Unified Messaging for Single Dealer Platforms

The advent of Single Dealer Platforms (SDP's) is advancing the state of the art of investment banking by integrating the delivery of pricing, risk and liquidity to traders. By aggregating information across asset classes and exposing it via intuitive web-based interfaces, sell side firms that implement SDPs can give their clients previously unachievable visibility into complex opportunities for profitable trades. SDPs can deliver value in all areas of securities, with their adoption most prevalent in fixed income (FI) trading and foreign exchange (FX).

Behind the scenes, SDPs have sophisticated messaging requirements – they demand high rate, ultra low latency messaging, guaranteed/transactional messaging between internal applications and real-time streaming to rich internet applications (RIAs) and mobile apps.

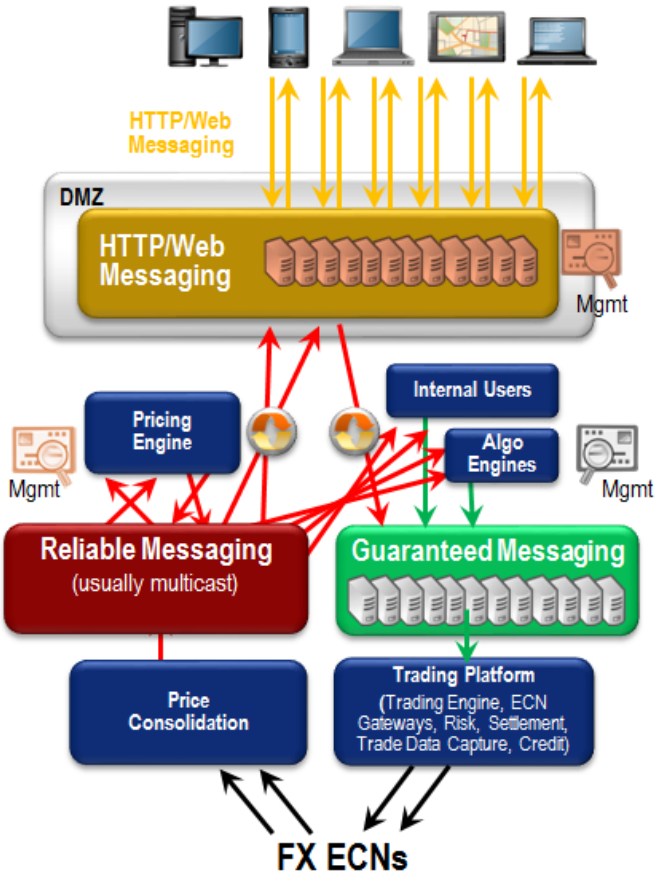
Solace message routers are the only solution that can meet all of these needs, and they offer exceptional performance, scalability, stability and manageability. This paper explains why Solace's unified platform solution is the ideal foundation for single dealer platforms.

 http://www.solace.com

SDPs using Legacy Technology

While SDPs can be comprised of many and varied application components, the required messaging infrastructure universally needs to support three kinds of messaging: real-time for quote distribution, persistent for order flow, and streaming over the web to rich internet applications. It has historically taken three separate messaging products and vendors to meet these needs, and the complication of bridging the environments and linking applications across platforms has made the design and deployment of SDPs a very complex and labor-intensive activity.

The diagram to the right shows the architecture and message flow of an SDP built with conventional messaging software.



1. Pricing streams are received from FX ECNs and passed through a consolidation and normalization layer.
2. This stream is fed to applications and users via a low latency pub/sub messaging system. This messaging has historically been multicast-based because that was the only way to achieve reasonable performance.
3. Orders from algos and internal users are sent to the trading platform for execution or routing back to the FX ECNs via gateways. Broker-based guaranteed messaging is used to guard against message loss, provide fault tolerance, and perform pub/sub distribution to other application components such as settlements, credit, and trade data capture.
4. Pricing engines receive the normalized prices and usually generate tiered currency spreads that are sent via the low latency messaging system to the web distribution layer in the DMZ. These servers provide connectivity to external applications which can be thick GUIs or Rich Internet Applications. The web distribution layer is typically a third technology with its own APIs, management and capabilities that must be integrated with both the low latency and the guaranteed messaging platforms.
5. External clients view streaming FX quote spreads along with other content and analytics and place orders which are carried through the web distribution layer and into the trading platform via one of the internal messaging systems.

