IBM Smarter Business and Technology Series

A headstart for the next decade

November 25, 2010

Ruby Ballroom, OneWorld Hotel

Building Smarter Analytics Information-in-Motion with IBM

Andrew Lim

Client Technical Professional

IBM

Agenda

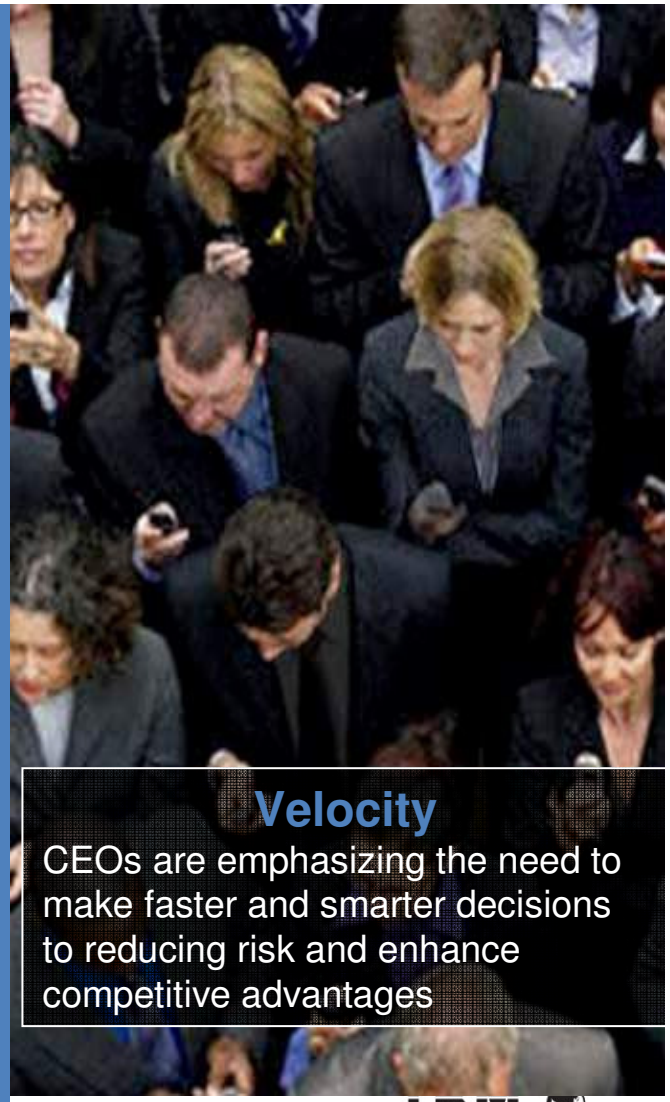
- Need for a smarter solution
- Product Overview
- Customer case studies
- Q/A

Volume

Every day, 15 petabytes of new information are being generated. Data volume are expected to double every 2 years.

Variety

80% of new data growth is from non-relational and non traditional data types like emails, documents, RFID feeds, multimedia etc



Velocity

CEOs are emphasizing the need to make faster and smarter decisions to reducing risk and enhance competitive advantages

The explosion of events requires extreme innovation

Traditional Computing

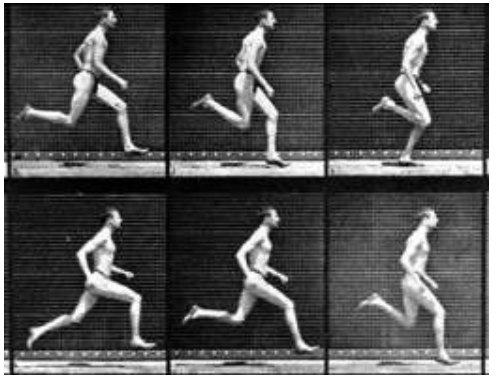
Historical fact finding with data-at-rest

Batch paradigm, pull model

Query-driven: submits queries to static data

Relies on Databases, Data Warehouses

Stream Computing



Real time analysis of data-in-motion

Streaming data

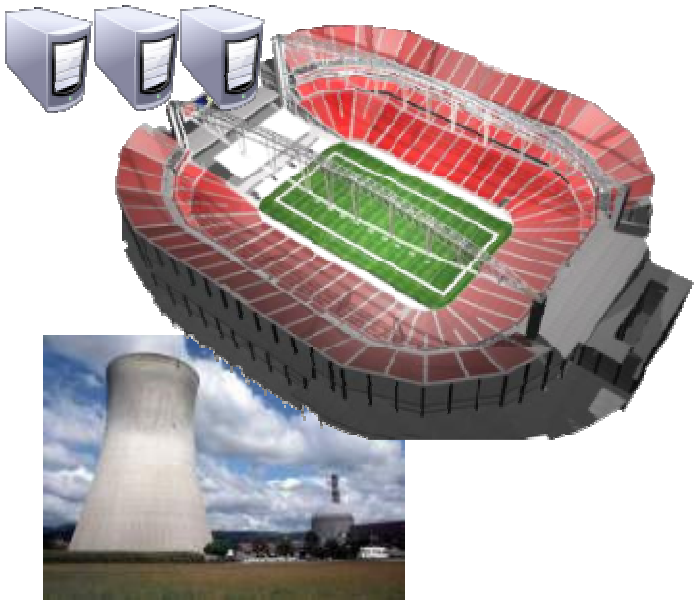
A stream of structured or unstructured data-in-motion

Stream Computing

Analytic operations on streaming data in real-time

Where incremental improvement is insufficient extreme innovation is required

When incremental improvement is impractical....



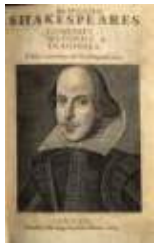
Invention is required

To match with performance of Blue Gene with traditional technology, the servers would occupy the space of several stadiums and required the power of a nuclear power plant.

Where incremental improvement is insufficient extreme innovation is required

When incremental improvement is impractical....

