

Dr. Abdel Labbi
IBM Research - Zurich

From Big Data to Smarter Decisions

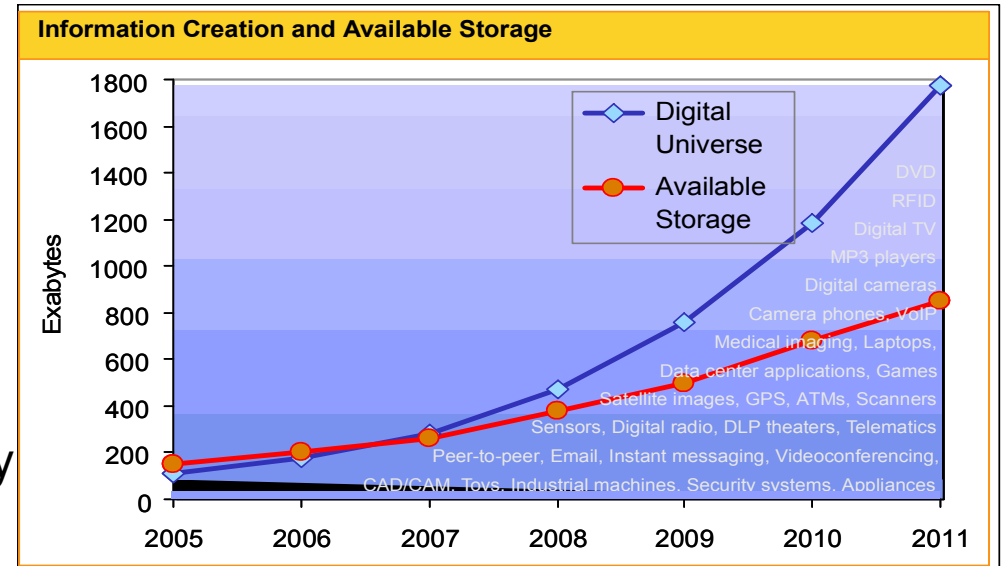


Dealing with the Information Deluge ...

Moore's Law:

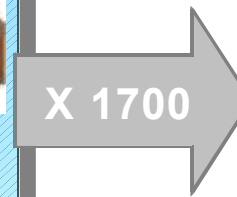
Every 18 months:

- New computing power = Σ existing power
- New storage capacity = Σ existing capacity



