



### 중점사항

- 복잡한 데이터 스트림을 실시간으로 분석
- 구조화 또는 비구조화된 방대하고 다양한 데이터를 지속적으로 분석
- 모든 크기의 컴퓨팅 클러스터에서 데이터 마이닝 기능을 통해 기존 애플리케이션을 원활하게 확장

## IBM InfoSphere Streams

### 버전 2.0 - 실시간 분석을 통해 정보의 신뢰성 향상

기업과 소비자들은 불과 몇 년 전에는 믿어지지 않았던 사상 초유의 속도로 데이터를 하루도 빠짐없이 생산해 내고 있습니다. 초대형 마트인 월마트(Walmart)는 시간당 백만 건 이상의 고객 거래를 처리하며, 2.5PB 이상의 데이터를 데이터베이스에 채웁니다. 이 데이터 규모는 미국 의회 도서관에 있는 모든 책 내용의 167배에 달하는 크기입니다.<sup>1</sup> 뉴욕 증권 거래소는 매일 증시가 시작될 때부터 끝날 때까지 1TB의 거래 정보를 캡처하며,<sup>2</sup> 구글(Google)은 하루에 24PB 정도의 데이터를 처리합니다.<sup>3</sup> BBC iPlayer는 매일 최대 7PB의 데이터를 전송하는 것으로 보고되고 있으며,<sup>4</sup> Facebook 사용자들은 20분마다 평균 270만 개의 사진을 FaceBook에 업로드하고 있습니다.<sup>5</sup>

전에는 겪어보지 못한 많고 다양한 정보에 직면한 기업들은 변화무쌍한 데이터로 인해 빠르고 현명한 비즈니스 의사결정을 내리는 데 어려움을 겪고 있습니다. 현명한 기업들은 데이터 마이닝 기법을 개선하여 오디오, 비디오, 이메일 같은 새로운 데이터 소스를 지원하는 것이야말로 기존 시스템의 가치를 높이고, 소비자 의식 동향에 좀 더 빠르게 대응하여 새로운 시장 기회를 발견하고 획기적인 신제품을 시장에 내놓을 수 있는 중요한 경쟁력을 확보하는 길임을 빠르게 실감하고 있습니다.

이러한 차세대 분석 처리 기법 중 하나인 IBM® InfoSphere® Streams는 급증하는 데이터 불륨을 실행 가능한 정보와 비즈니스 통찰력으로 바꿀 수 있는 최첨단 컴퓨팅 플랫폼을 제공합니다. InfoSphere Streams는 Big Data를 지원하는 IBM 플랫폼의 핵심 요소로서 뛰어난 확장성과 속도를 자랑하는 소프트웨어 인프라를 제공합니다. 이 인프라를 통해 기업들은 수천 개의 소스에서 실시간으로 유입되는 다양한 유형의 폭발적인 데이터를 실시간으로 분석할 수 있습니다. InfoSphere Streams를 통해 기업은 필요할 때 언제든지 핵심 비즈니스 데이터를 확보하여 데이터에 입각한 의사결정을 내릴 수 있습니다.



“미국 정부는 이질적인 실시간 데이터 스트림의 확장성과 복잡도를 수용하고 데이터 스트림을 빠른 속도로 분석할 수 있는 급진적인 데이터 분석 기법을 개발하기 위해 2003년부터 IBM Research와 협력해 왔습니다. 이 프로젝트를 위해 IBM이 구현한 기술이 바로 지금의 IBM InfoSphere Streams입니다. 이 프로젝트의 결과는 매우 성공적이었습니다. 미국 정부는 IBM의 분석 기법을 다른 기관에도 배치하여 앞으로 다양한 프로젝트에서 더 큰 성공을 거둘 수 있도록 할 예정입니다.”

