



Bringing Big Data to the Enterprise



## InfoSphere BigInsights의 가치

IBM Silicon Valley Laboratory  
업데이트: 2011년 9월

## InfoSphere BigInsights의 가치

여러분은 아마도 InfoSphere BigInsights에 대해 들어본 적이 있을 겁니다. InfoSphere BigInsights는 다양하고 많은 데이터 속에 숨어 있는 비즈니스 통찰력을 찾아내고 분석할 수 있도록 지원하는 IBM의 소프트웨어 플랫폼입니다. 여기서 말하는 데이터란 쓸모가 없거나 기존의 수단으로 처리하기가 어려워 종종 폐기 또는 무시되는 데이터를 의미하며 로그 레코드, 클릭 스트림, 소셜 미디어 데이터, 뉴스 피드, 이메일, 전자 센서 출력, 일부 트랜잭션 데이터까지도 포함할 수 있습니다. 이러한 데이터와 함께 점점 빠르게 증가하고 다양해지는 데이터들을 쉽게 처리할 수 있도록 IBM은 Apache™의 개방형 소스 Hadoop™ 프로젝트를 포함하는 BigInsights를 개발했습니다. 그럼 이 IBM 플랫폼이 왜 특별한지, 어떠한 가치가 있는지 함께 알아보겠습니다.

### 분석과 검색

BigInsights를 사용하면 내장된 처리 엔진과 어노테이터 라이브러리를 통해 대량의 텍스트 기반 정보를 쉽게 분석할 수 있습니다. 개발자들은 문서와 메시지에서 사람, 이메일 주소, 주소, 전화번호, URL, 공동투자 업체, 협력업체 등 관심 있는 항목을 빠르게 조회하고 식별할 수 있습니다. 이러한 기능으로 기업은 비구조화된 텍스트 데이터에 숨겨진 연관된 비즈니스 정보의 컨텍스트와 내용을 보다 쉽게 이해할 수 있습니다.

또한 프로그래머는 BigInsights의 Eclipse 기반 플러그인을 사용하여 자신만의 텍스트 분석 함수를 작성할 수 있습니다. 패턴 검색, 표현식 빌더, 테스트 환경이 내장되어 빠른 프로토타이핑이 가능하며, 테스트를 통해 특정 애플리케이션의 요구사항에 맞는 복잡한 텍스트 분석 함수를 작성할 수 있습니다.

비즈니스 분석가와 프로그래머 외의 사용자들은 스프레드시트형 데이터 검색 및 시각화 도구를 통해 많은 양의 데이터를 분석할 수 있습니다. 이 도구는 데이터를 수집 및 통합한 후 내장된 함수와 매크로로 데이터를 탐색하고 조작하는 기능과 차트를 작성하고 결과를 내보낼 수 있는 기능을 제공하는 단순한 GUI 인터페이스 도구입니다.

### 전사적 소프트웨어 통합

BigInsights와 “Big Data”에 대한 분석을 기존의 전사적 소프트웨어에 통합하는 것은 IBM의 핵심 이니셔티브 중 하나입니다. 이를 위해 BigInsights는 DB2, Netezza 같은 유명 데이터 웨어하우스 플랫폼에 대한 연결을 제공합니다.

특히, BigInsights는 Netezza와 DB2에 대한 JDBC 연결을 제공하기 때문에 개발자들이 데이터베이스의 기본적인 병렬 처리 기능을 사용하는 방법으로 각 데이터 소스와 데이터를 주고 받을 수 있어 효율성과 확장성이 보장됩니다. 예를 들어, BigInsights 개발자가 관계형 데이터베이스의 참조 데이터를 BigInsights에서 관리되는 데이터와 결합하면 보다 정교하고 확장된 분석이 가능합니다. BigInsights는 다른 관계형 데이터 소스에도 액세스할 수 있도록 일반 JDBC 커넥터를 제공합니다.

게다가, BigInsights는 배포 가능한 DB2 사용자 정의 함수 샘플도 제공하므로, DB2 사용자는 BigInsights에서 조회를 실행하고 DB2 데이터에 출력을 결합하여 결과를 DB2 사용자 및 애플리케이션에 제공할 수 있습니다.