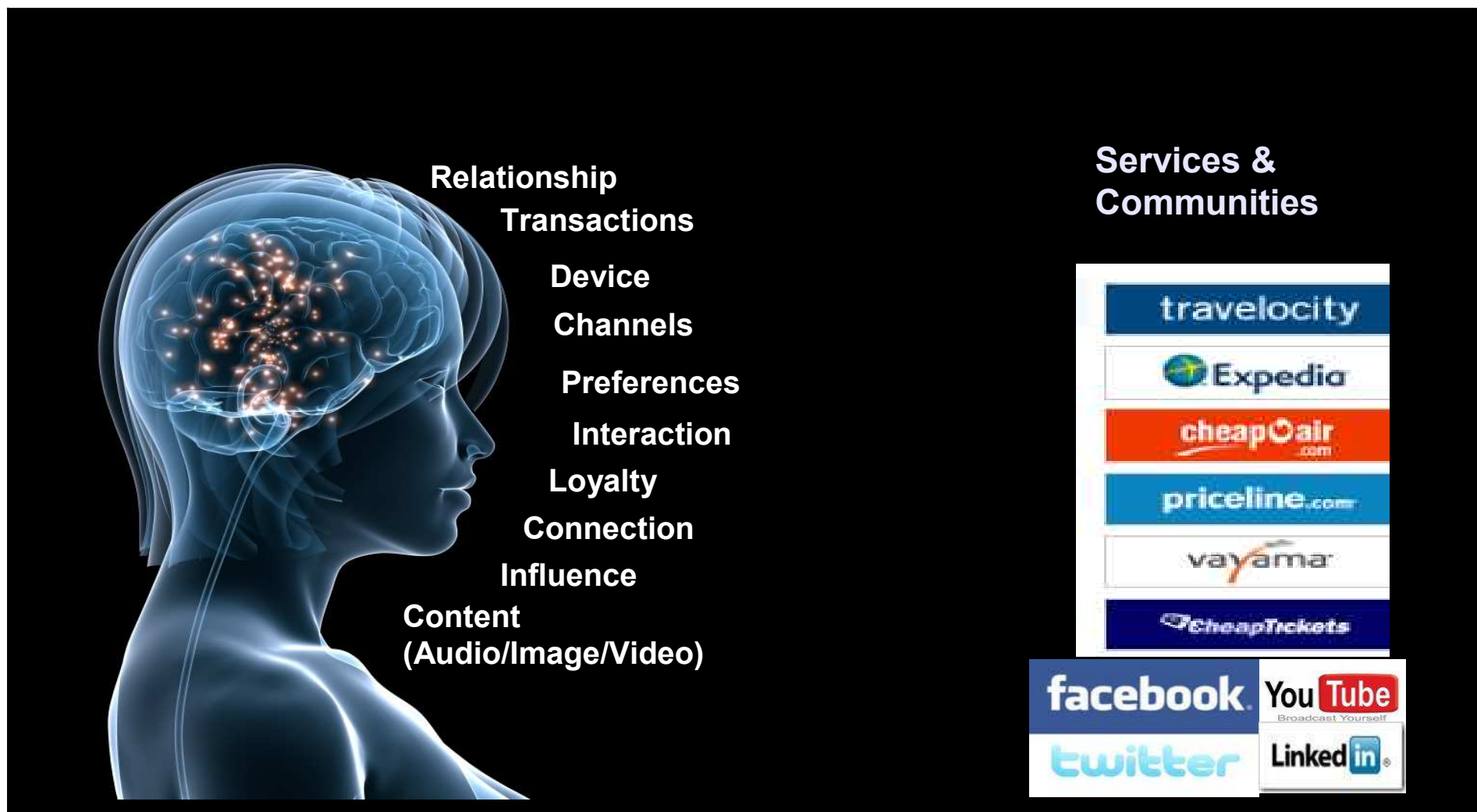
One Teraflop
is equivalent to:

