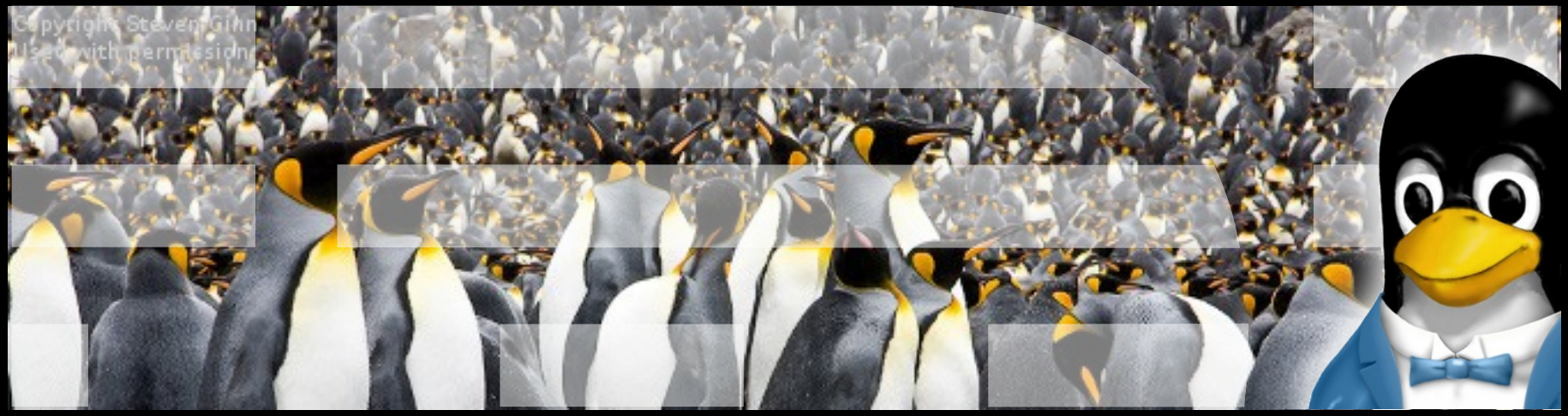
# Smarter Systems for High Performance Computing

## HPC Linux Financial Markets - April 4, 2011

Jean Staten Healy
Director, Cross-IBM Linux
IBM



# Agenda

- Watson Overview

- Panelist Presentations
  - Edward Epstein, IBM Research WATSON
  - Tom Befi, Insurance Services Office
  - Doug Beary, Datatrend Technologies
  - Vikram Mehta, IBM System Networking

- Panelist Q&A

# Today's Panelists

**Edward Epstein**
WATSON - Manager of Unstructured Information
IBM Research

**Douglas Louis Beary**
Account Executive, High Performance Computing
Datatrend Technologies, Inc.

**Tom Befi**
Vice President of Information Systems Services
Insurance Services Office, Inc.

**Vikram Mehta**
Vice President of Systems Marketing
IBM Systems & Technology Group

# Video - IBM Watson a System Designed for Answers

# Edward Epstein

*WATSON - Manager of Unstructured Information*
*IBM Research*

JEOPARDY! The IBM Challenge

www.ibmwatson.com

# Watson and HPC

IBM Research

# The Jeopardy! Challenge: *A compelling and notable way to **drive** and **measure** the technology **of automatic Question Answering** along 5 Key Dimensions*

**Broad/Open Domain**

**Complex Language**

**High Precision**

**Accurate Confidence**

**High Speed**

**$200**
If you're standing, it's the direction you should look to check out the wainscoting.

**$1000**
Of the 4 countries in the world that the U.S. does not have diplomatic relations with, the one that's farthest north

**$800**
In cell division, mitosis splits the nucleus & cytokinesis splits this liquid *cushioning* the nucleus

# The Big Idea: Evidence-Based Reasoning over Natural Language Content

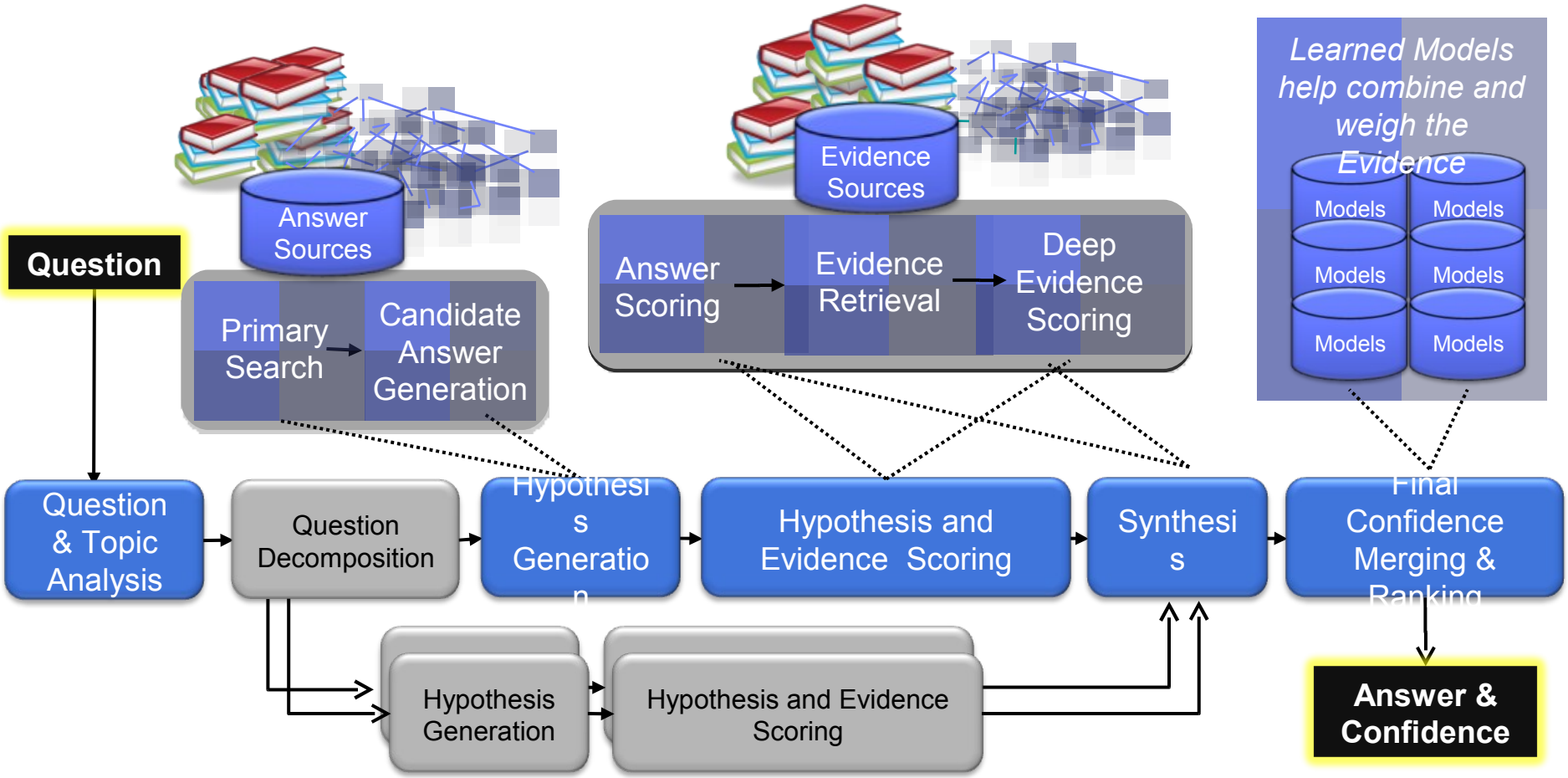
- **Deep Analysis** of questions AND content
- **Search for many possible answers** based on different interpretations of question
- **For each answer** find, analyze and score **EVIDENCE** from many different sources using many advanced NLP and reasoning algorithms
- **Combine scores** and compute an accurate **confidence** value for each possibility using statistical machine learning

