Delivering Scalable and Robust Data Infrastructures with DaaS in Financial Markets

Implementing Gartner’s responsive data management model for web-scale IT with high-performance tools

February 2017

In this issue

Welcome 2
Building Responsive Data Management for Web-Scale IT 4
Kx For DaaS 11
About Kx 15
Welcome
by Dan Seal, SVP, Kx For DaaS

The business challenges of data management, data validation and governance combined with the technical complexity of building and managing high-volume data infrastructures make delivering robust, scalable and performant Big Data solutions a truly daunting task.

It is no longer acceptable for technical teams to have a myopic view of ‘their bit of the problem’. Instead they must now fully understand the entire scope of systems and think through not only how to move data around the planet with the lowest latency for immediate consumption, but also consider how that data and supporting infrastructure must be managed for longer term storage and access. Real-time and historical data must both be available throughout the enterprise at all times if their real value is to be realized.

Financial markets are arguably the forerunner of the Big Data revolution. Since the mid-90’s data-driven decision making and the direct integration of historical and real-time analytics into the trading cycle have resulted in an industry where automated algorithmic trading is the norm and where data infrastructures have been scaled from meeting this initial use case to being the data backbone of the entire organization.

Today, for example, the Kx platform provides data for everything from pre-trade decision support tools for clients, to algorithmic trading, risk and surveillance monitoring, through to post-trade reporting and compliance, all from a single technology stack. In its delivery, Kx has employed the disciplines and techniques of recently coined DevOps practices, HTAP, web-scale IT and the eponymous ‘Big Data’ to ensure its scalability and continual availability throughout the enterprise.
Counting all twenty of the world’s largest trading firms among our clients, Kx and our kdb+ technology set the benchmark for delivering and managing scalable, sustainable Big Data infrastructures that deliver high-performance, low-latency solutions.

However, the world is changing. The hype cycle of Big Data is closing (a good thing) and industries and business verticals outside of finance are finally beginning to understand what Big Data really is, what the potential opportunities are for their business and also, the huge technical challenges they face in realizing their goals.

Kx itself is also changing. Big Data infrastructures for financial markets are mature and becoming commoditized. As a result, Kx no longer just offers kdb+ as a build technology, but now also offers it as a complete platform solution via Kx for DaaS.

In the capital markets environment DaaS provides both the market data and the order and execution data capture backbone, whose data and associated analytics can be served out directly to specific business solutions (e.g. Surveillance, Best Execution, TCA) without those solutions needing to recapture and persist their own copies of the data. Neither do they have to deal with the complexities of data cleansing and normalization, access methodologies, historical and real-time access or even streaming analytics. Kx for DaaS does the heavy lifting and centralizes these functions so that downstream solutions can focus on the real problem they’re trying to solve and not worry about repeating the complex and expensive task of data management.

The Kx for DaaS software platform 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. Kx is no longer just about financial markets.

The Gartner paper which follows outlines the considerations to keep in mind when creating distributed, responsive data management systems for today’s web-scale IT demands.

<table>
<thead>
<tr>
<th>Governance &amp; Control</th>
<th>Performance &amp; Scalability</th>
<th>Flexibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>A single, common API and access layer</td>
<td>True horizontal scalability – simply add hardware and distribute datasets</td>
<td>Handles both structured and unstructured data</td>
</tr>
<tr>
<td>Globally or locally distributed to ensure accessibility from any location</td>
<td>Co-location of data and compute resources to increase performance</td>
<td>HTAP design for simultaneously processing both real-time and historical data.</td>
</tr>
<tr>
<td>Sandbox environment for ad-hoc research and analysis environments</td>
<td>High redundancy and availability to ensure maximum uptime</td>
<td>Single access point to all enterprise data</td>
</tr>
</tbody>
</table>

Source: Kx
Web-scale IT requires new models of development and deployment, which must be supported by data management practices that adapt to uncertain and changing conditions. Data management leaders must create and support continually evolving information infrastructure to meet the needs of web-scale IT.

**Impacts**

- The dynamic, distributed environment of web-scale applications changes the required balance of availability, manageability and performance for information infrastructure.

- The iterative, faster pace of web-scale IT requires collaborative governance between data managers, application developers, and infrastructure and operations staff.

**Recommendations**

Data management leaders must:

- Focus on scaling operational capabilities and optimizing their information infrastructure decisions around manageability and operations, by adopting practices such as DevOps.

- Engage with vendors as engineering partners, not only as third-party support, and develop deep expertise in the products and services they deploy.

- Avoid building new infrastructure software. Software, like hardware, is trending toward commodity; software infrastructure should only be built if it provides differentiated value up the stack and you can stay ahead of the commoditization wave.
Unify information governance and management as integrated, easily consumable services built on the same stateless application infrastructure that is used by web-scale applications.

