

스폰서



**GEEK GUIDE**



혁신적인 앱 개발자가

속도가 빠른

**OSDBMS**를

좋아하는 이유



# 목 차

|   |           |
|---|-----------|
| 스폰서 소개 .....  | 4         |
| 개요 .....  | 5         |
| 혁신적인 앱 개발자가 직면한 과제 .....                            | 6         |
| 오픈 소스가 제공하는 이점 .....                                | 8         |
| MongoDB 문서 스토어 DBMS.....                            | 12        |
| EDB Postgres Advanced Server.....                   | 14        |
| Neo4j 그래프 데이터베이스.....                               | 16        |
| 인 메모리 데이터베이스 시스템.....                               | 18        |
| Redis.....  | 19        |
| Kinetica로 GPU 가속화 .....                             | 21        |
| <b>OSDBMS를 IBM OpenPOWER Systems 상에 구현하는 것이 올바른</b> |           |
| <b>선택인 이유 .....</b>                                 | <b>22</b> |
| MongoDB.....  | 24        |
| EDB Postgres Advanced Server .....                  | 24        |
| Redis.....  | 25        |
| Neo4j.....  | 26        |
| Kinetica.....                                       | 26        |
| <b>결론 .....</b>                                     | <b>27</b> |

Ted Schmidt는 제조 부문의 마케팅 및 전자 상거래 솔루션을 전문으로 하는 컨설턴트입니다. Ted는 애자일 움직임이 시작된 2001년부터 프로젝트 및 제품 관리 업무를 담당했으며, 20년 이상 소비재, 의료 장비, 전자제품 및 통신사를 위한 프로젝트 및 제품 납품을 관리했습니다. Ted는 제품을 개발하지 않을 때는 소설을 쓰고 <http://FloatingOrange.com>에서 소규모 그래픽 디자인 사업을 운영합니다. Ted는 PMI 컨퍼런스에서 발표한 경력이 있고 블로그 주소는 <http://FloatingOrangeDesign.Tumblr.com>이고 웹 사이트는 주소는 <http://FloatingOrange.com>입니다.



**GEEK GUIDES:**

전 세계 기술자를 위한 미션 크리티컬 정보.

**저작권 설명**

© 2017 *Linux Journal*. All rights reserved.

본 사이트/발행물은 *Linux Journal*의 승인 아래 작성되고 개발되거나 의뢰를 받고 발행된 자료를 포함합니다("자료"). 본 사이트와 이러한 자료는 국제 저작권 및 상표법에 따라 보호됩니다. 본 사이트와 이러한 자료는 국제 저작권 및 상표법에 따라 보호됩니다..

본 자료는 타인의 권리 비침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여 묵시적이든 명시적이든 어떠한 종류의 보증 없이 "현상태대로" 제공됩니다. 본 자료는 사전 통지 없이 변경될 수 있으며, *Linux Journal* 또는 해당 웹 사이트 스폰서의 약속을 대변하지 않습니다. 어떤 상황에서도 *Linux Journal* 또는 해당 스폰서는 본 자료에 포함된 정보를 사용하여 발생하는 어떠한 직접, 간접, 부수적, 특별, 대표적 또는 간접 손해를 포함하되 이에 제한되지 않고 본 자료에 포함된 기술적 또는 편집 오류 또는 누락에 대해 책임을 지지 않습니다.

1976년 미국 저작권법 107조 및 108조 의해 허용된 경우와 출판사의 서면 허가를 받은 경우를 제외하고, 본 자료의 어떠한 부분(텍스트, 이미지, 오디오 및/또는 비디오를 포함하되 이에 제한되지 않고)도 어떤 수단으로든 전체 또는 일부를 복사, 재생산, 재발행, 업로드, 게시, 전송 또는 배포할 수 없습니다. 개인적으로 비상업적 사용을 목적으로 단일 컴퓨터에 한 부를 다운로드할 수 있습니다. 이러한 사용의 경우 저작권 또는 기타 소유권 조항을 수정하거나 가릴 수 없습니다.

자료에는 타사의 자산인 상표, 서비스표 및 로고가 포함되어 있을 수 있습니다. 이러한 상표, 서비스표 또는 로고는 해당 회사의 사전 서면 동의 없이는 사용할 수 없습니다.

*Linux Journal*과 *Linux Journal* 로고는 미국 특허청에 등록되어 있습니다. 기타 모든 제품명 또는 서비스명은 해당 소유자의 재산입니다. 이러한 약관에 대해 문의하거나 *Linux Journal*의 라이선스 자료에 대한 정보를 요청하려면 [info@linuxjournal.com](mailto:info@linuxjournal.com)으로 이메일을 보내주시십시오.

## 스폰서 소개

### IBM

IBM은 뉴욕 아몽크에 본사를 둔 세계적인 통합 기술 및 컨설팅 기업입니다. 170여 개국에 진출한 IBM은 세계에서 가장 유능한 인재와 함께 기업, 정부 및 비영리 조직의 문제를 해결하고 경쟁 우위를 제공하고 있습니다.

혁신은 IBM 전략의 핵심입니다. 여러 시대의 기술과 산업의 변화를 거치면서 자기 혁신을 통해 고객을 위한 차별화된 가치를 만들어왔습니다. IT 산업이 그 어느 때보다 급격히 변화하고 있는 지금, IBM은 '하드웨어, 소프트웨어, 서비스' 기업 그 이상입니다. IBM은 이제 코그너티브 솔루션과 클라우드 플랫폼 기업으로 부상하고 있습니다.

클라우드가 제공하는 코그너티브 솔루션은 고객의 디지털 혁신의 열쇠입니다. 이러한 혁신을 위해서는 프로세서 및 컴퓨터 설계에서 스토리지, 네트워킹 및 통합 계층에 이르기까지 기업 IT 기반의 모든 수준에서 변화가 필요합니다. 개방형 기술로 구축되고 미션 크리티컬 애플리케이션을 위해 설계된 IBM Power Systems는 코그너티브 워크로드를 위해 설계된 인프라를 제공합니다.

# 혁신적인 앱 개발자가 속도가 빠른 OSDBMS를 좋아하는 이유

TED SCHMIDT

## 개요

소셜, 모바일 또는 IoT 애플리케이션 개발자라면 누구나 인정하듯이 레거시 관계형 데이터베이스 모델로는 더 이상 우리의 모든 요구 사항을 지원할 수 없습니다. 기존 데이터베이스 관리 시스템에는 아무런 문제가 없습니다. 다만 오늘날 디지털 환경에서 쏟아져 나오는 데이터의 다양성과 볼륨을 처리하도록 설계되지 않았을 뿐입니다. 심지어 우리 모두가 기대하듯이 유용한 데이터 중심의 기능을 제공하기 위한 데이터 처리 속도는 말할 것도 없습니다.

다행히 바로 이러한 데이터의 다양성과 복잡성을 처리하고 이를 유용하게 사용하는 데 필요한 속도로 데이터를 저장 및 분석하고 이에 따라 조치를 취할 수 있는 OSDBMS(오픈 소스 데이터베이스 관리 시스템)라는 전혀 새로운 환경이 개발되었습니다.