# **DeepQA**: The Technology Behind *Watson*

IBM

## *Massively Parallel Probabilistic Evidence-Based Architecture*

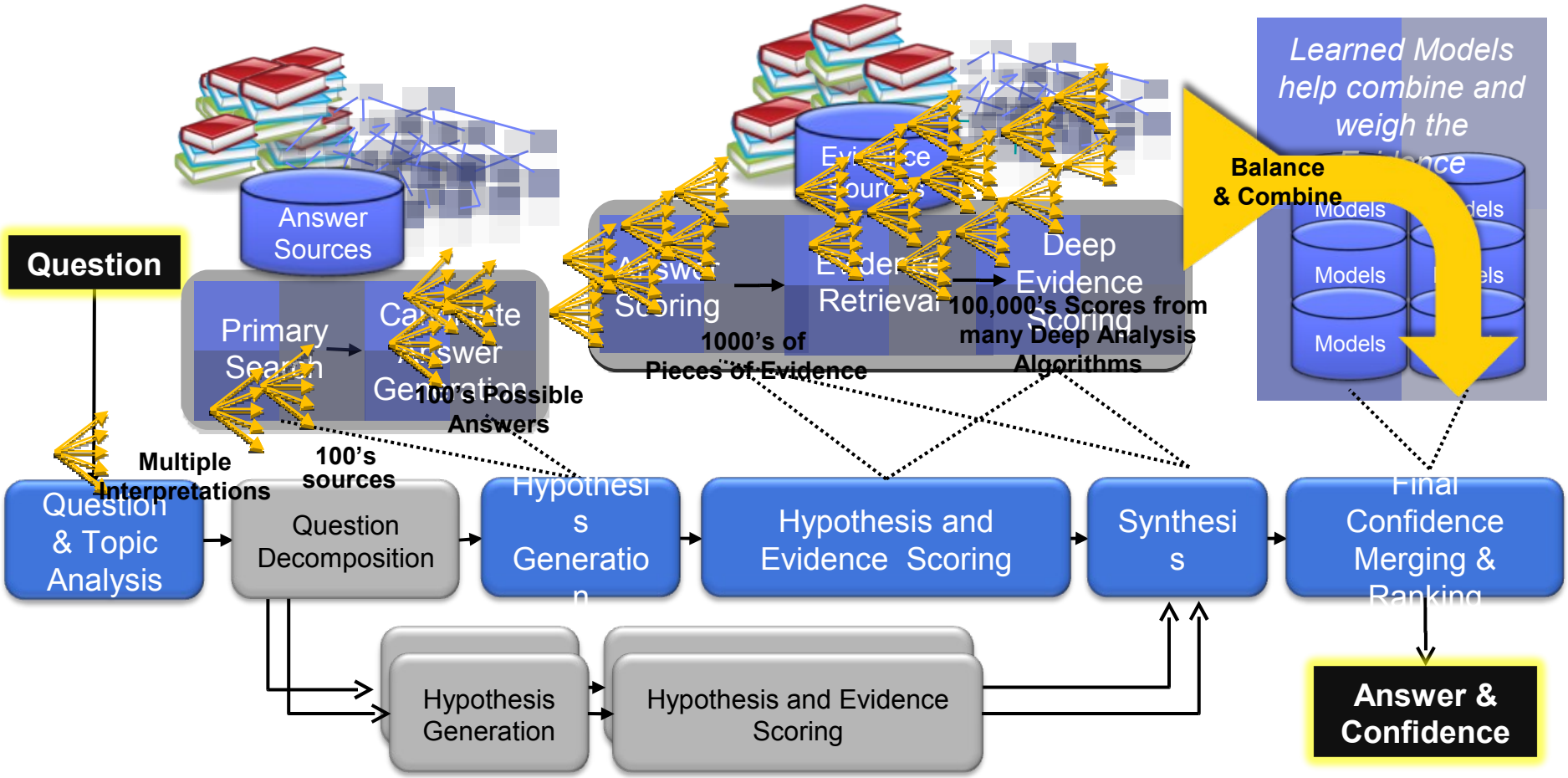
*Generates and scores many hypotheses using a combination of 1000's* ***Natural Language Processing***, ***Information Retrieval***, ***Machine Learning*** *and* ***Reasoning Algorithms.***
*These gather, evaluate, weigh and balance different types of* ***evidence*** *to deliver the answer with the best support it can find.*



# **DeepQA**: The Technology Behind *Watson*

## *Massively Parallel Probabilistic Evidence-Based Architecture*

*Generates and scores many hypotheses using a combination of 1000's* ***Natural Language Processing****,* ***Information Retrieval****,* ***Machine Learning*** *and* ***Reasoning Algorithms.***
*These gather, evaluate, weigh and balance different types of* ***evidence*** *to deliver the answer with the best support it can find.*

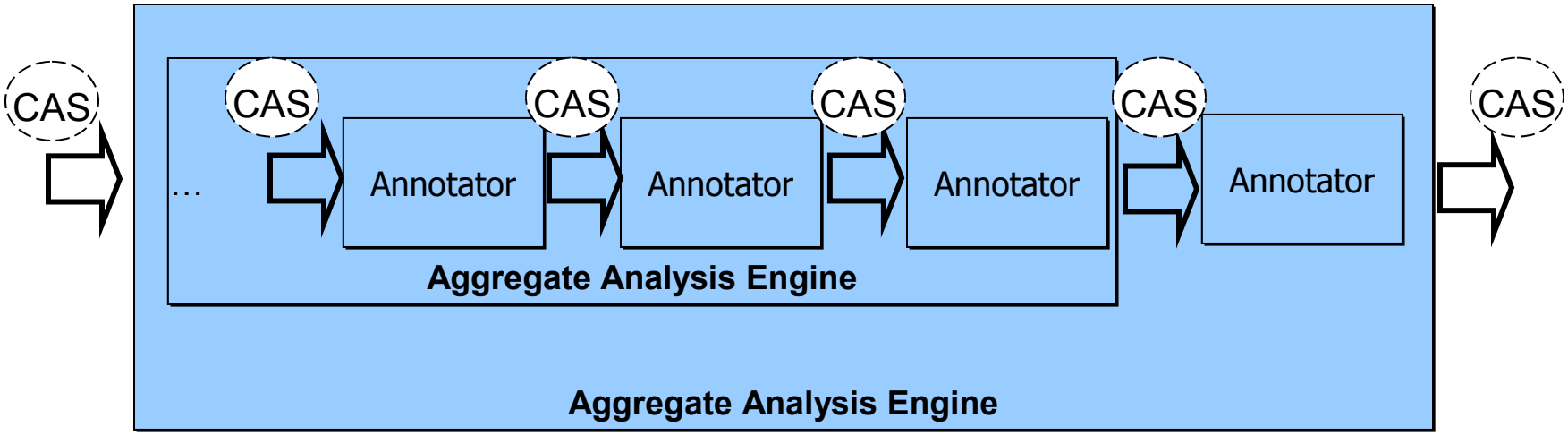


# Begin work on an Interactive System

- Starting point: 2 hours to process a single question
- Why was my team chosen for this work?
  - Core UIMA development team
  - **Apache UIMA** is heavily used for Watson analytics
    - Solves Interoperability
    - Solves Results Organization and Management
  - Previous 3 years had focused on UIMA scale out
  - Software engineers with history of optimizing complex analytics