- 287 Pentium 4s
- 10'000 Cray 1s
- 5'000 iPods
- 1 Game Console



IBM Roadrunner: World's Fastest Computer
1.7 Petaflops

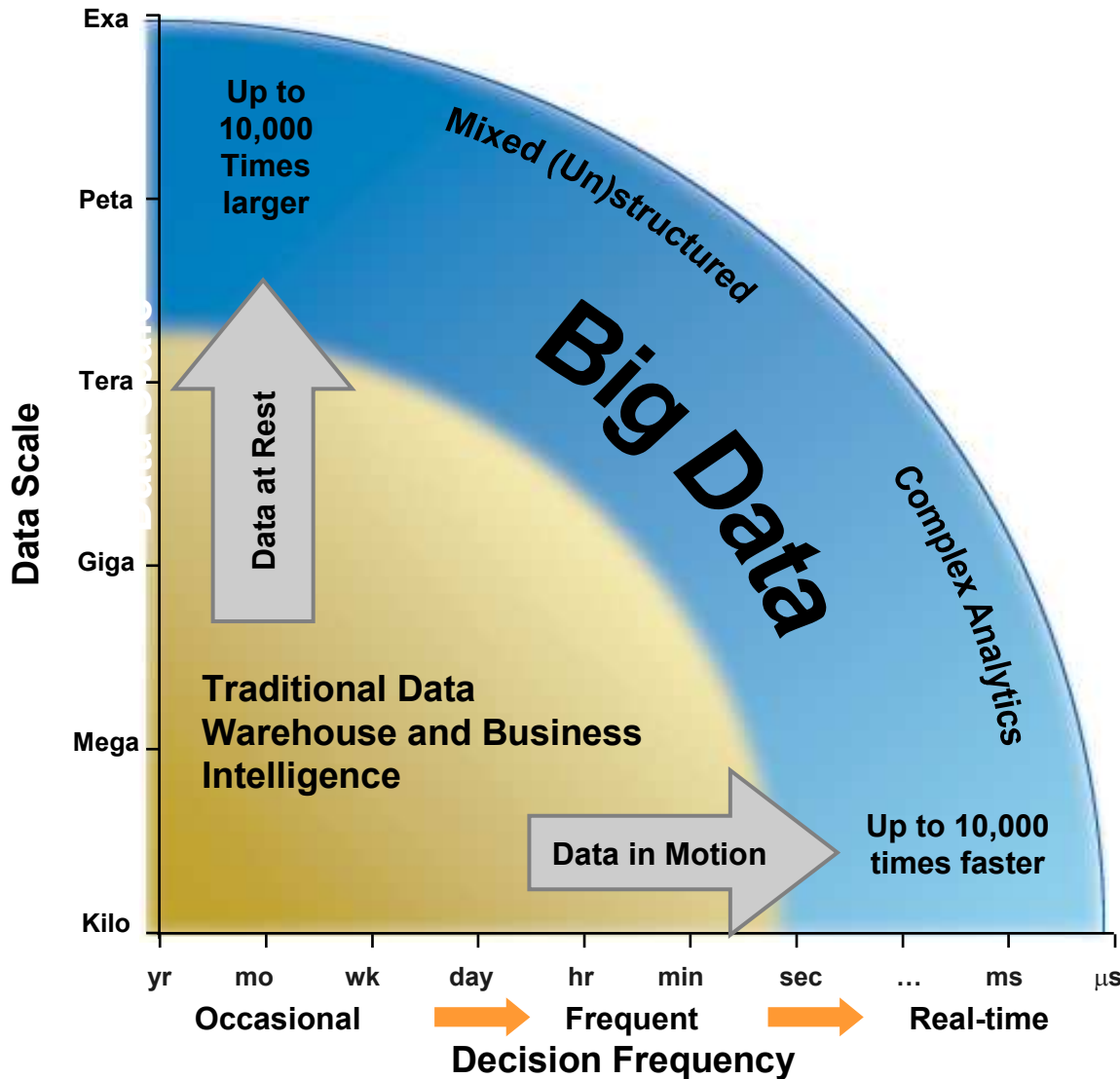
Dealing with Information Transparency ...



Advanced Analytics for the Masses ... against the Few!

The screenshot shows a Bing flight search for New York, NY (NYC) to Frankfurt, Germany (FRA) for the dates Mon, 6/11 - Mon, 6/18, 1 adult, Economy class. The current price is \$870. A 'PRICE PREDICTOR' overlay is present, showing a 'Tip: Wait' with a 75% confidence level, indicating fares are dropping or steady. The overlay also lists two scenarios: 'If We Are Right' (Lowest fares will stay within \$15 | 44% confidence and Lowest fares will drop \$31 - \$155 | 31% confidence) and 'If We Are Wrong' (Lowest fares will rise \$20 - \$68 | 25% confidence). A '7-day low fare prediction' box provides a tip to wait, noting that lowest fares are likely to drop or hold steady within the next 7 days, but price drops are sporadic and 50% do not last longer than 48 hours. A 'daily low fare history' chart shows the lowest fare observed each day from Dec 16 to Now, with a 120-day low of \$734 and an average low of \$881. The chart shows significant volatility, with a peak near \$1100 in late December and a low near \$700 in early January.