본 안내서에서는 출시된 몇몇 OSDBMS를 살펴보고 소셜, 모바일, IoT와 같은 소스의 새 데이터를 위해 혁신적인 앱을 개발할 때 직면하는 문제에 대해 OSDBMS가 제시하는 솔루션이 무엇인지 알아봅니다. 본 안내서에서는 출시된 모든 OSDBMS를 살펴볼 수는 없지만, 오픈 소스 SQL, NoSQL(그래프, 문서 및 키 값 스토어 포함), 인 메모리, GPU 가속화 제품을 비롯하여 주요 카테고리의 핵심 주자를 균형 있게 다룹니다.

새로운 앱이 생성하고 소비하는 빅데이터를 비롯하여 새로운 애플리케이션 환경의 주요 과제에 대한 논의로 시작하여, OSDBMS 환경에 대한 개요를 제공하고, 이어서 주요 OSDBMS 제품을 좀 더 자세히 살펴보겠습니다. 마지막으로 이러한 최신 DBMS가 최고 수준에서 작동하여 오늘날 가장 혁신적인 앱을 지원하기 위한 과제를 해결하고 전례 없는 속도, 처리량 및 규모를 제공하는 데 필요한 최적의 기술 플랫폼을 알아보겠습니다.

## 혁신적인 앱 개발자가 직면한 과제

세상은 '빅데이터' 및 이의 진화와 관련된 문제로 떠들썩합니다.

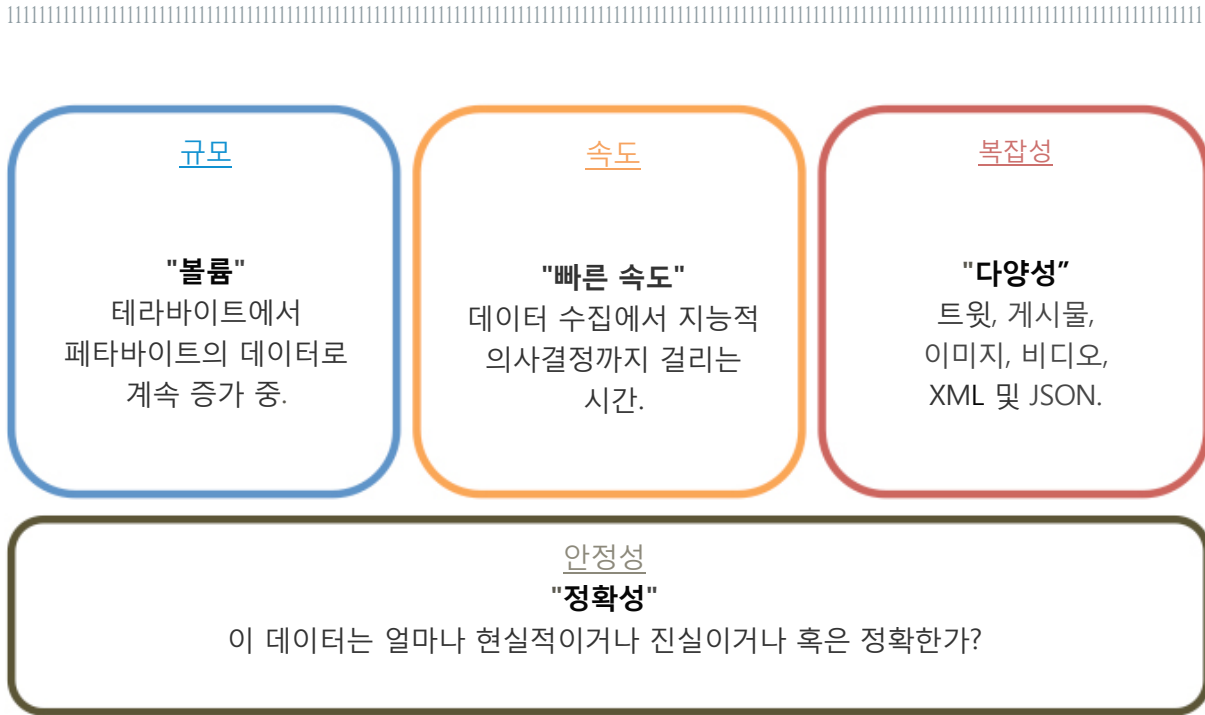


그림 1. 빅데이터의 4가지 요소

조직과 산업별로 빅데이터에 대한 요구 사항이 다르겠지만, 여기에서 제가 빅데이터를 이야기할 때는 대량의 데이터에 대한 즉각적인 액세스를 요구하는 세상에서 혁신적인 앱을 설계하고 개발할 때 직면하는 문제를 지칭합니다. 빅데이터의 외부 소스와 이러한 데이터를 캡처, 저장, 분석 및 시각화하는 데 따른 문제를 검토하고 있습니다. 혁신적인 데이터 중심의 솔루션을 제공하려는 앱 개발자는 규모, 속도, 복잡성 및 안정성이라는 빅데이터의 4가지 요소를 고려해야 합니다. 본 안내서에서는 속도와 복잡성 측면, 그리고 이를 해결하는 OSDBMS와 NoSQL 솔루션의 가치를 주로 다룹니다.

속도 요소는 데이터를 생성, 저장, 처리 및 검색하고 궁극적으로 분석에서 사용할 수 있게 될 때까지 걸리는 시간을 말합니다.



시간이 갈수록 기업은 데이터를 실시간으로 처리하고 비즈니스의 의사 결정 프로세스에 바로 스트리밍하길 원합니다. 교통 통제 시스템을 생각해봅시다. 속도에 대한 요구는 계속 증가하겠지만, 이러한 요구에 부합하기에는 몇 가지 제약이 따릅니다.

분명히 지연 시간(데이터가 수집된 시점과 사용할 수 있는 시점 간의 차이)이 속도에 영향을 줄 것입니다. 지연 시간과 더불어 무어의 법칙이 불안 요소로 작동합니다. 무어의 법칙은 2년마다 전반적인 컴퓨팅 처리 성능이 2배로 증가한다는 법칙입니다. 한동안 이러한 현상이 유지되었지만, 지금은 실리콘 칩이 제공할 수 있는 물리적 한계가 있다는 것이 일반적인 주장입니다. 이는 물리적으로, 점점 더 늘어나고 복잡해지는 데이터를 실시간으로 처리하는 데 필요한 속도를 계속해서 확보해야 하는 앱의 역량에 직접적인 영향을 줍니다.

다시 말해 개발자인 우리는 이제 더 빠른 속도에 대한 해답을 찾는 데 무어의 법칙에만 의존할 수 없으므로 더 큰 속제를 떠안게 되었습니다. 우리는 좀 더 혁신적이 되어야 합니다. 즉, 유용한 기능과 인사이트를 최대한 빨리 제공하기 위해 그 대용량 데이터를 처리하려면 기존 데이터베이스를 계속 사용하는 것이 아니라 오픈 소스 솔루션의 이점을 살펴봐야 합니다.