A trader would need to…



Read the entire works of Shakespeare 10 times in one second

and

Identify and execute a trade faster than a hummingbird wingflap

Invention is required

… to match the performance of
Algorithmic Trading with Stream Computing

For US equity electronic trading brokerage
1 millisecond = $4M in annual revenue

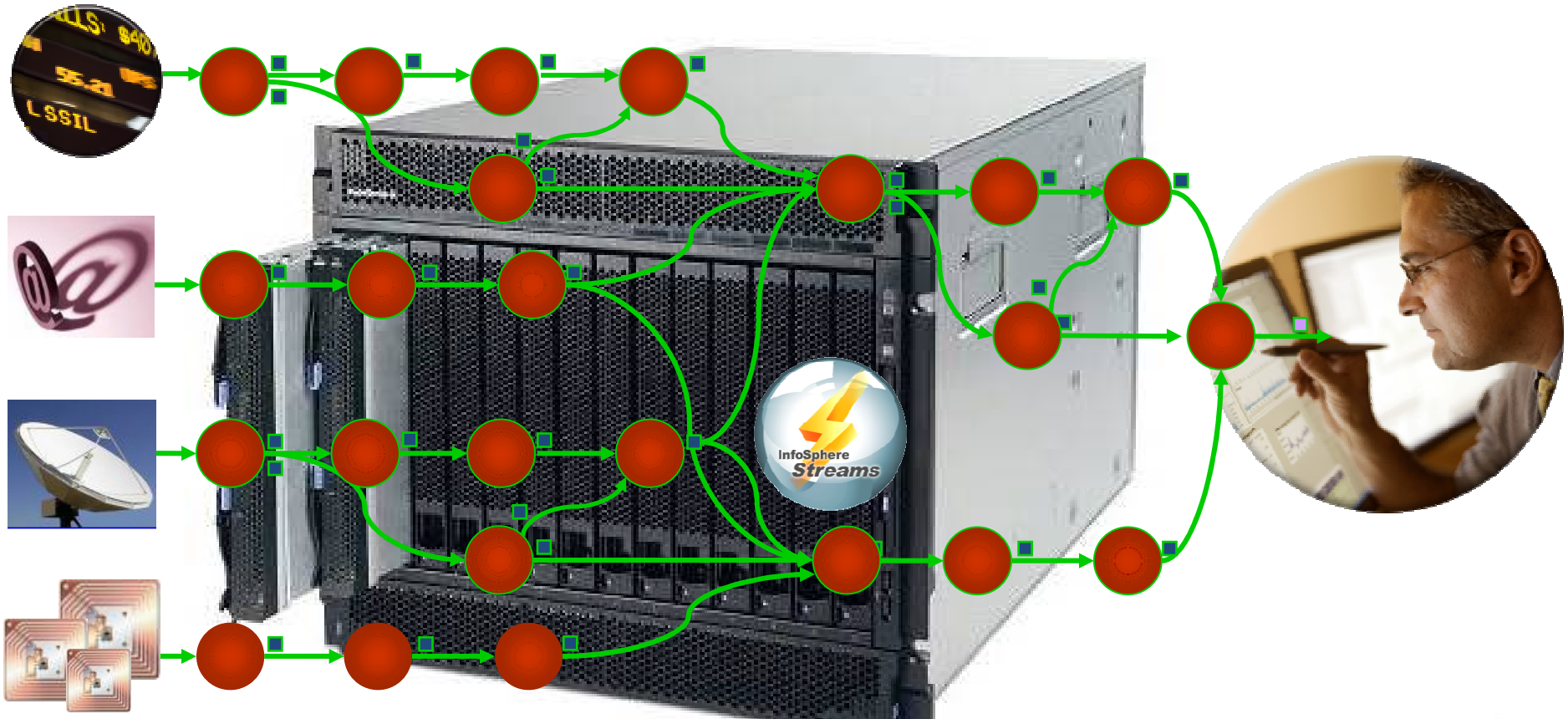
Source: Tabb Group

Extreme Computing for Extreme volumes and speeds

Stream Computing:

A new paradigm for ultra low latency and high throughput in-motion analytics

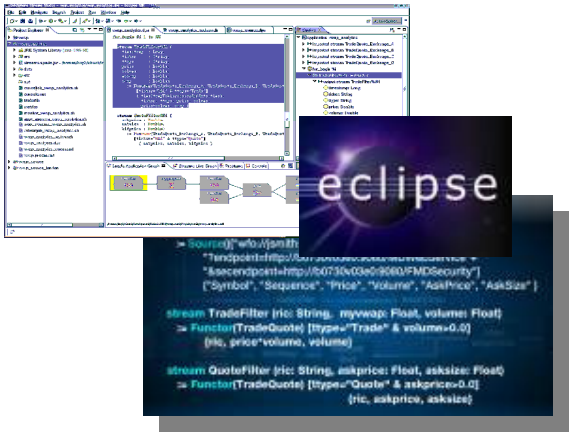
Continuous Ingestion → **Continuous Queries /Analytics on data in motion**



IBM

The InfoSphere Streams 1.2 Platform

Development Environment

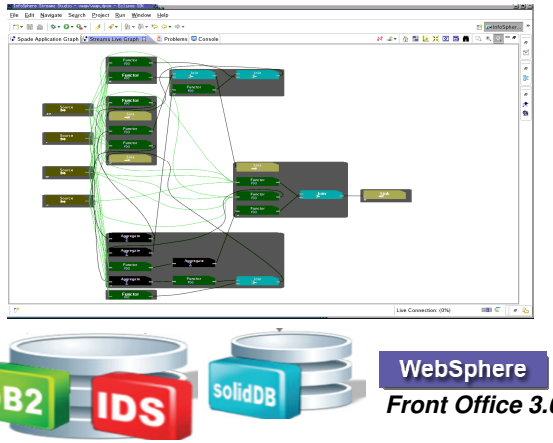


Streams Studio
Eclipse IDE for SPADE

Runtime Environment

Scalable stream processing runtime

Tools and Technology Integration



Front Office 3.0

Streamsight,
Built-in Stream Relational Analytics,
Adapters
Toolkits

Supported on x86 hardware, RedHat Enterprise Linux 5.3 and 5.4

Key technical capabilities of Streams

Language built for Streaming applications:

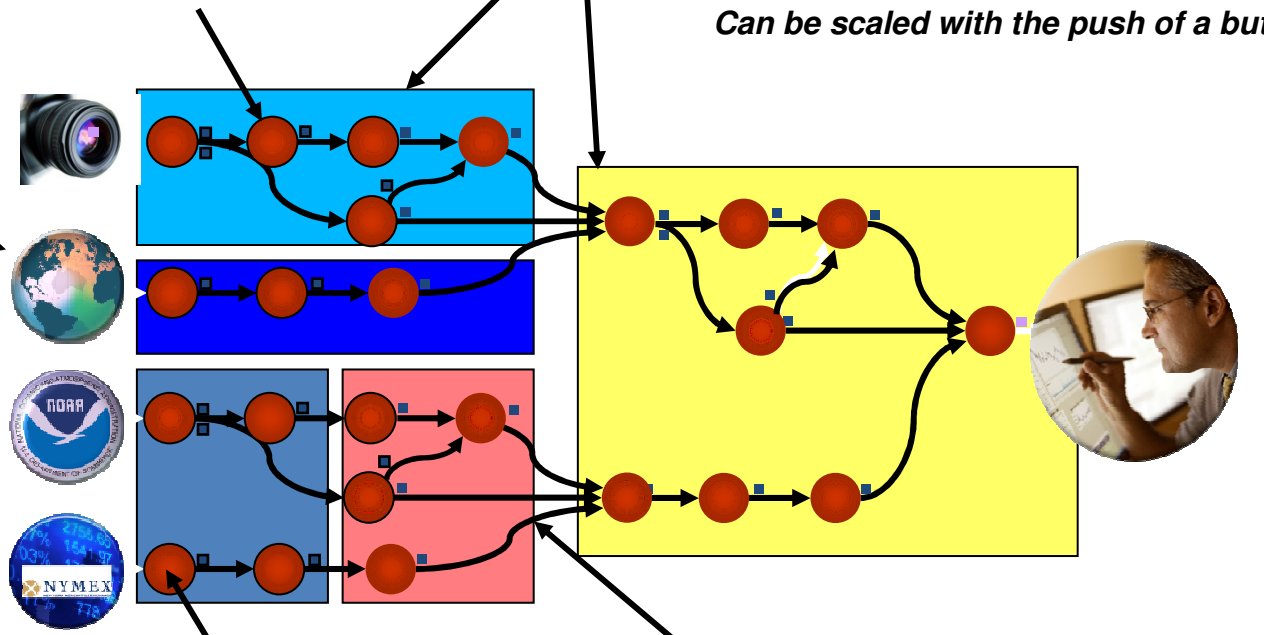
Reusable operators
Rapid application development
Continuous "pipeline" processing

Compiling groups of operators into single processes enables:

Efficient use of cores
Distributed execution
Very fast data exchange
Can be automatic or tuned
Can be scaled with the push of a button

Use the data that gives you a competitive advantage:

Can handle virtually any data type
Use data that is too expensive and time sensitive for traditional approaches



Easy to extend:

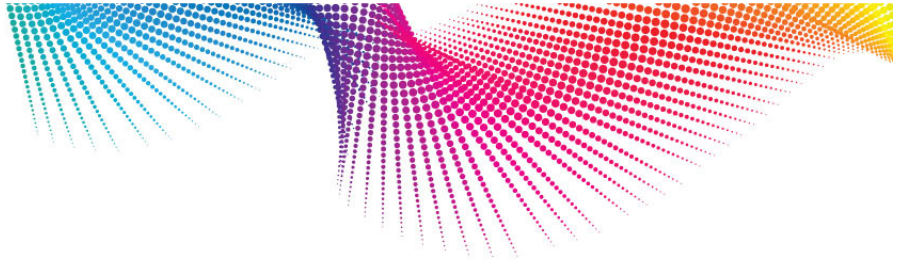
Built in adaptors
Users add capability using familiar and Java

C++

Extremely flexible and high performance transport:

Very low latency
High data rates

IBM Smarter Business and Technology Series

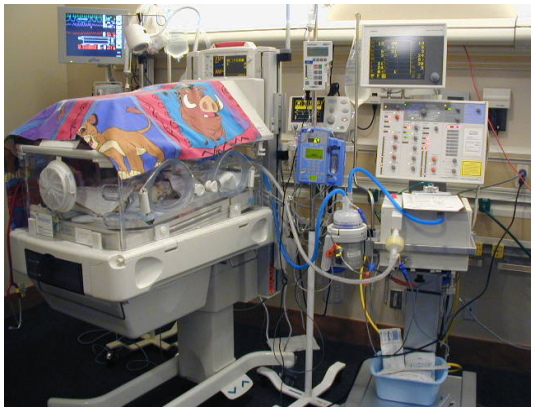


Customer Case Studies

IBM

‘Smart’ applications of InfoSphere Streams are emerging quickly

Neonatal Care



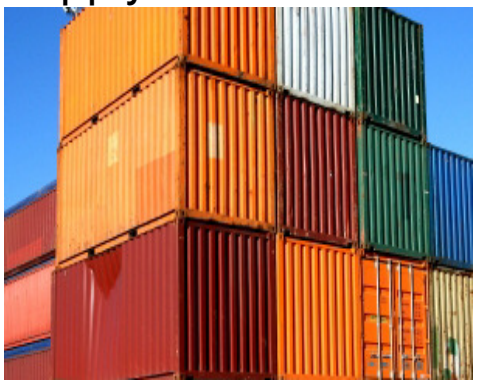
Trading Advantage

Environment

Law Enforcement

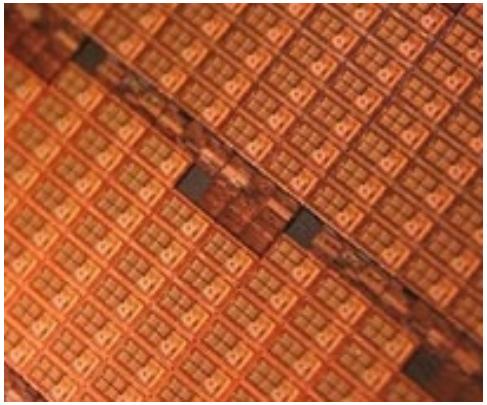
Radio Astronomy

Supply Chain



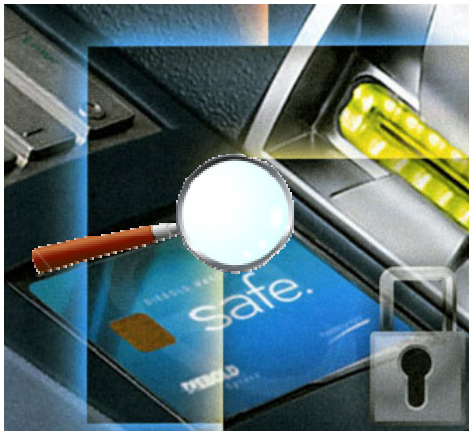
'Smart' applications of InfoSphere Streams are emerging quickly

Manufacturing

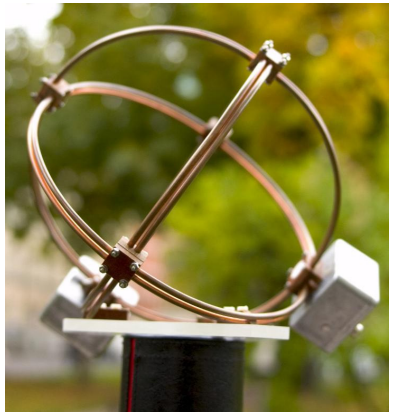


Smart Traffic

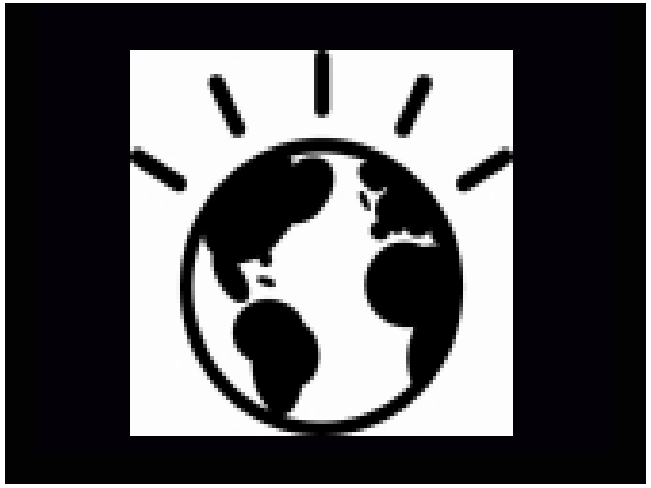
Fraud Prevention



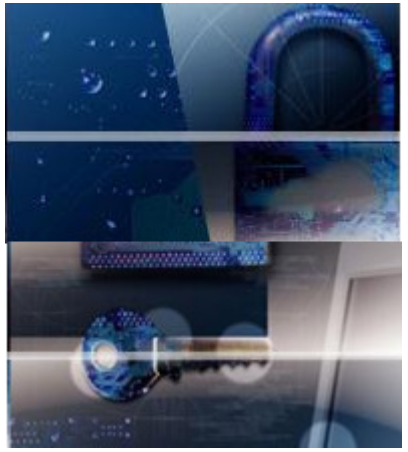
HF Radio



Smarter Way



Cyber Security



TD Bank

TD Bank Financial Group

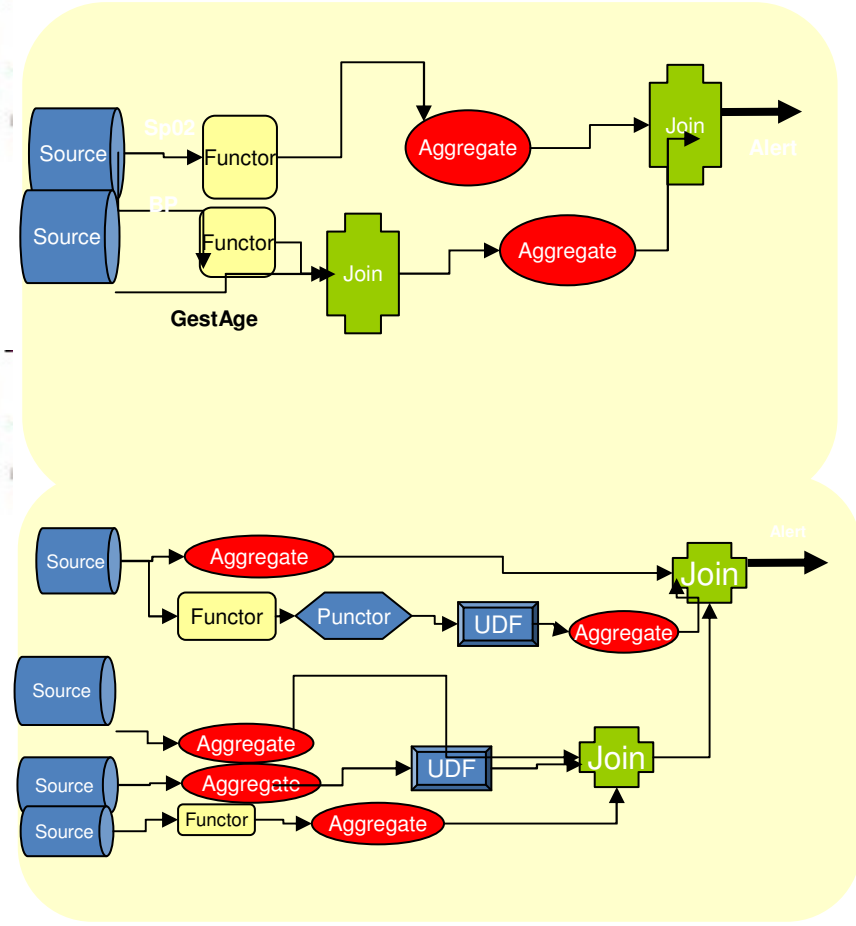
- Identify and execute trades
- Process over 5M events per second with average latency of 150 microseconds
- Expand to incorporate content feeds, news text, audio, video, to establish greater context for better decisions

