# IBM WebSphere MQ Low Latency Messaging
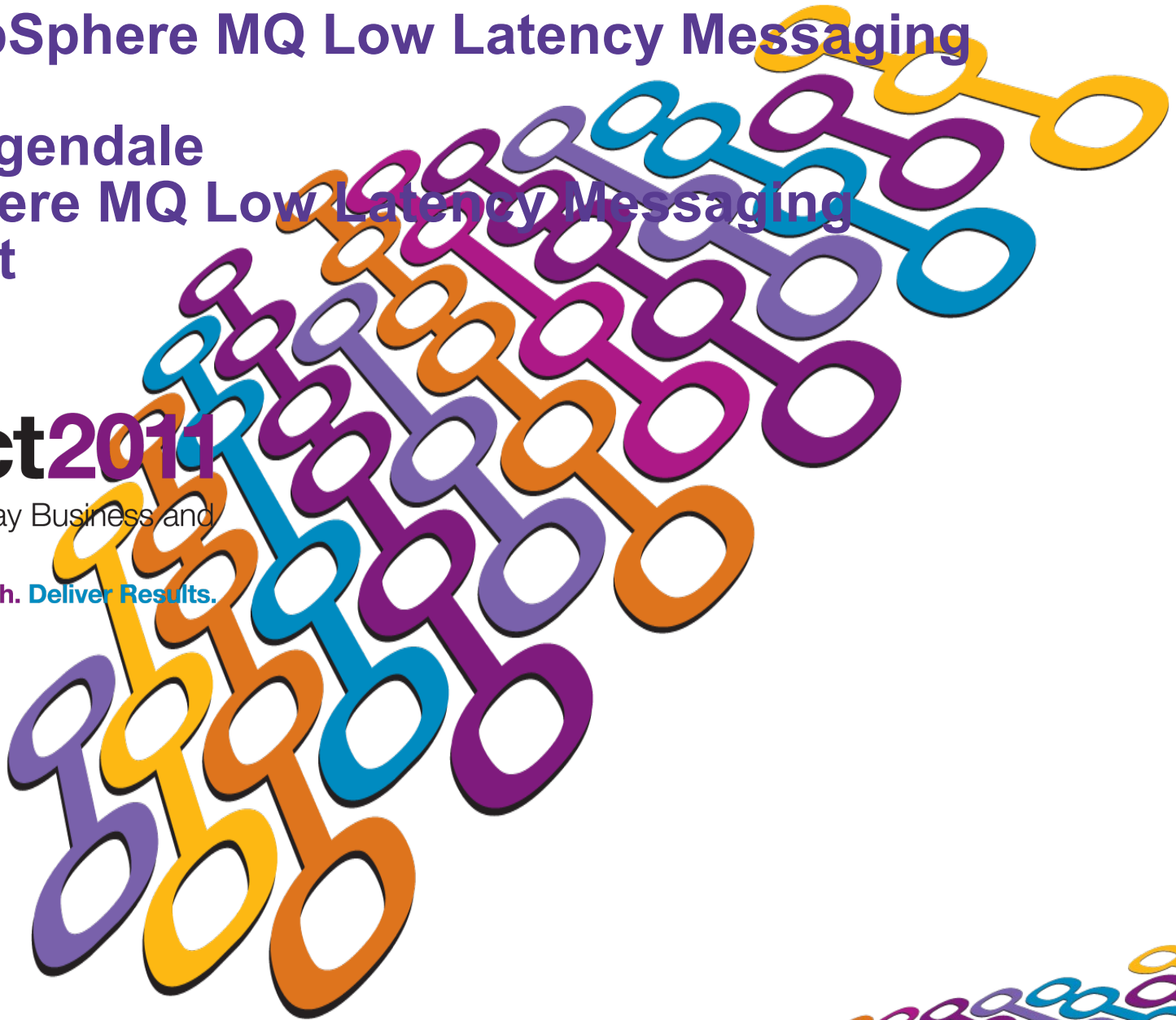
## Ken Borgendale
## WebSphere MQ Low Latency Messaging Architect

IBM Software

# Impact2011

Changing the Way Business and
IT Leaders Work

**Optimize for Growth. Deliver Results.**

# Overview of LLM

- IBM WebSphere MQ Low Latency Messaging (LLM) is ultra low-latency and high-throughput messaging primarily aimed at the financial services industry.  It provides support for front and mid office messaging.

- LLM uses direct transmitter to receiver messaging without any intervening message server.