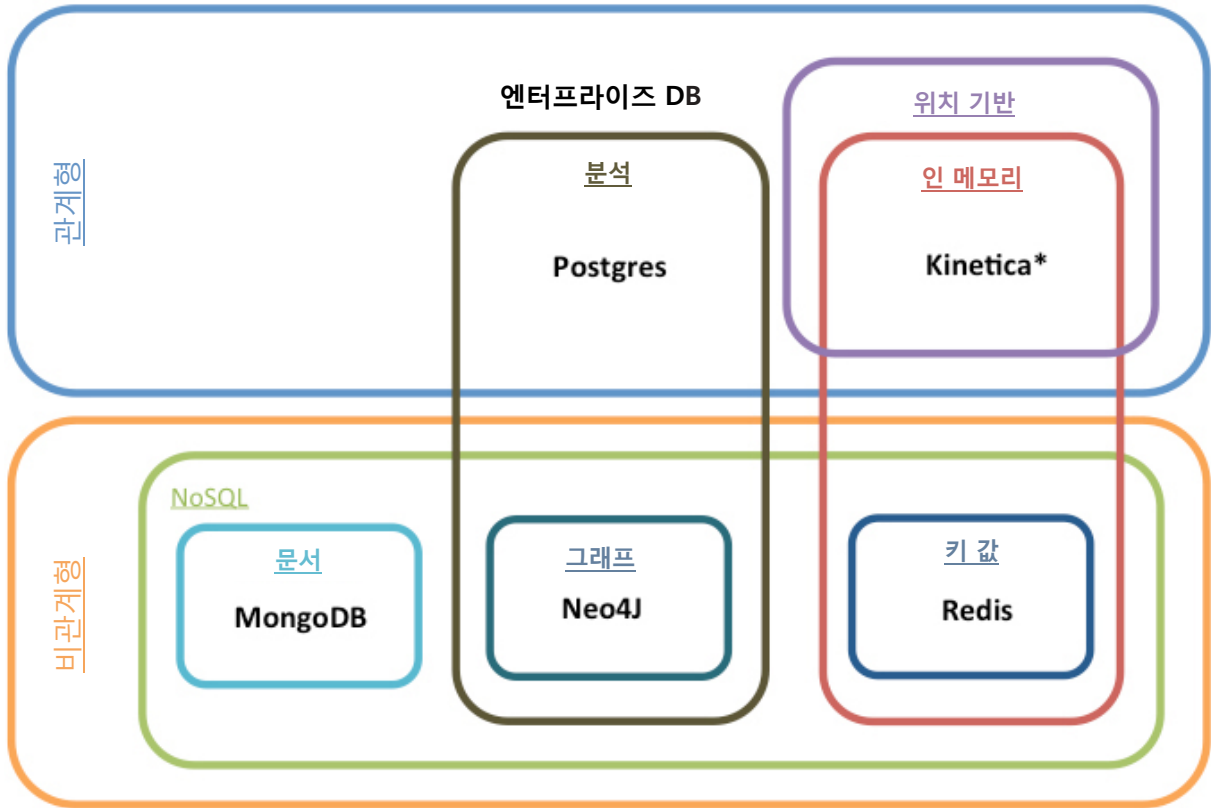
## 오픈 소스가 제공하는 이점

현재 수십 개의 오픈 소스 DBMS 솔루션이 있습니다. 여기에서는 혁신적인 신규 앱 개발과 관련하여 특별한 이점을 제공하는 5가지 솔루션인 MongoDB, Redis, Neo4j, PostgreSQL 및 Kinetica를 살펴보겠습니다. (설명: Kinetica는 오픈 소스가 아니지만, OpenPOWER Foundation에 회원으로 참여하는



개방형 기술 에코시스템의 일부입니다. 따라서 대규모 스트리밍 데이터 세트의 실시간 처리라는 특별한 이점 때문에 Kinetica를 논의에 포함하기로 결정했습니다. 본 안내서 뒷부분에서 언급하겠지만 OpenPOWER Foundation은 기본적으로 회원사가 POWER CPU를 혁신적인 새로운 방법으로 커스터마이징하도록 허용하여 속도와 파워를 최대한 높임으로써 무어의 법칙의 한계에 대응하고자 하는 개방형 기술 회원 조직입니다. <https://openpowerfoundation.org>에서 OpenPOWER Foundation에 대해 자세히 알아보는 것이 좋지만, 본 안내서에서는 혁신이 이루어지는 곳이라는 설명으로 대신합니다.)

OSDBMS로 이전하면 몇 가지 확실한 이점이 있으며, 관리 도구 세트가 엔터프라이즈 수준으로 발전함에 따라 이전에 존재하던 제약도 사라지고 있습니다. 독점 소프트웨어와 독점 하드웨어를 사용하지 않으므로 비용이 절감되는 것이 대개 오픈 소스의 가장 분명한 이점입니다. 하지만 이보다 큰 오픈 소스의 이점은 속도, 유연성 및 지능적 의사 결정입니다. 각 OSDBMS 솔루션에 가장 적합한 특정 사용 사례를 다룰 때 확인하게 될 것입니다. 오픈 소스가 혁신의 속도에 기여하는 것도 간과해서는 안 됩니다. 특성상 독점 솔루션은 변화하는 환경의 요구에 반응하는 속도가 좀 더 느립니다. 반면에 오픈 소스 솔루션은 실제 문제를 해결하기 위해 노력하는 다양한 커뮤니티의 많은 정보를 활용할 수 있습니다. 이처럼 오픈 소스 솔루션은 어떤 독점 솔루션보다 더 빠르게 진화하고 혁신합니다.



\* Kinetica는 오픈 소스가 아니지만, Open Tech 에코시스템의 일부입니다.

그림 2. OSDBMS 환경

OSDBMS 환경의 2가지 기본 카테고리인 비관계형(MongoDB, Redis, Neo4j 포함)과 관계형(EDB Postgres, Kinetica 포함)을 살펴보겠습니다.

비관계형 부분에는 NoSQL 키 값 데이터베이스인 Redis가 있으며, 빠른 속도와 간단하지만 명쾌한 데이터 운영이 중요한 게임 산업에서 특히 유용합니다. 다음으로 Neo4j는 NoSQL 그래프 데이터베이스이며 데이터 요소 간에 관계를 저장하는 데 특히 효과적입니다. 마지막으로 MongoDB는 범용 데이터베이스로 사용하기 적합한 NoSQL 문서 데이터베이스로서, 특히 스키마가 없어서 서로 다른 유형의

데이터를 매우 유연하게 저장할 수 있습니다.

