n processes for the RDB's
    o Historic data can be mounted n times and load balancing at the HDB layer.
- HDB's are also multi-threaded at the backend (disk level), RDB processes are single or multi-threaded.
- The Gateway provides a single entry point for the client and manages the query routing across the system.
- The combination of these approaches means there is theoretically no limit to how much data can be captured and processed.
- This flexible approach also ensures optimal usage of resources as we can deploy on anything from a single node to thousands. In addition Kx Data Refinery has native support for cloud solutions and can deploy on any cloud offering.

2.7 FAILOVER, REDUNDANCY AND SANDBOXING

Kx Data Refinery provides flexible deployment options to suit the level of redundancy required. It can run standalone, hot-warm, hot-hot (parallel) and in single or multiple data centers. The following diagram shows a typical Hot-Hot deployment in independent data centers with a T+1 sandbox environment.
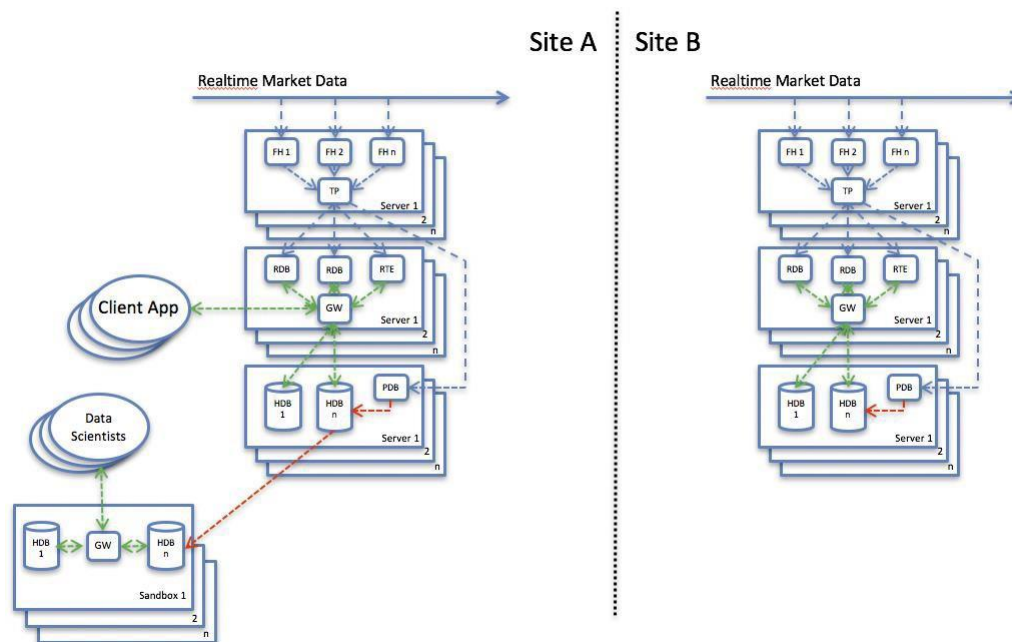
- Sites A and B can be located in separate data centers (or the same one).
- An independent data feed infrastructure upstream provides redundancy at the feed level as well as machine and process level within our own infrastructure.
- Clients maintain two connection handles, one to each site gateway. If the gateway or the data centre goes down, the client reconnects to the secondary.
- The gateways in each site maintain a secondary handle to all processes in the other site. If a handle disconnects e.g. through process failure then all queries are automatically re-routed to the secondary site. This provides seamless failover (for the client) in the event of process or machine level failure.
- Sandbox in this example is a T+1 copy but data can also be provisioned in real time if required.

Sandboxes Capabilities

The diagram above illustrates another powerful feature of Kx Data Refinery – the ability to sandbox users like Researchers and Quants who have markedly different usage profiles and requirements than those of applications and production systems.

- Applications are deterministic in their usage but require extremely high performance i.e. they use the system heavily but do 'standard things' very often requiring fast responses. These applications are often mission critical to the end user firm.
- Researcher and Quant needs are the opposite. Usage is non-deterministic, often requiring a high degree of customization and flexibility but outright performance is less of a concern and usage is not mission critical.