“Big Data” Requires New Analytics



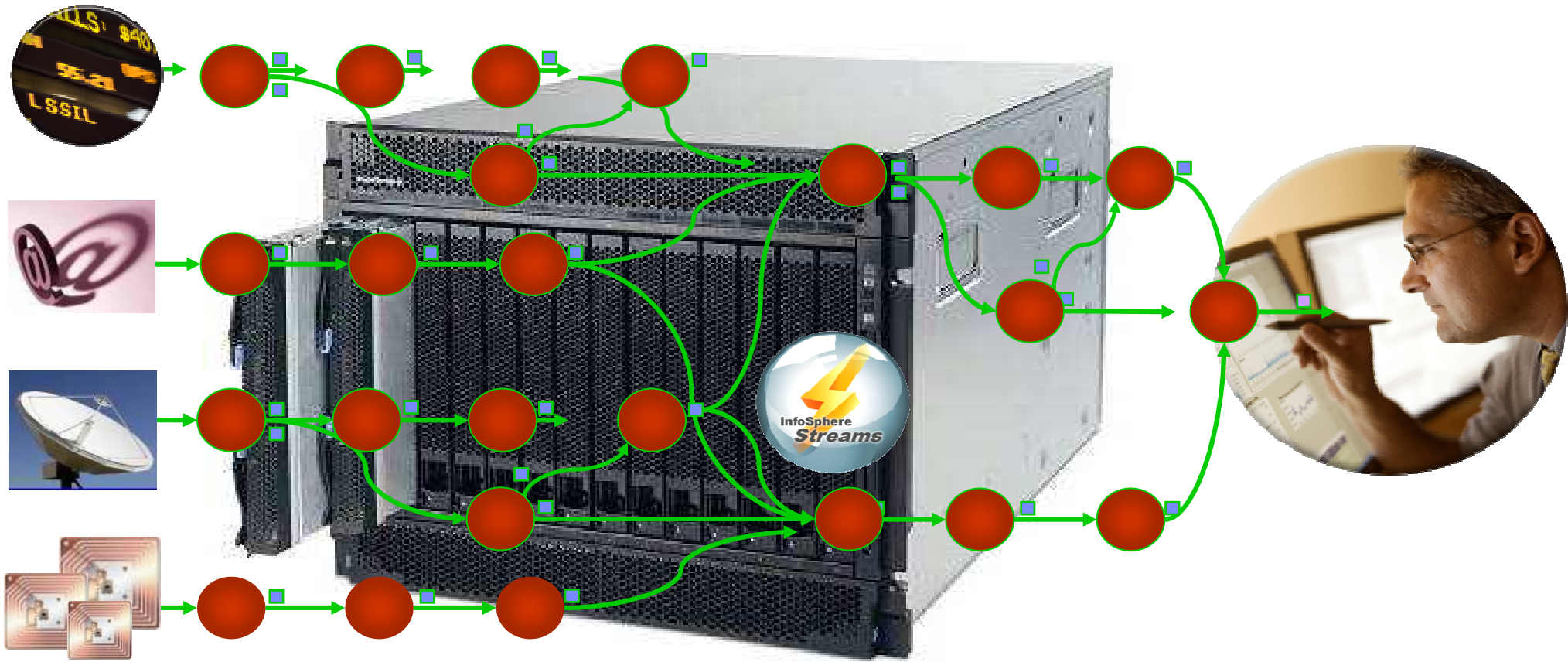
Telco Promotions
100,000 records/sec, 6B/day
10 ms/decision
270TB for Deep Analytics



Smart Traffic
250K GPS probes/sec
630K segments/sec
2 ms/decision, 4K vehicles

Stream Computing for Event Processing

➡ Continuous Ingestion ➡ Continuous Queries /Analytics on data in motion



Building next generation trading platforms ...



- ▶ Identify and execute trades
- ▶ Process over 5M events per second with average latency of 150 microseconds
- ▶ Expand to incorporate content feeds, news text, audio, video, to establish greater context for better decisions



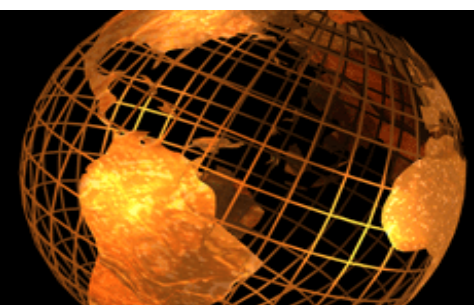
CIO TD Bank "TD Bank Financial Group worked with IBM Research to develop a first-of-a-kind architecture capable of consuming, analyzing and acting on real-time market data while maintaining sub-millisecond response times even under extreme data loads"

