

## Real Time Analytic Processing with IBM InfoSphere Streams v2.0

### Highlights

*Perform complex analytics on data in-motion*

*Handle massive volumes and variety of structured and unstructured data*

*Highly secure environment*

*Allows reuse of industry standard data mining scoring models*

*Incorporates support for high speed networks like InfiniBand*

*Runtime and developer tools facilitate application agility*

*".. A radical new approach that enables high speed, complex and scalable analytics on data streams in motion... will deploy additional installations to enable other agencies achieve greater success in various future projects", US*

**Government**

Data volumes are expected to double every two years primarily aided by growth in unstructured data such as audio and video.

Organizations are struggling to make 'truly' real time decisions to gain competitive advantage.

Traditional tools and technologies first require data to be recorded on a storage device and run queries after the fact to detect actionable insights. These tools also limited in their ability to support non-traditional data sources like audio, video, email etc that are often unstructured. Savvy organizations are fast realizing that time lost and data sources missed leads to missed opportunities.

IBM InfoSphere Streams addresses this gap by providing a state of the art platform for in-motion analytics on wide variety of relational and non-relational data types. It provides a highly scalable and agile infrastructure that can support incredible levels of data throughput. Streams achieves this by scaling from a single server to an unlimited number of nodes.

InfoSphere Streams enables a new paradigm for information processing and delivering many unique capabilities, including:

- Continuous analysis of streaming data with sub millisecond response times
- Extend existing applications without restarting the runtime or the application
- Reuse existing Predictive Model Markup Language (PMML) models with streaming data
- Support for InfiniBand
- Communications to Internet Protocol version 6 (IPV6) networks
- Strong support for developer and administrator productivity
- Standard Toolkit with relational, utility and adapter operators
- Database toolkit for persistence and enrichment of information streams
- Financial toolkit with built-in functions, inputs and analytics and sample applications



## IBM Information Management

*"Streams is a game changer"*

**Merv Adrian,**

**IT Market Strategy**

*"TerraEchos developers can deliver applications 45% faster due to the agility of Streams Processing Language."* **Alex Philp, CEO TerraEchos**

*"We think skyrocketing data volumes across various industries, a variety of data processing, and real-time processing needs can be improved by InfoSphere Streams, which possibly yields more than twenty to thirty percent profit to our customers."*

**Nobuyoshi Usami, Toshiba Solutions Corporation**

*"The Streams platform provides a unique combination of performance, agility, analytical sophistication, and the ability to handle effectively the wide range of unstructured contextual data required by complex situational awareness solutions."* **Jim Sharpe, CEO Sharpe Engineering**

InfoSphere Streams has achieved initial success with commercial and scientific applications across a wide spectrum of industries.

A communications company is using InfoSphere Streams to mediate over 100,000 call detail records per second and create simultaneous summaries. In Financial Services, an InfoSphere Streams based prototype application can analyze and correlate over 5 million market messages per second to execute algorithmic option trades with an average latency of 30 microseconds. In Healthcare, InfoSphere Streams is being used to correlate information from multiple sensors in a Neonatal Intensive Care Unit with a goal to detect medical conditions up to 24 hours earlier than experienced ICU nurses. In law enforcement and security, InfoSphere Streams is being used to correlate information from multi-modal surveillance systems and deliver real-time intelligence. Electric utilities are developing applications based on InfoSphere Streams to create a smart grid and to forecast electric generation. Other emerging applications on InfoSphere Streams include cybersecurity, intrusion detection, fraud prevention, realtime campaign management, space weather forecasting and crystalline structure analysis using x-ray diffraction techniques.

InfoSphere Streams delivers a wide array of analytic operators, an advanced set of development tools and an autonomic execution environment that scales to scores of nodes. Applications are built using the declarative Stream Processing Language, and the run-time environment manages execution across a cluster of computers. It

can handle on-the-fly processing including aggregations and correlations over time-based or count-based windows. Analytic operators supplied with the system can perform powerful mathematic and text functions. These can be extended with user defined operators, such as custom trading algorithms. Toolkit support allows Independent Software Vendors and central Information Technology departments to easily extend capabilities. Streams lets developers enforce necessary security policies within the application environment.