관계형 카테고리의 경우 Kinetica는 오픈 소스는 아니지만 빠른 속도가 제공하는 이점 때문에 포함시켰습니다. 위기 기반 관계형 데이터베이스인 Kinetica는 인 메모리 GPU 가속화 데이터베이스입니다. 대규모 스트리밍 데이터를 처리하는 전례 없는 속도의 핵심은 바로 GPU 가속화입니다(나중에 자세히 설명). 그리고 여기에서 검토할 또 다른 관계형 데이터베이스인 EDB Postgres Advanced Server는 EDB에서 PostgreSQL을 엔터프라이즈급으로 패키징한 것으로서 EDB가 적극적으로 참여하고 지원하는 커뮤니티에서 개발한 오픈 소스 제품입니다. 분석용으로도 매우 훌륭합니다.

비관계형 부문을 보면 또 다른 인 메모리 데이터베이스인 Redis가 있습니다. Redis는 NoSQL 키 값 데이터베이스로 게임 산업에 특히 유용합니다. Neo4j은 NoSQL 데이터베이스이자 그래프 데이터베이스로서, 데이터 요소 간 관계를 저장하는 데 특히 효과적입니다. 마지막으로 MongoDB는 NoSQL 데이터베이스이자 문서 데이터베이스입니다. 즉, 훌륭한 범용 데이터베이스이지만 최대 이점은 스키마가 없다는 것입니다. 스키마가 없으므로, 온도나 초당 회전수 증가와 같이 서로 다른 유형의 데이터를 저장할 수 있습니다.

중복되는 경우가 많은 각 OSDBMS의 이점은 궁극적으로는 사용자의 특정 사용 사례를 기반으로 하지만, 모든 데이터베이스가 개방형 기술 개발 모델의 일부라는 비슷한 이점을 공유합니다.

그뿐만 아니라 적절한 인프라에 배포되는 경우, 각 OSDBMS는 데이터 집약적 앱 개발에 진정한 혁신을 가져올 수 있는 모두가 열망하는 뛰어난 속도를 제공합니다.

## MongoDB 문서 스토어 DBMS

MongoDB는 문서 지향 데이터 모델이 적용된 오픈 소스 데이터베이스입니다. 다른 데이터 유형(int, long, date, floating point 등)을 포함하기 위해 JSON을 확장한 바이너리 JSON(BSON)에 데이터를 저장합니다. BSON 문서가 프로그래밍 언어 객체 구조를 따르기 때문에 이러한 BSON 문서를 사용하면 애플리케이션의 데이터를 데이터베이스의 데이터로 더 빠르고 쉽게 모델링할 수 있습니다. 모든 문서에는 여러 필드가 포함되어 있고, 각 필드에는 하위 문서 또는 배열과 같이 특정 데이터 유형의 값이 있습니다. 구조가 비슷한 문서는 컬렉션으로 함께 그룹화됩니다. RDBMS에서 컬렉션은 테이블, 문서는 열, 그리고 필드는 컬럼에 해당합니다.

문서 지향 데이터 모델에는 스키마가 없습니다. 따라서 널(Null) 값을 빈 필드에 저장하는 RDBMS와는 달리 MongoDB에서는 데이터가 없는 경우 데이터를 저장할 필드도 없습니다. 다시 말해 개발할 때 기존 스키마에 대한 변경을 걱정할 필요가 없으므로 끊임없이 변하는 비즈니스 요구 사항에 좀 더 민첩하게 대응할 수 있습니다. 애플리케이션을 훨씬 쉽게 개선할 수 있으므로 혁신으로 나아가는 기회를 제공합니다.

MongoDB의 문서 지향 모델에서는 정규화된 문서를 더 작은 테이블로 분할하지 않으므로 조인을 생성할 필요성이 감소합니다.

MongoDB는 JavaScript로 구축된 웹 앱을 신속하게 설치하거나, 다수의 실시간 카운터를 사용하거나, 대량의 이미지를 저장해야 할 때 특히 유용합니다.

조인 수가 줄어들면 확장성과 속도가 크게 개선됩니다. 다른 NoSQL 데이터베이스와는 달리 MongoDB의 좋은 점은 여러 컬렉션의 데이터를 통합하려는 경우 여전히 조인을 사용할 수 있다는 것입니다.

또한, MongoDB는 자동 샤딩 기능과 지리공간 애플리케이션에 대한 지원을 제공하므로 본 안내서에서 다루는 유형의 애플리케이션에 매우 적합합니다. 여러 개의 테이블과 조인으로 복잡한 RDBMS 파티셔닝과 비교하여 MongoDB의 샤딩은 키 공간을 파티셔닝하는 것으로 완료됩니다. 키가 문서 ID이고, 문서가 키 값 문서 스토어의 값이기 때문입니다. 모든 NoSQL 데이터베이스는 지연 시간을 줄이고 확장성을 개선하는 일종의 샤딩 또는 파티셔닝을 지원합니다.

자동 샤딩과 BSON 문서를 사용하는 MongoDB는 변화하는 비즈니스 요구 사항에 대한 좀 더 민첩하고 즉각적으로 접근할 수 있는 유연하고 속도가 빠른 데이터베이스를 제공합니다. MongoDB는 JavaScript로 구축된 웹 앱을 신속하게 설치하거나, 다수의 실시간 카운터를 사용하거나, 대량의 이미지를 저장해야 할 때 특히 유용합니다. IoT의 실시간 쿼리와 보고 요건을 충족하기 위해 사용할 때 놀라울 정도로 빠른 속도를 보입니다. 지리공간 기능을 지원하므로, 사용자 위치를 알거나, 사용자가 가야 할 곳을 보여주는 것이 중요한 애플리케이션에서

사용하기에 적합합니다.

개발자에게는 MongoDB가 제공하는 온디맨드 교육이 유용합니다. MongoDB는 서버 기반이 아니라 프로젝트 기반으로 개발 지원을 제공하는 데 이는 개발자와 운영 관리자에게 상당한 도움이 됩니다.

## EDB Postgres Advanced Server

EDB Postgres는 Enterprise DB에서 배포하는 오픈 소스 PostgreSQL 관계형 데이터베이스의 확장 버전이라고 할 수 있습니다. EDB Postgres가 PostgreSQL 뒤에 릴리스되긴 했지만, EDB는 PostgreSQL 커뮤니티의 핵심 구성 요소로서 적극적으로 참여하고 있습니다.

PostgreSQL가 제공하는 핵심 이점은 속도와 확장성이며, EDB는 풍부한 엔터프라이즈급 도구 세트로 이를 보완합니다. PostgreSQL 커뮤니티 에코시스템에 액세스할 수 있는 것이 분명한 장점으로 작용하며, 확장 가능한 고성능 데이터베이스에서 나오는 다른 장점들도 있습니다.

PostgreSQL은 사용자 정의 데이터 유형(예: XML), 테이블 컬렉션 및 Varray를 비롯하여 다양한 데이터 유형을 지원합니다. 텍스트 데이터, 인덱싱 및 검색을 지원하며, 여러 버전의 동시 제어 기능을 사용하여 중단 없이 읽기/쓰기 작업이 실행되도록 허용합니다. 마지막으로 저장된 프로시저가 C/C++, Java, JavaScript, Python, Perl 및 Ruby와 같은 언어로 작성될 수 있으므로 상당한 유연성과 선택권을 제공합니다.