From Corporate to Industry Information Platforms ...

www.orx.org

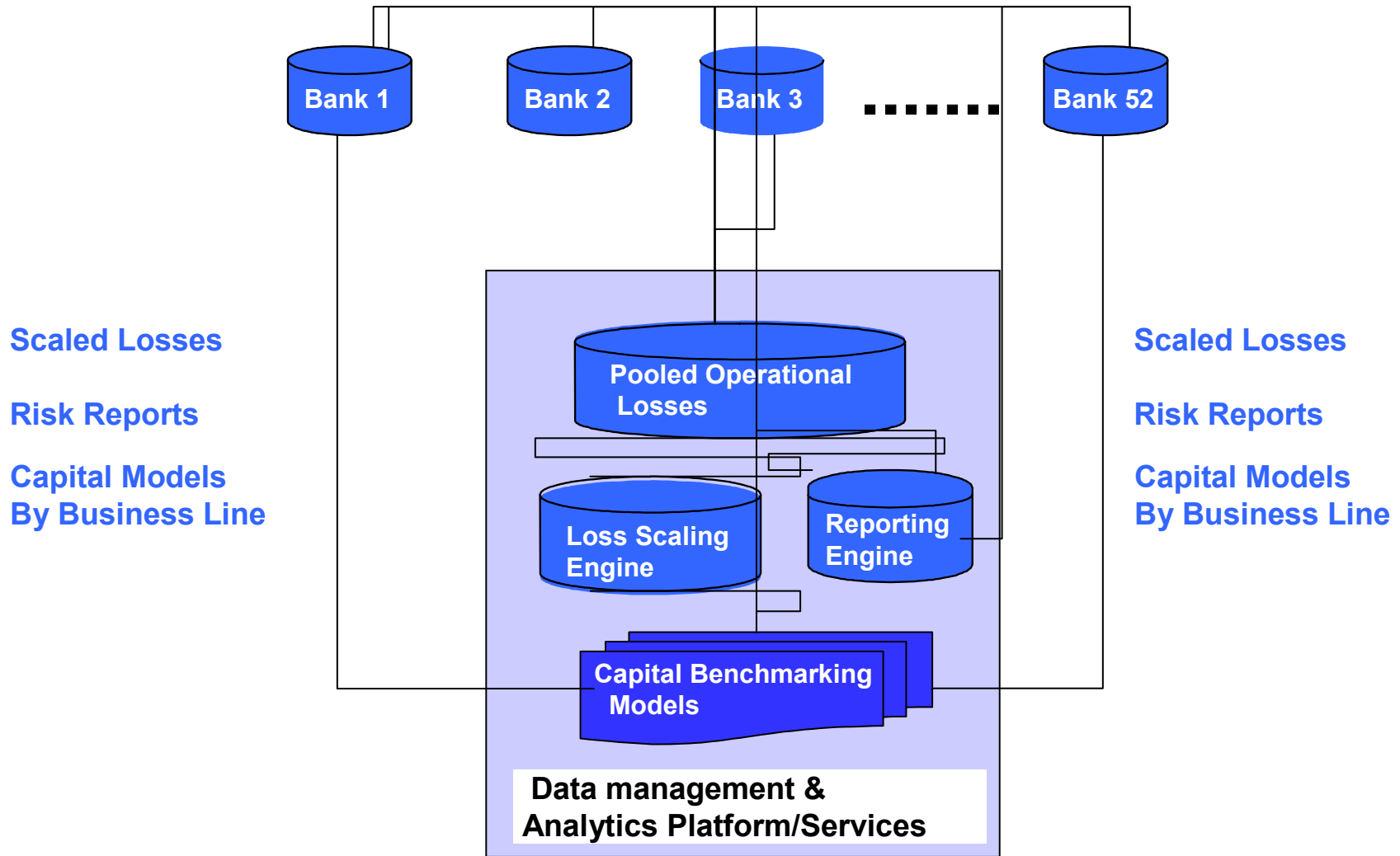
The Operational Riskdata
eXchange Association

An association of 50 leading
banks from 18 countries...