These differences necessitate separation of the physical system and the provision of a dedicated sandbox environment that provides full functionality and data access yet ensures that production system are not compromised by speculative and non-standard usage patterns that may overwhelm resources. Kx Data Refinery sandboxes provide such environments where required.

2.8 ENTERPRISE FEATURES

Kx Data Refinery provides all the control and management features crucial to an enterprise application. These are provided within Control for Kx and include functionality for configuring, deploying, maintaining and extending the Kx Data Refinery solution.
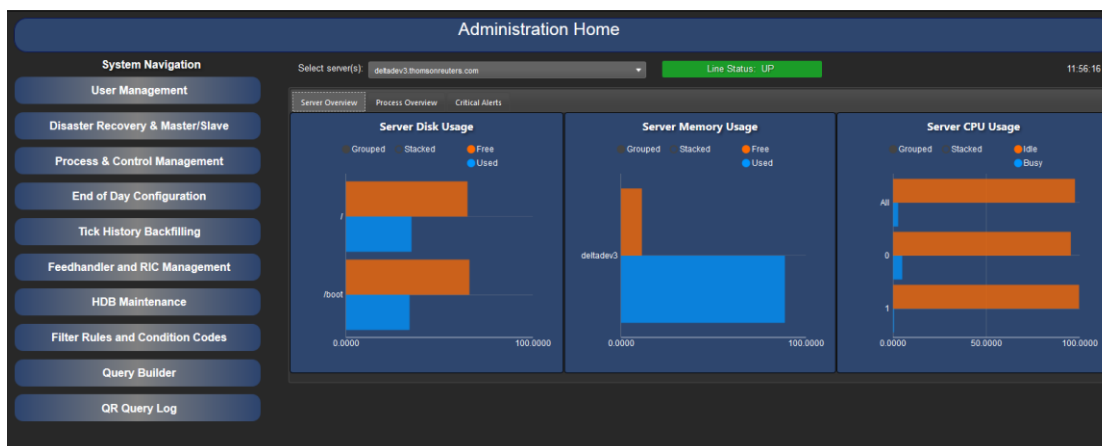
User Management, Authentication and Entitlements

Administrators can define entitlements and permissions at both user and group level. Entity Groups can be defined that group together sets of Dashboards or individual Dashboard Apps. User or user group entitlements can then be assigned to confer no access, full access or read only access. Interfaces are available for integration with LDAP and other security mechanisms for user authentication.

System Parameter and Control

Administrators can define and control parameters determining the codes, filters exchange etc.. on which queries are executed

An comprehensive audit log record all amendments and who made them

Additional features include:

Analytics Library	Failover and Replication	Monitoring
A set of customizable, modifiable analytics functions to modify and extend system behavior	Enabling Hot- Hot, Hot-Warm and Hot- Cold failover configurations	Early detection and notification of infrastructure and business level errors

3 KX DATA REFINERY CASE STUDY

Kx Data Refinery was selected by Thomson Reuters to provide the underlying technology for the next generation of its time series and analytics platform – Velocity Analytics 8.0 (VA8). The solution makes Thomson Reuters data easily accessible for client, reduces development overhead and costs.

Strategic Partnership with Thomson Reuters for Streaming Analytics

First Derivatives (FD) announces a strategic partnership for the use of its Kx technology with Thomson Reuters to power the latest version of their financial Data as a Service platform, Velocity Analytics. The agreement follows a comprehensive review of real-time and in-memory technologies by Thomson Reuters.

Velocity Analytics combines real-time and reference data with Thomson Reuters' deep tick history database to provide clients with market data for use by quantitative analysts, traders and compliance departments across the financial services industry. It provides both pre-built and custom analytics on enriched market data from more than 70 exchange, broker and vendor data sources. This enables clients to create their own data repository for analytics, compliance and other use cases.

A growing pipeline of buy-side and sell-side firms are currently evaluating the service, with several strategic customers engaged in an early adoption programme. The service will be delivered from the world's largest private financial markets Cloud network, Thomson Reuters Elektron.

Mike Powell, managing director for enterprise capabilities, Financial & Risk, at Thomson Reuters, commented: "Given Thomson Reuters' global customer footprint and the demand across the industry for streaming analytics and market insight driven by increasingly quantitative trading strategies and regulatory reporting requirements, demand should be strong for this solution. We are pleased to be working with FD to power Velocity Analytics using Kx technology."