The Problems with Piecemeal Architecture

As mentioned above, it usually takes at least three different messaging products to satisfy all the messaging requirements of an SDP:

a) **Real-time messaging for quote distribution**. These systems have historically been multicast-based to satisfy requirements for high rate, low latency and fanout.

b) **Persistent messaging for order flow**. These systems are typically broker-based to provide the persistence they require. JMS brokers are often used, but they suffer from poor performance, rigidity and inconsistent latency.

c) **Internet data streaming to RIAs**. These are typically gateways and not actual messaging systems. Some are simple bi-directional HTTP byte pipes, while others provide basic messaging capabilities. In either case, this product must be integrated with the two other messaging systems from the point of view of API/transport integration, mapping of topics, subscriptions, message contents and this work needs to be done by the SDP development team.

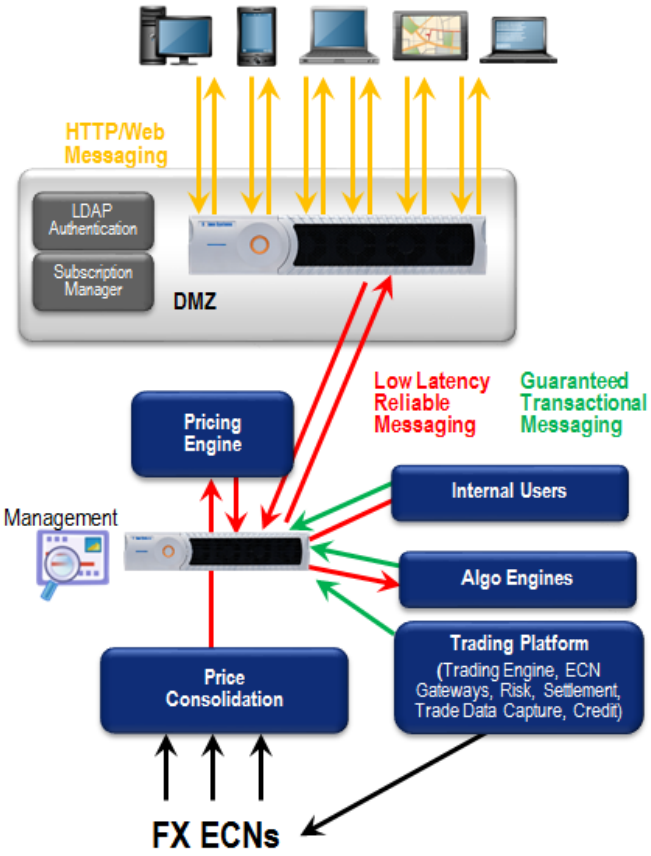
The integration of these systems introduces technical and operational challenges including:

1. Each messaging system has its own API and interaction idiosyncrasies that must be mastered by the various developers on the project – often requiring applications to use two or more messaging APIs in the same application. This adds cost and complexity to the initial development and ongoing support of applications.
2. The resulting system, once built, is complex and fragile due to the integration points and information mappings.
3. Performance typically suffers as a result of gluing together these various infrastructure components.
4. Scalability can be a greater challenge in some cases. For example, if STOMP or a similar protocol is used to communicate to the RIAs and each STOMP connection to a GUI is mapped to a JMS or other connection into the internal messaging system, then the internal messaging system must also scale along with the number of external internet clients.
5. The operations challenges are also significant: provisioning, monitoring, troubleshooting and capacity planning systems must be developed and the operations team must be proficient at managing at least three different messaging systems – some of which may be multicast, some are broker based. Then there is the qualification and rollout of periodic upgrades to each of these systems.
6. Dealing with internal users and applications is now completely different than dealing with external users and RIAs unless internal users also connect via the web distribution layer, which then places additional scale and performance demands on the web distribution layer.
7. Every aspect of the infrastructure, such as the difference in high availability architectures, security and monitoring, multiplicatively increases system complexity since it is typically completely different for each of the three messaging systems.