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"***

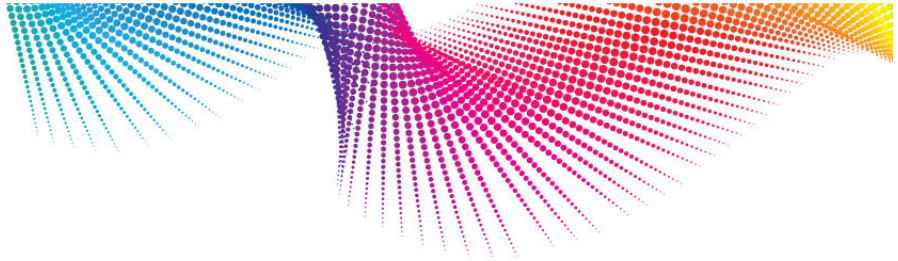
University of Ontario Institute of Technology

A Research Project to monitor premature infants at ICU

- Correlating SpO2 with Mean ABP to predict "Baby crashing"
  - Sp02<85%
  - BP< GestAge for 20 sec
- Nosocomial Infection Prediction
  - Monitoring HR variability
  - Monitoring CIS info
  - **Fusing data to predict sepsis**
  - **6 to 24 hours earlier than experienced ICU Nurses**



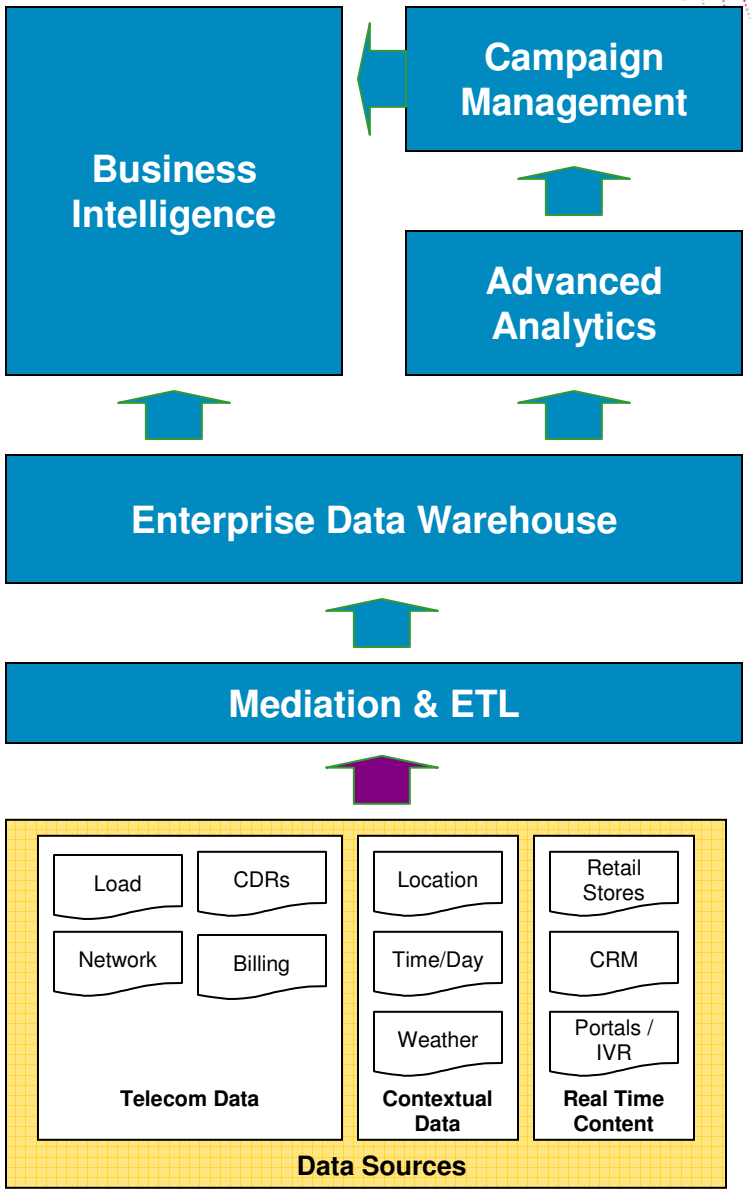
IBM Smarter Business and Technology Series



Smarter Promotions

IBM

Traditional Model of Data Collection to Marketing & Business Functions Execution



High Amount of Calls

Good customer ! does high amount of Calls

Notifies Customer

Analyze daily, weekly, monthly call volumes and generates trigger points

Auto Activate Promo
Free Landline Minutes
Unlimited calls to Landline ###
Free Trial for X Days

Dynamic Plan allocation & settings

IBM

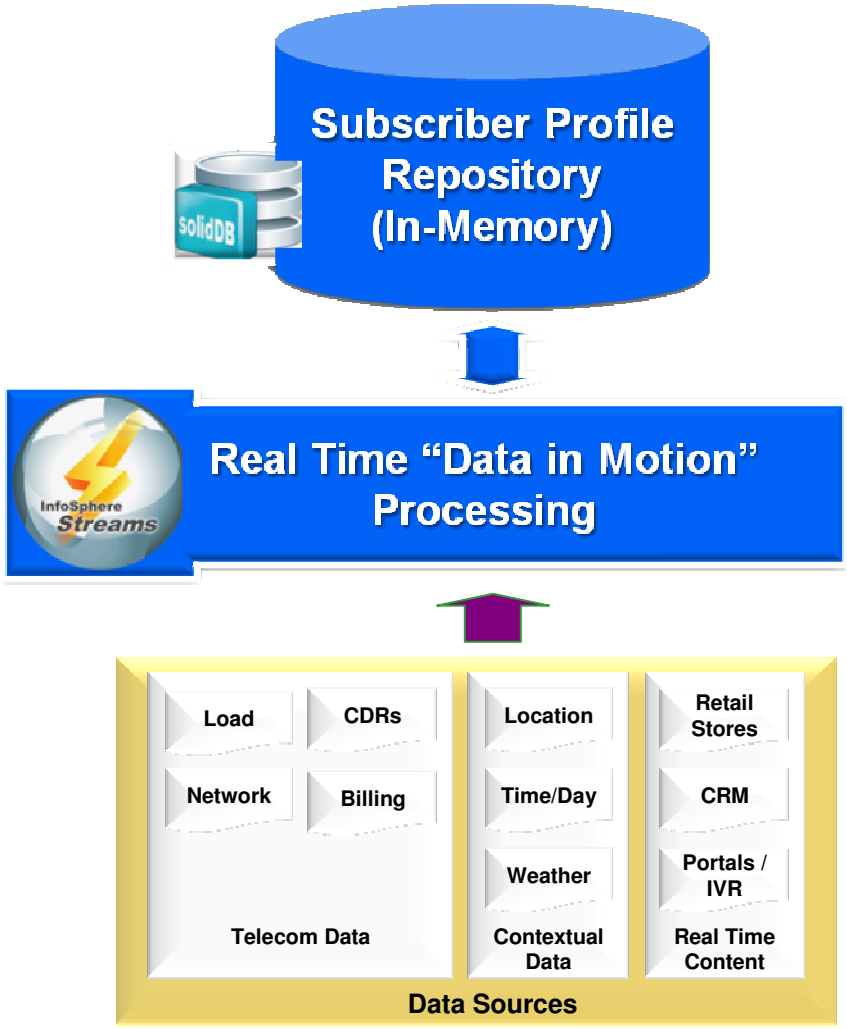
Real Time Processing & Execution

I hate Call drops # this is 5th in last 1 hour

We are sorry! We detected call drops on your connection, pls enjoy 20% discount on calls During the rest of the day

correct personalized / targeted response that will be executed in real time with retention promos

