

In-Motion Analytics with InfoSphere Streams

Krishna Mamidipaka

IBM Software Group



IBM Information
ON Demand 2010



INFORMATION-LED
TRANSFORMATION



LEAD
THE WAY



IBM

January 21 - SINGAPORE • January 26 – MALAYSIA • January 28 - THAILAND

Agenda..

→ What

- What is the need for Stream Computing?
- What is meant by Stream computing?

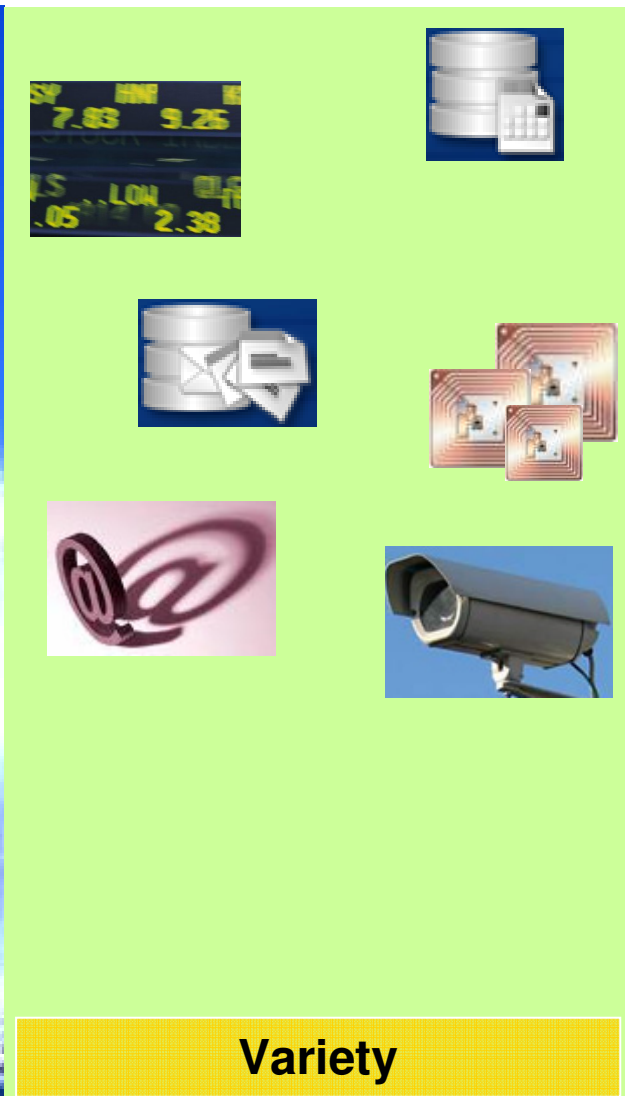
→ How

- How is it different from existing technologies?
- How does it 'fit' within existing information infrastructure?
- How does it work?

→ Where

- Where is it currently deployed (references / use cases etc)?
- Where to get further information?

What is driving the need for Real Time Analytic Processing



Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

Where incremental innovations is insufficient...

When incremental improvement
is impractical....



Invention is required

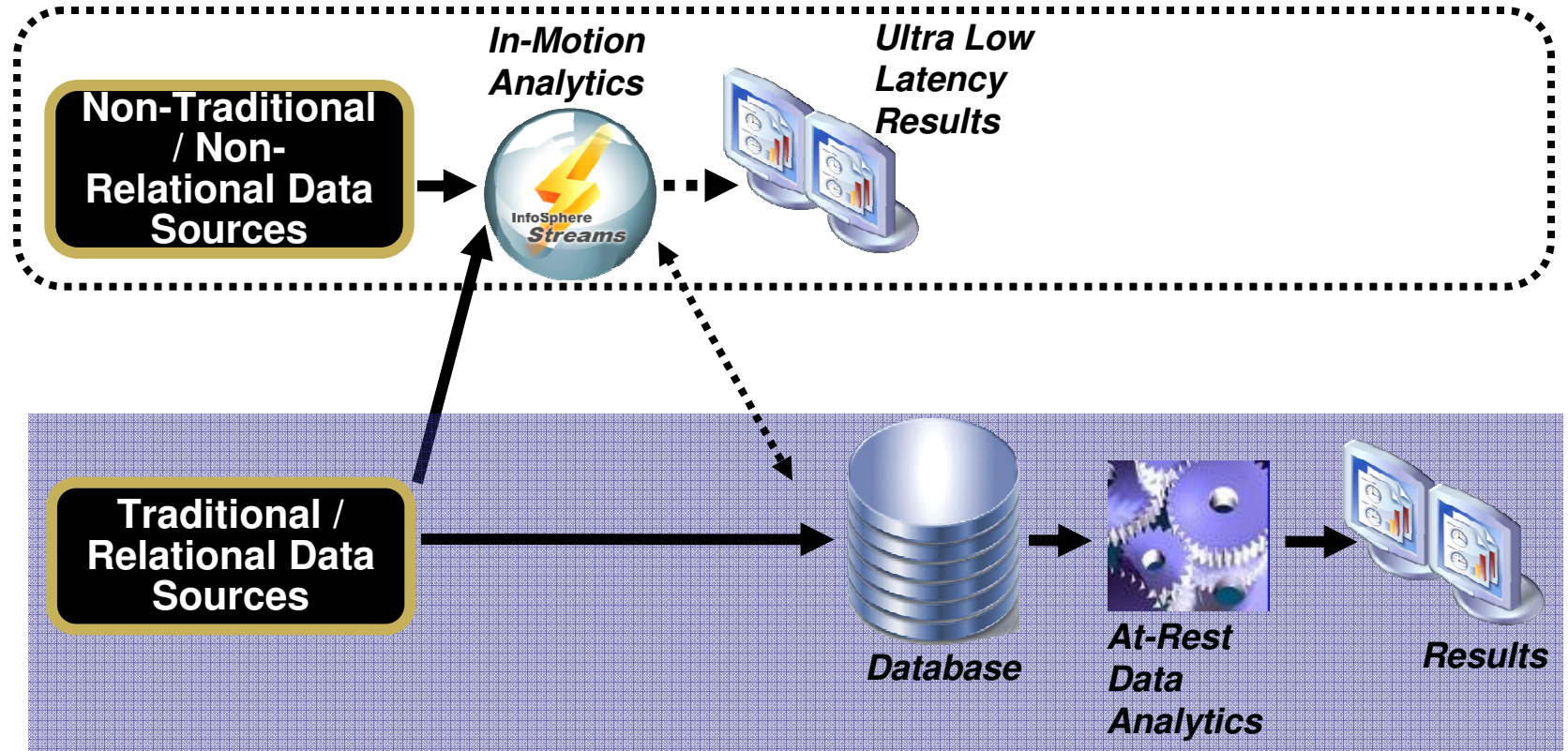


To match with performance of Blue Gene with traditional technology, the servers would occupy the space of several stadiums and required the power of a nuclear power plant.