IBM의 접근 방법은 기업의 “Big Data”와 기존 데이터를 서로 격리시키지 않습니다. 오히려 기존의 전사적 소프트웨어 플랫폼의 가치를 최대화하고, 기존 플랫폼에 맞지 않거나 쓸모 없는 분석 워크로드에는 BigInsights를 활용하여 통합적인 방식으로 기업의 비즈니스 분석 기능을 확장시킬 수 있도록 지원합니다.

## 관리 및 플랫폼 개선사항

IBM의 웹 기반 관리 콘솔은 기업의 BigInsights 환경을 실시간으로 보여줍니다. 이 콘솔에서는 노드를 시작 또는 중지하고, MapReduce 작업 상태를 확인하며, 로그 레코드를 검토할 수 있습니다. 또한 전체 플랫폼 상태를 평가하고, 선택적 컴포넌트를 시작 또는 중지하며, 분산 파일 시스템을 탐색하는 등 다양한 작업을 수행할 수 있습니다.

또한 BigInsights는 내장 웹 콘솔에 LDAP 인증을 적용하여 강력한 보안까지 제공합니다. 관리자는 LDAP 및 역방향 프록시 지원을 활용하여 승인되지 않은 사용자의 접근을 제한할 수 있습니다.

워크로드 관리가 고민이라면 유연한 작업 스케줄링 메커니즘을 사용하여 각 작업의 실행 시간에 따라 할당할 자원을 미세하게 조정할 수 있습니다. 예를 들어, BigInsights 스케줄러를 활용하여 작은 작업에 가장 많은 자원을 할당하도록 설정하면 해당 작업을 빠르게 완료할 수 있고, 평균 응답 시간 목표도 달성할 수 있습니다. 이 작업 스케줄링 옵션 외에 Hadoop의 선입선출(FIFO) 방법과 “공정” 스케줄링 방법도 사용할 수 있습니다.

이 외에도, BigInsights는 텍스트 기반의 압축 데이터를 효과적으로 처리하거나 런타임 환경 및 워크로드에 알맞은 방법으로 특정 애플리케이션 태스크를 처리할 수 있는 기능 등 여러 가지 성능적 장점을 지니고 있습니다.

## 운영 위험 감소

BigInsights에는 IBM의 표준 소프트웨어 라이선싱과 지원 계약이 적용되기 때문에 특정한 개방형 소스 프로젝트의 이용 약관에 대한 법적 사안을 고려하지 않아도 되며, 기술 지원 연락처가 없는 소프트웨어를 배치하므로 운영상의 위험 초래로 인한 염려가 줄어듭니다.

여러 개방형 소스 프로젝트로 구성된 플랫폼에 필수 애플리케이션을 배치하기 위해서는 코드 베이스에 대한 고도의 지식을 갖추고, 잠재된 버그나 결함을 수정할 수 있어야 합니다. 커뮤니티 기반의 지원 그룹과 포럼은 매우 유용할 수 있지만, 서비스 레벨 계약을 준수할 의무가 없거나 긴급한 문의에 응답하지 않을 수도 있습니다. 따라서, 각 기업이 플랫폼 문제를 직접 진단하고 해결하며, 자체적으로 플랫폼 업그레이드를 수행하여 첨단 기술을 유지 보수해야 합니다. 하지만 BigInsights가 있으면 이러한 복잡함과 운영상의 위험을 방지할 수 있습니다.

## 빠른 설치와 배치

BigInsights는 널리 사용되는 개방형 소스 및 IBM 기술 컬렉션을 포함하는 사전 구성된 플랫폼을 제공하므로 고객은 Big Data 프로젝트를 더욱 빠르게 시작할 수 있습니다.