- LLM is the fastest messaging in the industry

- LLM is a member of the IBM WebSphere MQ family specializing in fast and high volume messaging.

# IBM messaging vision

**Web 2.0** **J2EE / JCA** **API driven** **XMS** **JMS** **W**
**REST** **App Servers** **ser**

and Quality of Service levels to meet the varying business needs

- Augment WMQ's transports beyond its classic rock-solid "back-end" messaging style

- Extensive support and exploitation of computing platforms and runtime environments

- Underpin and extend ESB architectures

# Uses of LLM

- LLM is based on technology from IBM Research Lab in Haifa

- LLM is integrated into other IBM products:
  - IBM WebSphere Front Office for Financial Markets (WFO)
  - IBM WebSphere DataPower XM70 appliance
  - IBM InfoSphere Streams
  - . . .

# LLM customer use cases in financial services

- **Market data**
  - Rapid delivery from exchanges to market-data consumers
  - Managing message traffic including congestion, slow consumers, retransmissions, resources

- **Exchange order matching**
  - Reliable communication of positions, orders and confirmations with execution infrastructure
  - High availability for loss-less failover
  - Auditable without compromising latency & throughput
  - Sophisticated execution routing with advanced filtering

- **Order execution**
  - Event notifications for systems monitoring, risk analytics and compliance applications
  - Enabling Complex Event Processing with flexible filtering
  - Integration with existing Back Office systems
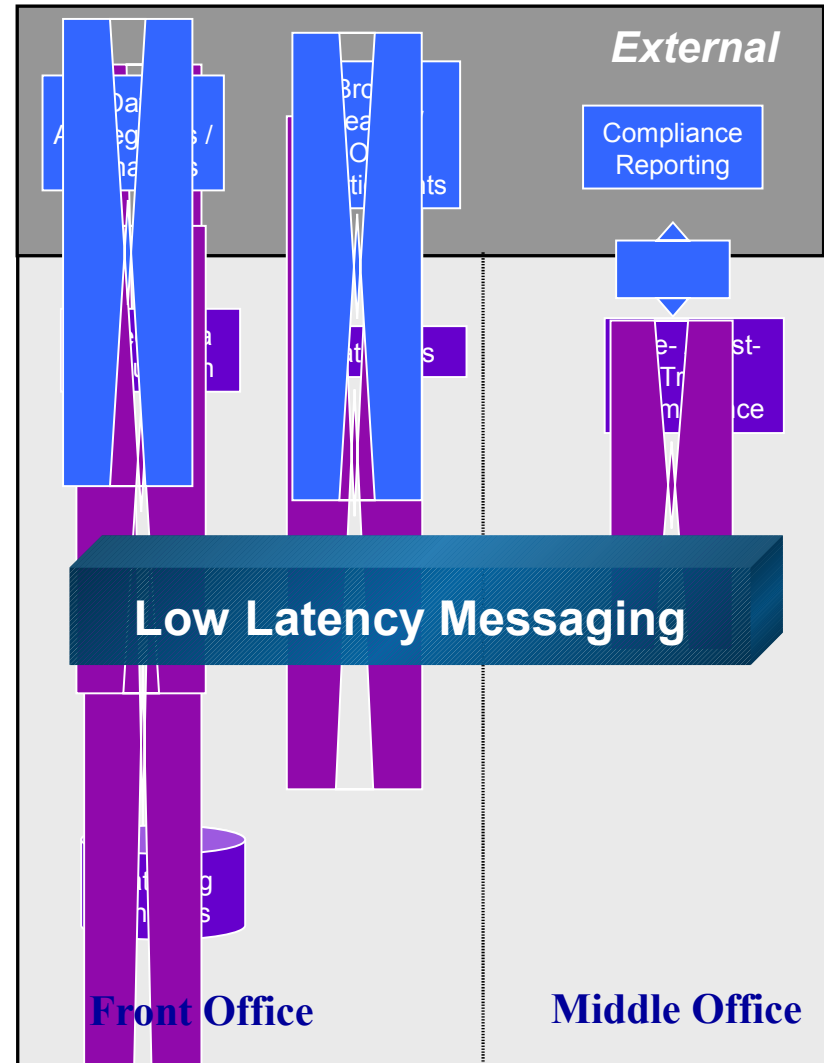
# LLM differentiating features

- Fastest messaging in the industry
  - LLM is optimized to deliver ultra low latency at high throughput
  - Performance, performance, performance

- Single API for all network protocols
  - IP (multicast, udp, tcp)
  - OFED verbs (native InfiniBand, RoCE and iWarp)
  - Shared memory
  - Multicast by unicast

- Best of breed reliability and high availability
  - No performance compromise
  - Sub-second failover with zero message loss
  - Support for replicated network hardware (IP bonding, IB failover)
  - Message persistence with recovery

# Low latency and high throughput

- LLM is optimized to deliver ultra low latency at high throughput

- Native OFED verbs support for InfiniBand and 10Ge

- Single digit microsecond single hop latencies for InfiniBand and 10Ge RDMA

- Throughput limited by network capacity for all network types to QDR InfiniBand (40Mbps).