EDB Postgres는 확장된 암호 프로필을 비롯한 엔터프라이즈급 보안 기능을 제공합니다. 오픈 소스 PostGIS 플러그인을

사용하면 EDB Postgres를 지리 정보 시스템(GIS)을 위한 백엔드 공간 데이터베이스로 구현할 수 있습니다.

또한, EDB Postgres에서는 트리거와 저장된 프로시저를 생성 및 디버깅할 수 있는 GUI 도구를 제공합니다. EDB Postgres는 하위 시스템 잠금에 대한 확장성 개선과 수직 확장 최적화를 제공하여 성능을 향상합니다. 따라서 다양한 데이터베이스(관계형, 문서 및 키 값 데이터베이스 지원)에서 통합이 가능하고 정형, 비정형 및 트랜잭션 데이터를 결합할 수 있습니다. 또한, 빠른 속도가 관건인 읽기 전용 앱에도 사용할 수 있습니다. DBA가 프로세스 전체에서 I/O와 CPU 사용에 대한 우선순위를 선택적으로 지정할 수 있으며, Oracle과 호환됩니다.

EDB는 Postgres 전문 지식, 기술 동영상, 수많은 설명서 및 매우 강력한 커뮤니티에 바로 액세스할 수 있는 Postgres 개발자 등록(Postgres Developer Subscription)을 제공합니다. 또한, 마이그레이션, 통합 및 관리에서 흔히 발생하는 부담을 덜어줄 수 있는 엔터프라이즈급 도구 세트도 제공합니다.

EDB Postgres가 등록 DBMS로 판매된다는 것도 주목할 만합니다. 등록에는 소프트웨어와 더불어 모든 업그레이드, 유지 관리 및 지원이 포함되어 있습니다.

특히 최적화된 서버 아키텍처(뒤에서 자세히 설명)에서 실행할 때 안전하고, 확장 가능하며, 유연하고 빠른 EDB Postgres Advanced Server는 훌륭한 최신 관계형 데이터베이스 솔루션입니다. Postgres Advanced Server는 공급업체와 관련된 큰 라이선스 비용 없이, 오픈 소스에서 제공하는 혁신과 함께 엔터프라이즈급 RDBMS에서 기대하는 모든 것을 제공합니다.

## Neo4j 그래프 데이터베이스

처음에는 그래프 데이터베이스에 특별히 놀라운 점이 없는 것처럼 보일 수 있습니다. 그래프는 노드와 관계라는 2가지 기본 요소로 구성됩니다. 모든 노드는 데이터 조각(엔티티)을 나타내고, 모든 관계는 2개의 노드가 서로 어떻게 관련이 있는지 나타냅니다. Facebook이나 Tumblr와 같이 사용자가 서로를 팔로우하는 소셜 네트워킹 사이트가 이러한 아이디어의 전형적인 예입니다. 사용자가 노드이고 "팔로잉"이 노드 간의 관계입니다.

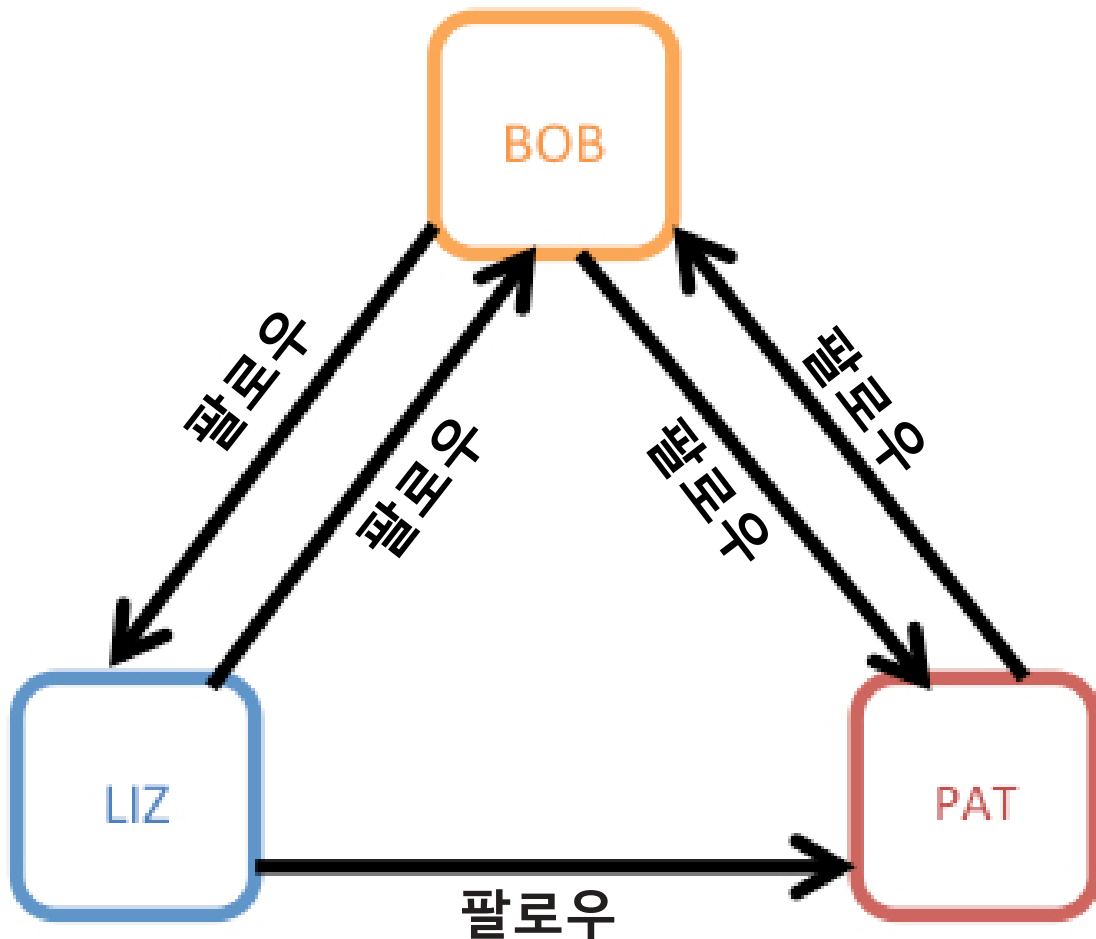


그림 3. 노드와 관계





그래프 데이터베이스에서는 관계가 가장 우선순위가 높습니다. 즉, 데이터 모델이 좀 더 간단하고 좀 더 명확하며, 외부 키와 같은 것에 신경을 쓸 필요가 없습니다. 또한, 2개의 노드 없이는 관계를 만들 수 없으며, 관계를 삭제하지 않고 노드만 삭제할 수는 없습니다.