# Apache UIMA

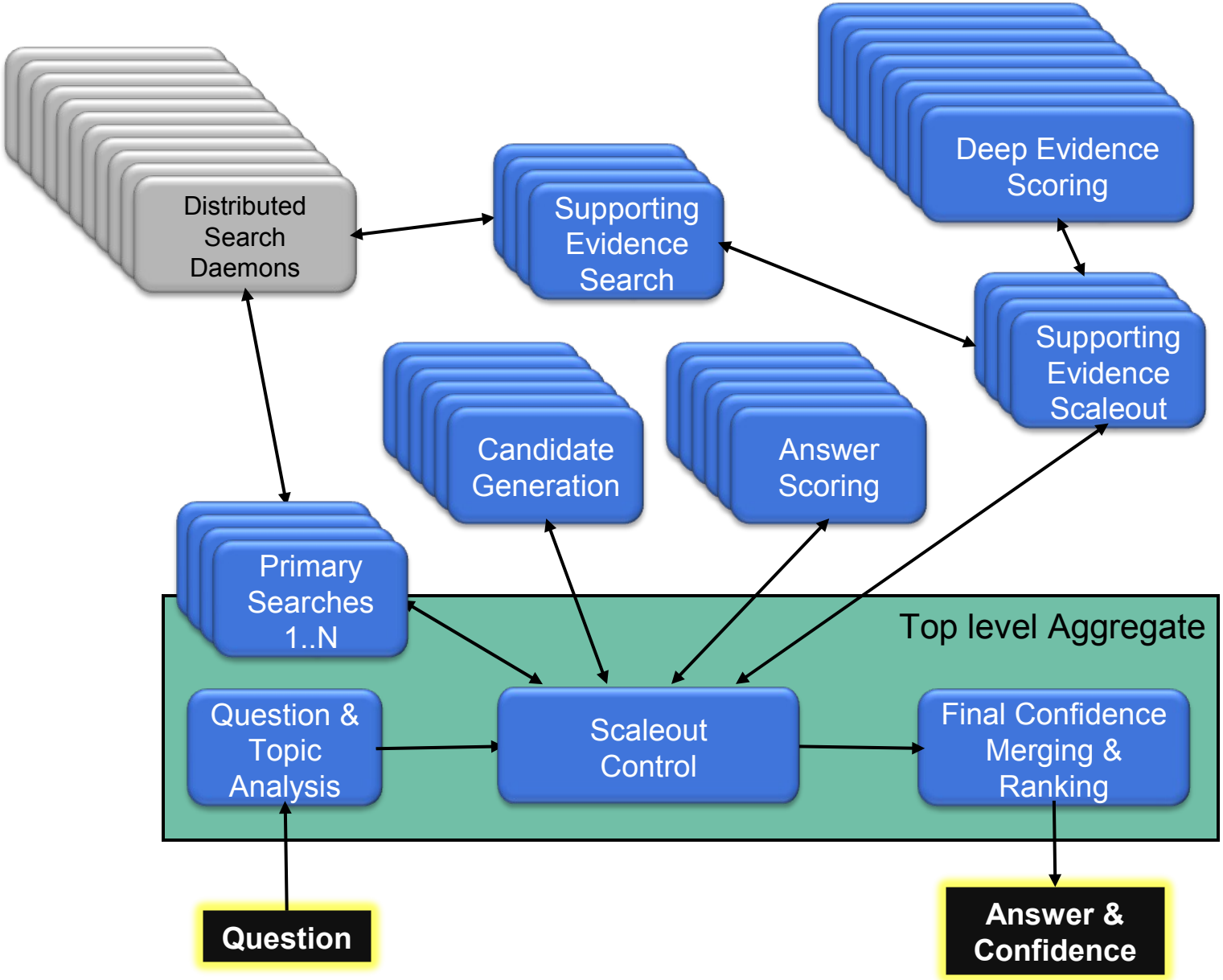
- Open-source framework and tools for building NLP applications
- Key Concepts
  - *Common Analysis Structure (CAS):* Container for Inputs & Outputs in user-defined data model
  - *Annotator:* Pluggable component (Java or C++, among others) that reads and writes a CAS
  - *Aggregate Analysis Engine:* Collection of Annotators



# Initial Scale Out Effort

- Move everything into RAM
- Scale out components with UIMA-AS
- Distribute search