- Fastest messaging in the industry
  - Reflector tests
  - STAC M2 benchmarks

# Single API for all protocols

- Write the application once and decide later on the network configuration
  - UDP multicast using PGM retransmission
  - UDP unicast
  - TCP unicast
  - OFED native verbs UD multicast and unicast
    - Supports InfiniBand and RoCE
  - OFED native verbs RC unicast
    - Supports InfiniBand, RoCE, and iWarp
  - Shared memory transport for co-located clients
- Supports multicast by unicast with mixed protocols
- Supports InfiniBand HDA and subnet failover
- Tested with RDMA kernel bypass libraries

# IBM LLM drives down latency and increases throughput

- ## Highest throughput

- ## Lowest latency

| Low Latency | | Single hop | |
|---|---|---|---|
| Network | Message size (bytes) | Transmission rate (msgs/sec) | |
| | | 10K | 100K |
| 1 GbE Ethernet | 120 | 29 µs | 34 µs |
| 10GbE Ethernet* | 120 | 4 µs | 4.5 µs |
| InfiniBand | 120 | 2 µs | 3 µs |
| Shared Memory | 120 | 1 µs | 1 µs |

| High Throughput 10 Gigabit Ethernet | |
|---|---|
| Message size (bytes) | Message Rate (msgs/sec) |
| 12 | 75,914,578 |
| 45 | 25,253,255 |
| 120 | 9,724,107 |
| 1200 | 985,846 |
| 12,000 | 98,225 |

# Both high throughput and low latency

- Consistent low latency as throughput increases
- Consistent low jitter as throughput increases
- Tests with 45 byte messages:

| Rate (msgs /sec) | Avg | 50% | 99% | Std. Dev. |
|---|---|---|---|---|
| 1,000 | 2.5 | 2.5 | 4.5 | 1.0 |
| 10,000 | 2.5 | 2.5 | 3.5 | 0.5 |
| 100,000 | 2.5 | 2.5 | 3.5 | 0.5 |
| 1,000,000 | 2.5 | 2.5 | 3.5 | 1.0 |
| 1,500,000 | 2.5 | 2.5 | 4.5 | 1.0 |

# Performance – compare protocols

- Reflector test with various transports with 120 byte messages

- RMM TCP is new in LLM V2.5 and shows about 10% increase in latency and decrease in throughput relative to UDP.

- 98 million messages per second on InfiniBand and shared memory with 12 byte messages

- New 10Ge RDMA show performance close to RoCE

| Transport<br><br>120 byte msg | Normal latency (μs) | | | Throughput (msgs / sec) |
|---|---|---|---|---|
| | Mean | 50% | 99% | |
| IP 1Ge UDP | 27 | 27 | 35 | 972,000 |
| IP 1Ge TCP | 29 | 29 | 36 | 845,000 |
| IP 10Ge | 12 | 12 | 14 | 9,700,000 |
| RDMA 10Ge (RoCE) | 4 | 4.5 | 6 | 8,300,000 |
| InfiniBand | 3 | 3 | 4 | 13,110,000 |
| Shared memory | 1 | 1 | 5 | 19,900,000 |

# Performance – STAC M2 benchmark

- The STAC M2 benchmark tests the ability to handle real-time market data under a variety of conditions.

- Measures latency and throughput

- Tests conditions such as slow consumer

- Recently published 10Ge with Solarflare and Juniper

- Recenty published 10Ge RoCE with Mellanox and IBM-BNT

- LLM has been used as the software component of most STAC M2 audited benchmarks

| STAC M2 Product and transport | | Normal latency (µs) | | | Out lier (µs) | Rate (Kmsg/ sec) |
|---|---|---|---|---|---|---|
| | | Avg | 50 | 99 | | |
| LLM | InfiniBand 2009 | 8 | 8 | 11 | 37 | 1000 |
| | 10Ge RoCE 2011 | 6 | 6 | 9 | 48 | 1400 |
| | 10Ge Solarflare and Juniper | 9 | 9 | 11 | 23 | 1500 |
| | 1Ge IP | 33 | 33 | 36 | 64 | 1000 |
| 29 West | 10Ge RDMA Solatflare and Juniper | 14 | 14 | 17 | 36 | 1300 |

# How fast is light?

- One of our competitors calls their ultra performance messaging FTL for Faster Than Light.