그래프 데이터베이스가 앱 개발에 제공하는 이점을 이해하기 위해서는 그래프 데이터베이스의 2가지 핵심 특성을 살펴보는 것이 중요합니다. 먼저, 네이티브 그래프 스토리지와 비네이티브 그래프 스토리지의 차이입니다. 관계형 또는 객체 지향 데이터베이스를 그래프 데이터베이스처럼 사용하는 것과는 달리 Neo4j와 같은 네이티브 그래프 데이터베이스는 그래프를 저장하고 관리하기 위해 특별히 설계되었습니다. 비네이티브 그래프 데이터베이스는 관계형 또는 객체 지향 데이터베이스를 데이터 스토리지로 사용합니다. 따라서 데이터 볼륨과 쿼리 복잡성이 증가하면 비네이티브 그래프 데이터베이스는 지연 시간이 대폭 증가합니다.

Neo4j의 두 번째 특성은 그래프 처리 엔진입니다. 네이티브 그래프 처리에서는 연결된 노드가 직접 서로를 가리킵니다. 이를 인덱스 없는 인접성(index-free adjacency)이라고 부르며 그래프 데이터베이스에서 데이터를 처리하는 가장 효율적인 방법입니다. Neo4j의 네이티브 그래프 처리 엔진은 일관된 실시간 성능을 제공합니다. 비네이티브 데이터베이스에서 수행해야 하는 비용이 많이 드는 인덱스 조회가 필요 없기 때문입니다.

이러한 특성은 사용자와 권한을 빠른 속도로 추적해야 하는 자격 증명 및 액세스 관리 애플리케이션 또는 전자 상거래 제품과 맞춤형 앱을 지원하는 실시간 추천 엔진에 적합합니다. 앞에서 언급했듯이 소셜 앱 데이터의 실시간 분석이 필요한 경우

이러한 특성은 필수적입니다.

Neo4j는 적응형 데이터 모델 덕분에 특히 민첩합니다. 즉, 기존 기능에 미치는 영향에 대해 걱정할 필요 없이 데이터 모델을 변경하여 새로운 비즈니스 요구에 대응할 수 있습니다. Neo4j는 빠른 속도를 내도록 설계되었습니다. 노드 간의 명시적 관계 때문에 Neo4j에서는 데이터 세트가 증가하면 필연적으로 발생하는 지연을 방지할 수 있습니다.

또한, Neo4j는 광범위한 문서 라이브러리, 기술 자료, 샌드박스 환경에 액세스, 앞에서 설명한 다른 OSDBMS와 같은 강력한 커뮤니티 지원 시스템 등 훌륭한 온라인 개발자 지원을 제공합니다.

## 인 메모리 데이터베이스 시스템

Redis 및 Kinetica와 같은 인 메모리 데이터베이스 시스템(IMDBS)은 데이터를 영구 매체에 저장하도록 설계된 전통적인 데이터베이스 시스템과는 달리 데이터를 주 메모리에 저장합니다. 기술적으로는 전통적인 데이터베이스를 RAM에 넣을 수 있지만, 디스크 스토리지를 위해 설계된 시스템의 오버헤드를 계속 떠안게 됩니다. IMDBS가 데이터를 메모리에 저장함으로써 I/O 작업과 캐싱 오버헤드를 피하므로, 전통적인 DBMS와는 비교도 안 되게 빠릅니다. IMDBS는 설계가 아주 단순하므로 메모리 및 CPU 요구 사항도 훨씬 적습니다.

데이터에 대한 아주 빠른 액세스와 조작이 필요한 애플리케이션은 IMDBS를 사용하기에 아주 적합한 대상입니다. 실시간 임베디드 시스템, 금융 시장 애플리케이션, 전자 상거래 및 소셜 애플리케이션의 경우, IMDBS가 최적의 선택입니다.

IMDBS의 경우 전통적인 DBMS 솔루션의 모든 성능 이점을 그대로 유지하면서 테라바이트 이상의 규모로 확장되는 경우가 종종 있습니다.

빠른 속도에서 진정한 이점을 확보할 수 있기 때문입니다. IMDBS는 속도가 빠를 뿐만 아니라 확장성도 뛰어납니다. IMDBS의 경우 전통적인 DBMS 솔루션의 모든 성능 이점을 그대로 유지하면서 테라바이트 이상의 규모로 확장되는 경우가 종종 있습니다.

키 값 데이터베이스 또는 스토어인 **Redis**는 연관 배열(associative array)을 저장, 검색 및 관리하도록 설계된 데이터베이스입니다. 연관 배열은 모든 키가 컬렉션에서 단일 값과 연결되어 있는 간단한 데이터 모델입니다(키 값 페어라고 부르는 관계). 파일 이름이나 해시, URI와 같은 임의 문자열은 이러한 각 키 값 페어에 있는 키를 나타냅니다. Blob로 저장되는 값은 이미지나 문서 등 어떤 유형의 데이터도 될 수 있습니다. 값이 blob로 저장되므로, 사전에 데이터를 모델링하거나 스키마를 정의할 필요가 없습니다. 또한, 성능 개선을 위해 데이터를 인덱스할 필요도 없습니다. 값이 opaque 데이터이므로, 요청에서 반환된 것을 값을 기준으로 필터링하거나 제어할 수 없습니다.

키 값 스토어는 쿼리 언어 대신 get, put 및 delete 명령을 사용합니다. 즉, 메모리에 있는 객체에 대한 직접 요청이 바로 데이터 검색 경로입니다. 데이터 간 관계를 계산할 수 없으므로 최적화 오버헤드가 없습니다. 인덱스를 저장할 위치, 네트워크 속도 또는 분산 시스템의 밸런싱을 걱정할 필요가 없습니다.

이러한 단순성 덕분에, 키 값 스토어는 매우 빠르고, 유연하며, 사용이 간편하고, 확장성과 이동성이 뛰어납니다. Redis는 데이터베이스, 캐시 및 메시지 브로커로 사용되는 오픈 소스(BSD 라이선스) 인 메모리 키 값 데이터 구조 스토어입니다.

Redis를 사용할 때의 장점 중 하나는 LPUSH, LTRIM 및 LREM과 같은 Redis 기본 명령어를 사용하는 데서 시작됩니다. 이러한 기본 명령어를 사용하면 전통적인 데이터 스토어로는 수행하기 힘들고 느린 작업을 훨씬 쉽게 완료할 수 있습니다. 예를 들어 웹 앱에서는 LREM을 사용하여 삭제된 글을 캐시에서 제거하거나, LPUSH를 사용하여 키에 저장된 목록의 헤드에 콘텐츠 ID를 삽입하여 홈페이지에서 최신 항목 목록을 보여줄 수 있습니다. LTRIM을 사용하여 목록의 항목 수를 제한할 수 있습니다. Redis의 이러한 간단한 명령은 개발자의 작업을 훨씬 쉽게 해줍니다.

이러한 단순성, 속도 및 짧은 지연 시간 때문에, 사용자 프로필과 기본 설정을 효율적으로 저장하는 데 관심이 있다면 Redis는 전자 상거래 앱 개발용으로 탁월한 솔루션입니다. 예를 들어 사용자가 보고 있는 것을 토대로 제품을 추천하거나 고객의 구매 습관에 맞게 개별화된 실시간 광고 및 쿠폰을 보여줄 수 있습니다. 모든 데이터가 메모리에 있으므로 데이터를 찾느라 시간이 지연될 일이 없습니다. 이는 매우 빠른 성능으로 이어집니다. 예를 들어 Redis를 다른 DB 앞에 캐시로 사용하면 속도가 놀라울 정도로 빨라집니다.