Analysis

Users’ expectations for today’s digital business applications are radically different from those of previous eras. Downtime is unacceptable; even degraded performance is preferable to downtime. Today’s applications must always be available for a global, mobile-first audience. Application availability is further complicated by unpredictable user growth and usage. Mobile-first or Internet of Things (IoT) applications may have exponential growth in data traffic, storage and processing demands, and these demands may also be uneven or sporadic.

Enterprises pursuing digital business strategies are discovering that they face the same performance, availability and scalability challenges their digital platform contemporaries have dealt with for years. However, these challenges aren’t insurmountable just because you have a late start. Enterprises can achieve the same results as cloud companies by adopting and adapting the same practices and technologies. Gartner refers to this discipline as “web-scale IT”.

Gartner defines web-scale IT as a system-oriented architectural pattern that enables the rapid and scalable development and delivery of web-based IT services by leveraging agile, lean and continuous principles. Web-scale IT is a pattern of global-class computing that delivers the capabilities of large cloud service providers within an enterprise IT setting by rethinking positions across several dimensions.

Figure 1 illustrates the key elements of web-scale IT.

Figure 1. Key Elements of the Web-Scale IT Value Chain

Source: Gartner (September 2016)

To date, application development and architecture has attracted much of the attention for web-scale IT. Application developers cope with these demands by architecting applications to be as stateless as possible. Removing or reducing state at the application layer allows applications to scale horizontally, delivering linear or near-linear scalability increases. This architecture also requires the data management layer to have similar scaling characteristics. This research explores how data management can be
responsive to the dynamic demands of web-scale IT and Figure 2 summarizes the impacts and our top recommendations for data management leaders.

**Impacts and Recommendations**

**The dynamic, distributed environment of web-scale applications changes, for data management leaders, the required balance of availability, manageability and performance for information infrastructure**

The performance and availability requirements of web-scale applications mean that the applications and their data must be managed in a distributed, parallel fashion. Applications and data can no longer reside in a single data center. The distributed nature of web-scale applications introduces two challenges for data managers:

- First, this distributed characteristic introduces the need for a compromise in the transaction models used by the underlying information infrastructure. This compromise is between ensuring all instances of replicated data are the same everywhere (strong consistency) and ensuring that users can always access data, even if it's out of date. The problem with strong consistency in a distributed environment is that it limits application throughput: when a change is made, all copies of the data must be in agreement with the change. With network and hardware latency, reaching this agreement can push applications out of acceptable SLAs. Today, many applications favor availability over consistency — in order to deliver a better user experience — while setting appropriate user expectations around the timeliness or accuracy of data.

---

**Figure 2. Impacts and Top Recommendations for Data Management Leaders**

<table>
<thead>
<tr>
<th>Impacts</th>
<th>Top Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>The dynamic, distributed environment of web-scale applications changes the required balance of availability, manageability and performance for information infrastructure.</td>
<td>Focus on scaling operational capabilities and optimize your information infrastructure decisions around manageability and operations.</td>
</tr>
<tr>
<td>Web-scale IT requires collaborative governance between data managers, application developers, and infrastructure and operations staff.</td>
<td>Unify information governance and management as integrated, easily consumable services built on the same stateless application infrastructure that is used by web-scale applications.</td>
</tr>
</tbody>
</table>

*Source: Gartner (September 2016)*
Second, managing a distributed information infrastructure is significantly more complicated than managing centralized systems. Basic operations such as software updates and schema changes are vastly more complex as systems become distributed. The number of nodes in a cluster may expand or contract as the application load changes or data volumes shift.

When faced with these challenges, early web-scale pioneers reacted quickly. Some, like Facebook and Twitter, developed new information infrastructure technologies more suited to the web-scale environment. Others adapted existing technologies by introducing new storage engines for existing DBMS products. The overall theme is that these companies sought to simplify their data management environment by targeting fewer problems to solve and by removing extraneous features. Based on the history of the technologies coming out of the web-scale natives, two trends are apparent:

- Newly created technologies have a very narrow scope for the enterprise. They are specialized for certain types of data, such as log data, time series or messaging; and certain types of availability, such as read-centric or write-centric. Despite their narrow scope, the distributed element of these systems still makes them extremely complicated to implement.

- Operational capability is paramount. The information infrastructure supporting web-scale IT must be always available. When applications suddenly stop responding at the worst possible time (which is always 3:00 a.m.), the ops team must be able to find, diagnose and resolve the problem as quickly as possible.

This engineering-heavy approach works for some, but most enterprises can’t be expected to employ hundreds (or even thousands) of developers to create new technology for mission-critical applications. Relying on purely open-source solutions is an option, but these projects often lack the security and manageability expected from enterprise end users. Open-source projects frequently come with higher levels of complexity than their commercial counterparts; however, commercial software doesn’t resolve the support issue for data managers. Remember, downtime is unacceptable; data managers must be capable of troubleshooting and resolving problems with little to no assistance. In short, data managers must become end-to-end experts in the information infrastructure they deploy.