- So how fast is light?

- Light travels 300 million meters per second, or 186 thousand miles per second.

- That means that light travels 30cm or about a foot per nanosecond

- Light in a fiberoptic cable takes about 1.5 times longer, so it takes about 5 µs per KM (8 µs per mile).

- In the time it takes LLM to send a message over InfiniBand, light would make it from here to the Houses of Parliament.

- We have not yet seem the STAC benchmarks from FTL yet, so we do not yet know if it is faster than light.

# Reliability – replication (RCMS)

- High availability based on component replication
- No performance penalty for using high availability
- Total order is maintained with multiple streams to ensure all replicas are identical
- Fast and automatic detection and failure recovery
- Application state sync and intra-tier communications
- Support for redundant networks
- Can use multicast by unicast
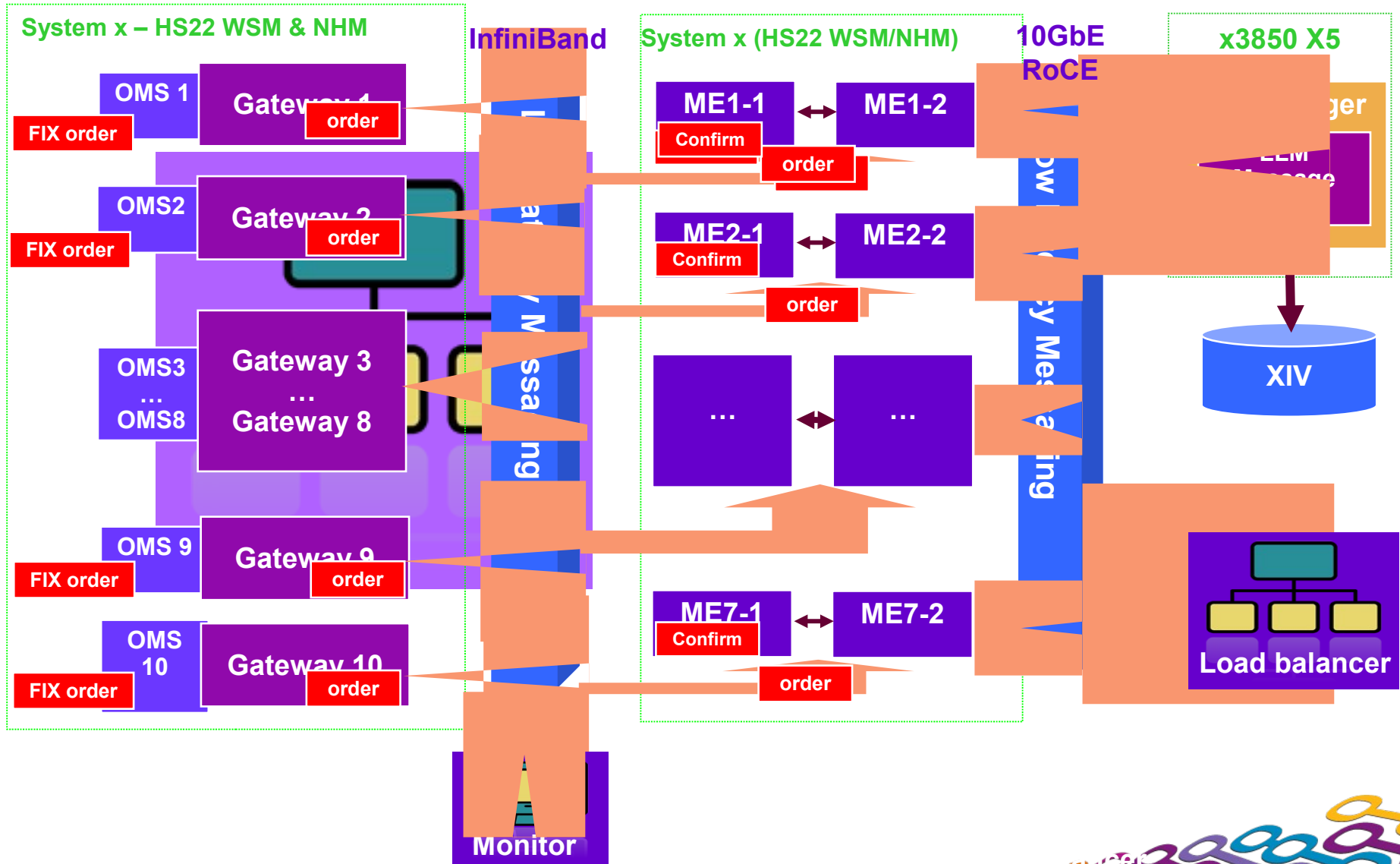- Split-brain prevention and detection with data checkpoints