개인적으로 Redis가 제공하는 온라인 지원을 높게 평가합니다. 여러 자습서, 관리자 안내서, 다른 개발자 리소스와 더불어 상세한 프로그래밍 안내서에 전체 명령 목록이 포함되어 있습니다.

Redis에서 제공하는 수많은 이점에 대해 자세히 알아보려면 <https://redis.io>를 참조하십시오.

**Kinetica로 GPU 가속화** 다시 말하지만 Kinetica는 오픈 소스 데이터베이스가 아니며, OpenPOWER Foundation의 개방형 하드웨어/소프트웨어 에코시스템의 일부입니다. Kinetica는 GPU(Graphics Processing Units)를 통해 가속화되는 분산 인 메모리 데이터베이스입니다. GPU는 메모리를 신속하게 변경 및 조작하여 표시할 이미지 생성을 가속화하도록 설계된 회로입니다. CPU에 다수의 코어와 대량의 캐시 메모리가 있지만, GPU에는 수천 개의 코어가 있으므로 경우에 따라 CPU의 100배가 넘는 속도를 내기도 합니다. GPU는 대규모 데이터를 가져와서 동일한 작업을 반복해서 수행하는 데 탁월하므로 원래 3D 게임 렌더링을 위해 만들어졌습니다. 최근에는 GPU가 금융 모델링, 연구 조사, 에너지 탐사 및 인공지능과 같은 분야의 컴퓨팅 워크로드를 가속화하는 데 사용됩니다. 최대 100배 더 빠른 처리 속도를 내는 병렬 처리 아키텍처 때문에 Kinetica는 대규모 스트리밍 데이터를 분석하는 데 적합합니다. 따라서 AI 워크로드를 정의하는 예측 분석에 안성맞춤입니다.

Kinetica는 GPU 처리 성능을 활용하여, 전통적인 데이터베이스보다 훨씬 적은 하드웨어 점유 공간과 아주 짧은 시간으로 대규모 데이터 세트(특히 스트리밍 데이터)를 관리합니다. 이는 IoT 애플리케이션과 지리공간 가상화에 특히 유용합니다. 대용량 데이터를 렌더링할 수 있는 가상화 도구를 제공하며 데이터를 분석하기 전에 스키마를 준비할 필요가 없습니다. Kinetica는 SQL과 완벽히 호환되고 쿼리가 간편하므로, 트랜잭션 시스템, 데이터 웨어하우스 및 데이터 레이크의 훌륭한 파트너 도구가 됩니다.

REST, JSON, Java, JavaScript, C++, Python 등을 지원하는 개발자 친화적인 솔루션입니다. 다시 말해 새로운 쿼리 또는 프로그래밍 언어를 배우거나 새로운 데이터 모델을 구축할 필요 없이 대규모 데이터 세트를 더 빠르게 탐색할 수 있습니다.

지금까지 OSDBMS 환경의 몇 가지 데이터베이스 제품을 살펴보았습니다. 비관계형 카테고리에는 문서 지향 데이터 모델인 MongoDB, 그래프 모델인 Neo4j, 그리고 키 값 접근 방식의 Redis가 있습니다. 관계형 카테고리에는 탁월한 분석 데이터베이스인 EDB Postgres와 인 메모리 옵션의 Kinetica가 있습니다. 이 모든 데이터베이스 관리 시스템의 공통적인 이점은 바로 빅데이터 분석 애플리케이션 개발에 관해 이야기할 때 가장 중요한 요소인 속도입니다.

이제 마지막으로 이러한 OSDBMS와 그 위에 구축된 앱을 호스팅하기에 적합한 플랫폼인 OpenPOWER LC 서버를 살펴보겠습니다. OpenPOWER LC 서버는 IBM과 OpenPOWER Foundation 파트너가 처음부터 빅데이터를 위해 설계했습니다.

## OSDBMS를 IBM OpenPOWER Systems 상에 구현하는 것이 올바른 선택인 이유

OSDBMS의 기능을 활용하여 진정으로 혁신적인 솔루션을 만들기 위해서는 처리 속도뿐만 아니라 협업이 필요합니다. 본 전자책 앞부분에서 OpenPOWER Foundation에 대해 언급했었습니다. OpenPOWER Foundation은 IBM, Google, NVIDIA, Mellanox Technologies, Tyan, Xilinx 및 Canonical 등 IT 분야의 세계 최고 기업을 비롯하여 250개 회원사로 구성된 컨소시엄입니다.



그림 4. IBM POWER8이 OSDBMS에 최적인 이유

OpenPOWER Foundation은 IBM의 Power 프로세서 아키텍처를 기반으로 한 시스템을 설계하기 위해 수년간 협력해오고 있습니다. 이러한 협업의 최신 상용 버전 중 하나를 IBM의 OpenPOWER LC 서버에서 찾아볼 수 있습니다.

IBM POWER8 프로세서 기술이 탑재된 OpenPOWER LC 서버는 여기에서 다른 OSDBMS 솔루션을 비롯하여 빅데이터 워크로드를 위해 설계되었습니다. IBM POWER8은 범용 플랫폼보다 4배 많은 프로세서 캐시, 메모리 대역폭 및 멀티스레딩을 제공합니다.

IBM POWER8은 Red Hat, SUSE 및 Canonical의 업계 표준 Linux를 구동합니다. 따라서 x86 Linux 애플리케이션을 Power로 이전하는 것이 좀 더 매력적이고 그 어느 때보다 간단합니다. Power 기반 Linux는 개발자가 데이터 집약적 앱을 위해 OSDBMS의 역량과 규모를 활용하는 데 필요한 혁신적인 플랫폼을 제공합니다.

POWER8 설계에서는 컴퓨팅 파워, 메모리 대역폭, I/O 처리량을 조합하여 빅데이터 및 분석 워크로드에 필요한 속도를 구현합니다. POWER8은 2소켓 서버에서 최대 2테라바이트를 제공할 수 있는 스케일 아웃 시스템과 최대 16테라바이트를 제공할 수 있는 엔터프라이즈 스케일 업 서버를 통해 범용 인프라보다 4배 많은 코어당 스레드와 4배 많은 메모리 대역폭 및 더 많은 메모리 용량을 제공하도록 설계되었습니다.

**MongoDB** 모든 데이터에 대한 통합된 실시간 뷰를 제공하는 플랫폼이 있다는 것은 MongoDB에 아주 좋은 소식입니다. IBM에 따르면 POWER8 기반 MongoDB는 Intel Xeon보다 40% 높은 서버당 성능을 제공합니다. 이는 데이터 센터 서버 스프롤에 아주 적합한 솔루션이며, 설치 비용까지 고려한다면 POWER8 기반 MongoDB는 x86 기반 시스템과 비교하여 2배의 달러당 성능을 제공합니다. 미래의 혁신을 위해 운영 자금을 절감할 방법을 찾고 있는 사용자에게 반가운 소식입니다.