InfoSphere Streams can leverage strong synergies with existing IT infrastructure and investments. For example streaming data such as stock market quotes can be enriched with an existing high speed in-memory database such as solidDB. By adding information to stock market ticks such as industry or interested traders, more complex analysis and customized alerting can be achieved. Streams can help determine in real time which parts of the incoming data are most relevant to be stored in a database for subsequent analysis. This can reduce storage and administration costs.



## IBM Information Management

### InfoSphere Streams includes:

- Streams Studio: An Integrated Development Environment (IDE) based on Eclipse 3.6.2 to rapidly develop, test and debug streaming applications and Streams Live Graph to visually display jobs and job components executing on the runtime cluster
- Streams runtime: Single server or a cluster of servers without limit to cluster size. High Availability features include ability to detect failing process elements, relocate, restart and optionally restore state.
- Toolkits, adapters and samples:
  - Standard Toolkit with relational, file, communications and utility operators
  - Internet Toolkit with HTTP, HTTPS, FTP, FTPS, RSS and file source operators
  - Database Toolkit with ODBC drivers and a high speed solidDB driver for stream enrichment
  - Mining Toolkit for Predictive Model Markup Language (PMML) scoring, ,
  - Financial Toolkit with WebSphere Front Office v3.0.2.1, Financial Information eXchange (FIX), QuantLib, Library functions that compute equity option derivative values like delta, theta, rho, vega etc
  - Over 50 sample applications and examples of Streams Processing Language artifacts

### Runtime Operating Environment:

- Operating systems supported: Red Hat Enterprise Linux (RHEL) Version 5.3, RHEL Version 5.4 or RHEL Version 5.5. Security Enhanced Linux (SELinux) capability is also supported.
- IBM™ Java SE v6.0-9.0 SDK
- Oracle® Java SE v6.0.x SDK
- Hardware supported: Intel/AMD x86 architecture (32 or 64 bit)

### Development Operating Environment:

- Operating systems supported: Red Hat Enterprise Linux (RHEL) Version 5.3 or RHEL Version 5.4
- Eclipse Platform SDK v3.6.2
- IBM Java SE v6.0-9.0 SDK
- Oracle Java SE v6.0.x SDK
- Hardware supported: Intel/AMD x86 architecture (32 or 64 bit)

### Supported Database Environments:

- IBM DB2® for Linux, Unix and Windows v9.5, 9.7 or higher
- IBM Informix® v11.5
- IBM Netezza® TwinFin
- IBM solidDB® v6.5
- Oracle 11gR2
- Microsoft® SQLServer® 2008
- MySQL™ v5.1

### Supported Browser Environments:

- Firefox v3
- Internet Explorer 6, 7, 8



## IBM Information Management

### **For more information**

To learn more about IBM InfoSphere Streams, please contact your IBM marketing representative or IBM Business Partner, or visit [ibm.com/software/data/infosphere/streams](http://ibm.com/software/data/infosphere/streams)

© Copyright IBM Corporation 2011

IBM Software Group

Route 100

Somers, NY 10589

Produced in the United States of America

April 2011

All Rights Reserved

IBM, the IBM logo, AIX, InfoSphere, Streams, solidDB, WebSphere are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both. Microsoft, Excel, Windows, Windows Server and Windows Vista are registered trademarks of Microsoft Corporation in the United States, other countries or both. Intel is a registered trademark of Intel Corporation or its subsidiaries in the United States, other countries or both. Linux is a registered trademark of Linus Torvalds in the United States, other countries or both.

Other company, product or service names may be trademarks or service marks of others. References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. Offerings are subject to change, extension or withdrawal without notice. All statements regarding IBM future direction or intent are subject to change or withdrawal without notice and represent goals and objectives only.