In-Motion Vs Traditional Analytics

**Stream
Analytics**

RTAP



**Traditional
Analytics**

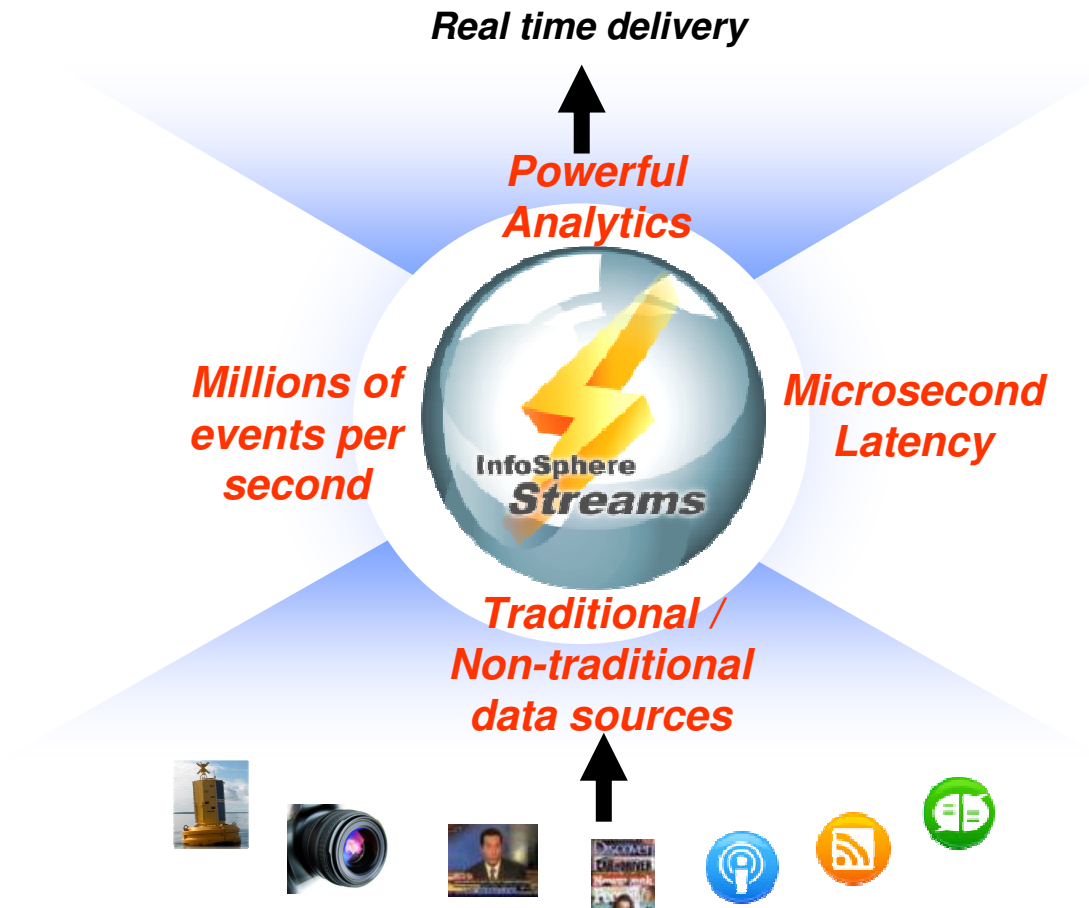
OLAP / OLTP

Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

Delivering 'Continuous Intelligence' with Powerful Analytics

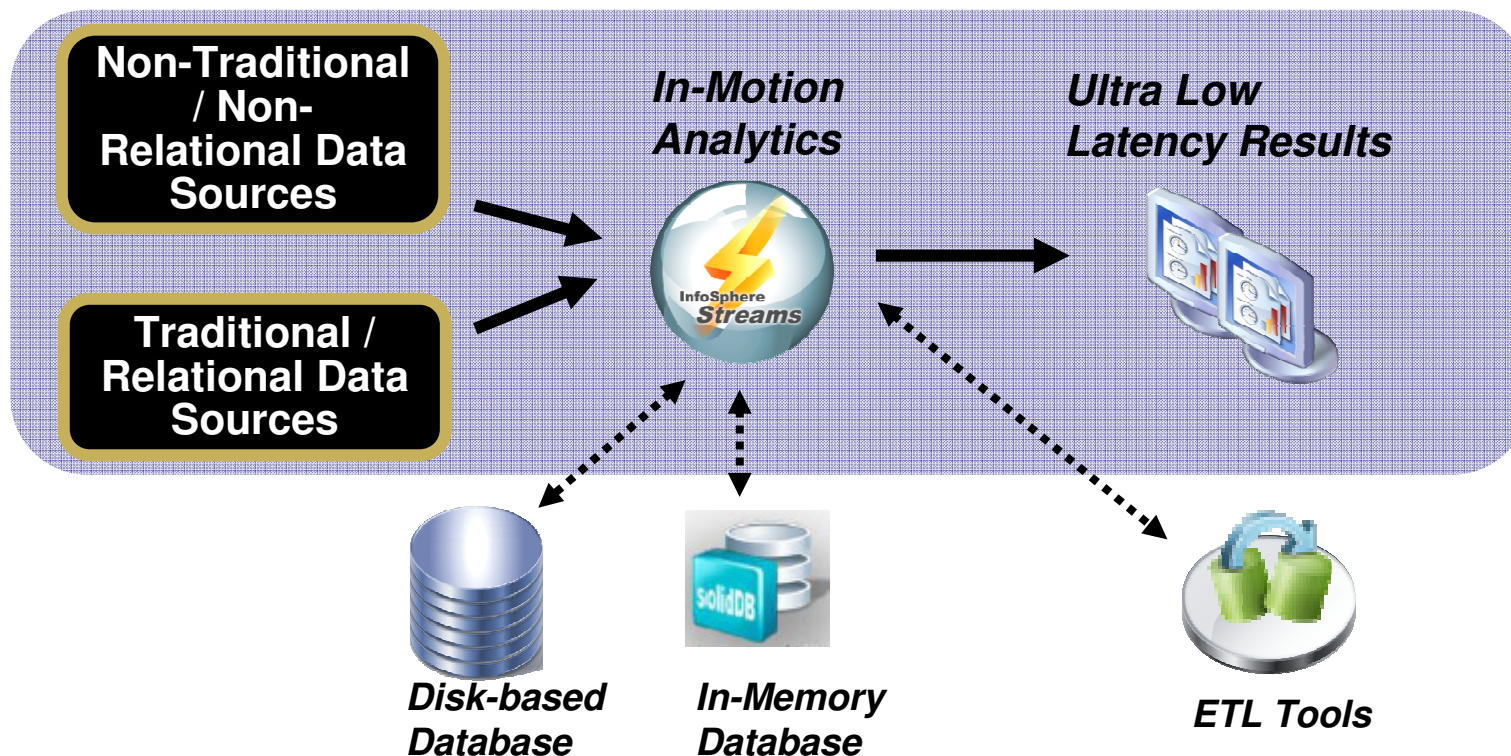
First of a Kind
Security Trading
application at TD
Securities:

- Peak throughput of 5 million messages per second
- Average end-to-end latency of 150 micro seconds