Real Time Processing for Campaign Execution



Event ID	5th Drop
MSISDN	60129844662
CustomerProfile	Standard
SubStatus	Active
SubStatusDuration	24
Location	OneUtama
AvgDailySpend	15
Longevity	24
PhoneType	iPhone 3G
Balance	200
Daily Voice MOU	5
Daily SMS Count	20

Gaining tactical agility with "Campaign Accelerator"

Challenge

Globe Telecom, the number two provider of mobile communications services in the Philippines, realized that it needed to reach a new level of agility in the creation and management of promotional service offerings, if the company was to thrive in its intensely competitive market.

Solution

IBM designed and built a SOA-based Smarter Promotion execution platform that enables Globe to rapidly and cost-effectively create marketing promos from reusable service components and automate the process. Hundreds of simultaneous targeted promos, enabled by the integration of customer intelligence, behavior segmentation, profit optimization and promotion execution drive revenue improvement and are delivered through an integrated and automated solution.

IBM® Service Provider Delivery Environment (SPDE) Framework
IBM WebSphere®
IBM Tivoli®
IBM Rational®
IBM Information Management®

What Makes it Smarter

Globe

Leveraging information gathered from customer & network activity, Globe is able to identify the optimal service promotion for each customer—and the best time to deliver it.

Less than one-year payback period and more than 95% reduction in time and cost of developing new promotions

600% increase in promotion effectiveness

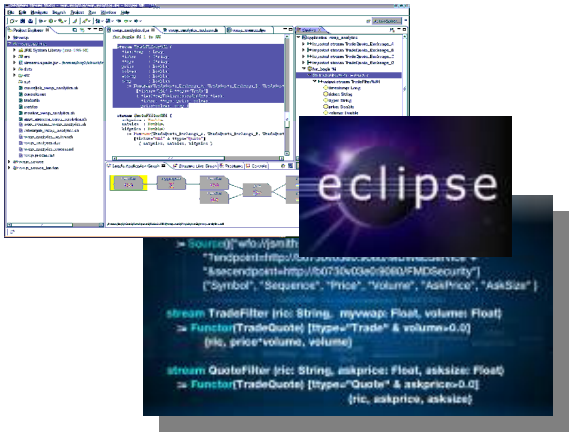
"We can react very quickly to promotional opportunities when they arise. Just as important, we can detect in near real time whether the mechanics of our promotion are working—and if they're not, we can change them almost instantly."

— Mario Domingo, Head of Product Design and Creation, Globe Telecom

InfoSphere Streams

The InfoSphere Streams 1.2 Platform

Development Environment

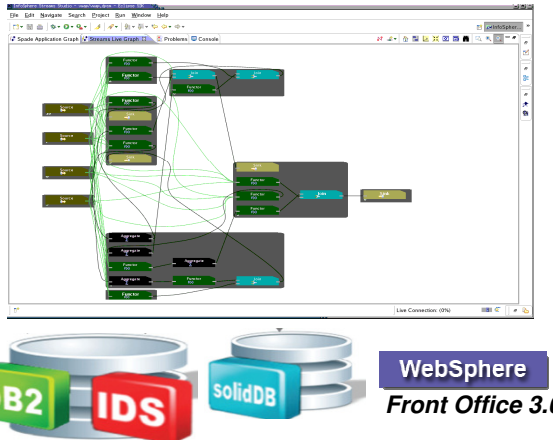


Streams Studio
Eclipse IDE for SPADE

Runtime Environment

Scalable stream processing runtime

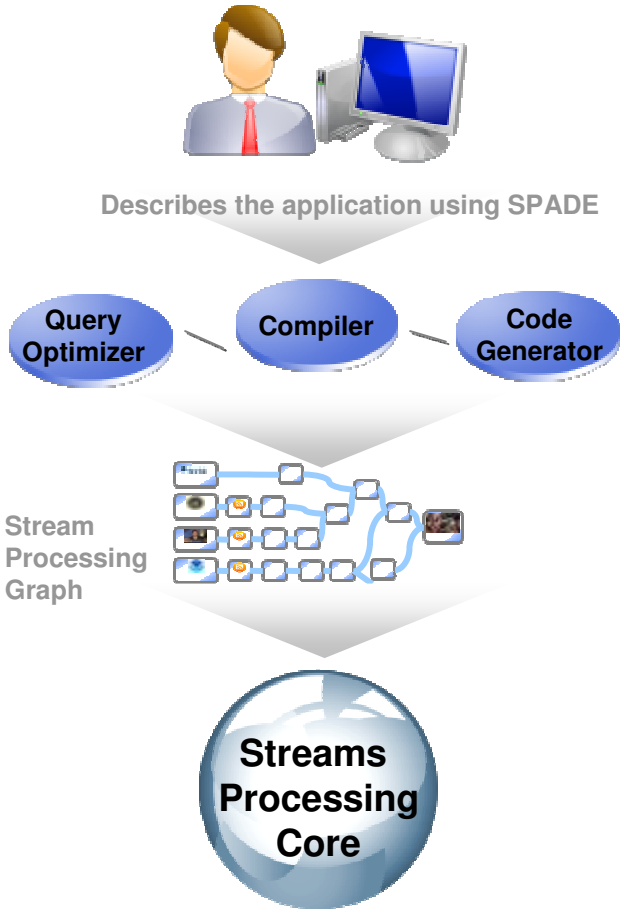
Tools and Technology Integration



Streamsight,
Built-in Stream Relational Analytics,
Adapters
Toolkits

Supported on x86 hardware, RedHat Enterprise Linux 5.3 and 5.4

SPADE: Stream Processing Application Declarative Engine - A language optimized for stream analytics



Key Features

- Programming language and a compilation infrastructure
- Efficient mapping to a wide variety of target architectures