Lastly, being responsive means knowing what to respond to. As a data manager, you must monitor data traffic constantly. You must always know how and when data is queried, moved and transformed. This allows you to identify patterns and to adjust infrastructure accordingly. Just as downtime is unacceptable, not knowing the state and disposition of your information infrastructure is equally unacceptable.

**Recommendations:**

- Avoid defaulting to data availability over consistency. Understand how applications use data, and their targeted SLAs, before selecting your information infrastructure.
Focus on scaling operational capabilities and optimize your information infrastructure decisions around manageability and operations by adopting practices such as DevOps.

Engage with vendors as engineering partners, not only as third-party support, and develop deep expertise in the products and services you deploy.

Avoid building new infrastructure software, because software, like hardware, is trending toward commodity. Only build if you can build differentiated value up the stack and stay ahead of the commoditization wave.

DevOps is a common organizational theme for companies native to web-scale IT. The teams creating production software also support it. Aligning infrastructure production responsibility with the development task creates incentives to automate as much as possible.

While aspects of DevOps are appealing to enterprises, especially the automation element, many organizations are unsure how to reconcile established information governance requirements with agile development approaches. Information governance, when done at all, is typically defined as a top-down, one-size-fits-all approach. This puts data managers on the defensive in agile environments that frequently require equally dynamic information governance policies and management capabilities.

You must engage with your application development and infrastructure and operations (I&O) peers on the best way to govern data in your web-scale environment. Start the conversation using Gartner’s Three Rings of Information Governance, presented in Figure 3.

Each information governance ring requires a different approach to how tightly linked governance and management must be. Governance is what you have to do, the policies that must be adhered to, while management is how you’re going to enforce those policies. The most critical data (the first ring of information governance) must have a linked governance and management layer. Data as a service (DaaS) approaches can be an ideal fit, with API-based...
intermediaries providing access to regulated data sources through a centrally managed governance tier. An API-based information governance management implementation offers the following benefits:

- **Flexibility.** Information governance shifts from being static to dynamic. This flexibility more accurately reflects the rapidly changing nature of data consumers.

- **Scalability.** APIs defining information governance policies and enforcement must be built in the same stateless, horizontally scalable style as the applications that use them. This allows the policy element of data management to scale as readily as the infrastructure.

- **Shared culture.** The digital world has moved to code. Engaging with application developers in a way that they understand and encourage brings you into their culture. This facilitates communication and creates a shared mechanical sympathy for the systems and applications created through collaboration.

Applications using data in the second ring of information governance take a “trust but verify” approach. Second-ring data is always treated as external data and must be verified relative to data quality or some master data rationalization. This can be implemented in an API-based data quality bus, linked with any necessary master data management hubs.

**Figure 3. Gartner’s Three Rings of Information Governance**

1. Most critical content — ”master” data, content or analytics; commonly referenced; centrally governed
2. Second most critical content; shared across applications or suite; regionally governed
3. Least critical content; specific use in single application or process; locally governed

Source: Gartner (September 2016)
Finally, in the third ring of information governance, applications producing and consuming data locally have maximum governance flexibility. Any required governance can be built into the application and the supporting data management infrastructure. There is no need to overreach on information governance at this level; any effort here has no impact outside the application, so only adopt extensive governance when it makes sense to do so.

Recommendations:

- Get involved early in the software development life cycle (regardless of development methodology) in order to understand how web-scale applications create and use data, and their corresponding impact on information infrastructure and governance.

- Prepare to adopt new technologies (such as emerging DBMSs) and practices (such as DevOps) for success in web-scale IT. Your existing data management infrastructure may not meet the expected requirements for performance, agility or scale.

- Unify information governance and management as integrated, easily consumable services built on the same stateless application infrastructure that is used by web-scale applications.

Evidence


The fact base for this research is primarily derived from interviews with web-scale companies. Additional sources for this research include:


- M Callaghan, “Incremental vs Rewrite From Scratch — Biased Guide to a Web-Scale DBMS.” Facebook, April 2015.


Source: Gartner Research Note G00300944, Nick Heudecker, 16 September 2016
Kx For DaaS

Born from financial markets where exchanges, regulators, data vendors and financial institutions seek to gather, store, process, distribute and ultimately monetize their data, Kx for DaaS is a data capture and analytics platform for the modern era. It provides a complete suite of tools for managing data from ingestion through to consumption by multiple parties in a consistent, controlled manner.

This is represented in Figure 1 on page 12, where data from multiple sources is ingested into the DaaS layer where it is validated, normalized and stored. This corresponds to the first layer of Data Governance recommended in Gartner’s Three Rings of Information Governance. This layer also includes 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 London Stock Exchange may come with a condition ‘A’ but on NYSE with an ‘R’).