Streams also enables efficient Integration with Stored Data

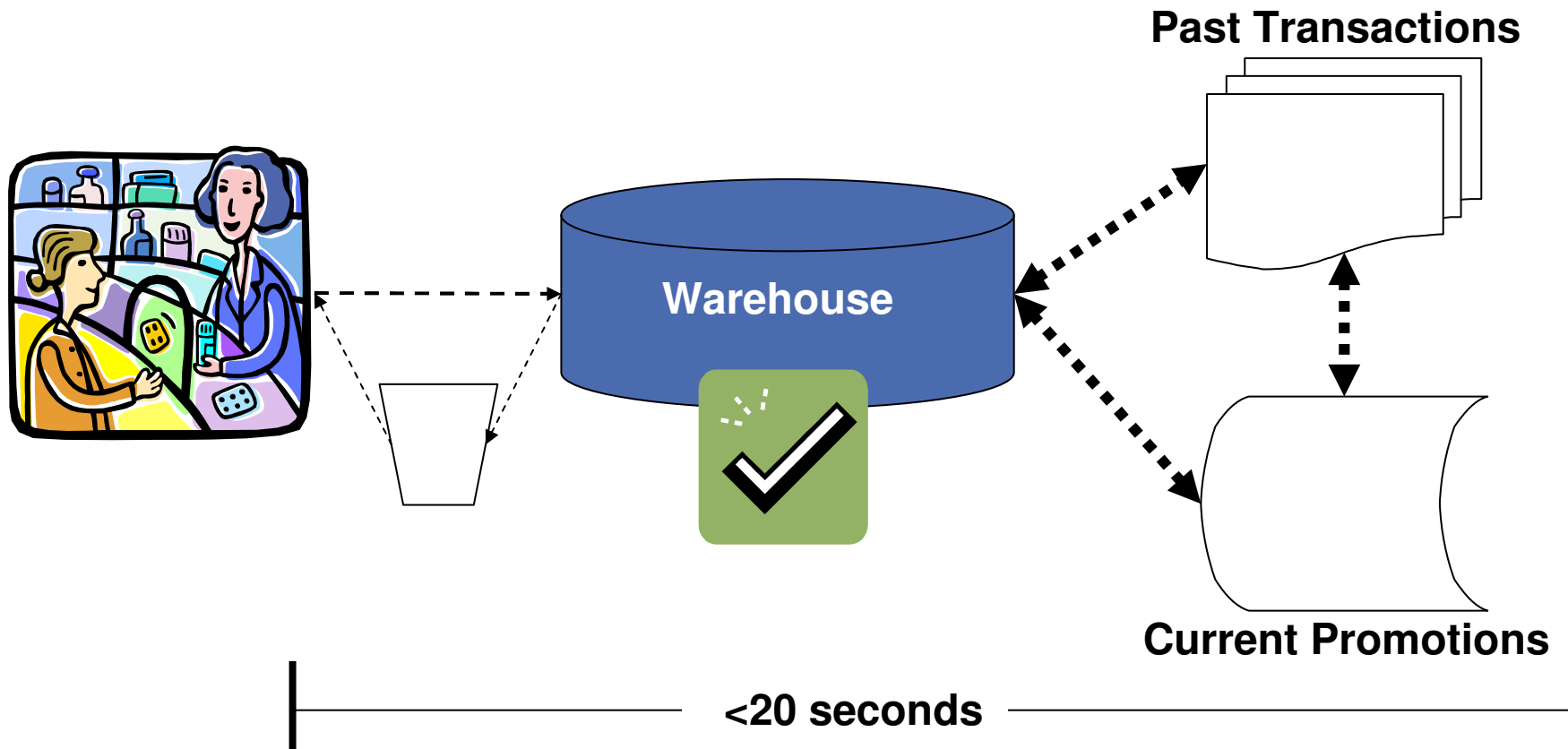
- Integration enables enrichment and persistence
- Support for disk based databases and in-memory databases like solidDB
- Supports standard JDBC/ODBC connect as well as Solid Accelerator API



Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

The need for Streams – a simple example

Business need: a retailer wants to do near real time analytics (on live POS transactions for the purpose of cross-sell and up-sell.



The need for Streams – a simple example

But, what if you wanted to add a lot more inputs (real time, non-traditional)??



Live Weather

Live Traffic

How crowded it is ?

Data from other stores

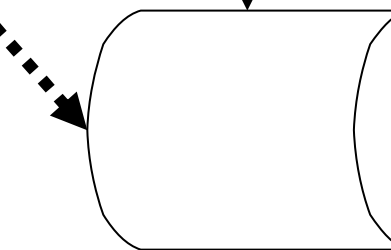
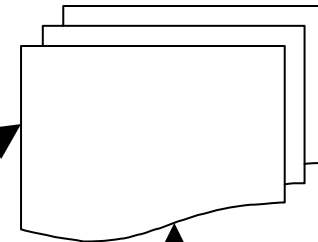
TV News

Stock Market

Industry Source Data



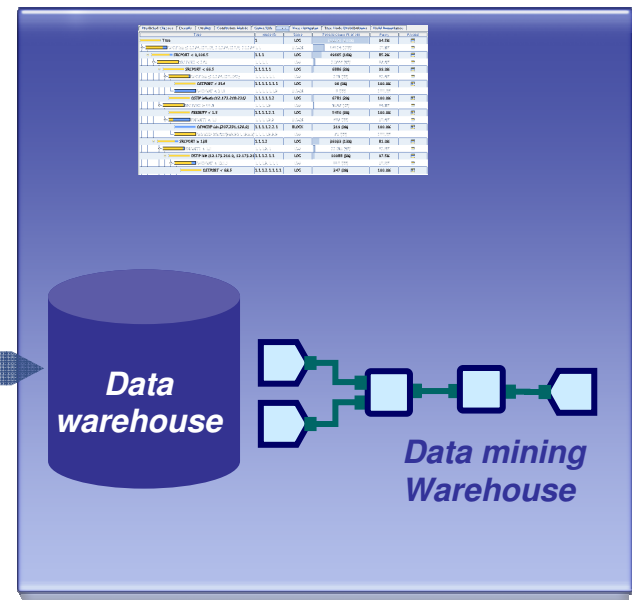
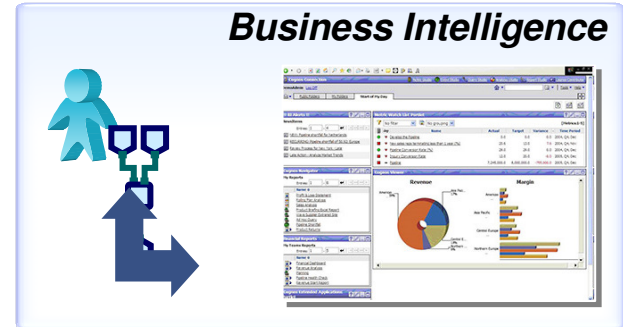
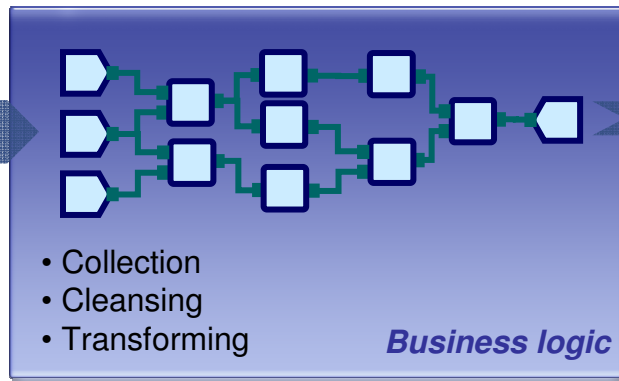
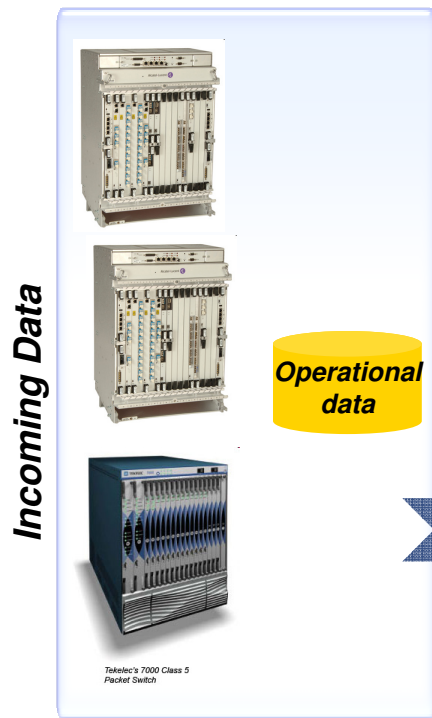
Past Transactions