Brian Conlon, Chief Executive Officer of FD, commented: "FD is pleased to partner with Thomson Reuters, using the ability of our Kx technology to enable real-time analytics on very large datasets to power the Velocity Analytics service. This agreementsignificantly increases our channel to market and the service will be rapidly deployed to customers through Thomson Reuters Elektron, which will accelerate our growth in this market."

This article is modified from a press release issued by FD on 27 April 2016

Out-of-the-box Analytics
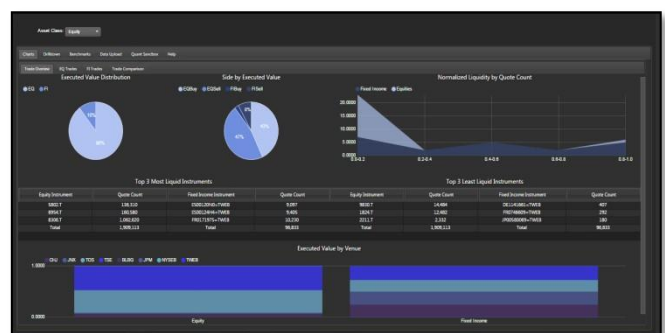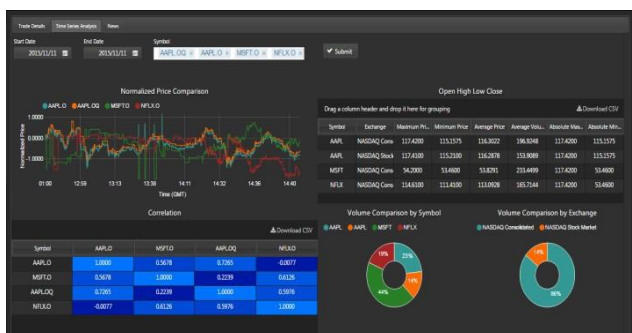
Trade Statistics (VWAP, TWAP, Min, Max, Close ..)

Quote Statistics (Spread, Mean, Count,..

Sample Analytics and Reports

- TCA Analysis
- Best Execution Analysis (RTS 27 & 28)
- Equity Pattern Analysis
- EBBO/NBBO
- SI Determination

Seamless real time and historical data access

- Real Time Market Data Feed
- Historical Tick Data
- Security/Reference Data
- Corporate Actions

Velocity Analytics 8 Functionality

- Real-time data capture for all financial markets asset classes.
- Integration with Thomson Reuters key data platforms – Elektron, Datascope Select and TRTH.
- Managed Data Quality – TR Staff will be using our administration and monitoring capabilities to provide feed handler configuration updates as a service to ensure data quality is maintained over time.
- Named Filters – An abstraction layer to enable dynamic filtering for the myriad of trade condition codes published from Exchanges.
- Real-time and Historical market data analytics.
- FX Normalization
- Timezone normalization
- Symbology mapping between different market and vendor instrument identifiers
- Multi-Source data support
- Holiday calendar integration
- Corporate action adjusted time series
- Streaming pub/sub calculations and analytics.
- Custom analytical filtering for outlier removal
- Visualization options via HTML5 dashboards and Java based IDE's.
- JSON based Webservice access or direct API based IPC communication.
- Data and functionality permissioning.
- The service is available as a hosted solution or deployed onsite.
-

The following calculations have been implemented within Velocity Analytics 8 for Thomson Reuters. All calculations listed are available across asset classes.

Trade Analytics	
lastPrice	firstPrice
firstVolume	lastVolume
minPrice	`minVolume
maxPrice	maxVolume
MavgPrice	avgVolume
sumPrice	sumVolume
VWAP, TWAP	turnover
tradeCount	firstInsertTime
lastInsertTime	firstExchangeTime
lastExchangeTime	medPrice
medVolume	openPrice
closePrice	Return

