



NoSQL DB 인기도 조사, 하둡 진영과 몽고 DB 의 경쟁 양상

○ 비즈니스 인텔리전스(BI) 전문기업인 제스퍼소프트(Jaspersoft)가 NoSQL DB 의 인기를 알 수 있는 빅데이터 지수(JBDI)를 발표

- 빅데이터는 대규모의 정형 및 비정형 데이터를 분석하는 것이므로, 정형 데이터를 SQL 쿼리로 관리하는 관계형 데이터베이스관리시스템(RDBMS)과 달리 SQL 외에 다양한 쿼리 시스템을 사용할 수 있는 NoSQL(Not-Only SQL) DB 를 필요로 함
- 대표적인 NoSQL DB 로는 구글의 빅테이블(BigTable) 클론프로젝트인 하둡(Hadoop) HBase, 페이스북이 만들어 공개한 카산드라(Cassandra) 등이 있으며, 뛰어난 확장성과 분산처리가 가장 큰 특징
- 대부분의 NoSQL DB 는 작업량을 복수의 머신에 분배하고, 자동적인 파편화(Sharding)를 지원하며, 파편화는 데이터를 여러 조각으로 분리하고 각 조각을 담당하는 머신들을 동기화한 후, 질의가 들어오면 해당 파편을 처리하는 머신에 실행을 지시하는 것
- 제스퍼소프트의 BI 소프트웨어는 NoSQL DB 들을 포함하여 다수의 데이터 소스들과 연결되므로 어떤 DB 가 빈번히 연결되느냐를 통해 인기를 측정
- 2011 년 1 월부터 2012 년 3 월 사이에 제스퍼소프트의 ‘NoSQL 커넥터’ 다운로드 수를 통해 조사한 인기를 보면 하둡 진영, 몽고 DB, 카산드라 순서

○ JBDI 의 조사 결과의 특징은 하둡 진영의 인기가 높아지는 반면, 대표적인 NoSQL 데이터베이스로 알려진 카산드라의 인기가 매우 낮다는 사실

- 단일 제품으로 보면 몽고 DB 의 인기가 가장 높으나, 하이브(Hive), H 베이스(HBase), 에이브로(Avro) 등 하둡 진영의 인기를 합하면 몽고 DB 보다 높게 나옴
- 하둡 진영 3 개 제품의 인기를 구분해 보면, SQL 과 보다 유사한 하이브가 H 베이스 보다 50% 가량 더 많이 연결된 것을 알 수 있음(3,682 대 2,360)
- 이는 빅데이터 응용분야의 인기가 점점 높아지고 하둡 생태계를 접하는 사람들이 늘

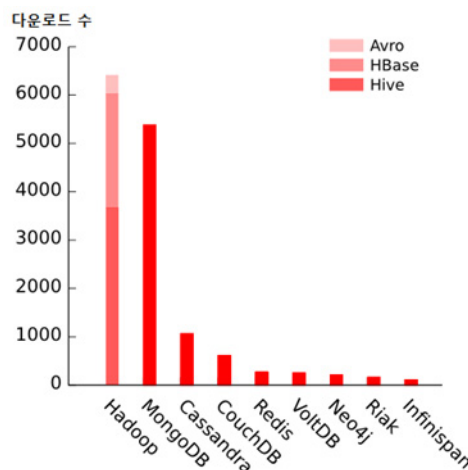
* 본 내용과 관련된 사항은 정보서비스팀(☎ 042-710-1771)과 ㈜크로센트 박종훈 수석 아키텍트(☎ 02-2078-2088)에게 문의하시기 바랍니다.