Current Promotions

<20 seconds

Traditional Information Architecture



Adding Stream Processing...

Context

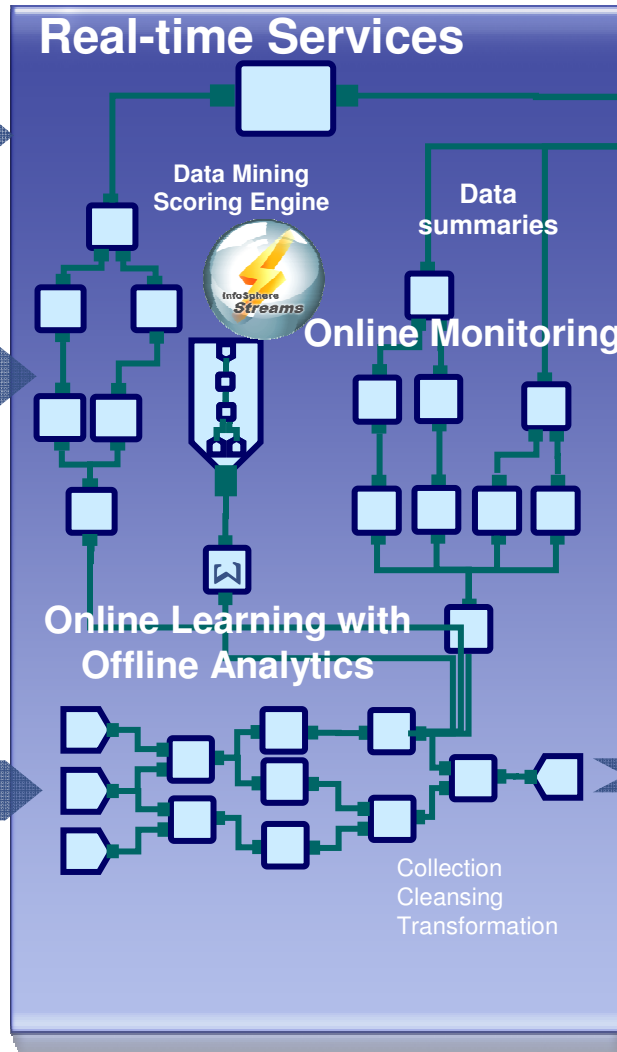


GPS Location, Transactions, Personal Health Monitor etc.

Incoming Data

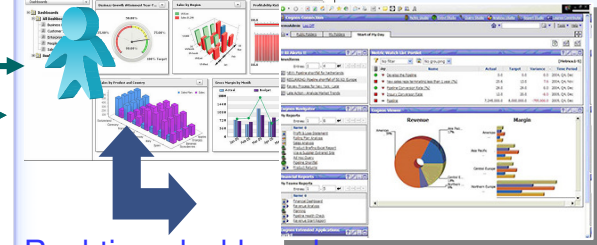


Telebit's 7000 Class S Packet Switch

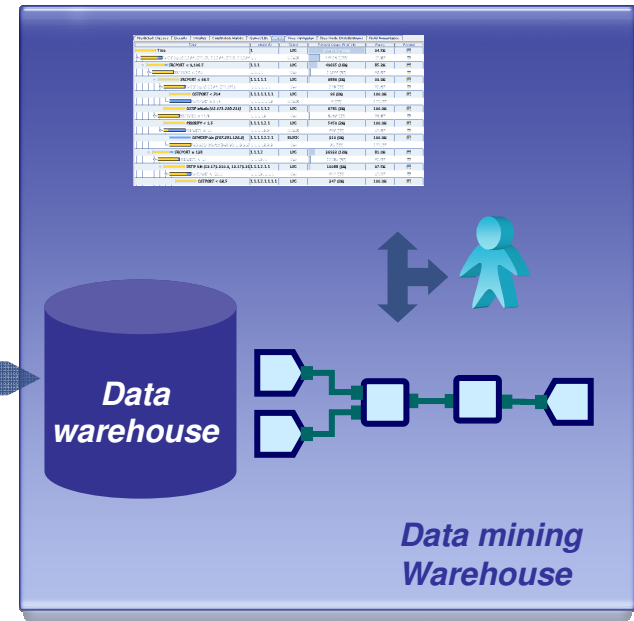


Network Monitoring

Business Intelligence



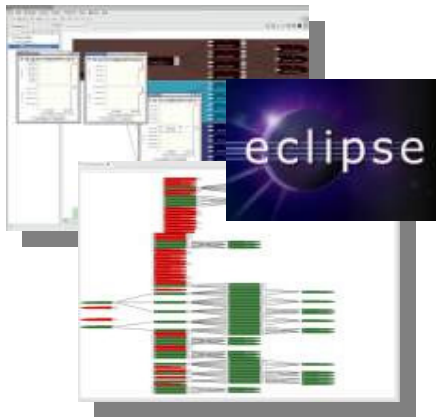
Real-time dashboards



Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

IBM InfoSphere Streams v1.2

Development Environment



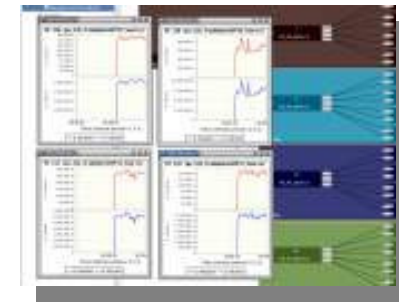
- Eclipse based IDE
- StreamSight
- Stream Debugger

