



InfoSphere Streams - Streaming Analytics for Big Data

- Built to analyze data in motion
 - Multiple concurrent input streams
 - Massive scalability
- Process and analyze a variety of data
 - Structured, unstructured content, video, audio
 - Advanced analytic operators
- Enables Adaptive Real-Time Analytics
 - With Data Warehousing
 - With Hadoop Systems



Stream Computing Represents a Paradigm Shift



Traditional Computing



Historical fact finding

Find and analyze information stored on disk

Batch paradigm, pull model

Query-driven: submits queries to static data



Stream Computing



Current fact finding

Analyze data in motion – before it is stored

Low latency paradigm, push model

Data driven – bring data to the analytics

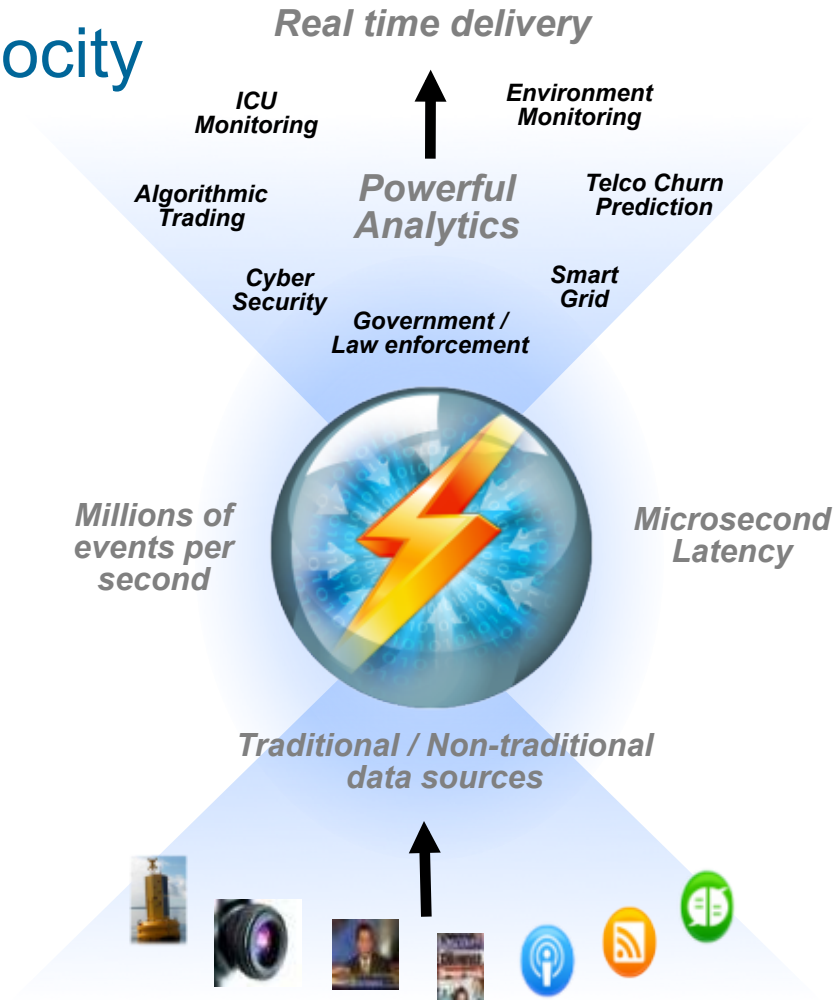


InfoSphere Streams Delivers Analytics for Big Data In-Motion

Key Big Data Challenge – Velocity

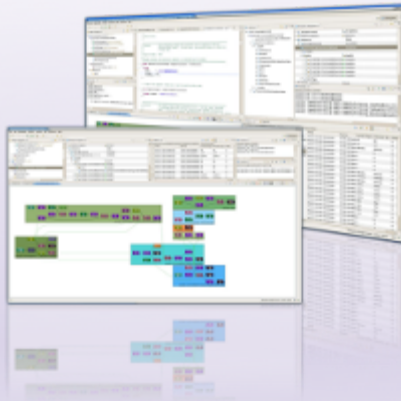
- Volume** Terabytes per second
Petabytes per day
- Variety** All kinds of data
All kinds of analytics
- Velocity** Insights in microseconds

*Example Streaming Data Sources:
Video, audio, networks, social media*



IBM InfoSphere Streams – Key Components

Agile Development Environment



- Eclipse IDE
- Streams Live Graph
- Streams Debugger

Distributed Runtime Environment



- Clustered runtime for massive scalability
- RHEL v5.3 and above, CentOS v6.0 and above
- x86 & Power multicore hardware
- Ethernet & InfiniBand

Sophisticated Analytics with Toolkits & Adapters



- Analytic Accelerators
 - Database
 - Mining
 - Financial
 - Standard
 - Internet
 - Big data (HDFS)
 - Text
 - User-defined
- Over 50 samples



