◆ **influxdata**®

How CJC Built a Big Data Visualization Platform for Their Capitol Market Customers

Steve Moreton

Senior Technical Director, CJC

CJC

APRIL 2018

Company in brief

Founded in 1999, CJC specializes in managed services, consultancy and products for global capital markets firms. CJC solutions leverage cloud technology to optimize real-time data distribution systems, as well as enhance compliance processes and overall commercial control.

CJC's mosaicOA is a big data visualization service that enables capital markets firms to visualize and analyze any significant data set, and thereby benefit from managing both the risk and opportunity that this data can provide.

With offices in London, New York, Hong Kong and Singapore — and with global teams experienced in a variety of mission-critical data and trading applications, infrastructures, processes and compliance requirements — CJC can quickly and cost-effectively help clients optimize their market data operations.

Case overview

CJC wanted to build a big data visualization platform for its capital market clients for ITOA (IT Operations Analytics). As cost optimization is a primary concern for IT leaders in the financial management space, access to ITOA helps provide visibility into system infrastructure performance, accelerate decision-making and improve planning. CJC created Mosaic Open Analytics (mosaicOA) for real-time capital market data visualization and analytics, using InfluxDB to store metrics from hundreds of specialist servers, networking and middleware systems, with some firms doing over 1 million writes per minute.

Delivering on-demand visibility and advanced analytics
while driving time and cost savings

> *"It's great to work with InfluxDB as a partner because we can say: "Look, this is what we need. Can you factor this in, potentially, into the next release?" And it's like: "Sure. That's great. Let's do it."*
>
> **Steve Moreton,** *senior technical director*

The business problem

CJC wanted to support real-time data infrastructures inside the capital markets across data centers globally. Capital market data comes from exchanges (such as the New York Stock Exchange and London Stock Exchange) which typically send their own data rounds straight to the clients, and also comes from well-known market data vendors such as Thomson Reuters and Bloomberg, who have millions of instruments all updating at very high rates per second. A huge number of industries and individuals around the world need that data at very low latency all at the same time. CJC needed to respond to its clients' demand for visualizations of their IT system health and their need to view all infrastructure, have incredible granularity, and visualize algorithms to their servers.

Keeping tabs on such a critical and complex IT infrastructure is a major challenge. Existing solutions can be expensive and often not fit for purpose. Furthermore, it's tough to get management buy-in for the investment required to ensure the smooth-running of trading data systems. In this context, the concept of ITOA has gained traction among financial institutions in the past few years as they strive to gain control over their critical processes. ITOA is an approach designed to retrieve, analyze and report data for IT operations — it applies big data analytics to large datasets where IT operations can extract unique business insights, thereby increasing revenue and achieving cost savings.

In fact, operational metrics are emerging as the Holy Grail of financial data professionals. For financial institutions active in multiple asset classes and across financial centres, execution venues and regulatory regimes, maintaining the data infrastructure to support the increasingly complex trading business is an ongoing headache. With so many moving parts, market data and transaction messaging systems are prone to system overloads, bottlenecks and other potential pitfalls that can disrupt data flows and — worse still — bring trading to a halt.

The technical problem

CJC has provided its clients with ITOA since 2012 as part of its standard managed services. Yet

feedback from its engineering team and clients revealed the need to go deeper and embrace the full 'truth' of client infrastructures. They needed to collect data on usage of standard components like CPU, RAM, OS, along with industry-specific application level data from systems such as TREP, Solace, Wombat and BPIPE. Client infrastructures can easily generate hundreds of thousands to millions of metrics — all updating at high frequency with some at sub-second intervals. This data has to be tracked, managed, stored and analyzed — not just for a day or a week, but for months and potentially years. This immediately led CJC to several challenges:

- Big Data analytics around IT system capacity can have vast capital expenditure — database storage represents one of the highest costs in the capital markets business.
- Traditional relational databases are not designed for big data analytics or capital markets data, which is time series data.
- Public cloud providers come with data gravity concerns, egress charging and unknown costs.
- Once information was stored, it presented the challenge of how to visualize it.