Runtime Environment



- RHEL v4 or v5
- SELinux
- x86 multicore hardware
- Up to 125 servers

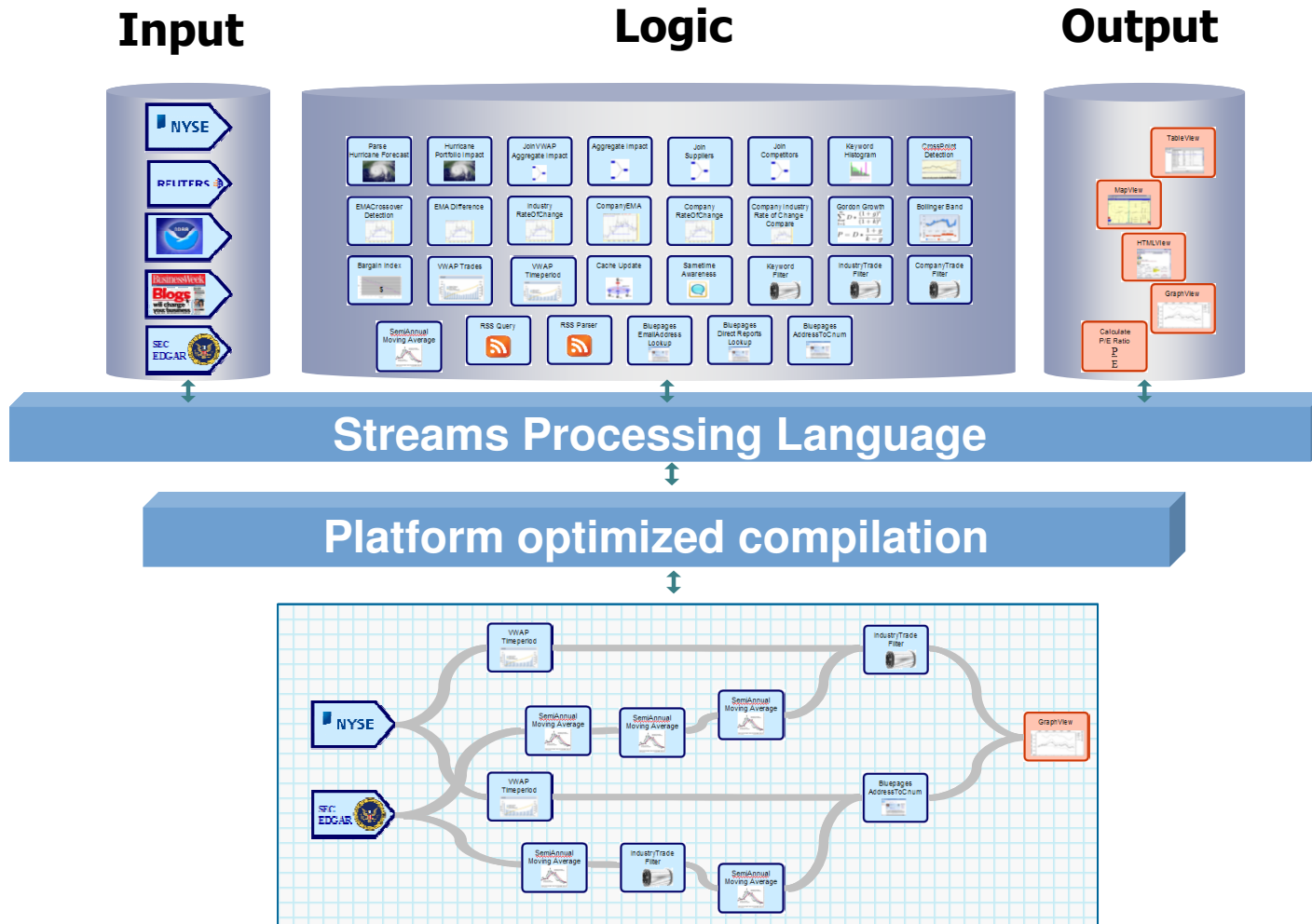
Toolkits & Adapters



- Financial Analytics Toolkit
- Connectors to data sources
- Math and Text functions
- Operator Library

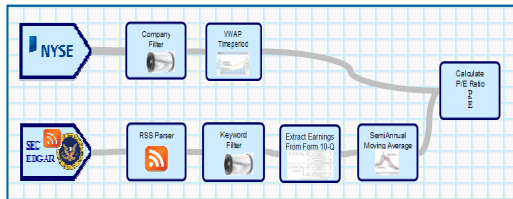
Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

Streams Programming Model



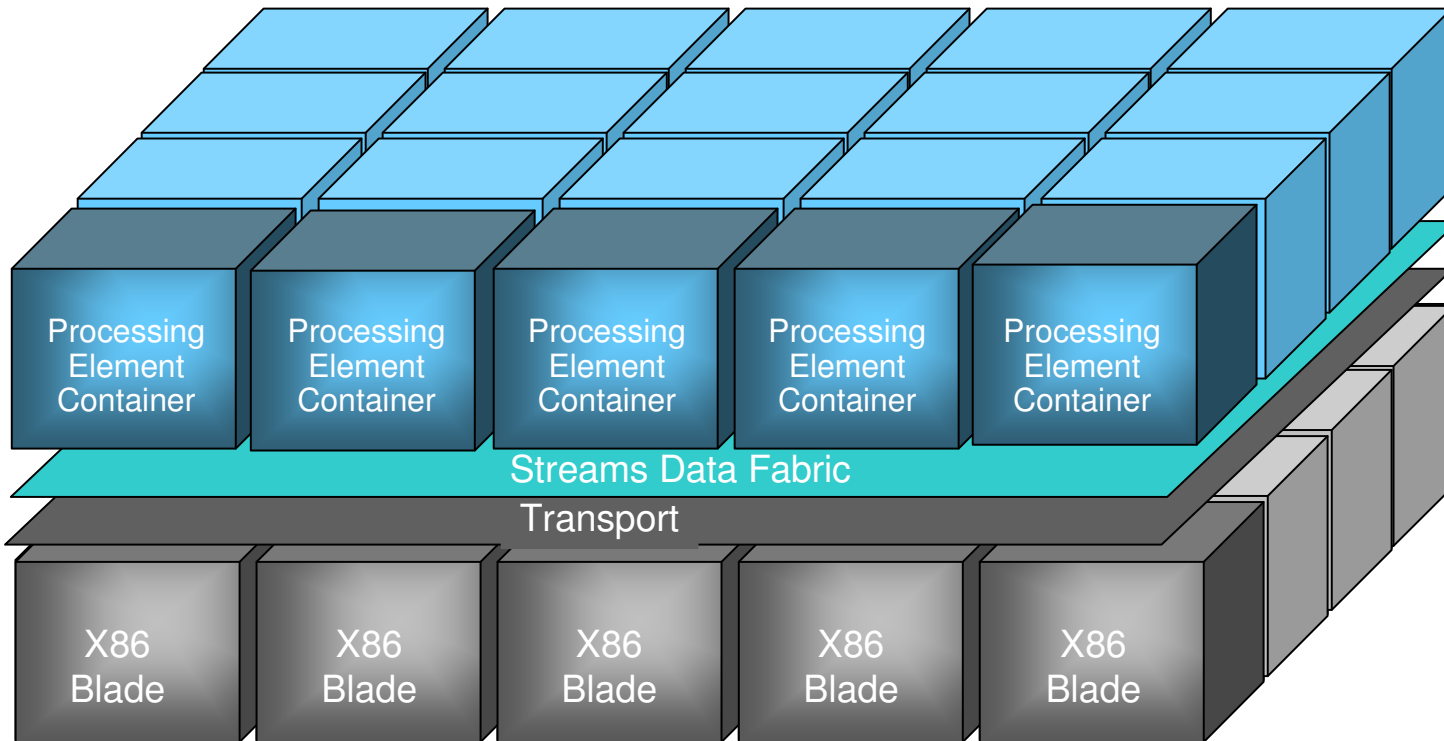
Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