SDPs using Solace

By providing an integrated messaging platform that handles all of the above-mentioned messaging requirements, Solace eliminates the need to integrate and manage multiple technologies. This reduces the development and operational cost, system complexity and risk associated with deploying and running the platform while significantly increasing its flexibility and robustness.

The figure to the right shows a Solace architecture for the same FX SDP shown above. Note how much less complex the system is in terms of architecture, infrastructure components, applications and administration.



Pricing data flows from ECNs to the internal and external traders via low latency reliable messaging represented by red arrows, and orders flow to the trading platform and back-office through the same Solace message routers using guaranteed transactional messaging represented by green arrows. Additional Solace message routers in the DMZ handle the delivery of pricing data and placement of orders.

Flexible support for client subscription authorization and per-user access to tiered pricing as well as user authentication are just some of the ways the Solace message router can be controlled by business logic specific to each deployment.

Solace's solution simplifies and reduces risk in three areas:

- **Architecture:** Handling all messaging requirements with a small number of high-capacity purpose-built devices eliminates integration points and reduces the number of moving pieces required for messaging functionality and high availability.
- **Applications:** The existence of a single, rich API in each language supporting all messaging functions and message exchange patterns greatly simplifies the development of applications, letting developers focus on core functionality without worrying about the receipt or delivery of information. And since hardware-based message filtering ensures that applications are only sent data they have subscribed to, they don't need to deal with identification, filtration and deletion of unwanted information.
- **Administration:** Solace gives administrators a single view of the operations and management of the entire platform, including granular per client metrics and alerts that software-based messaging can't offer.

Solace's Key Capabilities and Advantages

When used for intranet messaging, the Solace message router provides unmatched value as a messaging infrastructure due to its performance, scale, robustness and management capabilities. In use cases such as SDPs where messaging is also required over the internet to RIAs, Solace meets those additional requirements without introducing the need to stitch together disparate technologies.

Enterprise Messaging

Unified Platform

Solace's high capacity devices and unified API mean that a compact, architecturally simple Solace deployment can meet all of the information distribution requirements of a typical SDP: the real-time streaming of market data, guaranteed flow of transactions, WAN synchronization of distributed datacenters and web messaging capabilities, described in the next section. The scalability, performance and features of Solace's solution result in a simple, robust platform that enables application architects and developers to focus on their business rather than the plumbing.

Solace's high capacity devices and unified API mean that a compact, architecturally simple Solace deployment can meet all of the information distribution requirements of a typical SDP.

Performance

Solace's message processing is embedded in silicon so there's no operating system in the datapath, no OS interrupts, context switching or data copies between kernel and user space. Solace delivers messages using TCP connections instead of multicast, which eliminates the operational complexities of multicast and means every client receives exactly the messages they need so they don't waste CPU identifying and discarding unwanted messages.

- **Shared Memory IPC:** Solace supports Inter-Process Communication using shared memory for processes running on the same server. Solace IPC provides latency as low as 480 nanoseconds at millions of messages per second.
- **Reliable Messaging:** Solace can route millions of messages per second with latency under 25 microseconds.
- **Guaranteed Messaging:** Solace can route hundreds of thousands of guaranteed messages per second with low, consistent latency.
- **WAN:** Solace accelerates the distribution of data over the WAN between datacenters using distributed topic routing protocols as well as techniques found in dedicated WAN optimization appliances.

Virtualization

Solace message routers can be "virtualized" into logical partitions, or logical brokers, giving applications their own messaging environment within a shared infrastructure. This technique can be used to perform two common functions in an SDP. The first function is to provide a private message bus among processes of a common application, such as the trading platform that is separate from the quote distribution or the "public" order flow message bus. The second function is to separate internal and external pricing data so the pricing engine can receive internal prices over one messaging VPN and publish tiered spreads via a separate "external pricing" VPN.

Monitoring and Management

Solace message routers enable management and monitoring on a per-client basis via the SolAdmin GUI or as a component of umbrella management systems from our partners

ITRS and Nastel. The ability to monitor performance metrics such as per-client queue depths, message rates, packet loss is built into the Solace message router and available either programmatically or via the SolAdmin GUI. This is available for all client applications whether connected over the intranet or the internet. Solace also partners with TS-Associates to provide non-intrusive performance monitoring of all real-time components of an SDP.