- US 미국 정부

### 정보 처리의 새로운 패러다임

미국 정부가 IBM Research과 함께 새로운 분야를 개척하는 데 핵심적인 기반이 된 InfoSphere Streams는 끊이지 않는 데이터 스트림의 방대한 볼륨을 수용, 필터링 및 분석하여 연관된 애플리케이션을 개발하고 실행할 수 있는 개발 플랫폼과 런타임 환경을 제공합니다. InfoSphere Streams는 구조화 또는 비구조화된 데이터 소스에서 유입되거나, 아래와 같이 매우 다양한 디지털 정보가 포함된 데이터 스트림을 지원합니다.

- 텍스트 파일, 스프레드시트, 이미지, 비디오 및 오디오 자료
- 이메일, 대화, 인스턴트 메시징, 웹 트래픽, 블로그, 소셜 네트워킹 사이트
- 금융 거래, 고객 서비스 기록, 무전기 데이터, 시스템 및 애플리케이션 로그
- 위성 데이터, GPS 데이터, 센서 로그, 출근 기록계 및 액세스 기록

InfoSphere Streams는 복잡한 데이터를 신속하게 분석할 수 있는 강력한 컴퓨팅 플랫폼에 이러한 이기종 데이터 유형을 융합시킵니다.

단 한 대의 서버에서 무한대에 가까운 노드를 수용할 수 있는 이 플랫폼은 기존의 인프라와 애플리케이션에 통합할 수 있는 수 많은 분석 연산자와 고급 개발 도구 세트를 제공합니다. InfoSphere Streams 기술을 소유한 기업은 현재의 IT 투자를 확대하고, 이미 익숙하게 사용하는 애플리케이션에 새로운 수준의 비즈니스 인텔리전스를 통합할 수 있습니다.

예를 들어, 금융 서비스 기업은 InfoSphere Streams를 통해 내부 데이터베이스의 데이터(예: 산업 거래자 또는 관심 있는 거래자)에 실시간 주식 시세를 통합할 수 있습니다. 주식 시장 틱에 컨텍스트를 추가하면 보다 복잡한 분석을 수행하고, 경보를 사용자 정의할 수 있습니다. 이와 동시에, InfoSphere Streams는 유입되는 데이터 중 추후 분석에 유용한 부분을 실시간으로 판별하여 해당 부분만 데이터베이스에 저장할 수 있도록 지원하므로 스토리지 비용과 관리 비용을 절감할 수 있습니다.

### 민첩한 개발 환경을 지원하는 고급 도구

개발자와 관리자는 InfoSphere Streams를 통해 보다 강력한 데이터 마이닝 기능을 이전보다 더 쉽게 구축하여 애플리케이션 및 워크로드의 성능을 개선할 수 있습니다. InfoSphere Streams는 특정 업계나 기능에 적합한 솔루션을 쉽게 개발할 수 있는 통합 툴킷과 샘플 애플리케이션을 제공합니다. 여기에는 가장 자주 사용되는 연산자가 포함된 데이터 마이닝 툴킷, 금융 서비스 툴킷, 표준 툴킷 등이 포함됩니다.

InfoSphere Streams는 Eclipse 기반의 대화식 개발 환경(IDE)인 Streams Studio도 제공합니다. Streams Studio에서는 편집기와 마법사를 사용하여 InfoSphere Streams 애플리케이션을 개발할 수 있으며, InfoSphere Streams 애플리케이션의 애플리케이션 구조 그래프와 런타임 모니터링을 사용할 수 있습니다. InfoSphere Streams는 많은 스트리밍 애플리케이션을 전달해 온 IBM의 풍부한 경험을 바탕으로 매우 정교한 데이터 분석 기능을 제공하기 때문에 고객은 애플리케이션의 기능을 쉽고 빠르게 확장할 수 있습니다.

## Big Data 분석 측면의 전환

InfoSphere Streams를 도입한 기업은 이미 지나간 동향과 이벤트에 뒤늦게 대응하기 보다는 현재의 동향과 이벤트를 실시간으로 확인할 수 있습니다. 빠르게 움직이는 거대한 데이터 스트림에 의미를 부여하는 기능 덕분에 이 기술은 이미 다양한 업계에서 성공을 거뒀습니다. 예를 들어, 현재 InfoSphere Streams를 사용하는 한 통신업체는 초당 20만 건의 통화 세부 기록을 중재하는 동시에 이러한 데이터를 요약할 수 있습니다. 일반적으로 12시간이 소요되는 요약 프로세스를 거의 즉시 완료할 수 있기 때문에 고객은 비즈니스 영향을 개선하는 동시에 실시간 통화 정보를 토대로 중요한 의사결정을 내릴 수도 있습니다.