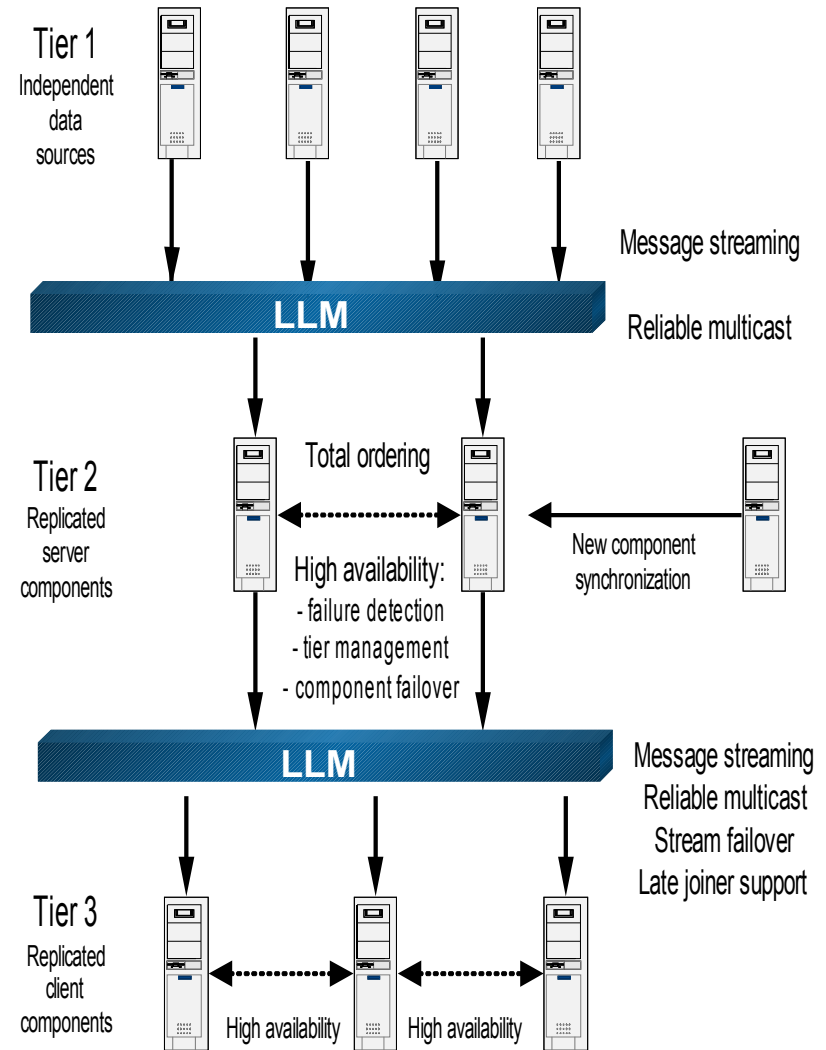
# Reliability - persistence

- LLM provides a Message Store service within the Coordination Manager to store and retrieve messages.

- Can be used for audit/logging of messages, or for durable subscriptions.

- Extremely efficient asynchronous disk write

- Control over assured delivery:
  - ACK after write
  - Route after write
  - Maximum time to write

- Retrieve with message filtering over time or message range

- Does not slow down transmitter to receiver times

- Automatic recovery of receiver with durable subscription

# SXRA SIFMA demo data flow

**System x – HS22 WSM & NHM**

**InfiniBand**

**System x (HS22 WSM/NHM)**

**10GbE RoCE**

**x3850 X5**

OMS 1 — FIX order — **Gateway 1** — order

OMS2 — FIX order — **Gateway 2** — order

OMS3 ... OMS8 — **Gateway 3 ... Gateway 8**

OMS 9 — FIX order — **Gateway 9** — order

OMS 10 — FIX order — **Gateway 10** — order

**ME1-1** ↔ **ME1-2** — Confirm — order

**ME2-1** ↔ **ME2-2** — Confirm — order

... ↔ ...