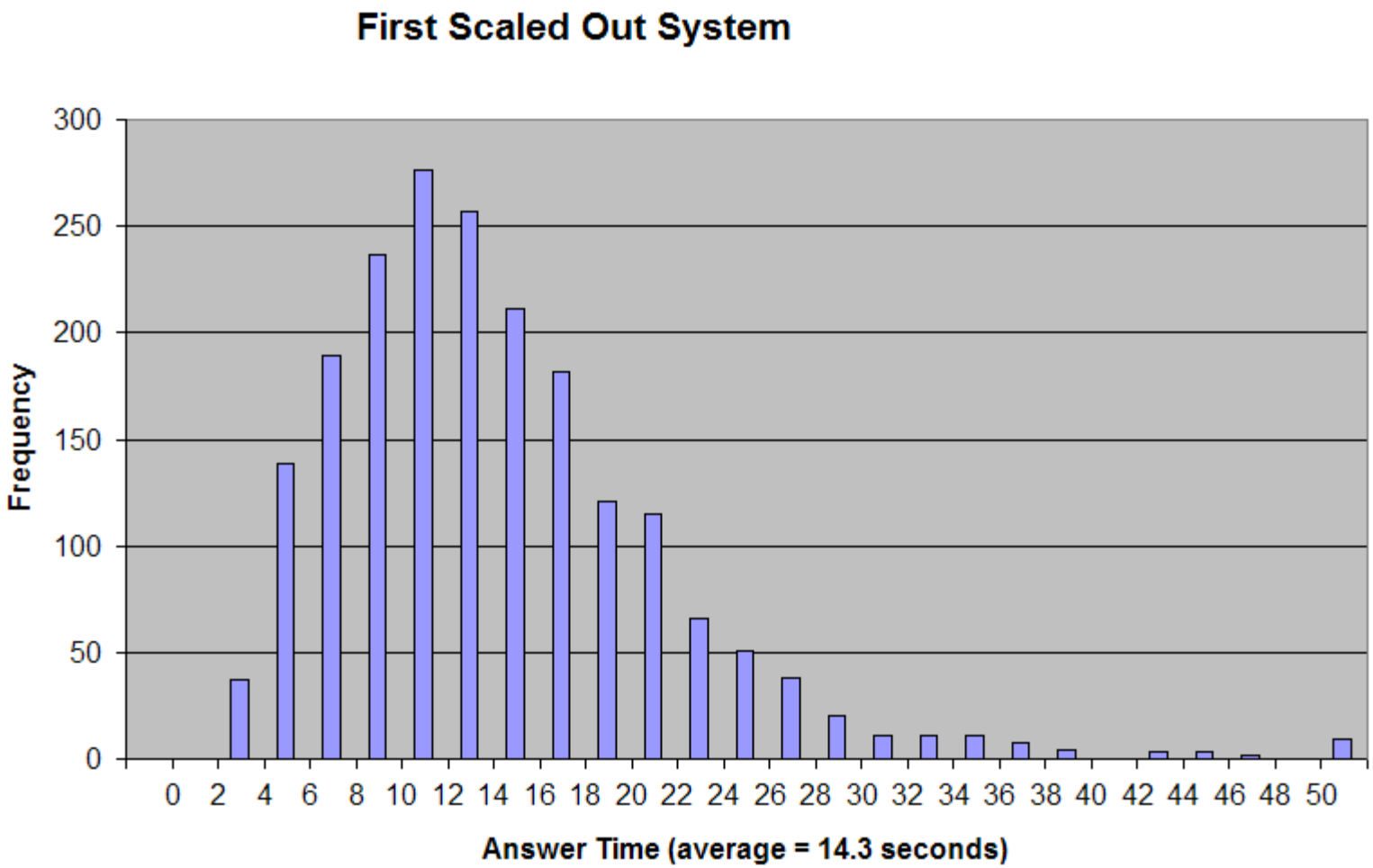
# Selected Watson Processes



# Characteristics of Watson Application

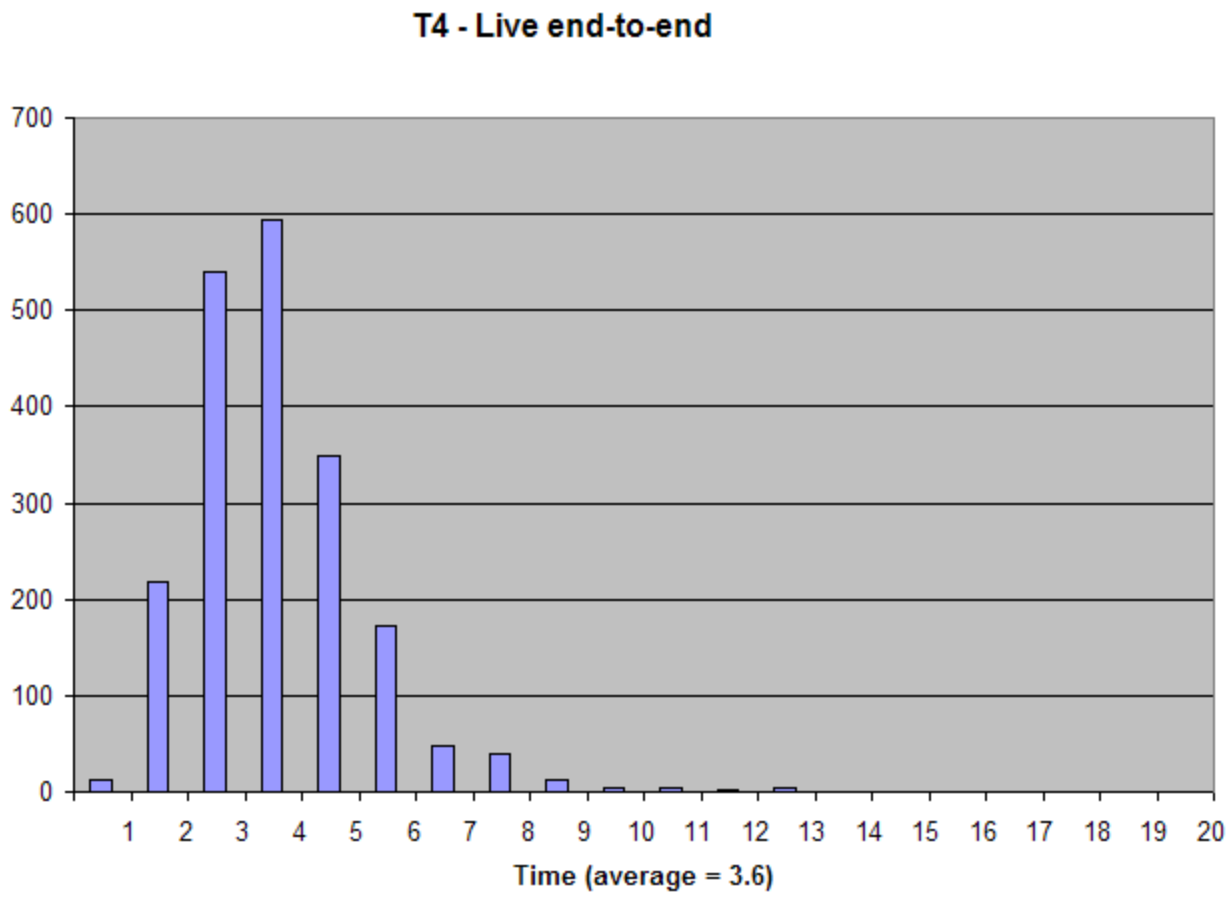
- ~200 Java processes
  - Most with 30 GB Heaps
  - Some with 10s of GB in filesystem buffers
- ~200 C++ processes
  - 2 GB resident

# After first 8 months of Scale Out Work …



# 4 more months of Scale Out Work …

- Pre-compute deep NLP analysis of entire text corpus
- Hammer on every computation outlier



# Let the Games Begin

- Test against internal contestants
- Demo to JPI
- Begin “Sparring” matches with former Jeopardy! contestants

# Next 12 Months

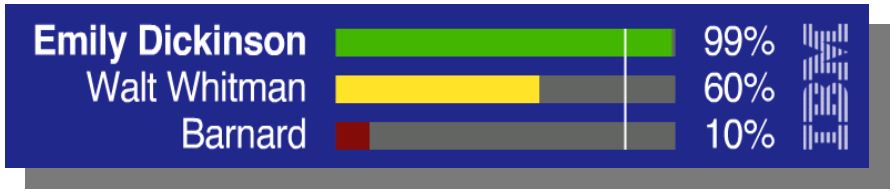
- Improve Accuracy
  - Add missing analytic components
  - Add new analytic components
  - More than double the knowledge source
- Further speed improvements
  - Migrate production system to Power 750 servers
- Merge development and production source code
- Final Sparring matches against Tournament players

# Power 750 is a Good Fit for Watson

- High performance CPUs
  - Essential to meet speed requirements
- 32 real CPU cores per node
  - Far fewer nodes needed
- Large shared memory per node
  - More flexibility (e.g. very large memory training tasks)
- High memory bandwidth
  - Enabled full CPU utilization

# Precision, Confidence & Speed

- **Deep Analytics –** We achieved champion-levels of *Precision* and *Confidence* over a huge variety of expression




- **Speed** – By optimizing Watson's computation for Jeopardy! on **POWER7** processing cores we reduced **average answering time below 3 seconds** – fast enough to compete with the best.


- **Results –** in 55 real-time sparring games against former **Tournament of Champion Players last year**, Watson put on a very competitive performance in all games, winning 71% of the them!