WAN Optimization

Solace provides efficient, dynamic message routing over the WAN using our distributed routing protocols to send messages only where they should go while providing location transparency for loose coupling. Solace also features various capabilities typically found in WAN optimization devices such as streaming gzip compression implemented in hardware and striping data across multiple TCP connections to gain significant efficiency over high round trip time (RTT) connections. With such an efficient, dynamic infrastructure, applications in one datacenter can communicate with applications in another datacenter in a completely transparent manner.

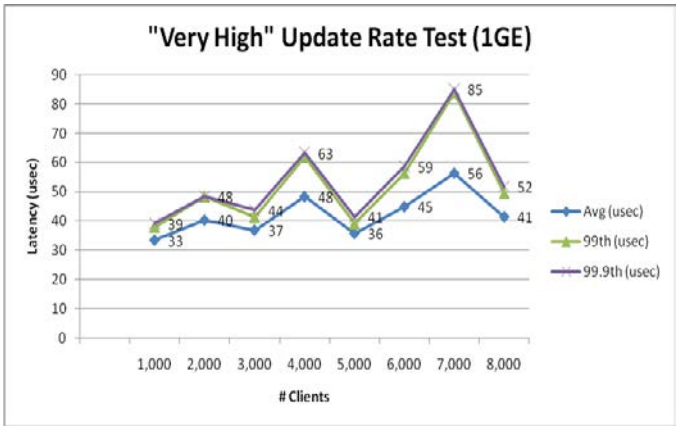
Web Messaging Capabilities and Advantages

At the web distribution layer, Solace offers many advantages over traditional products.

Performance

In a variety of tests covering different message rates, numbers of connected clients, and network environments, Solace's solution demonstrated vastly superior performance in terms of raw latency and consistent of latency. The chart below exemplifies the kind of advantage Solace offers. The results of exhaustive testing are fully documented in a 16-page whitepaper that is available via our web site at http://solacesystems.com/webmessaging.

This test shows that Solace's solution introduces nominal latency into the equation – double digit microseconds compared to the many milliseconds introduced by competitive solutions. That means Solace is 50-200x faster. The results also demonstrate the tight latency distribution of Solace's solution, with low standard deviation even when considering the 99.9th percentile.



The test was run with 20,000 topics, and each message consisting of a 50 Byte payload and 5 Byte topic. Each client was configured with 100 subscriptions, and set to receive messages at a rate of 100 messages per second. The network was 1 Gigabit Ethernet.

No Need to Integrate Web Distribution and Internal Messaging

With conventional products, the platform team must integrate the web distribution layer with the internal messaging systems. This entails API/transport level integration, topic translations, subscription management and data format translations—and system management and monitoring added on top of that.

With Solace, the web distribution layer connects seamlessly with the Solace internal messaging layer using TCP as a transport with dynamic subscription and routing protocols ensuring that the only messages flowing over the internal DMZ boundary are those for which there are subscribers. As RIA clients dynamically subscribe/unsubscribe to flows,

subscriptions are automatically distributed by the message router in the web messaging layer to the internal message routers to start/stop the flow of messages. There is no need to perform adaptation between API/transport layers, data formats, topics or subscriptions since this information is uniform across the web and internal messaging systems.

Solace's web distribution tier places very little incremental load on the internal messaging system compared to other solutions, as the routing connections made between the message routers are the only connections – not one connection per web client as in other systems. This means Solace's solution scales in the DMZ without transferring scalability challenges onto the internal messaging system.

Solace places much less load on internal messaging systems than other solutions.

Messaging API Capabilities for RIAs

Solace's RIA APIs provide full messaging capabilities to the application developer:

- Client authentication that the Solace message router validates against an LDAP, Radius or database, and which can easily be integrated with Single Sign on Servers.
- The ability to subscribe to hierarchical topics with wildcard support.
- The ability to send and receive individual, well-formed messages.
- Support for publish/subscribe and request/reply message exchange patterns with unique, system-assigned "reply-to" addresses to allow point-to-point communications.
- Machine- and runtime-independent data types to ensure interoperability across languages, OSs and RIA containers.
- Connection management to the Solace web fanout message router.