전체 소프트웨어 플랫폼을 이루는 개방형 소스 프로젝트를 하나씩 다운로드 하여 구성하고 테스트할 필요 없이, BigInsights에 내장된 GUI 기반의 설치 도구를 호출하여 단 몇 분 안에 플랫폼을 시작할 수 있습니다. 대부분의 개방형 소스 프로젝트는 다른 컴포넌트의 특정 버전과 호환되지 않는 소프트웨어를 사전 설치하도록 요구하는 경우가 많은데 이러한 경우에 특히 이 도구는 매우 유용합니다. 또한 일부 개방형 소스 오픈링을 사용할 경우 코드를 컴파일하고 설치한 후 추가 작업이 필요하기 때문에 가동하고 실행하는 데 많은 시간이 필요하지만, 이 설치 도구를 사용하면 이러한 작업에 소요되는 시간을 절약할 수 있습니다.

BigInsights의 유연한 설치 도구를 통해 설치할 선택적 구성요소와 플랫폼 구성 방법을 지정할 수 있습니다. BigInsights 설치 진행 상황은 실시간으로 자세히 보고되고, 내장된 “상태 검사” 기능으로 설치가 올바르게 완료되었는지 자동으로 확인할 수 있습니다.

BigInsights가 가진 다양한 기술에 대한 자세한 사항은 IBM의 공지 문서와 BigInsights InfoCenter를 참조하십시오. Hadoop, Pig, Hive, HBase, Jaql, Lucene, Zookeeper, Flume, Oozie, Avro, 웹 관리 콘솔, 텍스트 분석 (스프레드시트형 분석 도구) 등에 대해 기술한 IBM 배포판이 포함된 Enterprise Edition을 바로 찾으실 수 있습니다.

## IBM: Big Data 소프트웨어와 전문 지식의 원천

이제 여러분도 IBM과 함께 다양한 전사적 소프트웨어, 하드웨어, 서비스와 관련된 폭넓은 전문 지식을 바탕으로 자신감과 확신을 가지고 Big Data 프로젝트를 시작할 수 있습니다. IBM이 제공하는 자세한 사항을 <http://www.ibm.com/bigdata>에서 지금 바로 확인하십시오.



© Copyright IBM Corporation 2011

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

한국아이비엠주식회사  
고객만족센터

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

2011년 10월

All Rights Reserved

본 문서 또는 본 문서의 일부를 IBM의 사전 동의 없이 다른 언어로 번역하거나 어떤 형식이나 방법으로든 복사 또는 복제할 수 없습니다.

IBM 및 IBM 로고는 미국 또는 기타 국가에서 사용되는 International Business Machines Corporation의 상표 또는 등록상표입니다. 기타 회사, 제품 및 서비스 이름은 타사의 상표 또는 서비스표입니다.

다음 단락은 현지법과 상충하는 영국이나 기타 국가에서는 적용되지 않습니다. IBM은 타인의 권리 침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여(단, 이에 한하지 않음) 묵시적이든 명시적이든 어떠한 종류의 보증 없이 이 책을 "현상대로" 제공합니다. 일부 국가에서는 특정 거래에서 명시적 또는 묵시적 보증의 면책사항을 허용하지 않으므로, 이 사항이 적용되지 않을 수도 있습니다.

본 문서에 포함된 모든 성능 데이터는 제한된 환경에서 산출된 것입니다. 따라서 다른 운영 환경에서 얻어진 결과는 상당히 다를 수 있습니다. 일부 성능은 개발 레벨 상태의 시스템에서 측정되었을 수 있으므로 이러한 측정치가 일반적으로 사용되고 있는 시스템에서도 동등하게 나타날 것이라고는 보증할 수 없습니다. 또한 일부 성능은 추정을 통해 추측되었을 수도 있으므로, 실제 결과는 다를 수 있습니다. 이 책의 사용자는 해당 데이터를 본인의 특정 환경에서 검증해야 합니다.

본 문서의 정보는 별도의 통지 없이 변경될 수 있습니다. IBM은 정보의 변경에 대한 통지 의무 없이 정보를 변경할 수 있습니다. IBM은 본 문서의 정보를 최신 상태로 유지할 것을 약속하지 않습니다.

여기서 IBM 제품 또는 서비스를 언급하는 것이 IBM이 영업하는 모든 국가에서 이들 제품 또는 서비스를 사용할 수 있다는 것을 의미하지는 않습니다.

C. M. Saracco ([saracco@us.ibm.com](mailto:saracco@us.ibm.com))