Streams Runtime Illustrated



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

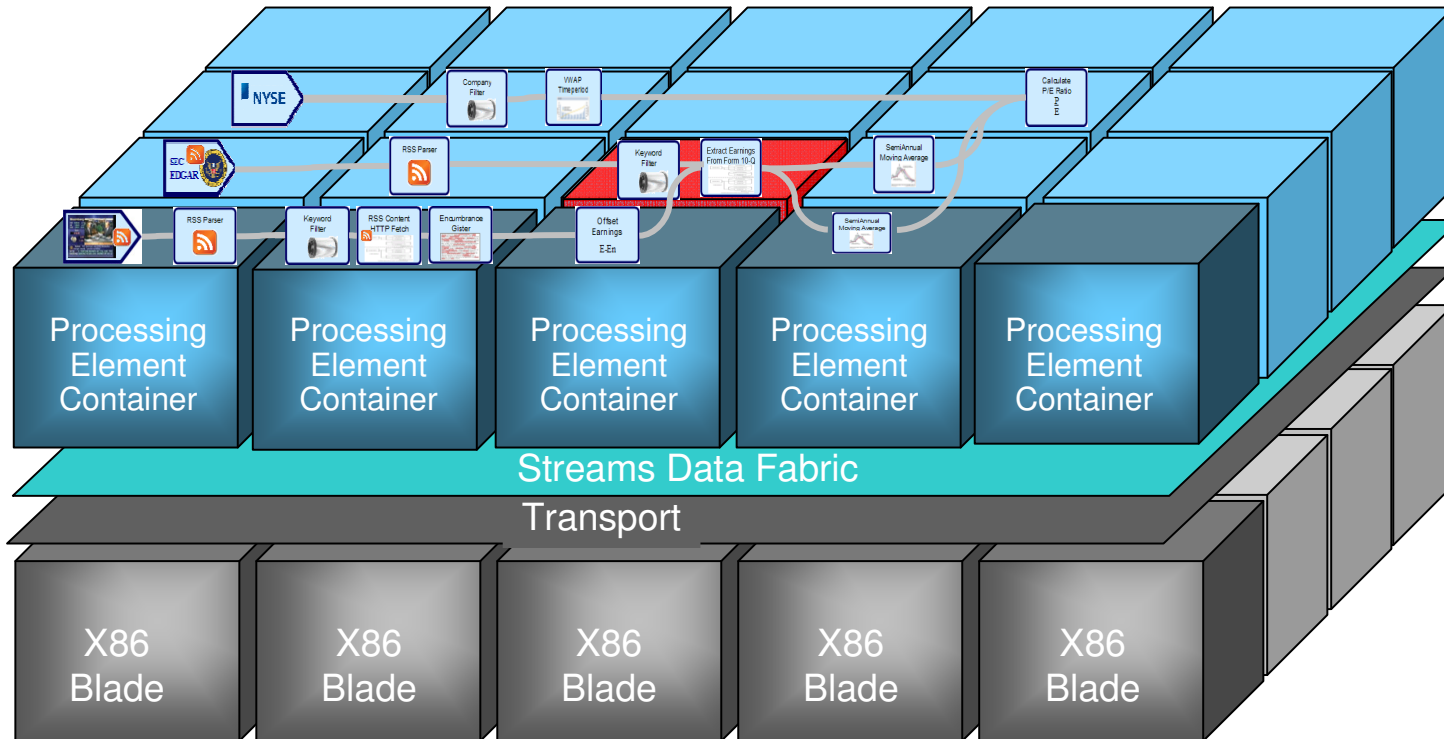
Runs on commodity hardware – from single node to blade centers to high performance multi-rack clusters



Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

Streams Runtime Illustrated

Can adapt to changes in resources, workload, data rates



'Smart' applications are fast emerging

Healthcare



Securities Trading



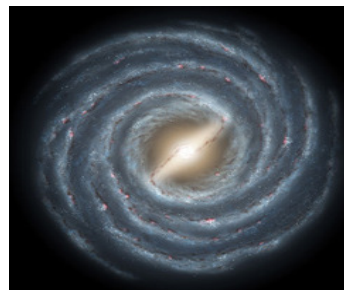
Environment



Law Enforcement



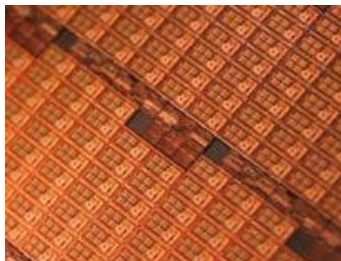
Radio Astronomy



Telecom



Manufacturing



Smart Traffic



Smarter Utilities



Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

Smarter Surveillance and Intrusion Detection

TerraEchos

- State-of-the-art surveillance system for US Navy based on Streams platform
- Acoustic signals from buried fiber optic cables are monitored, analyzed and reported in real time for necessary action
- Currently designed to scale upto '000s of parallel streams of raw binary data



Predictive Analytics using InfoSphere Streams in a neo natal ICU helps detect life threatening conditions upto 24hrs earlier

- Real Time analytics and correlations on physiological data streams
 - *Blood pressure, Temperature, EKG, Blood oxygen saturation etc.,*

- Early detection of the onset of potentially life threatening conditions
 - *Upto 24 hours earlier than current medical practices*
 - *Early intervention leads to lower patient morbidity and better long term outcomes*

- Technology also enables physicians to verify new clinical hypotheses



'World's fastest' options trading prototype

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"

Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

Telecom

- Real Time CDR Processing
- Real Time Geo Spatial analysis
- Applications:
 - Churn Prediction
 - Profitability Management
 - Real Time Promotions
 - Misuse and theft prevention
 - Value add services



What are analysts saying??

InfoSphere Streams Is A Game Changer

The future is here and it might be time to re-think the way we do business... by joining the stream.