# THANK YOU

# Video - Perspectives on Watson: Finance

# Tom Befi

*Vice President of Information Systems Services*
*Insurance Services Office, Inc.*

# ISO z/Linux Experience

THE SCIENCE OF RISK℠

# Who am I?

- Vice President Information Systems Services for ISO
  - Infrastructure
    - Data Center Operations
    - Systems Programming (Mainframe/Distributed)
    - Network Engineering (voice/data telecom)
    - Desktop Computing Environment
    - Infrastructure Architecture
  - Technical Support
    - Internal Technical Help Desk
    - Information Center

# Who We Are and What We Do

- Verisk Analytics provides Data, Analytics and Decision support products across multiple vertical markets in and around risk mitigation
- ISO is a member company of Verisk Analytics that operates in the P&C Insurance vertical

# Our Brands

Risk Analysis

P & C

Healthcare

Mortgage

# ISO Environment

- Historically a Mainframe Shop
- 2 Mainframes (z9 and a z196)
- 250 Suse Linux Servers on 7 IFL's
- 700 Distributed Servers
- 14 Million lines of Cobol
- CICS/DB2/MQ/Model204
- VB / Visual Studio

# Address the Issues

Business Issues (Circa 2003)

- High Cost of Ownership on Servers Due to:
  - Short support half life
  - Complexity
    - Environmental (sprawl, DR)
    - Application
  - Little re-usability
- Security Concerns
  - Hackers/Virus Target
- Availability and Reliability

Strategic Direction

- Technical realignment to Java
- Developing foundation architecture
- Build reusable Frameworks
- Shifting from point solutions to Enterprise-wide Development
- Consolidated Deployment on the mainframe

# The Whys?

Why JAVA?

- Platform Longevity
- Merry-go-Round
- 99 Person Years every 3-4 years
- Portability
- Platform Independence
- Open Source (Where applicable)
- Cobol Resource Issue
- Cobol not taught in most colleges
- Boomers retiring

Why Websphere?

- Best of Breed at the time & still is
- Supportable (IBM)
- Multi-Platform Support
- Leverage with IBM (Single Vendor)

Why Mainframe?

- Virus Attacks
- Security Vulnerabilities
- Availability, Reliability and Scalability
- Better Utilization of Hardware
- Simplify environment
  - Less tiers
  - Disaster Recovery
  - Better virtualization
- Economy of Scale
  - Less systems support staff required

# Expected Benefits

- Eliminate Technology Complexity
  - TIE (Tolerate/Integrate/Eliminate)
  - Java
  - DB/2 / Websphere / MQ

- Eliminate Software Development Fragmentation
  - Architecture
  - Alignment
  - Re-Use (Enterprise Frameworks)

- Eliminate Server Sprawl
  - Scalability/Reliability/Manage-ability
  - Better Security and protection (virus/hacker)
  - Less complexity (especially for DR)
  - More efficient utilization of hardware (Ex: Virtualization)

# Achieved Benefits

- Eliminate Technology Complexity
  - TIE (Tolerate/Integrate/Eliminate)
  - Java
  - DB/2 / Websphere / MQ

- Eliminate Software Development Fragmentation
  - Architecture
  - Alignment
  - Re-Use (Enterprise Frameworks)

Slowed Down

- ~~Eliminate~~ Server Sprawl
  - Scalability/Reliability/Manage-ability
  - Better Security and protection (virus/hacker)
  - Less complexity (especially for DR)
  - More efficient utilization of hardware

*Note:*

*Resource utilization increase on the mainframe commensurate with server consolidation*

# z/OS Growth Curve Analysis

Issues:

- Utilization growth above what was expected
- Corresponding expense growth especially non-relevant 3rd party software

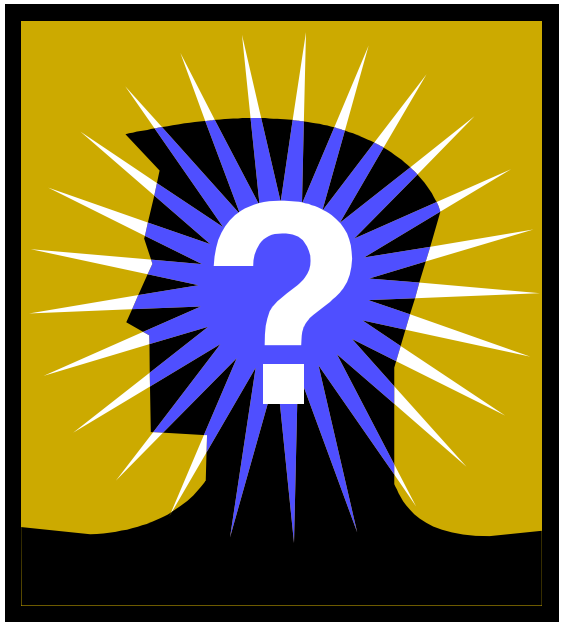
Actions:

- Worked with IBM
- Application Efficiency
- System Tuning
- Various platform alternatives reviewed
- Decision: Migrate to z/Linux

# Why z/Linux?

- Lower Cost
  - z/Linux software/hardware less expensive than z/OS
  - Put off z/OS upgrades (cost avoidance)
- Environmental Simplification
  - Simpler Allocation model
  - More flexible architecture
  - Easy/Quick to build additional environments
  - Simpler Disaster Recovery
- Better use of environment
  - Full use of H/W
  - Platform Independence
  - Unix Sys Admins instead of z/OS Sys Progs

# Questions?



# Thank You

Visit us online at

**www.aer.com**

**www.air-worldwide.com**

**www.hcinsight.com**

**www.iix.com**

**www.iso.com**

**www.veriskhealth.com**

**www.xactware.com**

# Douglas Beary

*Datatrend Technologies, Inc.*
*Account Executive, High-Performance Computing*

# Smarter Systems for Financial HPC

Breaking through latency barriers in high-frequency trading….

And other innovative IBM technology applications.



Doug Beary

Account Executive
High Performance Computing

Direct 919.961.4777

doug.beary@datatrend.com

Datatrend TECHNOLOGIES INC. Taking IT Beyond

*Raleigh, NC*

*www.datatrend.com*


Bank of America

BB&T

FannieMae.

WELLS FARGO

# Datatrend Snapshot

| | |
|---|---|
| Leading IT solution provider | • Best-in-class<br>  • Data center consulting<br>  • Server and storage solutions<br>  • Network infrastructure services |
| 24 Years Experience | • Founded in 1987 |
| National and International Reach | • Headquartered in Minneapolis<br>• Field offices in Florida, North Carolina & Washington, DC |
| Strong partnerships | • Industry leading hardware and software providers offering first class solutions<br>  • IBM, BMC Software, Oracle, Brocade, VMware |

# Top Performing Solution Provider

Recognized as a top solution provider by customers and partners, including:

**IBM Beacon Award Winner**
Multi-Year Winner

**IBM Beacon Award Finalist**
Multi-Year Winner

**2010 IBM Competitive Leadership Award**

**2008 IBM Innovation Leadership Award**

## Customer Recognition

**Wells Fargo Premier Solutions Provider Award**
Multi-Year Winner

**JCPenney Solutions Supplier Recognition Award**
Multi-Year Winner

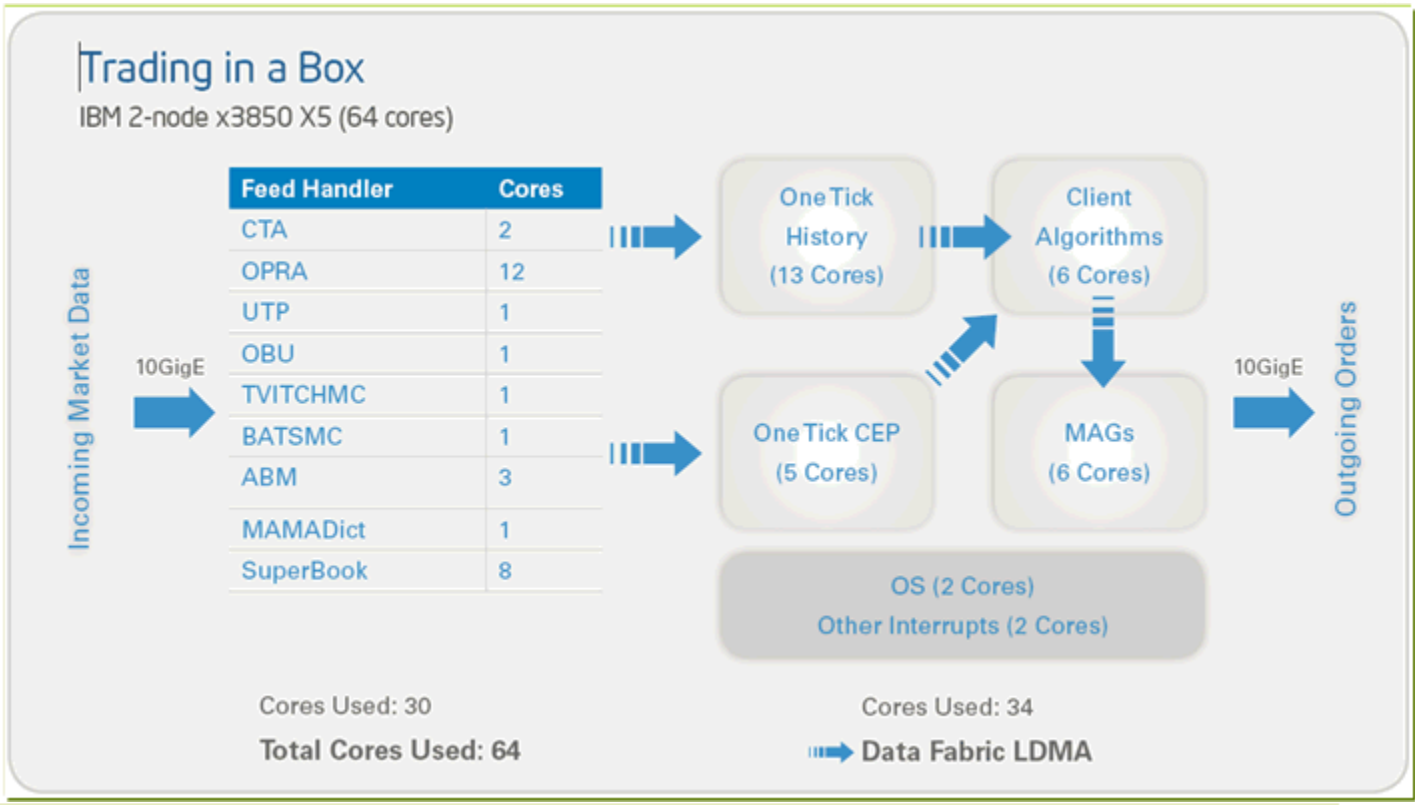
## Top Performing Partner

**One of BMC Software's Top 11 Strategic Partners**

**One of IBM's Top Performing Solution Providers**

**Top 10**
System x Solution Sales

**Top 25**
Power Systems & Storage Solution Sales




## OVERALL RESULTS

| Publisher | Consumer | Maximum throughput (aggregate) | Average latency | 99.9% latency | 99.999% latency |
|---|---|---|---|---|---|
| Constant rate (100,000 msg/s) transport latency | Single consumer | 100,000 msg/s | 0.8 µs | 2.0 µs | 38 µs |
| | 10 consumers | 1,000,000 msg/s | 2.0 µs | 3.5 µs | 25 µs |
| Arrowhead Feed (5,000 packets/s) platform latency | Single Consumer | 3,398 msg/s | 31.8 µs | 95 µs | 217 µs |
| | 6 Consumers | 22,237 msg/s | 34.9 µs | 119 µs | 243 µs |
| Arrowhead Feed (25,000 packets/s) platform latency | Single Consumer | 13,144 msg/s | 29.5 µs | 126 µs | 136 µs |
| | 6 Consumers | 85,588 msg/s | 29.9 µs | 98 µs | 179 µs |