At the time, CJC had very basic performance, user, and network metrics. They could only keep the data for a short period of time and were continually purging their relational database. Their clients had to store a small percentage of their metrics and perform guesswork as to which metrics to keep and visualize. Stored data was visualized in older tools like Excel.

CJC's Very Basic Server Capacity Report Prior to InfluxDB
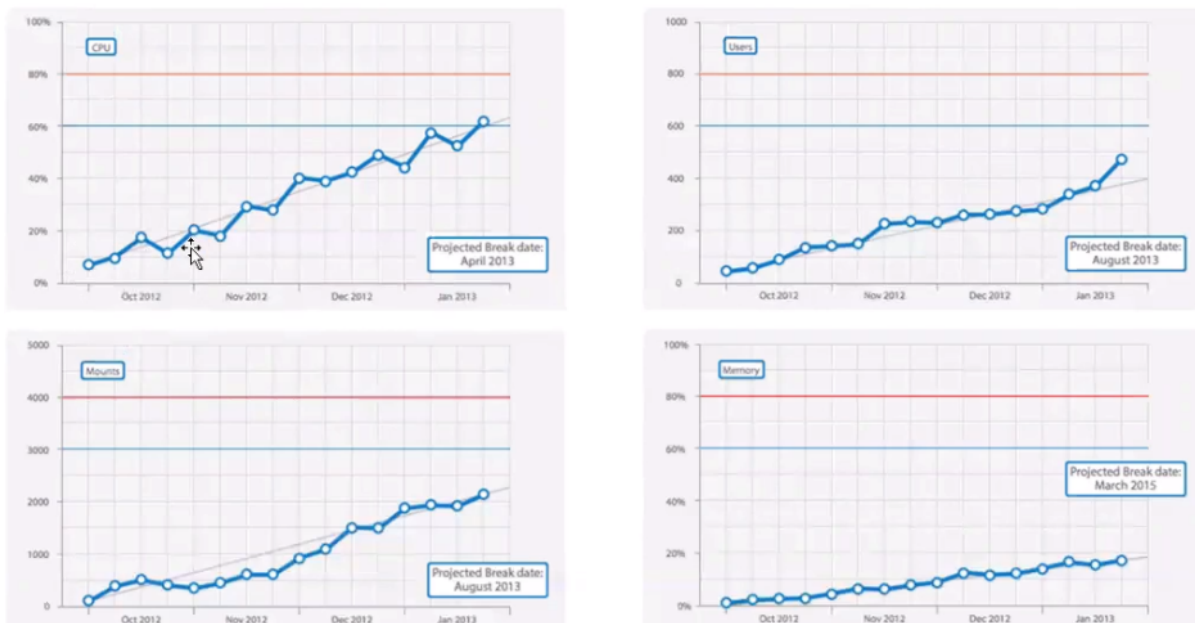
To support the visualization capabilities in demand by their clients, CJC quickly realized that old-fashioned databases weren't fit to purpose and that they needed a time series database.

The solution

> *"We have to keep cost down, so we were looking for something potentially open source. InfluxDB had a lot of things which could be of value here, especially knowing how big the data was, and being able to keep the shape of the data over a long period of time while being able to automatically drop the details through data retention policies also keeps costs down."*

Why InfluxDB?

InfluxDB topped CJC's database prospect list because they had to keep cost down and it is an open source database purpose-built for time series data. Using InfluxDB, they set out to build mosaicOA from scratch. The key InfluxDB feature that CJC liked is the data retention policy feature, which made it possible to jump between policies very quickly, and thereby identify data trends at different 'altitudes' of granularity (top-down and close-up views). As different volumes of data preserved can generate radically different trend line perspectives, InfluxDB retention policies are key to how mosaicOA works, with the below summarizing data handling for 180 days:

- For the first 60 days, CJC keeps every piece of information that comes in.
- After 60 days, CJC keeps data to minute intervals.
- For a further 30 days, CJC keeps the min, max, and mean for the minute.
- After 30 days, they only keep data at the hour interval.