** 본 내용은 필자의 주관적인 의견이며 NIPA 의 공식적인 입장이 아님을 밝힙니다.

- 어나고 있으며, 처음에는 진입장벽을 낮춰주는 보다 친숙한 도구를 찾고 있음을 시사
- 하둡 진영에서 에이브로의 인기도가 상대적으로 낮은 것은 에이브로가 데이터 집약적인 응용분야를 위해 설계된 하둡에 맞게 시리얼화된 포맷이어서 H 베이스나 하이브의 방법론에 비해 범용성이 떨어지기 때문
- 카산드라는 하둡 진영 및 몽고DB와 큰 격차로 3위에 랭크된 반면, 레디스(Redis)는 급부상하여 카산드라와 카우치DB(CouchDB)에 근접

○ JBDI에서 단일 제품으로 가장 높은 순위를 차지한 몽고DB는 2008년에 10Gen이 개발한 NoSQL DB로 최근 이용이 급증

- 10Gen은 벤처캐피탈의 후원을 받고 있는 신생기업으로 높은 기술력을 인정받아 최근 4년간 급성장하고 있으며, 몽고DB라는 이름은 'humongous(거대한)'의 철자에서 가운데 다섯 개를 따온 것으로 빅데이터를 의미
- 몽고DB는 오픈소스로 개발된 문서 지향적 스토리지로서 스키마가 없고 전체 색인을 통해 다이내믹 쿼리를 제공하며, 문서 저장에 가장 적합하기도 하지만 비디오와 이미지 등 여타 비구조화 데이터 처리도 가능
- 프로그래밍 모델은 데이터 교환을 위해 고안된 경량의 텍스트 기반 개방형 표준인 제이슨(Javascript Object Notation: JSON)을 바이너리로 코딩한 BSON
- 몽고DB의 또 다른 특징은 확장성 있는 아키텍처로서 자동적인 파편화를 이용하여

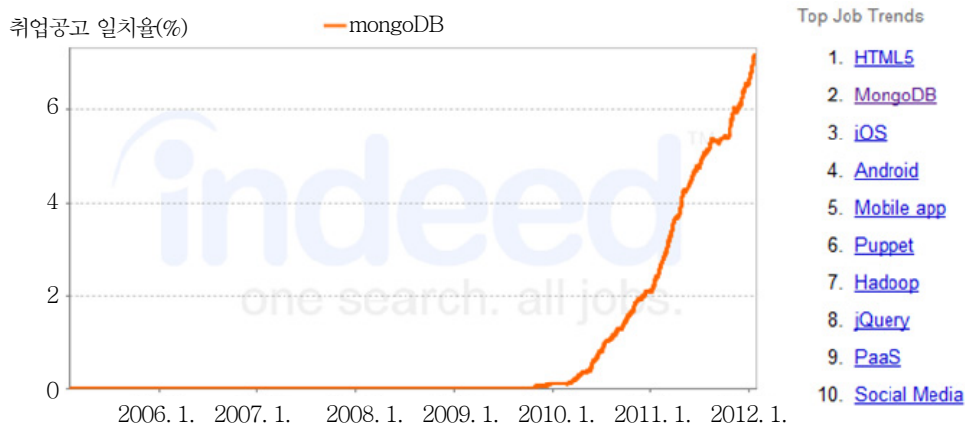


<자료>: Japersoft, 2012. 3.

(그림 1) 제스퍼소프트의 NoSQL 커넥터 인기도(2011.1.~2012. 3.)

수평적으로 규모를 확대할 수 있으며, 몽고DB는 고가용성, 복구, 퍼포먼스를 위해 마스터-슬레이브 또는 P2P 복제를 지원

- 몽고DB는 아마존 AWS, VM 웨어, 레드햇 등과 파트너십을 체결하면서 사용이 급증
 - 몽고DB는 아마존의 AWS를 경유한 클라우드를 통해 배포가 가능하며, 10Gen은 지원 서비스, 교육, 컨설팅을 통해 수익을 얻고 있음
 - 이 밖에 VM 웨어, 레드햇 등과도 제휴를 맺고 있으며, 모든 클라우드 플랫폼 제공업체들이 자신들의 고객들에게 옵션으로 몽고DB를 제공
 - 몽고DB는 현재 400개 이상의 고객을 확보하고 있는데, 특히 1,000만 명의 이용자를 보유한 위치기반 체크인 서비스 포스퀘어가 도입하면서 유명해지기 시작
 - 이밖에 크레이그스리스트 등 유수의 인터넷 기업들이 다수 이용하고 있으나 주 이용고객은 저가 하드웨어를 이용해 고가용성과 뛰어난 퍼포먼스를 추구하는 벤처기업들
 - 몽고DB에 관한 많은 교재가 출판되어 있고 개발자 커뮤니티 역시 빠르게 증가하는 추세로서, 미국의 대표적 IT 취업사이트인 인디드닷컴에 따르면, 몽고DB가 HTML5에 이어 현재 두 번째로 인기가 높은 키워드로 부상
- 그러나 최근 개발자 커뮤니티를 중심으로 몽고DB의 성능과 안정성에 대한 문제를 제기하기 시작
 - 2011년 말부터 기업용 소프트웨어 개발자 커뮤니티인 인포메이션큐(InfoQ)를 시작으로



<자료>: Indeed.com, 2012. 4. 1 현재.