- ▶ ABN AMRO
- ▶ Banc Sabadell
- ▶ Banco Bilbao Vizcaya Argentaria
- ▶ Banco Pastor
- ▶ Banco Português de Negócios
- ▶ Bank Austria - Creditanstalt
- ▶ Bank of America
- ▶ Bank of Nova Scotia
- ▶ Barclays Bank
- ▶ BMO Financial Group
- ▶ BNP Paribas
- ▶ Caixa Catalunya
- ▶ Caja Laboral
- ▶ Cajamar
- ▶ Commerzbank AG
- ▶ Credit Agricole
- ▶ Danske Bank A/S
- ▶ Deutsche Bank AG
- ▶ Dresdner Bank AG
- ▶ Erste Group Bank AG
- ▶ Euroclear Bank
- ▶ Fortis
- ▶ Grupo Banco Popular
- ▶ Grupo Banesto
- ▶ Grupo Santander
- ▶ Hana Bank
- ▶ HBOS plc
- ▶ HSBC
- ▶ ING
- ▶ Intesa Sanpaolo
- ▶ JPMorganChase
- ▶ Lloyds TSB
- ▶ National City Corporation
- ▶ Northern Trust
- ▶ Rabobank
- ▶ RBC Financial Group
- ▶ Royal Bank of Scotland
- ▶ Skandinaviska Enskilda Banken AB
- ▶ State Street
- ▶ TD Bank Financial Group
- ▶ US Bancorp
- ▶ Wachovia Corporation
- ▶ Washington Mutual
- ▶ WestLB

The ORX Operational Risk Analytics Platform



Challenges in Social Media Analytics

Identify Viral Potential

Who are the most influential people and how do they spread information

Discovering Topics

What are travelers talking about ...

Causal Influence

Who is influenced by who..



Sentiment Analysis

What are people's attitudes towards airline ...

Profiling Influencers

What are influencers profiles, status, value, hobbies, jobs,...