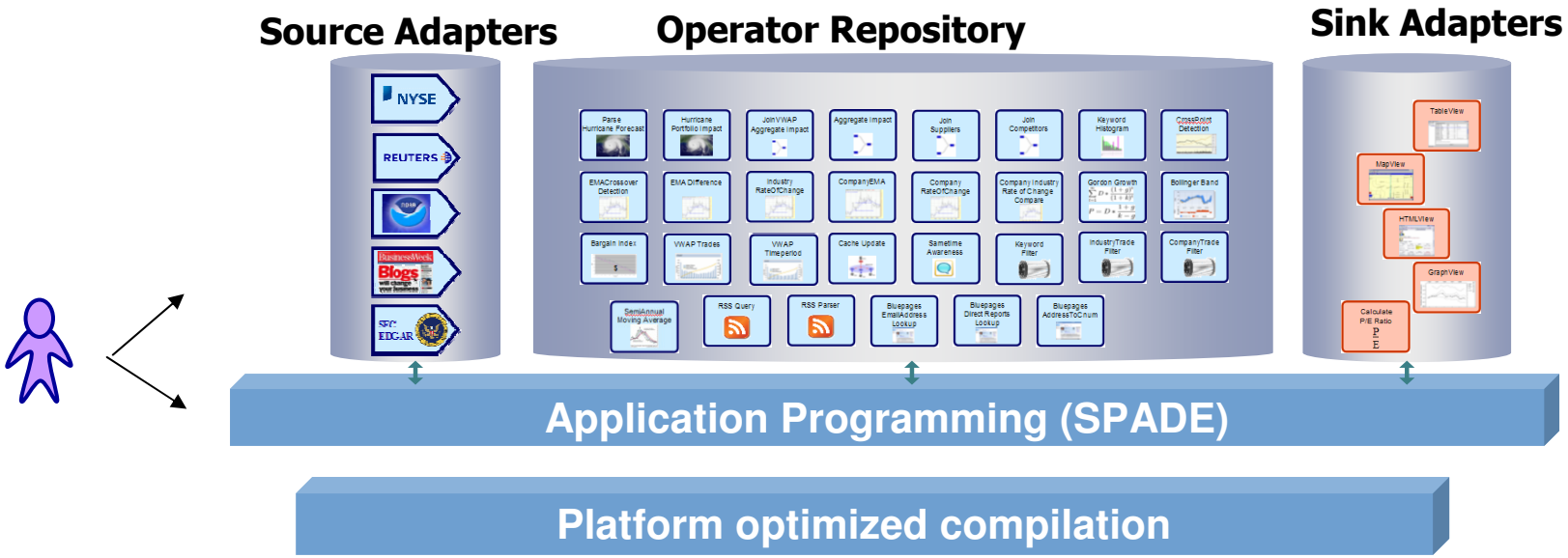
Key Components

- Toolkit of stream processing operators and execution directives
- Stream Relational Operators
- User Defined Operators
- Edge Stream Adapters
- Controls to effect placement, partitioning, etc

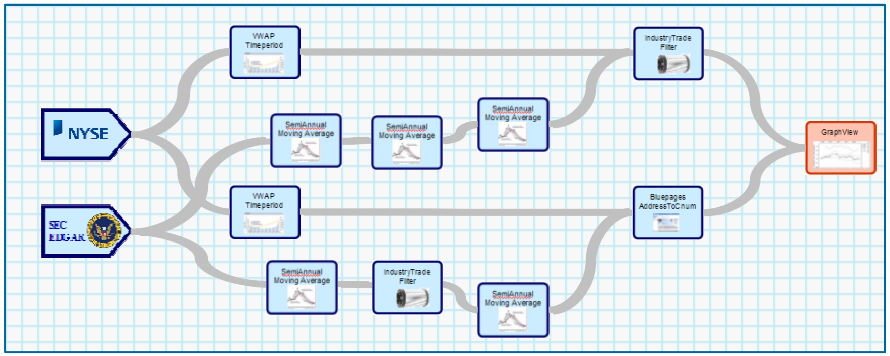
Key Benefits

- Optimized for event Analytics
- Faster time to deployment

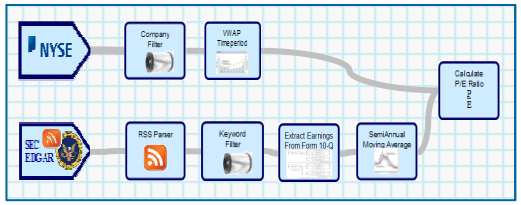
Programming Model



- Consumable
- Reusable set of operators
- Connectors to external static or streaming data sources and sinks

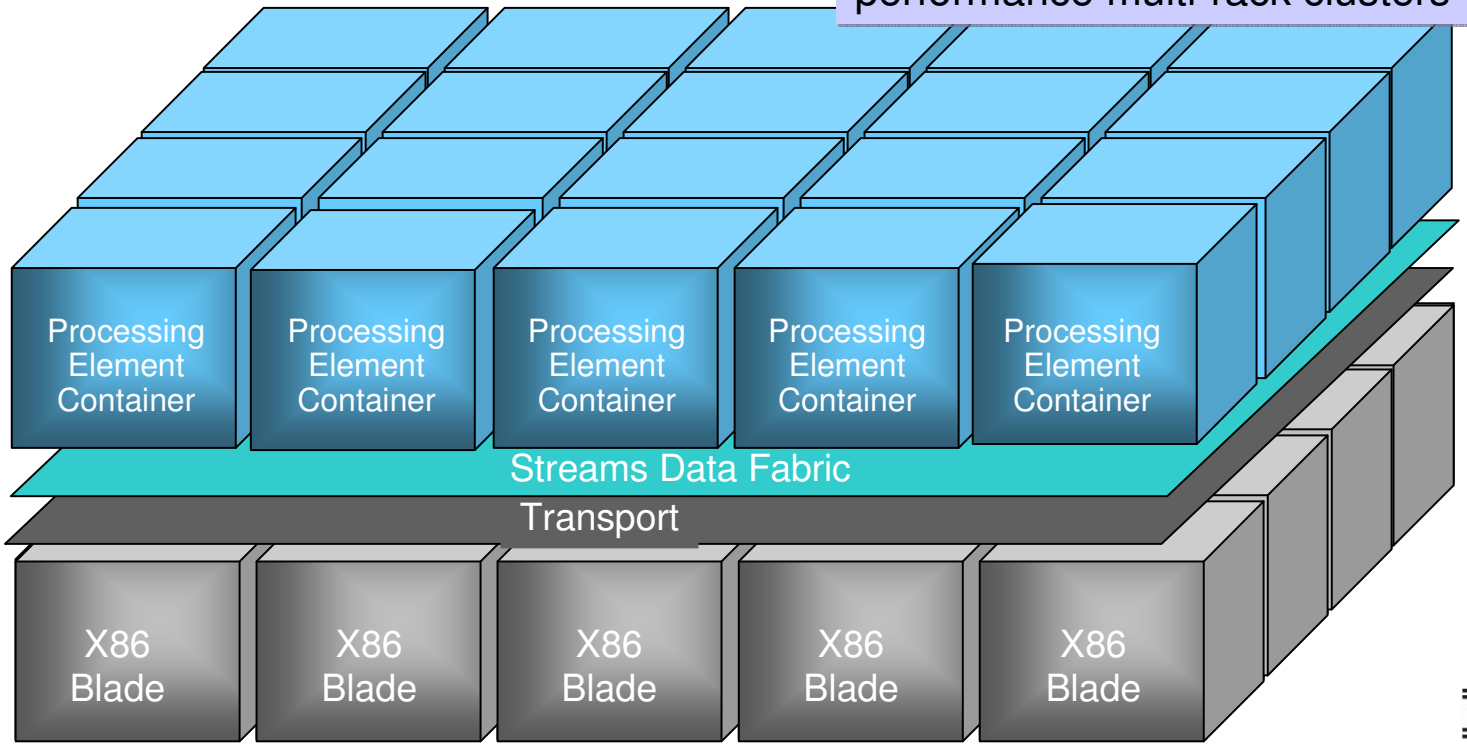


Streams Runtime



Optimizing scheduler assigns operators to processing nodes, and continually manages resource allocation

Runs on commodity hardware – from single node to blade centers to high performance multi-rack clusters

