

Real Time Analytic Processing with IBM InfoSphere Streams V1.2

Highlights

- Perform complex analytics on data in motion
- Handles massive volumes of multiple structured and unstructured data types
- Highly secure environment
- Allows reuse of PMML models
- Incorporates support for WebSphere LLM and Infiniband

".. 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 expected to double every two years primarily aided by growth in unstructured data,. 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 primarily unstructured. Savvy organizations are fast realizing that time lost and data sources missed leads to missed.

IBM InfoSphere Streams can address this gap by providing a state of the art platform to run in-motion analytics on wide variety of relational and non-relational data types. It provides a highly scalable and flexible infrastructure that can support incredible levels of data throughput. Streams achieves this by seamlessly scaling from single server to 125 nodes.

InfoSphere Streams radically extends the state of the art in information processing by simultaneously addressing several technical challenges, including:

- Sub millisecond response times
- Continuous analytics on data in motion
- Extend existing application at runtime without restarting the application
- Promotes reuse of existing Predictive Model Markup Language (PMML) models within streams domain
- Highly secure runtime environment
- Support for WebSphere MQ LLM and Infiniband
- Strong support for developer and administrator productivity
- A set of built-in functions, inputs and analytics routinely used for typical trading applications in the financial services domain

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

"Streams is a game changer" **Merv Adrian,**
IT Market Strategy

".. innovative technology solution to better manage high volumes of real time information.. a significant competitive edge", **CIO,**
TD Securities.

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 70 microseconds. In Radio Astronomy, InfoSphere Streams is being used to correlate information from multiple radio telescopes to deliver enhanced imagery. 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 to security personnel. In the Water Management arena, applications based on InfoSphere Streams are being developed to interpret signals from acoustic sensors to better understand and manage natural ecosystems. Other emerging applications of InfoSphere Streams are in areas such as wildfire monitoring and control, intrusion detection, fraud prevention, call detail record processing, and manufacturing test analysis.

InfoSphere Streams delivers a wide array of analytic operators, an advanced set of development tools and an autonomic execution

environment that scales to 125 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 windows. Analytic operators supplied with the system can perform powerful arithmetic functions and text analytics. These can be extended with user defined operators, such as custom trading algorithms. 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 store in a Database for subsequent analysis. This can reduce storage and administration costs.

InfoSphere Streams includes:

- Streams Studio: An Integrated Development Environment (IDE) based on Eclipse 3.4 which may be used by developers within the organization and deployed on desktop PCs
- Streams runtime: May be deployed on a single server or a cluster of servers
- Tools and adapters: Includes the Mining Toolkit for PMML scoring, StreamSight, ODBC drivers, high speed solidDB drivers, file I/O, TCP/IP communication adapters, generic and domain specific input adapters such as arithmetic and math operators, WebSphere Front Office, Financial Information eXchange (FIX), QuantLib, Library functions that compute equity option derivative values like delta, theta, rho, vega etc

Runtime Operating Environment:

- Operating systems supported: Red Hat Enterprise Linux (RHEL) Version 5.3 or RHEL Version 5.4. SELinux is also supported.
- 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
- Hardware supported: Intel/AMD x86 architecture (32 or 64 bit)

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

© Copyright IBM Corporation 2010

IBM Software Group

Route 100

Somers, NY 10589

Produced in the United States of America

March 2009

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.