These capabilities allow RIA developers to understand the design of the entire system, and to interact freely with applications inside the corporate intranet so they can focus on business issues instead of plumbing and messaging challenges.

Solace automatically uses the most efficient transport in each situation so developers can focus on the semantics of their application instead of worrying about HTTP long polling and connection management.

Transparent and Efficient Web Distribution

In order to provide efficient, bidirectional, asynchronous communication capabilities to the application developer, Solace supports various efficient transports between the RIA and the web distribution layer. The capabilities available depend on the constraints of the particular RIA runtime or web container, and the policies of web intermediaries such as HTTP Proxies, firewalls and NATs. The most efficient transport allowed in the particular environment is automatically discovered and used by the RIA APIs and the web distribution layer. This means application developers can focus on the semantics of their messaging application without worrying about the details of things such as HTTP long polling and HTTP connection management.

Scalability

The Solace message router performs large scale, high rate message distribution over TCP by embedding this functionality into purpose built hardware. Because message routing and forwarding is performed in hardware, a single instance of a message can be fanned out to any number of clients over the web without the scale, jitter and performance limitations that equivalent software-based products introduce – and in a much smaller footprint.

Dynamic, Customized, Rule-based Subscription Management

While the Solace message router can impose per-topic access controls on users via Access Control Lists, FX systems in particular often have fine grained and per-client access control requirements. In order to allow SDP developers to provide their own data access

controls, the Solace message router supports the ability for a 3rd party subscription manager to inject topic-based subscriptions on behalf of other applications, such as RIAs. The usage pattern is as follows:

- In order to subscribe to a particular real-time flow, the RIA sends a message to the well-known topic of the subscription manager requesting access to certain data.
- The subscription manager decides based on its own business rules whether to allow or disallow the request for this particular user and then adds one or more (e.g. for baskets) subscriptions to the Solace message router for the requested feed(s) for this particular user.

This feature gives application teams significant flexibility and control. For example, where different tiers of currency spreads are provided to external clients based on business arrangements, a user requesting a real-time feed for US-Canadian dollars could be configured with a subscription to a different Solace topic carrying this spread based on the tier of service allocated to that user (e.g. Bronze, Silver, Gold). The data formats in the updates remains the same as all other users, but the content of the spreads differs based on the business relationship. The RIA doesn't even need to know the actual topic being subscribed to, as the currency information is in the payload of the message. This makes the messaging system topic irrelevant to the RIA, eliminating a system dependency.

Per-user Rate Controls

Solace offers a feature called "message eliding" that can be applied individually to each user whether connecting to the message router over the web or over the corporate intranet. Eliding gives each user the option of defining a custom per-topic rate so they can, for example, receive updates at a rate of at most 5 per second per symbol even though the source is publishing updates at much higher rates. This allows a single stream of real-time updates to be sent throughout the infrastructure, adapting it user-specific needs at the edge on a per-client basis without any additional infrastructure.

Eliding is useful when serving data to people or to applications over high-latency WAN links or over the internet. In addition to throttling throughput, eliding ensures that only the most recent update for a given topic is provided to the end client, with newer messages overwriting older, queued messages for the same topic. If the client becomes slow or the network gets congested, eliding provides updates at an even lower rate while still ensuring that the most recent data is delivered. Messages sent from streaming applications are marked as either eliding eligible or not so quotes on a particular topic can be elided, but trades on that same topic, for example, are not.

Users can set update rates by topic so they can, for example, receive 5 updates per second per symbol even though the source is publishing at much higher rates. This means a single stream of updates can meet the needs of high-speed systems without overwhelming slow consumers.

Summary

The implementation of single dealer platforms can give investment banks and sell-side firms a leg up in their own trading activities, in the service they offer their clients, and their overall profitability and competitiveness. With the only solution that can meet all of the underlying messaging requirements of SDPs, Solace can help financial institutions take advantage of this important strategy much more quickly, cost-effectively and safely than with any other combination of technologies..

To learn more visit solace.com or call +1 613-271-1010.