# The IBM eX5 Portfolio

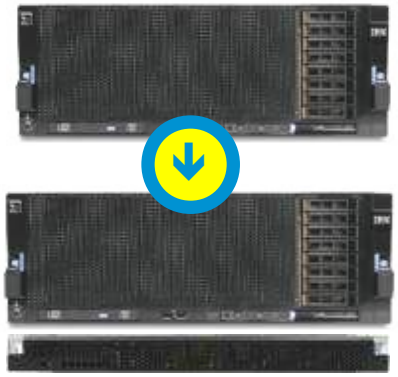
**System x3850 X5**

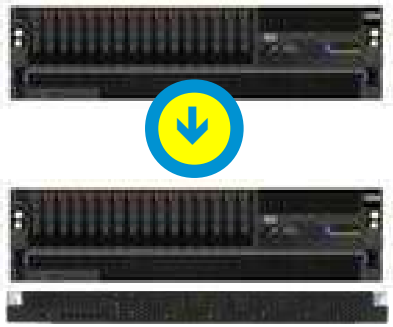
**BladeCenter HX5**

**System x3690 X5**

## MAX5

**Maximum memory scaling independent of processors**





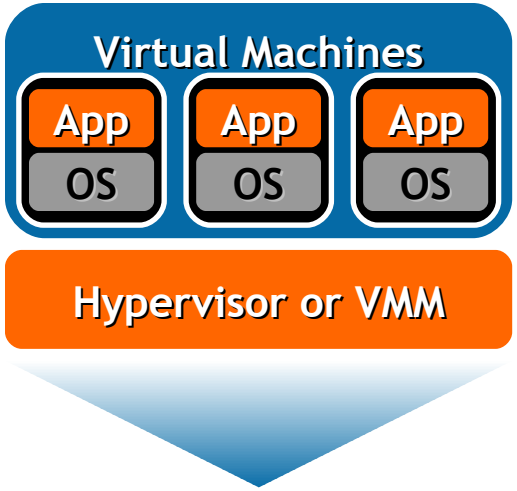
## eXFlash

**Extreme I/O Operations**
**Solid State Drive storage**

# Server Virtualization - Inside Out

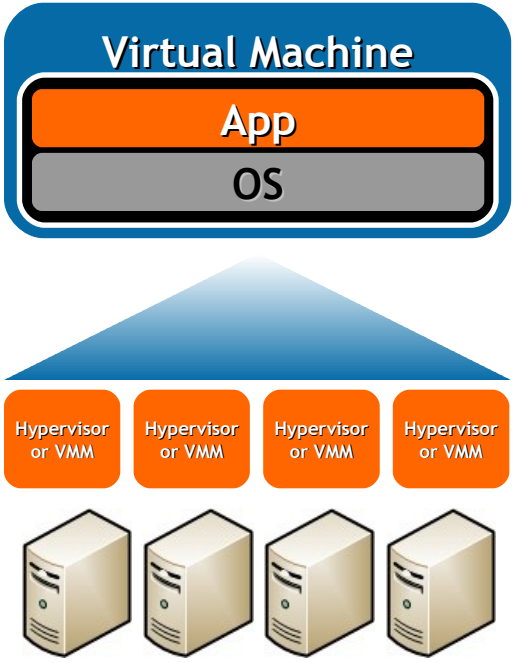
## PARTITIONING

Subset of the physical resource




## AGGREGATION

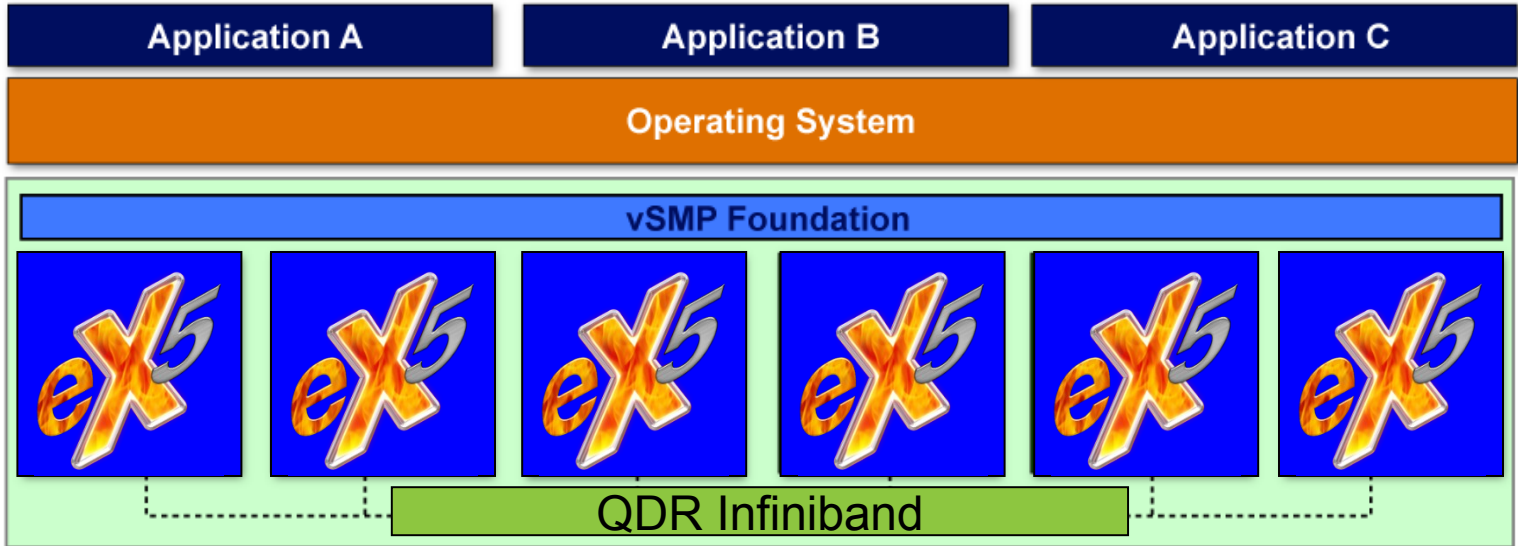
Concatenation of physical resources




ScaleMP™


*Large Memory Workloads*

Maximize Memory
Minimize Cost
Simplify Deployment




**Up to 160 Gbps Interconnect to Each Node**

**Up to 128 Nodes, 8192 Cores, 16,384 Threads, 64TB RAM**

***One System Image***

# Workloads for Large VMs

- *Large memory workloads*
  - Few cores (might be of single node), memory span across nodes
- *I/O intensive workloads*
  - Few cores (might be of single node), I/O span across nodes
  - Memory as a buffer
- *CPU demanding workloads, requiring shared-memory*
  - Threaded applications (OpenMP, Pthreads)
- *Throughput workloads* or multi-process with some communication (simplicity of execution)
  - MPI, *algorithmic trading*, etc.

Example: x3850 X5

- ***512 Cores***
- ***24.5TB Memory***
- ***One OS***

# Vikram Mehta

*Vice President of Systems Marketing*
*IBM Systems and Technology Group*

**“Speed doesn’t kill, being slow does”**

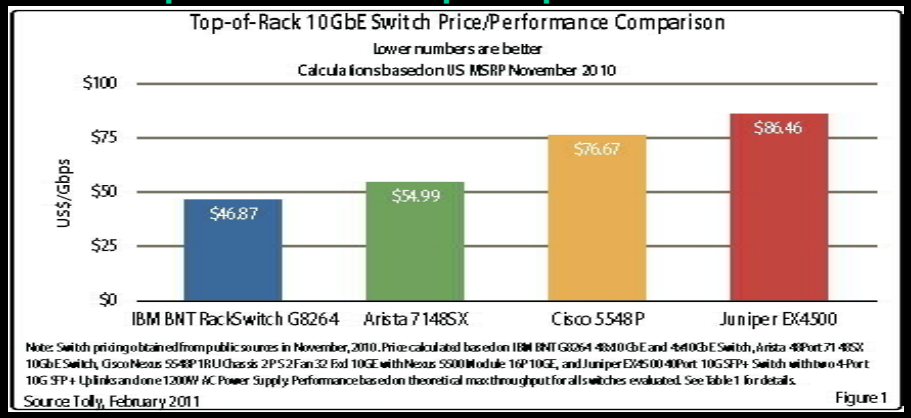
# Why?