**ME7-1** ↔ **ME7-2** — Confirm — order

**XIV**

**Load balancer**

**Monitor**

# Bolsa Comercio:
## New securities trading platform for Chilean Stock Exchange

**Objectives:**
- Position stock exchange for market environment of the future:
  - Higher velocity
  - Low latency
  - High transparency

**Actions:**
- Base new stock exchange securities trading platform on WS MQ Low Latency Messaging:
  - Lowest latency and highest throughput messaging platform in the industry

**Benefits:**
- Will improve flexibility, scalability, security of service to corporate clients
- Will allow exchange to optimize its electronic transaction management – increasing capacity to 3,000+ orders per second (100x greater than present) and reduce latency to two ms (100x lower than current level)
- Will position Chilean Stock Exchange in vanguard of world's regional exchanges, and allow it to leverage its strategic alliance with the Mexican Stock Exchange

"The new transactional engine of the Santiago Stock Exchange will process more than 3,000 orders per second with latency measured in microseconds and under the concept of total connectivity."
*- Jose Antonio Martinez, General Manager, Santiago Stock Exchange*

# Deutsche Börse AG:
## Next Generation Trading Infrastructure





Objectives:

· New *standardized global* technology infrastructure
  · Supporting future operations of Deutsche Börse's trading systems

Actions:

· *Build new architecture* for different markets and financial instruments using:
  · Open source components, including Linux
· Internally developed software
  · Vendor software
  · IBM WebSphere MQ Low Latency Messaging, for high speed transport

Benefits:

· Efficiency, speed, throughput for ISE, Eurex, Xetra users – across all asset classes
· Processing >1million quotes/sec at <1ms/quote
· Optimum support for future expansion

*"Our key objective is to deliver a leading-edge, accurate and reliable transaction backbone, building on our track record of innovation as a first-mover in electronic markets. Deutsche Börse is committed to augmenting its IT capacity with best-in-class partners."*

*- Dr. Michael Kuhn, CIO Deutsche Börse AG*

# Chi-X Global Technology:
## High performance trading technology and services



> "Chi-Tech selected WebSphere MQ Low Latency Messaging as a messaging backbone for our high performance, low latency matching technology, assuring that our platform continues to meet the speed and mission-critical reliability requirements of today's algorithmic and high frequency trading communities."
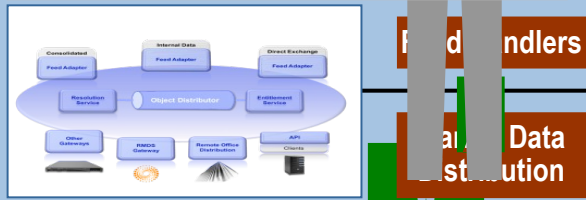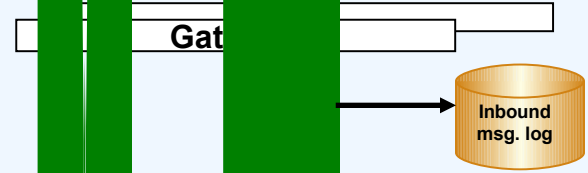> -- Richard Leung, Chief Technology Officer, Chi-Tech

# Exchange (NYSE, NASDAQ, etc.)

**Data Feeds**

Stock Exchange Reference Architecture (SXRA)

**Execution Venue**

Master File DB

Inbound Msg

Working DB

Trxn log

---
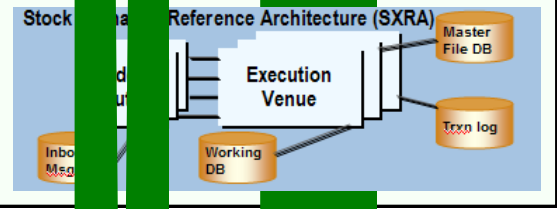
## Monitoring and Reporting

Lotus Expeditor

COGNOS
AN IBM COMPANY

## Front Office v3.0

WebSphere

Consolidated Feed Adapter | Internal Data Feed Adapter | Direct Exchange Feed Adapter

Resolution Service | Object Distributor | Entitlement Service

Other Gateways | RMDS Gateway | Remote Office Distribution | API Clients

**Feed Handlers**

**Market Data Distribution**

**Gateway**

Inbound msg. log

---

## IBM WebSphere MQ Low Latency Messaging

---

**Market data analytics**

## Analysis, Pricing and Order Generation

**InfoSphere** **Streams**

RSS Parser | Keyword Filter | Extract Format From ... 