The above data retention strategy preserves the shape of the data (key to creating thresholds) without causing data size to grow. Further, users can perform end data projections and assess projection accuracy through system growth visualizations. By seeing the trend line indicating potential appliance crashes, the client can take action to prevent them.

mosaicOA Enabling End Data Projections

Data views at different altitudes and degrees of granularity

When CJC designed mosaicOA, they wanted a 'Google Earth approach' to the data, enabling an accurate depiction of the data for the altitude users are looking at, and views at different degrees of granularity.

Zoom-In and Zoom-Out Data Visualization Capabilities

With InfluxDB, CJC is looking for predictive algorithms, and for that purpose, have put in place their Correlated Input Matrix Process (CIMP), which runs every five hours to split up highly granular data and identify data which goes over a certain threshold.

Event Analysis Supported by CJC's CIMP Process

- All the data is visualized in one line which makes it easier to spot outliers and enables data scientists to predict when events are going to occur.
- CJC's CIMP has the capability to visually eliminate Saturdays, Sundays, and all non-market hours, resulting in more accurate growth reporting. CJC is able to do that in a few seconds, through a Snapshot Request (where the server is asked to bring in all the market data information as a snapshot up to the current point).
- mosaicOA enables not only storage but visualization of Snapshot Request increase over specific time intervals and across the entire client state.
- The CIMP process, apart from being a time-saving tool, results in more accurate reporting.

> *"The beautiful thing about the data retention policies is that you get the data as granular as you like."*

mosaicOA platform launch

mosaicOA was launched with its initial use case focusing on ITOA, where the service processes real-time infrastructure data, such as CPU and memory data, and performance metrics from industry-specific components such as Solace, TREP, BPIPE, and Corvil, in order to deliver a clear view of

the health and performance of real-time market data systems. CJC's Proof of Concept was very successful. After an engineering cycle that took 18 months, CJC launched mosaicOA to capture and visualize metrics, and thereby gain the ability to analyze the performance of its market data infrastructure in real time.

**Pulling in Data from a Client's Multi-Region Infrastructure**

mosaicOA features designed for capital market data needs

- Private and secure cloud-based infrastructure offering elastic hardware scaling and high-performance storage – no data gravity issues
- Access via cross connect, Equinix cloud financial exchange / VPLS
- Connects to multiple database sources, Solace Statspump, ITRS (Open access or nanomsg via Kafka)
- Time series database designed for big data storage, real-time queries, analysis and projections
- Data scope vectoring provided by the CJC proprietary 'persistence engine'
- Browser GUI with advanced visualizations and features beyond open source alternatives
- Continual development of PaaS, IaaS and SaaS
- 24 x 7 full monitoring, management and support
- Client production hosted in London, US, Canada, APAC

Powerful Search Function within mosaicOA
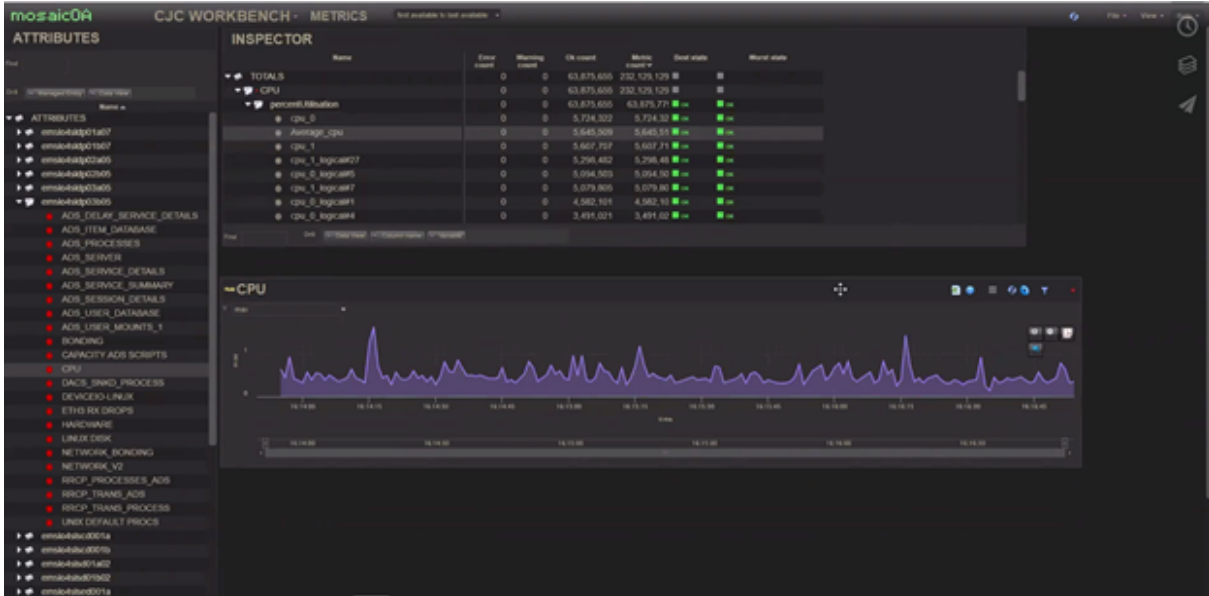
**Technical challenges of building mosaicOA**