IBM InfoSphere Streams Elevates CEP for Operational Intelligence

Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>



Q/A Session

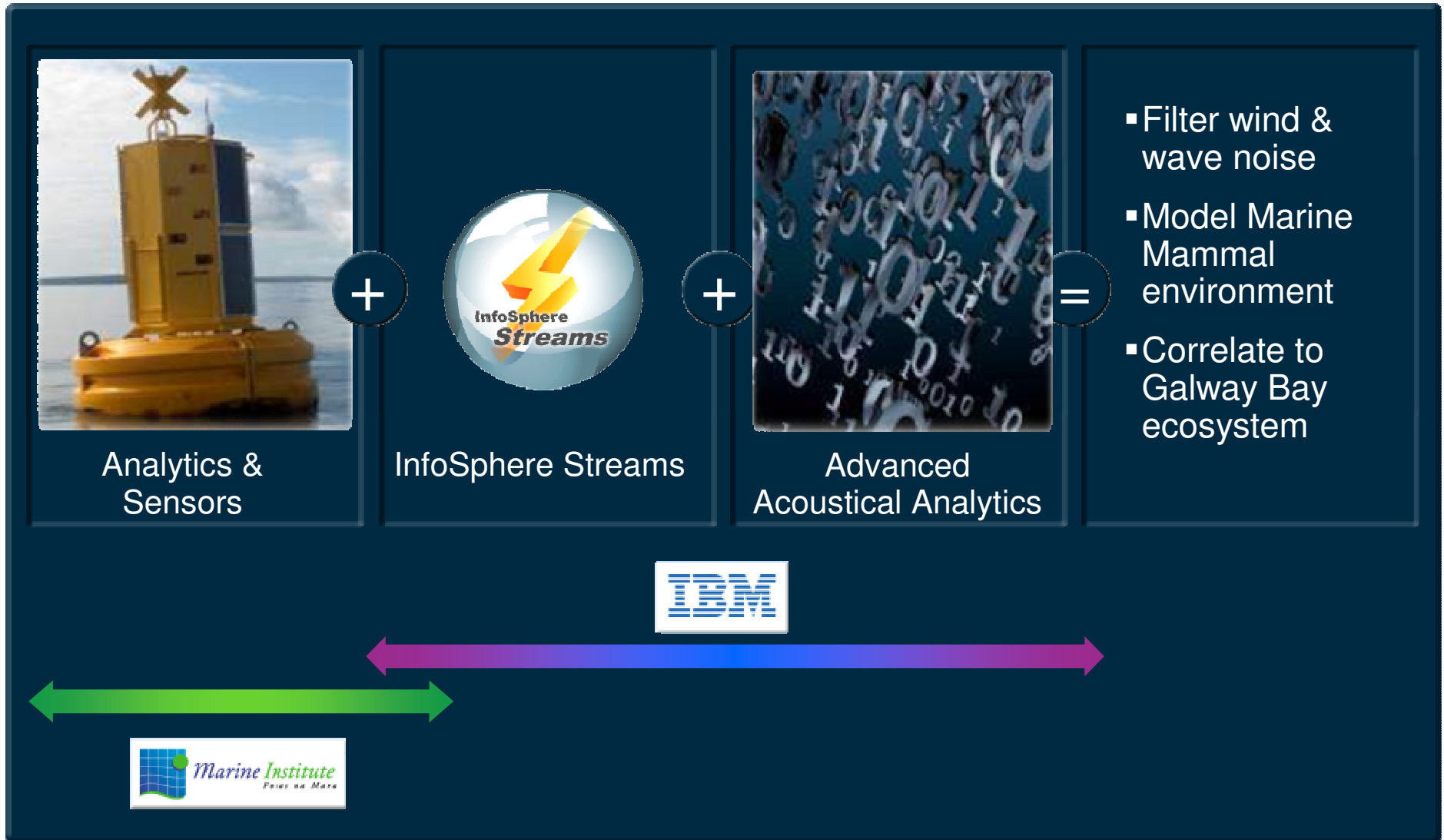
Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>



Back Up

Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

Real Time Marine Mammal Position and Behavior Modeling



Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

Real Time Detection and Management of Wildfires

■ Wildfire Management application

- Realtime US map of wildfire risk
- Detect wildfire smoke
- Task NOAA satellite and NASA UAV to monitor wildfire
- Generate health alerts



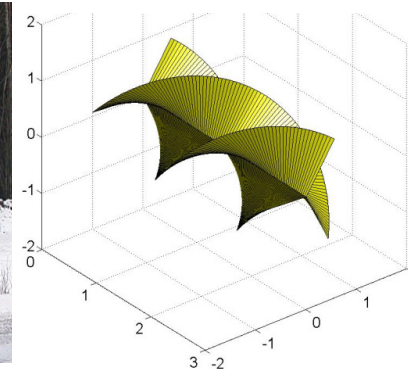
'Nowcasting' Solar Storms: Swedish Institute of Space Physics

■ Harmful effects of Solar storms

- Electric Grid failures
- Communication signal failures
- Other biological effects..

■ 8GB/Sec from each antenna needs to be analyzed

- Total = 1.3TB/sec and growing
- Data Storage is not an option



Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>

Traffic Control System in City of Stockholm: Version 2.0

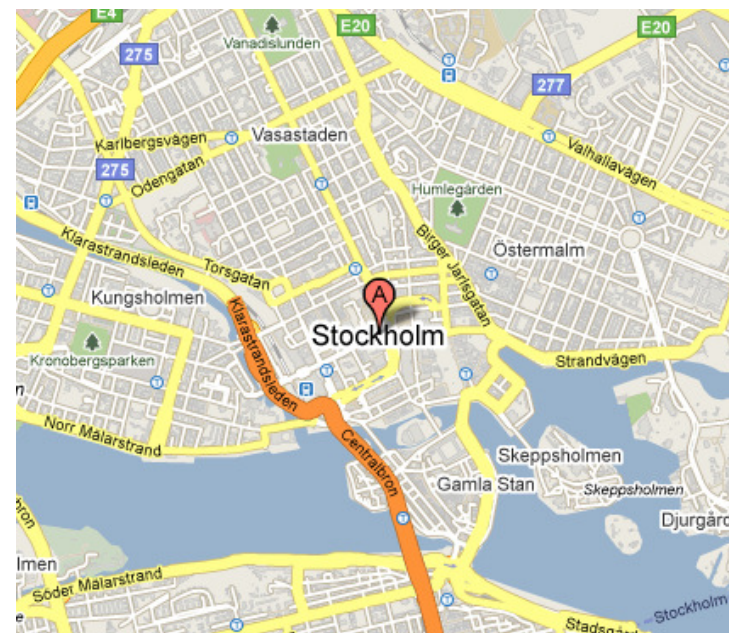
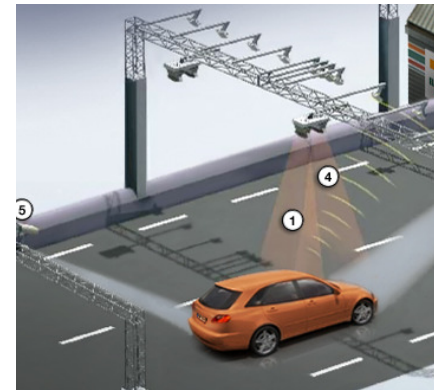
→ Data sources

- GPS from 1000's taxis
- Loop Sensors
 - Speed of traffic
 - Flow – density of traffic (cars per second)
- CCTV video inside tunnels
- Real Time Weather data