금융 서비스 부문에서 InfoSphere Streams 기반의 애플리케이션을 사용할 경우 초당 5백만 건의 시장 메시지를 분석하고 연관시킬 수 있으며, 평균 30 마이크로초 안에 알고리즘 기반의 옵션 거래를 실행함으로써 경쟁업체보다 더 빠르게 시장의 동향을 파악할 수 있는 민첩성을 확보할 수 있습니다. 의료 부문에서는 신생아 보호실에 있는 여러 센서 정보를 분석하는 데 InfoSphere Streams를 사용하여 환자의 건강 상태를 최대 24시간 전에 미리 예측하는 전 방향 예지 예측 기법을 구현함으로써 질병을 더욱 효과적으로 치료할 수 있습니다.

데이터를 빠르고 정확하게 분석하여 가장 최근의 정보를 토대로 의사결정을 내려야 하는 기업이라면 InfoSphere Streams를 통해 많은 혜택을 누릴 수 있습니다. InfoSphere Streams로 경쟁적 우위를 확보할 수 있는 방법에 대해 알아보려면 [ibm.com/software/data/infosphere/streams](http://ibm.com/software/data/infosphere/streams)를 참조하십시오.

*“TerraEchos의 개발자들은 민첩한 Streams Processing Language 덕분에 애플리케이션 전달 시간을 45% 단축할 수 있다.”*

- Alex Philp, CEO, TerraEchos

## InfoSphere Streams V2.0: 특징 및 개선사항

- 기존의 Java 또는 C++ 코드, PMML(Predictive Model Markup Language) 모델 재사용 지원
- IBM WebSphere® MQ Low Latency Messaging (LLM) 전송 기술 및 InfiniBand 지원
- Internet Protocol 버전 6(IPv6) 네트워크와 통신 지원
- Streams Studio: 스트리밍 애플리케이션을 빠르게 개발, 테스트, 디버깅할 수 있는 Eclipse 3.6.2 기반의 통합 개발 환경(IDE) / Streams Live Graph: 런타임 클러스터에서 실행 중인 작업 및 작업 구성요소를 시각적으로 표시
- Streams Runtime: 클러스터 크기에 제한이 없는 서버 클러스터 또는 단일 서버, 오류가 있는 프로세스 요소를 찾아내는 고가용성 기능 제공. 상태 전환, 재시작, 복원 가능
- 툴킷, 어댑터 및 샘플:
  - 관계, 파일, 통신 및 유틸리티 연산자가 포함된 표준 툴킷
  - HTTP, HTTPS, FTP, FTPS, RSS 및 파일 소스 연산자가 포함된 인터넷 툴킷
  - 스트림 강화에 필요한 ODBC 드라이버와 고속 IBM solidDB® 드라이버가 포함된 데이터베이스 툴킷
  - PMML 스코어링을 위한 마이닝 툴킷
  - 주식 옵션의 파생 가치를 계산할 수 있는 IBM WebSphere Front Office V3.0.2.1, Financial Information eXchange(FIX) 및 QuantLib 라이브러리 함수가 포함된 금융 툴킷
  - 50개 이상의 샘플 애플리케이션과 Streams Processing Language 아티팩트 예제

---

## InfoSphere Streams: 시스템 요구사항

### 런타임 운영 환경

- 지원되는 운영 체제: Red Hat Enterprise Linux(RHEL) 버전 5.3, RHEL 버전 5.4, RHEL 버전 5.5, Security Enhanced Linux(SELinux) 기능 지원
- IBM Java SE v6.0-9.0 SDK
- Oracle Java SE v6.0.x SDK
- 지원되는 하드웨어: Intel/AMD x86 아키텍처 (32 또는 64비트)