**EDB Postgres Advanced Server** EDB Postgres Advanced Server는 POWER8 기반 little-endian Linux에서 구동됩니다. EDB Postgres Advanced Server를 IBM의 OpenPOWER LC 서버에서 구동하면 고성능 멀티 스레딩, 더 많은 캐시, 데이터를 위한 더 큰 대역폭, x86 기반 시스템보다 2배 뛰어난 성능 가격비가 제공됩니다. IBM 벤치마크는 OpenPOWER LC 서버가 Intel Xeon보다 60% 뛰어난 코어당 성능을 제공한다는 것을 보여줍니다. 한 번 더 말하자면 이는 귀사에서 빅데이터 앱을 위해 더 적은 수의 코어에 더 많은 워크로드를 배정하고 인프라 설치 비용을 줄여 혁신을 위한 리소스를 충분히 남겨둘 수 있는 완벽한 기회입니다.



표준 Redis API를 변경하지 않고도 모든 Redis 클라이언트와 연동되는 이 솔루션을 통해, CAPI-Flash 가속화가 탑재된 단일 POWER8 서버가 1밀리초 미만의 지연 시간으로 초당 20만 건이 넘는 작업을 처리할 수 있습니다.

---

**Redis** Open POWER Foundation의 회원인 Redis Labs와 IBM Power Systems는 긴밀히 협력하여 IBM Data Engine for NoSQL(IBM 840 Flash System 상에 Redis를 실행)에 대한 Redis 지원의 일환으로 POWER8과 CAPI(Coherent Accelerator Processor Interface)에 최적화된 Redis 솔루션을 제공하고, RAM을 대신하는 IBM CAPI-Flash 카드 및 Redis Labs Enterprise Cluster(RLEC) for Flash 소프트웨어를 제공합니다. 표준 Redis API를 변경하지 않고도 모든 Redis 클라이언트와 연동되는 이 솔루션을 통해, CAPI-Flash 가속화가 탑재된 단일 POWER8 서버가 1밀리초 미만의 지연 시간으로 초당 20만 건이 넘는 작업을 처리할 수 있습니다. 정말 빠른 속도입니다. 또한, 테라바이트 규모 데이터 세트의 90%를 플래시에 저장하고 10%만 RAM에 저장할 수 있습니다. 순수 RAM 기반 Redis 솔루션과 비교하면 설치 비용을 70% 이상 절감할 수 있습니다. Redis 테스트에 따르면 IBM Open POWER LC 서버는 x86보다 성능이 87% 더 뛰어납니다.



**Neo4j** IBM OpenPOWER LC 서버 기반의 Neo4j는 매우 큰 그래프를 저장하고 처리할 수 있으며 세계에서 가장 확장성이 뛰어난 그래프 데이터베이스 플랫폼을 제공합니다. 대규모 그래프 처리의 주요 과제 중 하나는 실시간 성능을 저하시키지 않고 데이터 세트 크기를 처리하는 방법입니다. 위에서 설명한 것처럼 CAPI와 Flash를 통해 LC 서버에서 사용할 수 있게 된 56TB의 확장 메모리로 인해 실시간 쿼리의 크기가 증가합니다. 메모리에 저장할 수 있는 그래프의 크기가 증가했기 때문입니다. 하지만 POWER8이 탑재된 LC 서버 라인은 균형을 유지합니다. 각 코어는 동시에 8개의 하드웨어 스레드를 처리할 수 있으므로 이는 12코어 칩에서 총 96개의 동시 스레드에 해당합니다. 온칩 메모리 컨트롤러는 메모리와 시스템 I/O에서 높은 대역폭을 활성화합니다. IBM OpenPOWER LC 서버에서 CAPI 가속화가 활성화되면 Intel Xeon과 비교하여 거의 2배 높은 성능을 제공합니다.

**Kinetica** Kinetica에서 최고의 성능을 얻을 수 있을지는 전적으로 CPU와 GPU 간 데이터 이동 속도에 달려있습니다. Kinetica는 시스템 메모리를 활용하도록 설계되었기 때문입니다. 새로운 NVIDIA NVLink Technology와 IBM POWER8은 Kinetica를 통한 고성능 분석 제공에 대해 가장 경제적이고 가장 앞선 접근 방식을 제공합니다. NVIDIA NVLink와 같은 상호 연결 기술은 CPU와 GPU 간에 문을 활짝 열어 Kinetica가 시스템 메모리를 충분히 활용할 수 있도록 지원합니다. GPU 처리가 더는 I/O 하위 시스템을 통해 데이터가 이동하는 속도에 국한되지 않으므로, Kinetica에서 대규모 데이터 세트를 처리할 수 있습니다.

IBM에 따르면 Kinetica는 NVLink가 적용된 IBM OpenPOWER LC 서버에서 이와 유사한 x86 시스템 대비 처리량이 2.5배 증가합니다.

## 결론

크고 복잡한 데이터를 생성하는 크고 복잡한 세상에 살고 있습니다. 전통적인 데이터베이스, 개발 접근 방식 및 처리 플랫폼은 오늘날 요구되는 실시간 데이터 및 분석 기능을 제공하는 데 필요한 성능을 보장해 주지 않습니다.

OSDBMS는 혁신적인 개발자에게 폭주하는 다양한 데이터를 위한 성능과 유연성 등 여러 가지 이점을 제공합니다. 하지만 아마도 더 중요한 것은 변화와 현실 세계에서 발생하는 문제에 신속하고 효과적으로 대응하는 활발한 커뮤니티에 대한 액세스 등 개방형 개발 모델의 장점일 것입니다. OSDBMS에는 즉각적인 반응과 빠른 성능뿐만 아니라 엔터프라이즈급 도구와 독점 솔루션의 제약에서 벗어나는 자유가 있습니다. 이러한 장점으로 OSDBMS의 사용성과 액세스 가능성이 증가합니다.

앱 개발자에게 가중되는 요구가 계속 진화함에 따라 앱 자체가 이러한 요구에 대응하여 계속 혁신되어야 합니다. OSDBMS의 개방형 에코시스템은 아이디어와 혁신이 실제 사용할 수 있는 솔루션으로 유입되는 데 필요한 환경을 제공합니다. 긴밀한 협업을 통해 IBM과 OSDBMS 분야의 리더는 이러한 새로운 데이터베이스 기술을 최대한 활용할 수 있도록 업계 최고의 플랫폼을 제공하고 있습니다.

아래 링크를 클릭하여 POWER 기반 OSDBMS로 애플리케이션을 개발하면 어떠한 이점이 있는지 직접 확인하십시오. [https://www-01.ibm.com/marketing/iwm/dre/signup?source=mrs-form-12148&S\\_PKG=ov53321](https://www-01.ibm.com/marketing/iwm/dre/signup?source=mrs-form-12148&S_PKG=ov53321). ■