In building mosaicOA, CJC faced and overcame multifaceted technical challenges including:

- Realizing how big the data volume was
- Tuning the messaging systems coming from the monitoring systems and the collectors to ensure uninterrupted data flows between all the different sources
- Handling cases where clients re-purpose their server (a problem they solved by providing asset management capabilities)
- Tuning their infrastructure, including replacing their spinning disks with SSD then storage array technology, to optimize database writes

Technical architecture

"All important is InfluxDB time series database, designed for big data storage for the queries in our system projections. The data scope vectoring we do...that's something that we could've only done pretty much with InfluxDB."

InfluxDB Deployment in mosaicOA
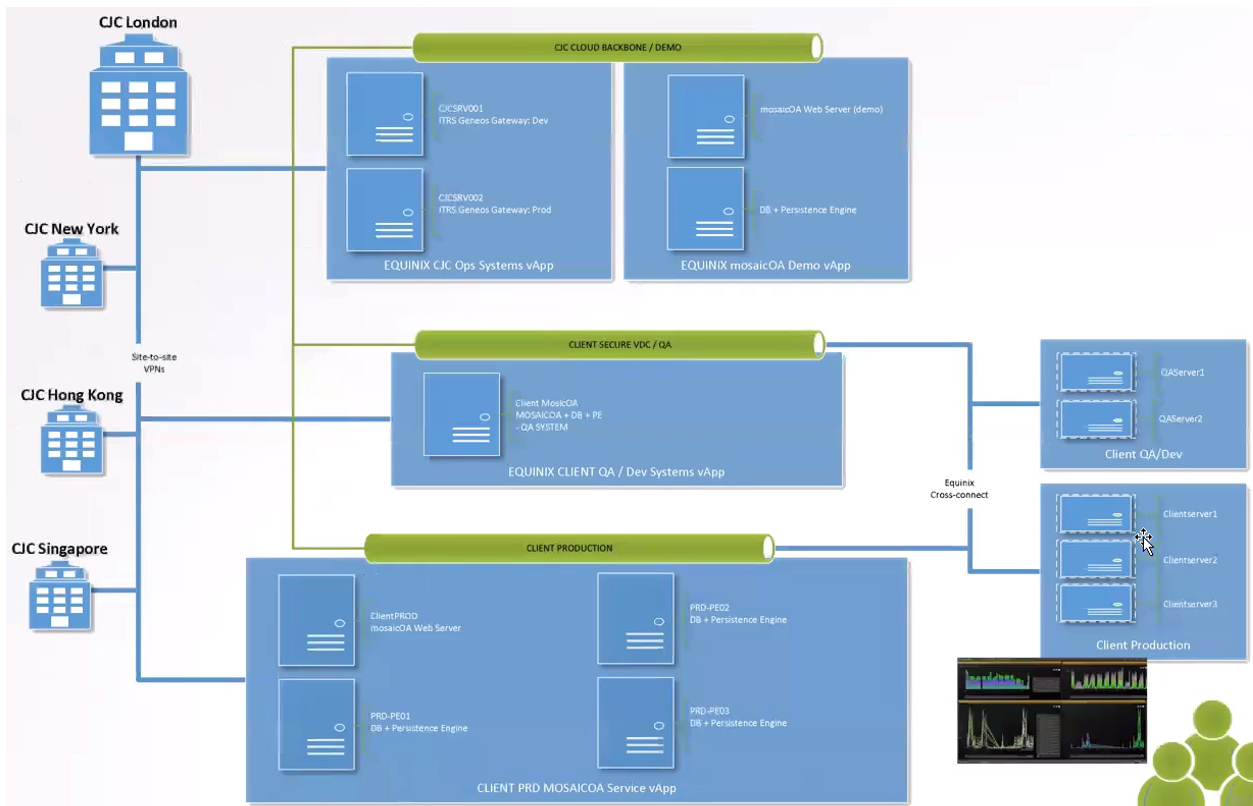
- As CJC clients are very sensitive about their data, CJC could not use a cloud platform provider like AWS or Google and had to connect with clients via cross-connect or VPLS.
- The data streams into mosaicOA via a messaging system from a third-party monitoring system that the client already has on their side.
- CJC use a middleware messaging platform called Solace to consolidate all the messaging into one messaging system.
- CJC has a number of collectors which stream the data directly into an InfluxDB database that resides in CJC's private cloud.
- The front-end queries the database. That query engine is CJC's proprietary persistence engine, enabling CJC to look at all the servers.
- The persistence engine blocks all the datasets that CJC identifies as unrequired datasets which regularly add unrequired costs. Generally, CJC have about three persistence engines running on the service.