=

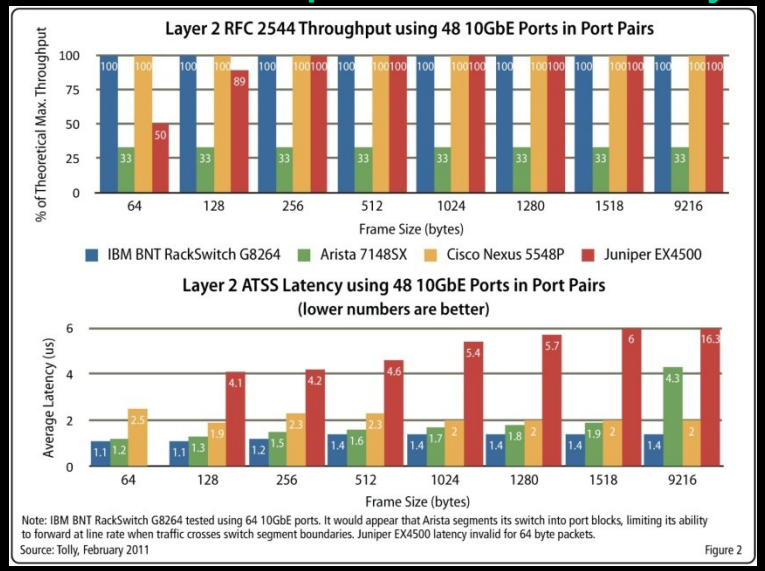
=

# IBM BNT RackSwitch G8264: Proven Performance

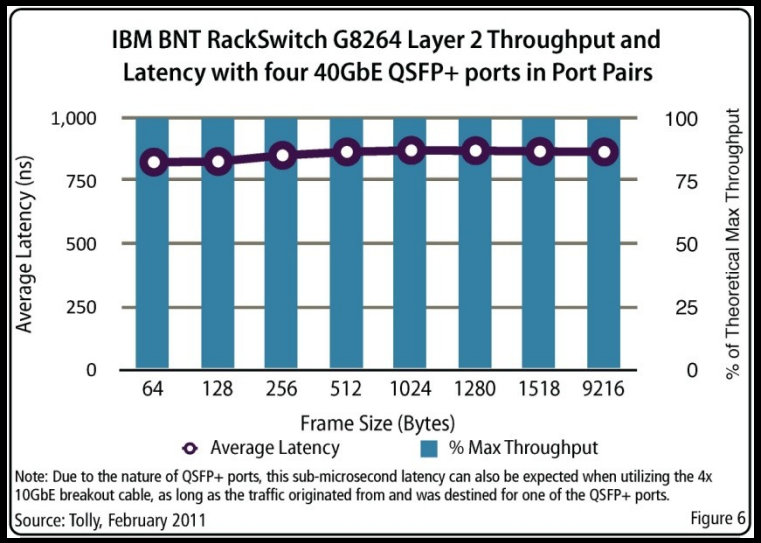
## Up to 84% better price/performance



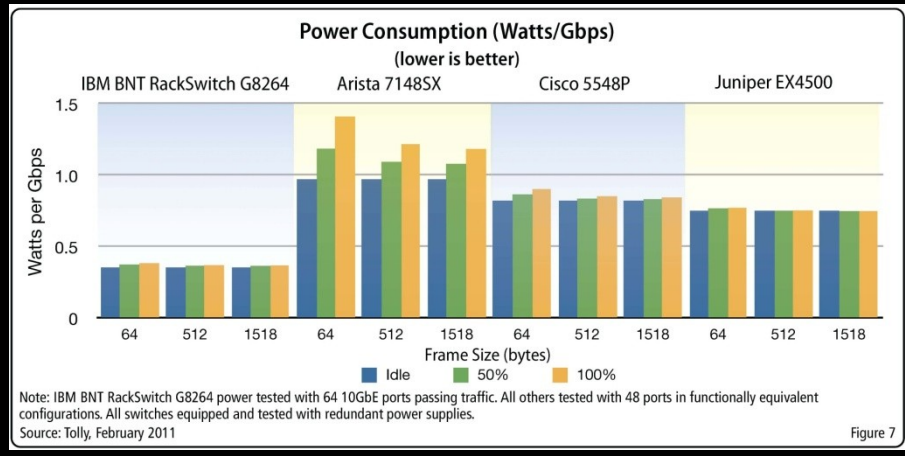

## Line-rate with up to 11.5x lower latency




## Line-rate 40G with sub-microsecond latency



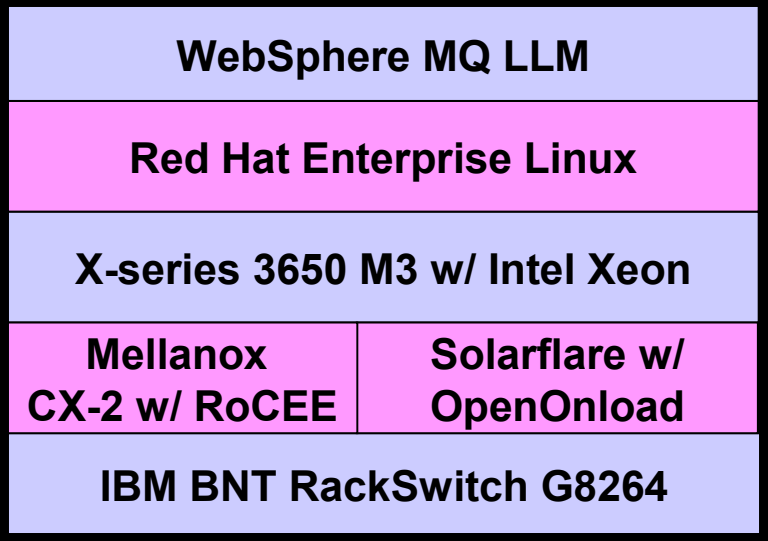

## Up to 71% less power consumption




Source: Tolly Group Competitive Performance Evaluation, #211108, March 2011

# HFT Messaging Solution: IBM BNT G8264 and WMQ LLM

| STAC-M2 Benchmark™ BASELINE Test Comparison (1 Producer, 5 Consumers) | Avg (µs) | Max (µs) | Std Dev (µs) | Highest Rate (Msg / Sec) |
|---|---|---|---|---|
| IBM LLM / IBM-BNT 8264 / Solarflare w/ OpenOnload | 9 | 21 | 0 | 1.5 Million |
| 29W LBM / Cisco N5010 / Solarflare w/ OpenOnload | 14 | 33 | 1 | 1.3 Million |
| 29W LBM / Cisco 4900M / Solarflare w/ OpenOnload | 15 | 30 | 1 | 1.3 Million |




- IBM BNT RackSwitch G8264 with LLM delivers the best (STAC™ Published) 10 GbE performance
  - Extremely Low Mean Latency 9 µSec
  - Deterministic Performance – Near Zero Jitter
  - Highest Supply Rate – 1.5 Million msg / sec

- IBM and its partners have demonstrated ultra low latency messaging solutions. IBM Offers:
  - IBM BNT G8264 10 / 40 GbE High Perf switch
  - IBM WebSphere MQ Low Latency Messaging
  - Choice of High Performance Network Adapters
    - Mellanox ConnectX-2
    - Solarflare 10GbE w/ OpenOnload
  - High Performance X-series servers

# Reflector Test Latency: IBM BNT RackSwitch G8264 and WMQ LLM

| LLM Latency using IBM BNT G8264 10/40 GbE and Solarflare SFN5122F OpenOnload | | | |
|---|---|---|---|
| Msg Rate [msgs/sec] | Single Hop | | RTT |
| | Average [µsec] | 99P [µsec] | Std Dev [µsec] |
| 10,000 | 5.95 | 6.5 | 0.80 |
| 100,000 | 6.24 | 6.5 | 0.83 |
| 1,000,000 | 8.72 | 10.5 | 1.43 |

| LLM Latency using IBM BNT G8264 10/40 GbE and Mellanox CX-2 RoCEE | | | |
|---|---|---|---|
| Msg Rate [msgs/sec] | Single Hop | | RTT |
| | Average [µsec] | 99P [µsec] | Std Dev [µsec] |
| 10,000 | 3.6 | 4.5 | 0.7 |
| 100,000 | 3.6 | 4.5 | 0.9 |
| 1,000,000 | 4.3 | 5.5 | 2.2 |

**Additional IBM Reflector Tests show a 10GbE solution from IBM and its partners delivers extremely low latency performance that scales to very high message rates.**

# Are you next?

- 65% of sell-side firms are in the process of upgrading to 10GE for their US equity business; 13% are fully converted.

- - TABB Group 2010

# Panelist Q & A

# Legal



**IBM Corporation**
**New Orchard Road**
**Armonk, NY 10504**
**U.S.A.**

**Produced in the United States of America**