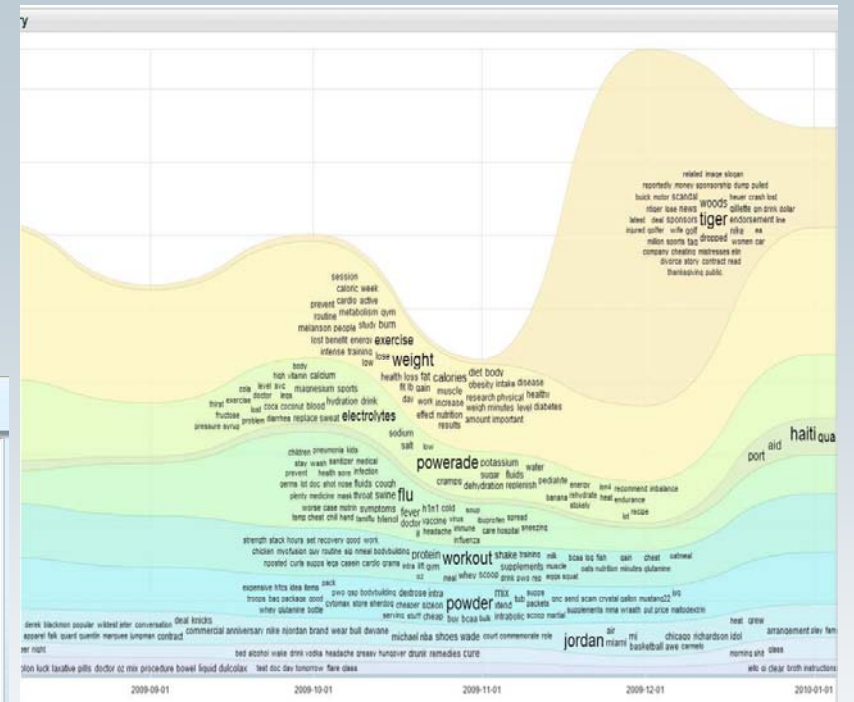
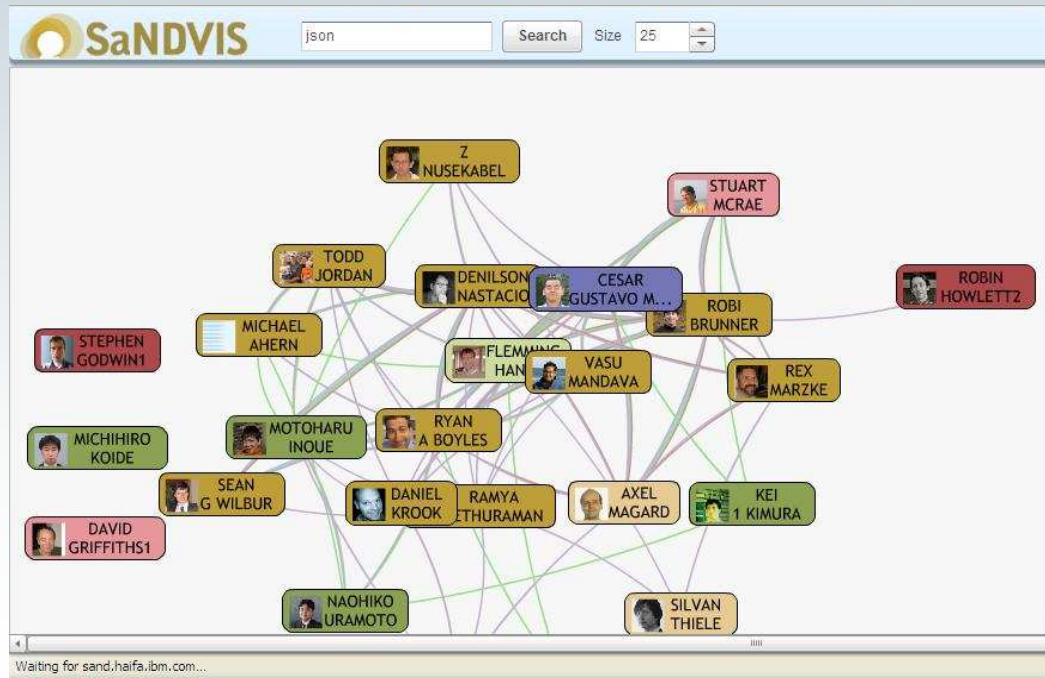
Automatically Discover Emerging Topics and Key Influencers

Identify emerging trends

- Process hundreds of millions of topic-document pairs

Identify key influencers

- Analyze millions of relationships



Precision, Confidence, and Speed – What’s Next?

The image shows a game show set with two contestants at podiums. The central screen displays a brain icon. The podiums show scores and handwritten questions. A data bar at the bottom shows the percentage of correct answers for three options: Emily Dickinson (99%), Walt Whitman (60%), and Barnard (10%).

Contestant	Score	Question	Answer
Left	\$24,000	Who is Stoker? (FOR ONE WELCOME OUR NEW COMPUTER OVERLOADS)	\$1,000
Center	\$77,147	Who is Bram Stoker?	\$17,973
Right	\$21,600	WHO IS BRAM STOKER?	\$5600

Option	Percentage
Emily Dickinson	99%
Walt Whitman	60%
Barnard	10%

The Future of Smarter Systems

Extend Watson-type of technology

- Moves beyond “question-in & answer-out” to always “learning” evidence-based decision support