(그림 2) 인디드닷컴의 IT 취업 키워드 트렌드 순위

- 몽고DB 개발자 커뮤니티와 해커뉴스 등에서 몽고DB 성능에 대한 문제 제기가 증가
 - 문제 제기는 주로 몽고DB의 데이터 안정성에 대한 것으로, 개발자들은 몽고DB가 명성에 비해 안정성이 취약하다고 지적
 - 몽고DB는 데이터 처리가 매우 빠르다는 점을 인정받고 있는데, 이는 데이터를 쓸 때 메모리에 데이터를 저장한 후 백그라운드 스레드를 통해 디스크에 기록하기 때문
 - 인메모리(In Memory) 데이터베이스 이용 방식은 빅데이터 처리에 있어 필수적인 기능으로 하둡 진영 역시 빠른 속도로 인해 최근 빅데이터 시장에서 주목받고 있음
 - 영국의 IT 전문지 더레지스터는 몽고DB를 실제 적용해보니 하둡처럼 대용량 시스템에 적합한 구조가 아닐 뿐더러 규모 확장에 한계가 있다고 지적
 - 몽고DB의 속도는 색인 크기와 메모리에 좌우되는데, 메모리가 가득 차 HDD로 내려가 데이터를 처리할 경우 처리 속도가 급감한다고 설명
- 이에 따라 NoSQL 데이터베이스 시장에서 향후 하둡 진영과 몽고DB의 위상에 변화가 생길지 관심
 - 더레지스터는 몽고DB 개발자를 원하는 시장의 수요가 많음에도 불구하고, 막상 개발자들은 몽고DB를 활용하려 하지 않으며, 몽고DB에서 카산드라로 이동하려는 움직임이 눈에 띈다고 보도
 - 또한 카산드라 외에 H 베이스, 하이브처럼 더 나은 데이터 처리 능력을 보이는 하둡 진영의 오픈소스 NoSQL들이 몽고DB의 인기를 대체할 것으로 전망
- 빅데이터 시장이 아직 초기임을 감안하면, 결국 시장의 전개 과정에서 데이터의 특성에 따라 NoSQL 제품들도 특화되며 전문화될 수 있을 것
 - RDBMS의 경우도 처리하려는 데이터의 크기와 모델에 따라 데이터베이스들의 장단점이 있으며, 결국 선택은 용도에 맞게 결정되는 경향
 - 빅데이터 시장은 빅데이터에 대한 정의조차 아직 정립되지 않았을 정도로 이제 막 시작되었기 때문에, 어떤 NoSQL DB가 가장 적합하고 우수한지는 빅데이터 시장의 전개와 성숙도에 따라 결정될 일
 - 빅데이터의 유형은 SNS의 데이터, 센서 데이터, 위치측정 데이터 등으로 다양하기 때문에 용도별 데이터 처리에 적합하게 NoSQL DB들이 분화해 갈 것으로 예상
 - 개발자 커뮤니티들은 최근 몽고DB의 인기도가 높았던 이유는 개발자들이 가장 쉽게

접할 수 있었기 때문으로 설명하며, 몽고DB 대신 카산드라와 H 베이스 등 하둡 진영에 대한 관심이 높아지고 있지만 기술적으로는 더욱 난이도가 높다는 점을 지적

- 또한 하둡과 같은 오픈 소스는 발전 단계에서 생각지 못한 변수가 항상 발생하기 때문에 카산드라든 하둡이든 지속적으로 기술 발전 내용을 따라잡으려는 노력이 중요하지 유행에 따라 기술을 선택하는 우를 범하지 말 것을 충고

(RedMonk, 3. 26 & SYS-CON & Bloter.net, 3. 29.)