Massively Scalable Stream Analytics

Linear Scalability

- Clustered deployments – unlimited scalability

Automated Deployment

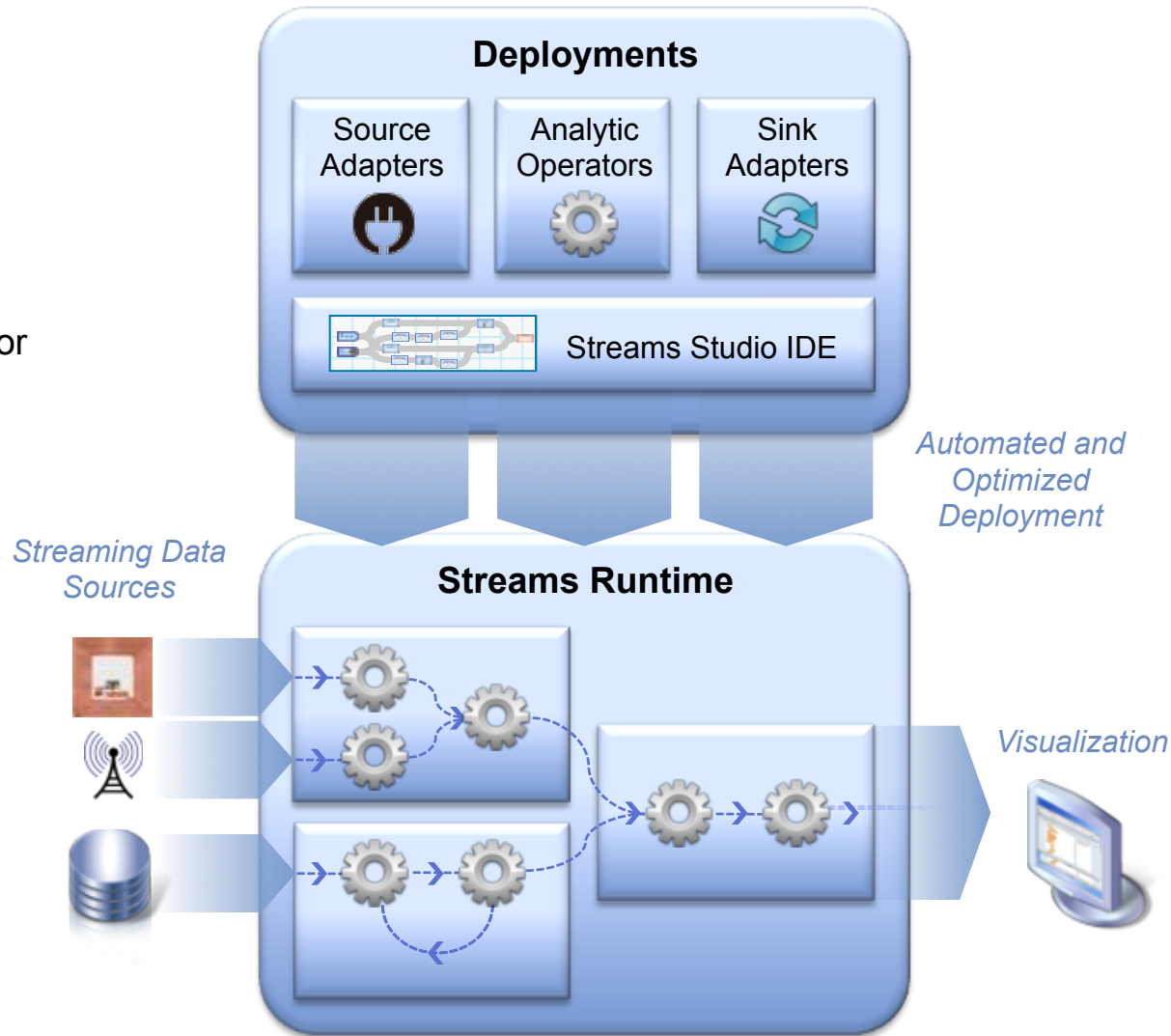
- Automatically optimize operator deployment across nodes