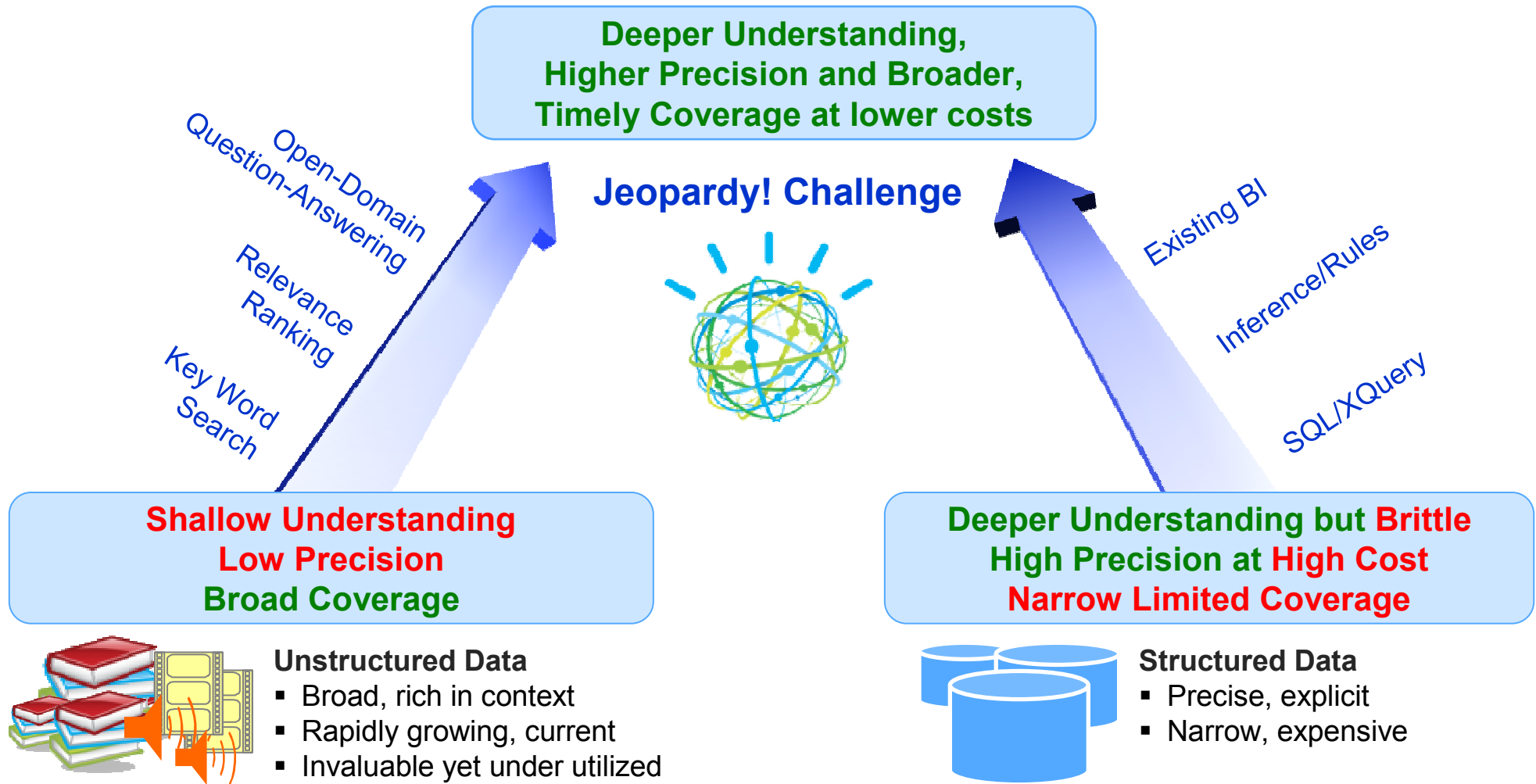
Lead in new domains

- Addresses the enterprise need to convert growing volumes of information into actionable knowledge
- Demonstrates business value in critical problem spaces, starting with Healthcare

Enable efficient adaptation

- Efficiently adapting and scaling Watson to new domains requires a novel blend of engineering and research

Smarter decisions over (un)structured content ... in real time!





Thank You!