What's next for CJC?

CJC is considering adopting the InfluxData platform more comprehensively in order to monitor and manage Kubernetes clusters.

Results

> *"When a client says, "Hey, we need this," it only takes us a short amount of time to build this. And a lot of it is just leveraging the power that InfluxDB gives us."*

mosaicOA, as part of the CJC Cloud Platform, provides a private secure cloud-based infrastructure offering elastic scaling, very high-performance storage, and real-time notifications of system health and activities:

- mosaicOA brings users a powerful front-end without requiring any infrastructure upgrades, while achieving cost savings on database sizes.
- CJC's data scope vectoring technique eliminates the need for expensive, high-cost, cloud technology and reduces the costs of long-term big data storage.
- mosaicOA delivers visibility into ITOA across multiple time intervals and levels of granularity and is thereby enabling early trend identification, forecast development, and faster, more informed decision-making.
- The mosaicOA browser-based GUI features powerful visualizers including drill-downs, pie/bar/column/3D charts, tree grids, tree maps, and HTML dynamic templates, all of which can be exported to PDF or csv.
- CJC can perform comparative analyses (such as comparing database writes to help scale the system) among clients being managed.
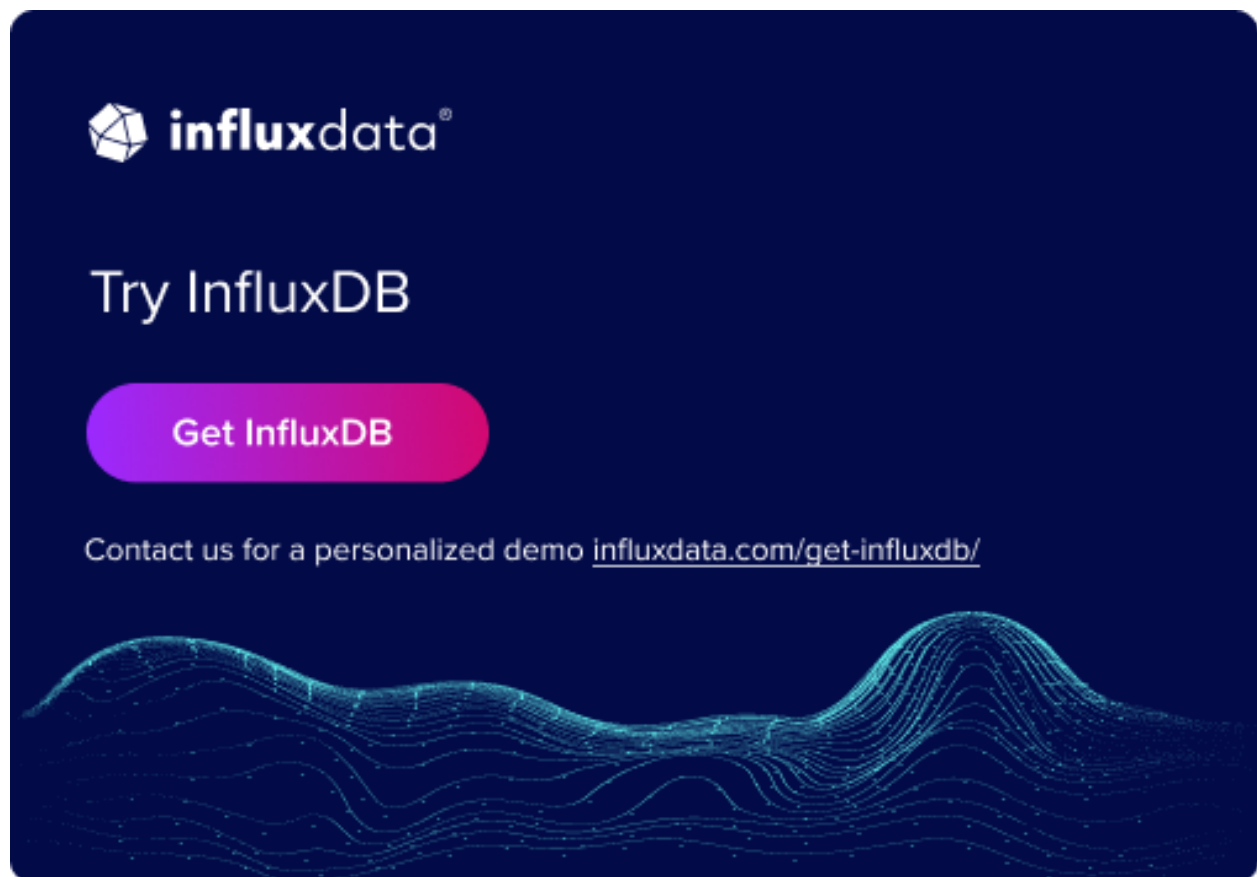
The benefits of a holistic view of market data and trade messaging system performance include proactive load management, advanced system planning, and avoidance of expensive downtime. Performance metrics can be collected to identify potential bottlenecks before they cause operational problems. Data outages can be averted, allowing market data teams to provide better services for internal teams and end users.

mosaicOA registered a major win for the CJC private cloud platform and enabled CJC to go head-to-head on price with major public cloud vendors, deliver a service from a secure private location and provide specialist add-on services outside of server orchestration. CJC's cloud platform eliminates the need for financial institutions to manage their own data centers, enabling them to turn on a tap and consume the data services they need while focusing on their core business. To demonstrate mosaicOA functionality and performance, CJC offer a demo environment from their private cjc.cloud platform. They release a new version of front-end every two weeks to respond to client demand for new ways of interrogating the data.

Using InfluxDB, CJC is fulfilling the big data visualization and infrastructure needs of its demanding capital markets IT clients globally.

About InfluxData

InfluxData is the creator of InfluxDB, the leading time series platform. We empower developers and organizations, such as Cisco, IBM, Lego, Siemens, and Tesla, to build transformative IoT, analytics and monitoring applications. Our technology is purpose-built to handle the massive volumes of time-stamped data produced by sensors, applications and computer infrastructure. Easy to start and scale, InfluxDB gives developers time to focus on the features and functionalities that give their apps a competitive edge. InfluxData is headquartered in San Francisco, with a workforce distributed throughout the U.S. and across Europe. For more information, visit influxdata.com and follow us @InfluxDB.