Sentiment | Calculate

**Analytics** → **Order Generation**

Market

history

IDS

Reference data

solidDB
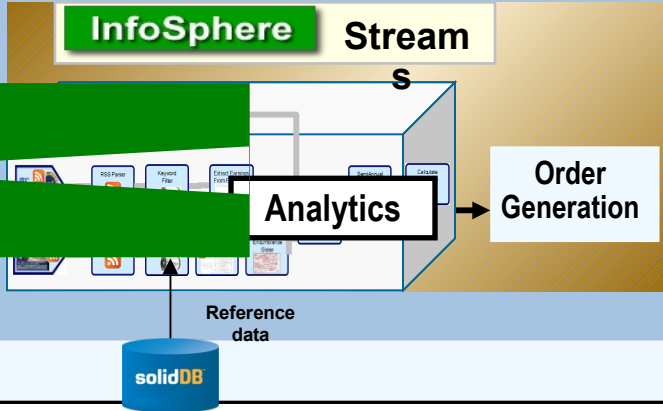
## Execution Strategy, Order Routing
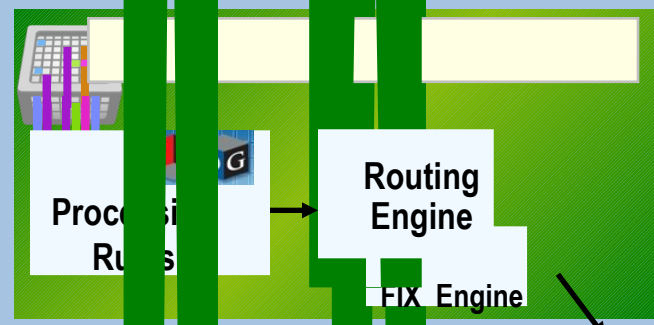
**Processing Rules** → **Routing Engine**

FIX Engine

"write behind" technology

Persisted Grid data
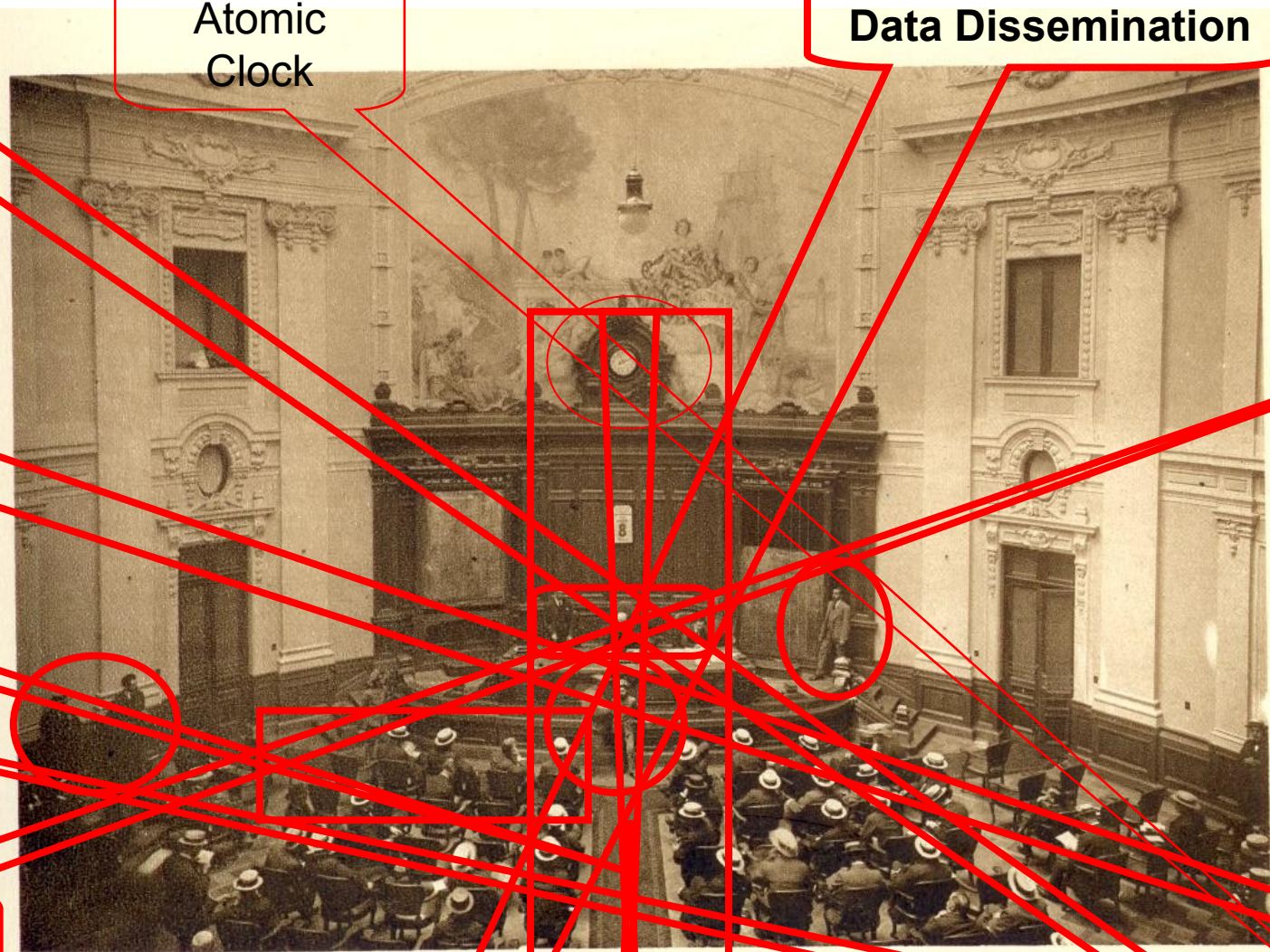
# Chilean Stock Exchange 1929