Kx for DaaS provides facilities to not only capture and store the required trade conditions but also provides an abstraction layer so data filtering can be managed server-side through a simple configuration, without downstream applications having to resort to code to implement the cleansing routines they need. Data filtering can then be managed centrally for the organization by a data management team, avoiding the issues of fragmented data processing logic being hard-coded into downstream applications.
In the business logic layer, applications can consume enterprise-certified data rather than having to individually source and process the information themselves, saving cost, accelerating development and ensuring enterprise-wide consistency. Moreover, additional governance can be applied on derived data by centralizing its calculation and distributing to users as certified. Customized results can then be overseen by data stewards rather than being independently evaluated within individual applications and without enterprise oversight.

A rich visualization layer then offers configurable HTML5 dashboards that provide powerful BI and reporting capabilities. Researchers and quants can access the same data using the same underlying APIs but can be sequestered into sandbox environments which emulate the full data environment but immunize production from non-standard usage or requests that may compromise performance. This third level of governance therefore enables flexibility, yet still imposes control.

**DaaS Design Principles**

The “Four V’s” of Big Data (volume, velocity, variety and veracity) continue to push 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 for DaaS 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 deploy on one machine in one location or on hundreds/thousands deployed globally meaning we have limitless scalability.
- **Velocity** – Kx for DaaS 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** – DaaS feed handlers provide connectivity to multiple sources and formats with the flexibility to extend to others with accompanying data management functionality. Our unified data access API means retrieving that data is simple and accessing data from another source is as simple as changing a single argument.

- **Veracity** – Kx provides managed services for the validation, cleansing and workflow management of data exceptions, a service for data queries and analysis along with an analytics and data cleansing framework that is easy to configure and adaptable to data from any source.

The DaaS architecture to achieve these design goals is illustrated below in Figure 2 along with the accompanying analysis, deployment and runtime infrastructure to support it.

---

**Figure 2: Kx for DaaS Technical Architecture**

Source: Kx
DaaS: Enabling Data Connectivity

Today’s world is about connectivity and collaboration; from social media and supply chain integration to cooperation to co-opetition, systems are becoming increasingly interconnected. It is ironic, therefore, that while many of the supporting in-house applications may share network connectivity, and possibly business logic in the form of shared libraries, they retain individual, isolated islands of data within the enterprise. Not only does it impose cost overhead for sourcing and maintaining multiple copies of the data, it also compounds the risk of inconsistency and dramatically slows down application development time.

If systems developers could concentrate on developing business logic secure in the knowledge that the supporting data was continually available, easily consumable and certifiably correct, the speed of development would accelerate to match the speed of requirements. Few businesses can respond at such speed but Kx for DaaS enables that transformation:

- Our HTAP architecture means you can cover all data at all velocities.

- Our horizontal scalability means platforms can be implemented on commodity hardware to whatever capacity is required for your organization, small or large, and our methodologies follow best practice, described by Gartner as web-scale IT.

- Our data-centric platform and associated management tools support strong data governance that enables a common, continuously available access point to enterprise data and data mastering for the entire organization.

- Finally, our modular design means true support for DevOps approaches whilst maintaining the integrity and performance of your core data capture and analytics backbone. Developers can use the same foundation libraries that Kx for DaaS is built on to rapidly develop their own applications and integrate these quickly with the data backbone. Researchers can make use of sandbox environments to use the full power of the analytics infrastructure whilst giving them the flexibility they need to find alpha in your data. Mission-critical applications can use a dedicated, highly redundant, performant and managed platform to provide the performance and availability you need to power your enterprise forward.

Now a part of the First Derivatives group, Kx can not only offer the software solution, but also provide expertise for implementing, supporting and maintaining your data infrastructure, allowing your own IT teams to focus on your core business, safe in the knowledge that the data and technology you rely on is in good hands.

Source: Kx
About Kx

Kx is a leader in high-performance, in-memory computing, streaming analytics and operational intelligence. The company’s focus is on delivering the best possible performance and flexibility for high-volume, data-intensive analytics and applications.

For more than 20 years the world’s largest banks, hedge funds, brokers and regulators have been relying on Kx technology. Kx software includes the kdb+ database and q programming language, as well as a rich suite of products and solutions built on top of kdb+ to implement business layer analytics and visualizations.

Kx Systems, Inc. is a subsidiary of First Derivatives plc, which is quoted on the London Stock Exchange (FDP:LN) and is a global technology provider with 20 years of experience. FD employs over 1,700 people worldwide and has operations in London, New York, Stockholm, Singapore, Hong Kong, Tokyo, Sydney, Palo Alto, Toronto, Philadelphia, Dublin, Belfast and its headquarters in Newry.

For more information about Kx please visit www.kx.com. For general inquiries, write to info@kx.com.