Quote Analytics		
firstBidPrice	firstBidSize	firstAskPrice
firstAskSize	`astBidPrice	lastBidSize
lastAskPrice	lastAskSize	minBidPrice
minBidSize	minAskPrice	minAskSize
maxBidPrice	maxBidSize	maxAskPrice
maxAskSize	avgBidPrice	avgBidSize
avgAskPrice`	avgAskSize	sumBidPrice
sumBidSize	sumAskPrice	sumAskSize
quoteCount`	avgSpread	firstInsertTime
lastInsertTime	firstExchangeTime	lastExchangeTime
lastSpread	firstSpread	maxSpread
minSpread	sumSpread	avgMidPrice
sumMidPrice	lastMidPrice	firstMidPrice
maxMidPrice`	minMidPrice	medBidPrice
medBidSize	medAskPrice	medAskSize
medSpread	medMidPrice	

4   MANAGED SERVICES

In addition to supplying the software required to run Kx Data Refinery, we also provide supporting services for its operation and support. These skills have been developed in implementing and supporting enterprise market data and reference data platforms for some of the world's largest financial services organizations. These services are particularly relevant for your clients who may have non-standard delivery formatting and data enrichment requests

Managed Service

Benefits

Industry best-practices, procedures and service levels

Access to experienced data management and data analysis professionals

On-demand capacity to Ramp-up / Ramp down resources as required

End-to-end SDLC provisioning with SLAs to mitigate risk project delivery risks

Implementation

Our team can work with your IT professionals to build a cost efficient infrastructure to deliver data to customers. These services include:
- The provision of consulting services around the choice and location of data centers, hardware and network
- Capacity planning
- Messaging protocols
- Building Feedhandlers for data integration
- Virtualization techniques
- Storage and compression optimization
- Familiarity with the latest developments in unstructured data
- Stress testing
- Resource monitoring
- Business continuity planning - failover, resilience and disaster recovery

Data Scientists

We have teams of data scientists available to work with your customers to extract value from the infrastructure. Kx Systems can also white label these services on a revenue share basis. As well as expertise in the latest Big Data technologies our team has deep capital markets domain expertise. Our services include:
- Development of new canned analytics
- Optimization of analytics
- Creation of customized analytics for customers
- Building customized dashboards to visualize data
- Creating customized alerts and workflows
- Integrating customer private data with data from Kx Data Refinery
- Building of custom APIs
- Integration with other technologies such as R, Python. Java, C++ and Matlab

Production Support

We can provide a global support team to work with your customers to provide a white labeled quality service to your customers.
- 24/7 production support
- Local support in major global financial centers - London, New York.
- Industrialized processes for logging and resolution of issues
- Scrubbing and cleansing of data
- Maintenance of Feedhandlers

Kx also provide data management and software development services covering requirements gathering, design, development, testing, deployment through to post-delivery support and maintenance. With a long history of delivering consulting services to the capital markets Kx has assembled experts in each of these areas and taps their collective knowledge and experience to deliver superior quality software and data management support. In delivering its services it incorporates formal project management methodologies and best practices in terms of account management and governance.

kx

EMEA

Head Office
3 Canal Quay,
Newry,
BT35 6BP
N. Ireland
Tel: +44 (0)28 3025 2242

Belfast
11-13 Gloucester Street,
BT1 4LS
N. Ireland
Tel:+44 (0)28 9023 3518

Dublin
Fleming Court,
D04 N4X9
Rep. of Ireland
Tel: +353 (0)1 630 7700

London
Cannon Green Building,
1 Suffolk Lane,
EC4R 0AY
United Kingdom
Tel:+44 (0)207 3371210

Americas

New York
45 Broadway,
New York,
NY 10006
USA
Tel:+1 (212) 447 6700

Toronto
31 Lakeshore Road East,
Mississauga, On,
L5G 4V5
Canada
Tel: +1 (289) 329 0636

Ottawa
300 Terry Fox Drive,
Kanata, On,
K2K 0E3
Canada
Tel: + 1 (613) 216 9095

Palo Alto
#375
555 Bryant Street,
CA 94301
USA
Tel: +1 (650) 798 5155

APAC

Sydney
22 Pitt Street,
Sydney,
NSW 2000
Australia
Tel: +61 (0) 2 9236 5700

Singapore
One Raffles Quay,
North Tower,
#30-03
048941
Singapore
Tel: +65 6592 1960

Hong Kong
Two Exchange Square,
8 Connaught Place,
Central
Tel: +852 2168 0715

Tokyo
Sanno Park Tower,
2-11-1 Nagata-cho,
Chiyoda-ku,
100-6162
Japan
Tel:+81 (0)36 205 3494

www.kx.com