→ Output

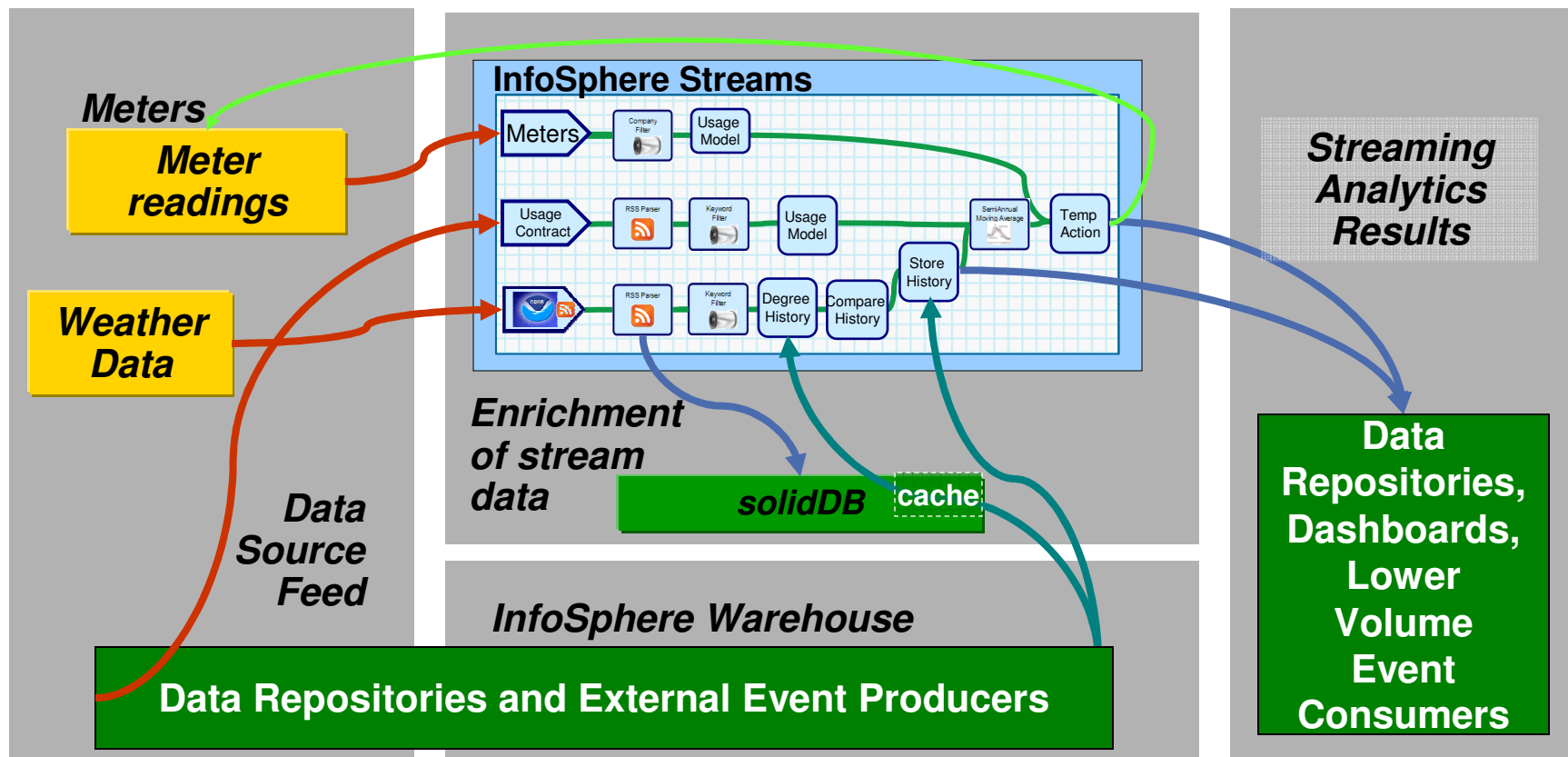
- Travel time forecasts
 - Via SMS
 - Now, In 30 minutes, 1 hour, 2 hours etc

→ Integrate with existing system



Smarter utilities

- Analysis of data from multiple sources
- Realtime decisions – lower voltage on high degree days

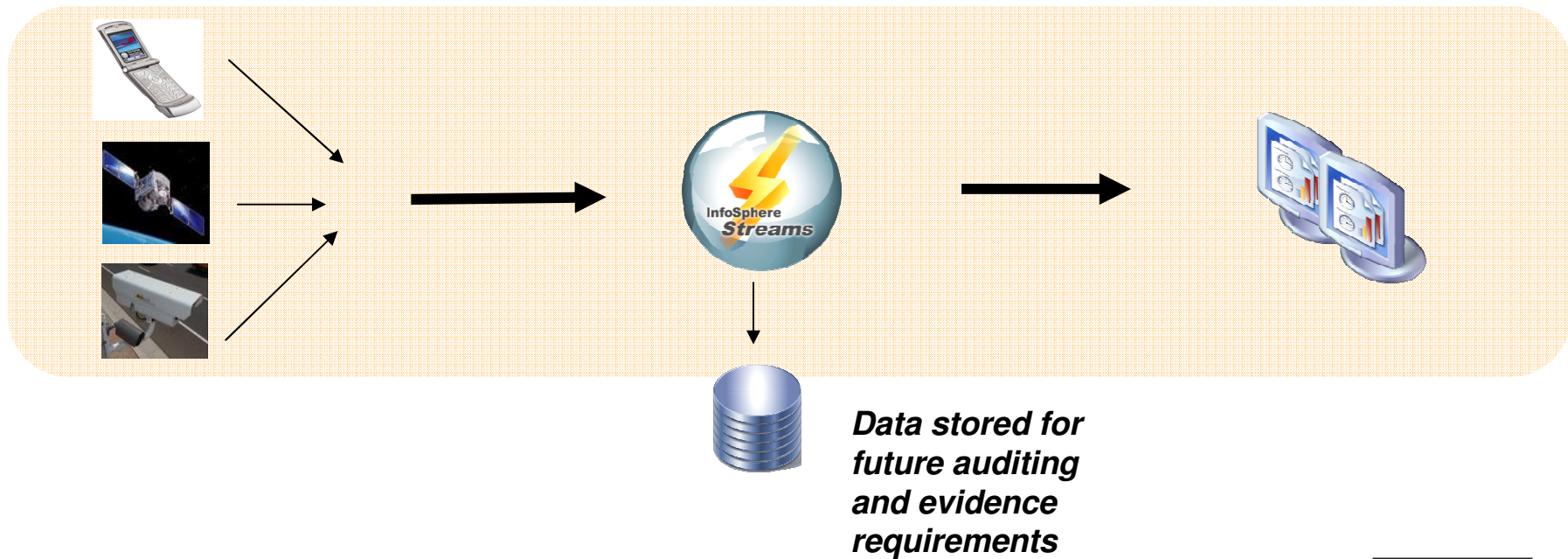


Government and Law Enforcement: e911 Support

Data from 911 calls, satellite feeds, imagery from city traffic cameras

Streams defines the geo spatial location of the call by running powerful analytics in real time using satellite communication link and draws in city camera feeds from around the area

Real time support for 911 dispatcher and field personnel



Join us on TWITTER! - <http://twitter.com/IOD2010ASEAN>