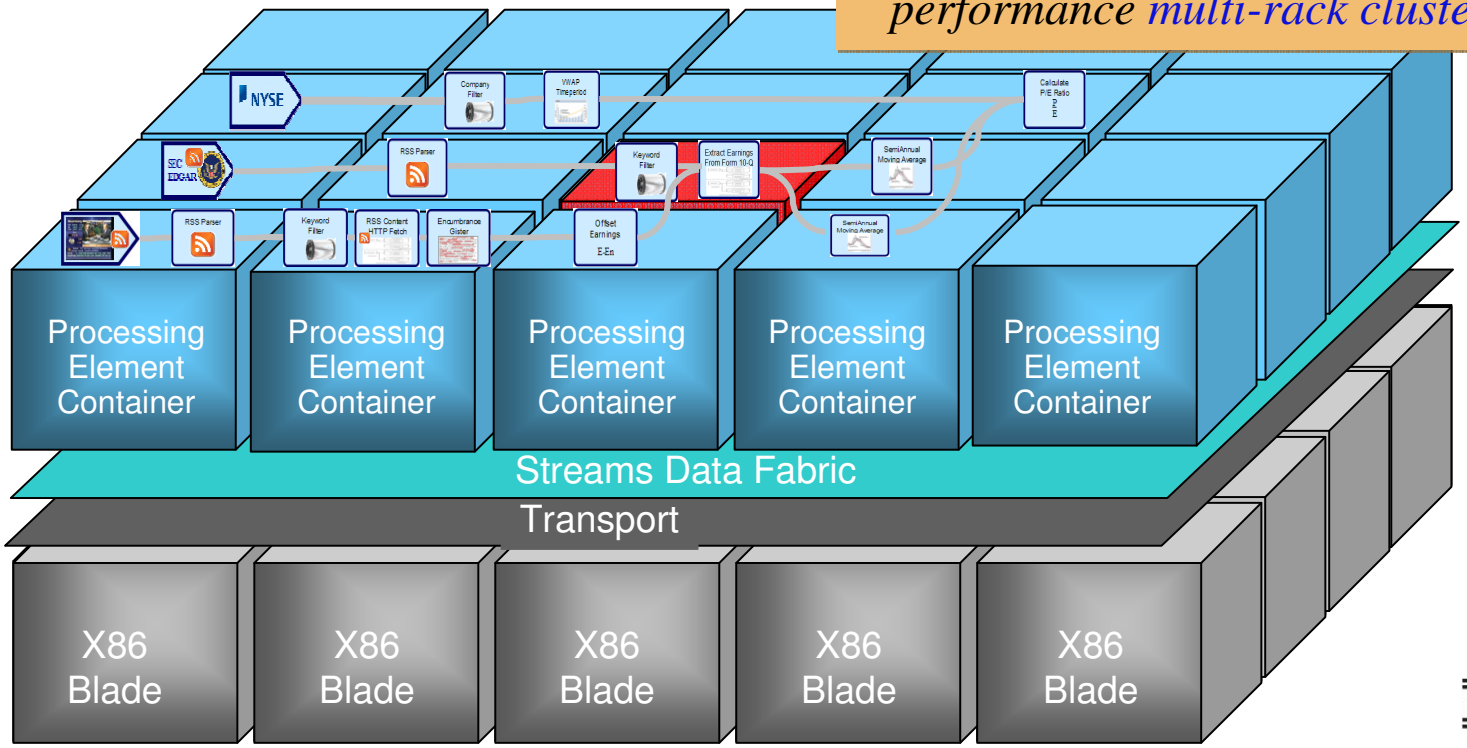


Streams Runtime

Optimizing scheduler assigns operators to processing nodes, and continually manages resource allocation

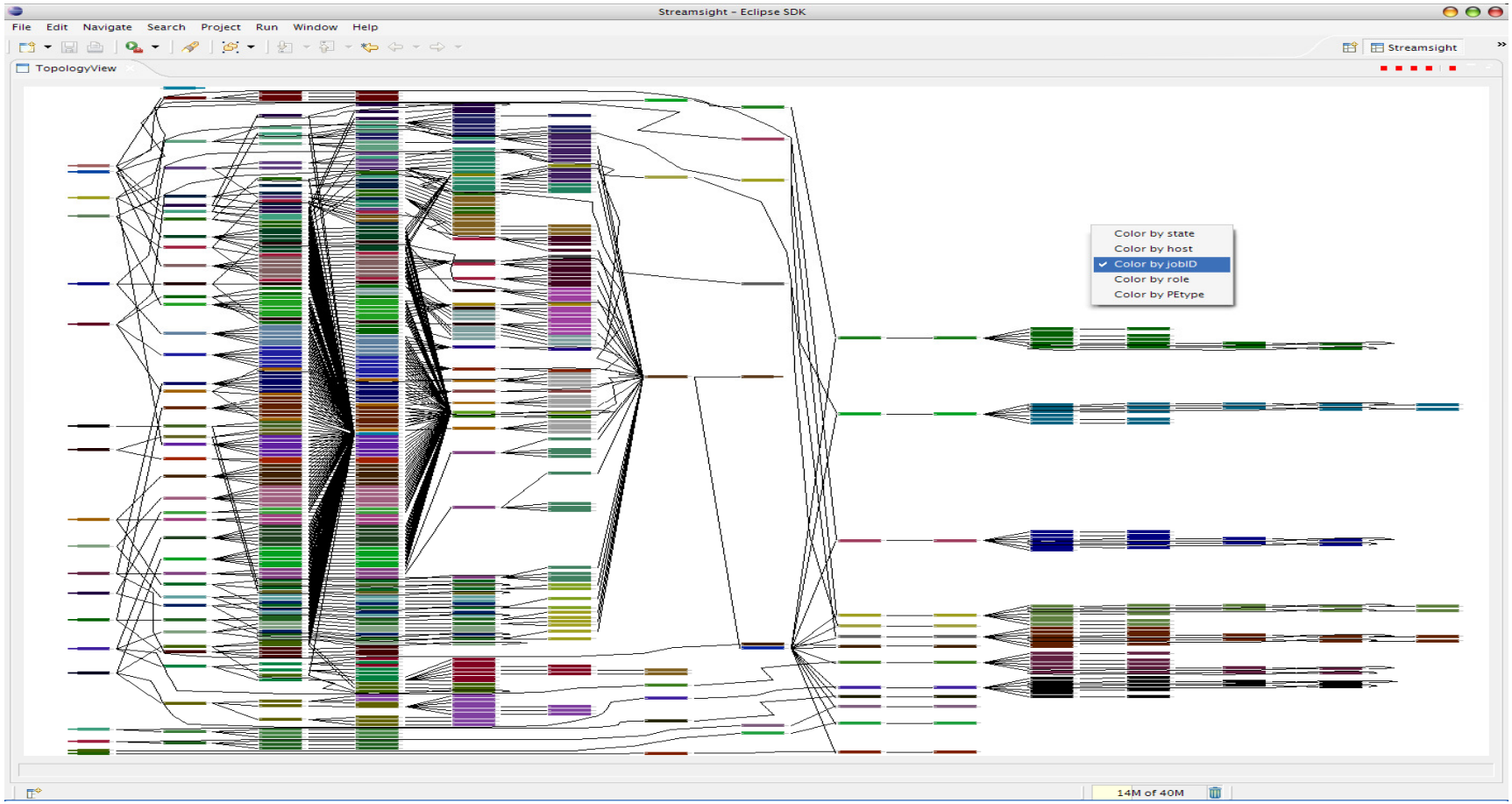
Adapts to changes in resources, workload, data rates

Runs on commodity hardware – from single node to blade centers to high performance multi-rack clusters