### 개발 운영 환경

- 지원되는 운영 체제: RHEL 버전 5.3, RHEL 버전 5.4
- Eclipse Platform SDK v3.6.2
- IBM Java SE v6.0-9.0 SDK
- Oracle Java SE v6.0.x SDK
- 지원되는 하드웨어: Intel/AMD x86 아키텍처 (32 또는 64비트)

### 지원되는 데이터베이스 환경

- IBM DB2® for Linux, UNIX 및 Windows 9.5, 9.7 이상
- IBM Informix® v11.5
- IBM Netezza® Twinfin®
- IBM solidDB V6.5
- Oracle 11gR2
- Microsoft SQL Server 2008
- MySQL v5.1

### 지원되는 브라우저 환경

- Firefox 3
- Microsoft Internet Explorer 6, 7, 8



© Copyright IBM Corporation 2011

(135-270) 서울시 강남구 도곡동 467-12  
군인공제회관빌딩

한국아이비엠주식회사  
고객만족센터

TEL: (02)3781-7114  
www.ibm.com/kr

2011년 10월

Printed in Korea  
All Rights Reserved

IBM, IBM 로고, ibm.com 및 InfoSphere는 미국 또는 기타 국가에서 International Business Machines Corporation의 상표 또는 등록 상표입니다. 이와 함께 기타 IBM 상표가 기재된 용어가 상표 기호(® 또는 ™)와 함께 이 정보에 처음 표시된 경우, 이와 같은 기호는 이 정보를 발행할 때 미국에서 IBM이 소유한 등록상표 또는 일반 법적 상표입니다. 또한 이러한 상표는 기타 국가에서 등록상표 또는 일반 법적 상표입니다. 현재 IBM 상표 목록은 웹 "저작권 및 상표 정보" ([www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml))에 있습니다.

Netezza 및 Twinfin은 IBM 회사인 Netezza Corporation의 상표 또는 등록 상표입니다.

Intel, Intel 로고, Intel Inside, Intel Inside 로고, Celeron, Intel Xeon, Intel SpeedStep, Itanium 및 Pentium은 미국 또는 기타 국가에서 Intel Corporation 또는 자회사의 상표 또는 등록 상표입니다.

Java 및 모든 Java 기반 상표 및 로고는 Oracle 및/또는 그 계열사의 상표 또는 등록 상표입니다.

Linux는 미국 또는 기타 국가에서 Linus Torvalds의 등록 상표입니다.

Microsoft, Windows, Windows NT 및 Windows 로고는 미국 또는 기타 국가에서 Microsoft Corporation의 상표입니다.

UNIX는 미국 또는 기타 국가에서 The Open Group의 등록 상표입니다.

기타 회사, 제품 및 서비스 이름은 해당 회사의 상표 또는 서비스 표입니다.

1 "Data, data everywhere." The Economist, February 25, 2010.  
[www.economist.com/node/15557443?story\\_id=15557443](http://www.economist.com/node/15557443?story_id=15557443)

2 Henschen, Doug. "New York Stock Exchange Ticks on Data Warehouse Appliances." *Information Week*, May 16, 2008.  
[www.informationweek.com/news/software/bi/207800705](http://www.informationweek.com/news/software/bi/207800705)

3 Dean, Jeffrey and Ghemawat, Sanjay. "MapReduce: simplified data processing on large clusters." *Communications of the ACM*, Volume 51, Issue 1, January 2008.  
<http://portal.acm.org/citation.cfm?doid=1327452.1327492>

4 Lanxon, Nate. "iPlayer uncovered: What powers the BBC's epic creation?" Crave/CNET UK, <http://crave.cnet.co.uk/software/iplayer-uncovered-what-powers-the-bbcs-epic-creation-49302215>

5 NW Linux: Linux and Technology News, December 31, 2010, <http://nwlinux.com/facebook-data-and-post-statistics-over-20-minutes-time>



재활용 가능, 재활용 하십시오.