Performance Optimization

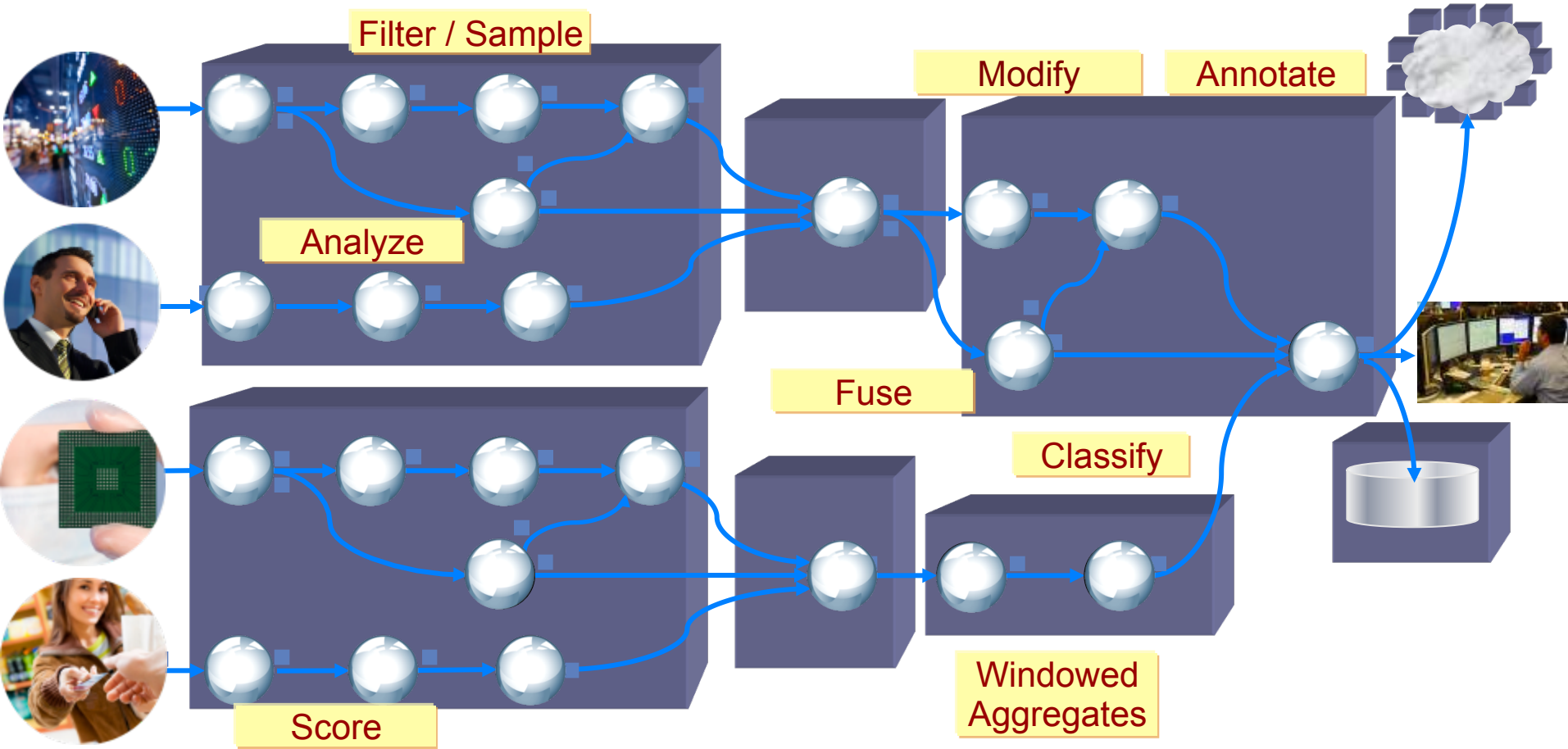
- Parallel & pipeline operations
- Efficient multi-threading

Analytics on Streaming Data

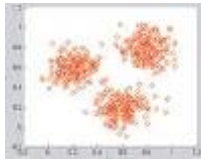
- Analytic accelerators for a variety of data types
- Optimized for real-time performance



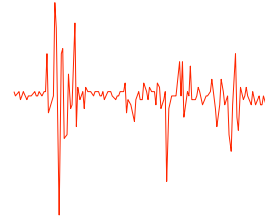
Big Data in real-time with InfoSphere Streams



Analytic Accelerators Designed for Velocity (and Variety)



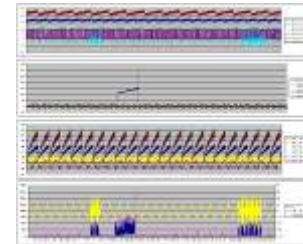
Mining in Microseconds
(included with Streams)



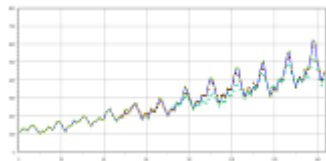
Acoustic
(IBM Research)
(Open Source)

Text
(listen, verb),
(radio, noun)

Simple & Advanced Text
(included with Streams)
(IBM Research)
(Open Source UIMA)



**Advanced
Mathematical
Models**
(IBM Research)



Predictive
(IBM Research)



$$\sum_{\text{population}} R(s_t, a_t)$$

Statistics
(included with
Streams)



Geospatial
(IBM Research)



Image & Video
(Open Source)



Putting it all together ...end-to-end big data solution