Streamsight to manage running applications

- Logical and physical view of apps



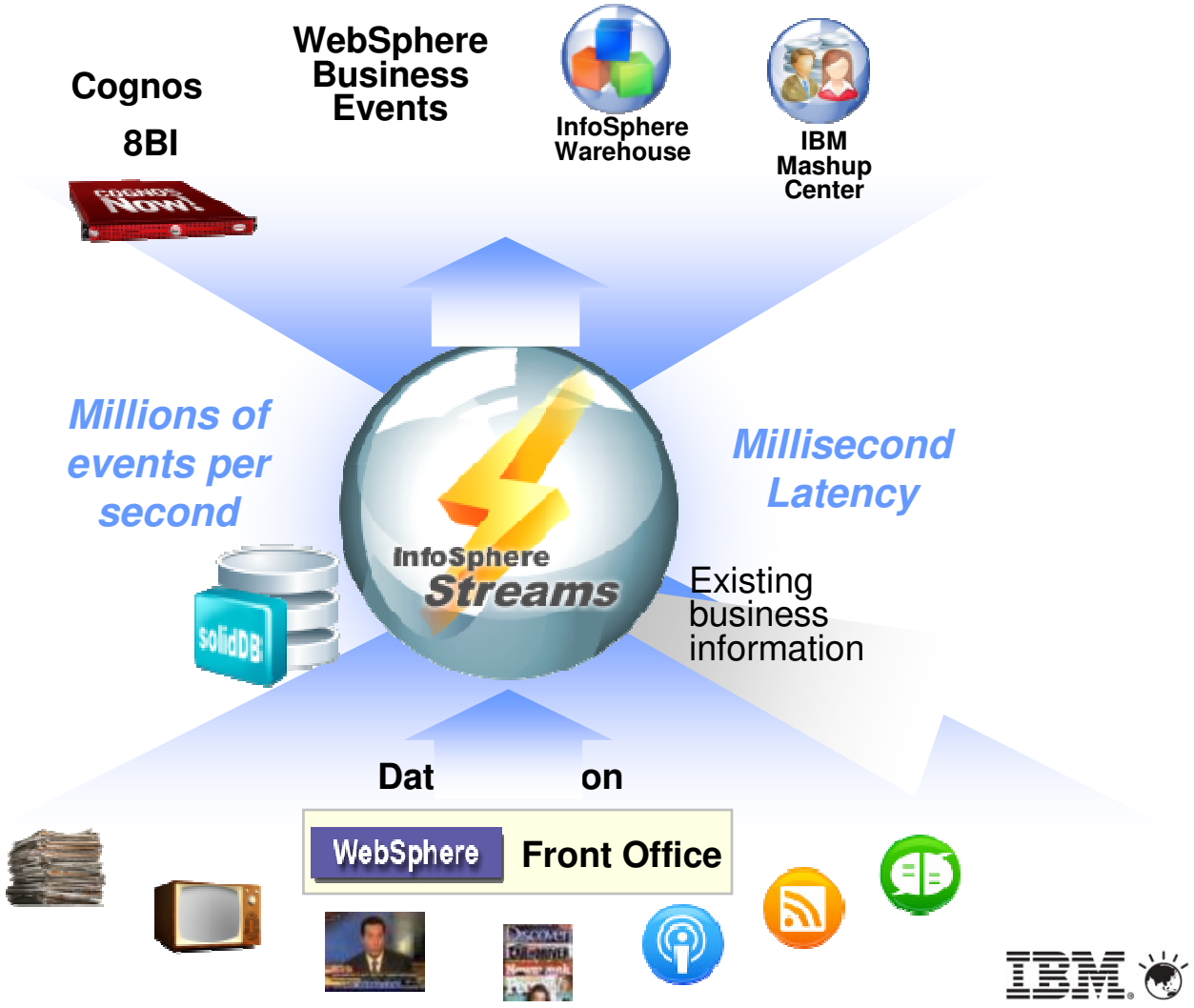
IBM InfoSphere Streams directions

Output Adapters
Cognos Now
WebSphere MQ
Mashup Center
WebSphere Business Events

Data Enrichment
solidDB
DB2
IDS
InfoSphere Warehouse

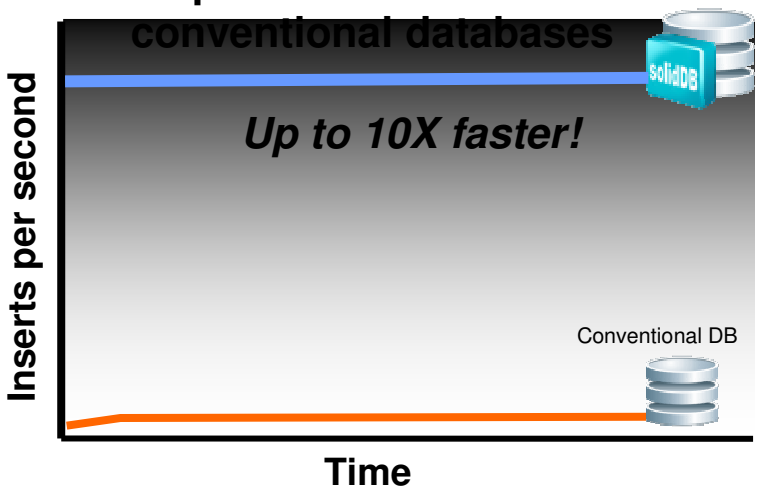
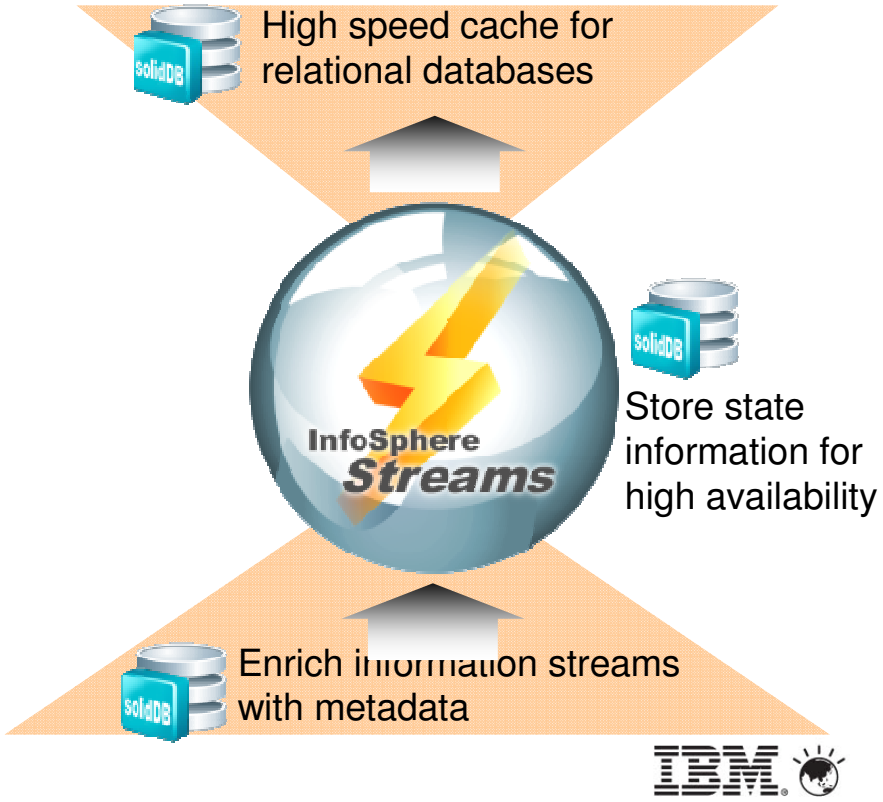
Tools
IBM Data Studio
Warehouse analytics
Financial analytics
Video/audio analytics

Source Adapters
WebSphere Front Office
IBM Mashup Hub
WebSphere MQ
RSS Feeds



InfoSphere Streams and solidDB for extreme data speeds

Up to 10x faster than conventional databases



Provides high speed database for event processing