The matching engine

Atomic Clock

Data Dissemination

Distribution Gateway
(Primary and Backup)

Monitoring

Proximity or Co-location

Data BUS

Al fondo se ve el cuadro alegórico, notable obra de D. Pedro Subercaseaux

# Additional information

- Financial Markets Analytics is the IBM portfolio of offerings that address high velocity trading and analytics in the front office.

- Faster Data:
  ibm.com/financialmarkets/fasterdata

- Financial Market Analytics:
  ibm.com/financialmarkets/analytics

- WebSphere MQ Low Latency Messaging:
  ibm.com/software/integration/wmq/llm
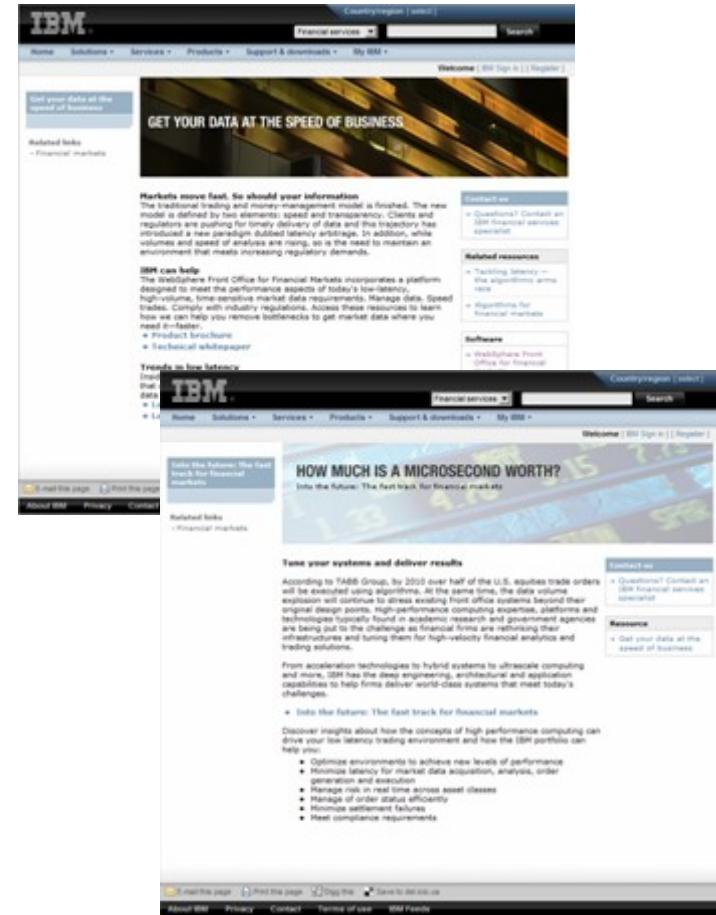
# Copyright and Trademarks

© IBM Corporation 2011. All Rights Reserved.