




*Netezza의 TwinFin이 출시된지 1년도 채 되지 않아, 100여 고객이 어플라이언스를 채택했으며 다음은 고객사 일부 목록입니다:*

- 코리아크레딧뷰로
- 우리 캐피탈
- NICE신용평가정보
- Blue Cross Blue Shield of Massachusetts
- BlueKai
- Catalina Marketing
- Con-way Freight
- DataLogix
- Epsilon (얼라이언스 데이터 업체)
- interCLICK
- IntercontinentalExchange
- Japan Medical Data Center
- Kelley Blue Book
- Marshfield Clinic 및 Marshfield Clinic Research Foundation
- MediaMath
- MetroPCS
- MicroAd
- MyLife.com
- NYSE Euronext
- Pacific Northwest National Laboratory
- Premier, Inc.
- Safeway
- The Nielsen Company
- Wind Telecom

	<b>목차</b>	<b>페이지</b>
<b>1</b>	<b>소개</b>	<b><u>4</u></b>
<b>2</b>	<b>온라인 트랜잭션 처리(OLTP) 및 데이터 웨어하우징</b>	<b><u>6</u></b>
<b>3</b>	<b>쿼리 성능</b>	<b><u>8</u></b>
<b>4</b>	<b>작업 간편성</b>	<b><u>16</u></b>
<b>5</b>	<b>가치</b>	<b><u>23</u></b>
<b>6</b>	<b>TwinFin을 통한 Exadata 성능 향상</b>	<b><u>27</u></b>
<b>7</b>	<b>결론</b>	<b><u>28</u></b>



IBM Netezza는 큰 데이터를 쿼리 및 분석하도록 설계된 기술에 초점을 맞춥니다. Netezza의 혁신적인 데이터 웨어하우스 어플라이언스는 시장에 큰 바람을 일으키고 있습니다. 낮은 운영 비용 및 TCO로 데이터를 분석하려는 많은 고객들은 Oracle로부터 데이터 웨어하우스를 Netezza로 전환하였습니다. 그 후 Oracle은 Exadata를 출시했습니다. Exadata는 외관상으로는 TwinFin의 기능을 모두 가지고 있으며 온라인 트랜잭션도 처리하는 것처럼 보입니다. Exadata의 강점은 Oracle 고객들의 온라인 트랜잭션 처리(OLTP) 시스템 비용을 간소화하며 줄인다는 점입니다. Exadata는 여러 시스템을 통합하며, 중복이 되지 않도록 하고, 복잡했던 구성을 하나의 확장 가능한 고성능 OLTP 환경으로 대체해 줍니다. 그런데 기업들이 분석 데이터베이스를 Exadata에 통합하려고 하게 되면 이러한 혜택은 줄어들게 됩니다. Netezza는 그동안 Oracle 고객들에게 많은 가치를 제공해 왔습니다. Oracle OLTP 데이터베이스에서 분석 워크로드를 없애으로써 200여개가 넘는 기업들이 두 제품 모두의 최고가치를 이용하고 있습니다.

데이터 웨어하우스 플랫폼으로서 Exadata와 TwinFin을 시험한 이 보고서는 Netezza의 관점에서 작성되었지만, 신뢰성 보장을 위해 Bloor Research의 연구 책임자인 Philip Howard와 Monash Research의 사장 Curt Monash의 자문을 참조하였습니다.

혁신적이 되려면 다른 방법으로 사고하고 행동하며 새로운 접근법을 통해 문제를 해결해야 합니다. Netezza는 데이터 웨어하우징에 대해 고객이 필요로 하는 것들에 집중합니다. TwinFin은 당사 고객의 데이터 웨어하우스 쿼리에 대해 탁월한 성능을 제공합니다. TwinFin은 고객에게 간편성을 제공합니다. SQL 및 Linux에 대해 기본적인 지식을 가지고 있는 분들은 누구나, 역동적으로 변하는 워크로드 중 일관된 서비스 수준을 유지하기 위해 필요한 몇 가지 관리 업무 수행 능력을 가지고 있습니다. 간편성을 자랑하는 TwinFin의 성능은 데이터 웨어하우스 유지 및 운영 비용을 절감해 줍니다. 그리고, 당사의 고객들은 이전에 시도가 불가능한 것으로 여겼던 분석 어플리케이션을 활용하여 새로운 비즈니스 가치를 창출합니다.

**"Netezza는 Exadata에 영감을 주었습니다. Teradata도 Exadata에 영감을 주었습니다"라고 Larry Ellison은 2010년 1월 27일에 인정했습니다.**

**"우리의 경쟁력을 자극하여 하드웨어 비즈니스 산업을 시작할 수 있도록 한 점에 있어서 그들에게 감사합니다."**

"Netezza는 Exadata에 영감을 주었습니다. Teradata도 Exadata에 영감을 주었습니다"라고 Larry Ellison은 2010년 1월 27일에 인정했습니다. "우리의 경쟁력을 자극하여 하드웨어 비즈니스 산업을 시작할 수 있도록 한 점에 있어서 그들에게 감사합니다."<sup>1</sup> Larry Ellison의 말 속에 숨겨진 다른 의미는, Oracle은 최고에만 관심을 가진다는 점입니다. Exadata는 Oracle의 전략적 방향성을 나타냅니다. 즉, OLTP 데이터베이스 관리 시스템을 채택하고 Sun의 초병렬 스토리지 시스템과 제휴하는 것을 말합니다. Oracle은 온라인 트랜잭션과 분석 쿼리 모두를 고성능으로 처리하기 위한 Exadata V2를 출시했습니다. Oracle이 OLTP에 탁월하다는 점은 기정 사실입니다. 하지만 데이터 웨어하우징 및 분석은 OLTP와는 매우 다른 요구사항을 소프트웨어 및 하드웨어적으로 수용하여야 합니다. Exadata의 데이터 웨어하우징은 간편성과 가치면에서 심도있는 검토가 필요합니다.

본 백서는 데이터 웨어하우스의 온라인 트랜잭션 처리와 쿼리 및 분석 처리의 차이를 검토하는 것으로 시작합니다. 그 다음, 쿼리 성능, 작업 간편성, 가치의 관점에서 Exadata와 TwinFin에 대해 알아봅니다.

데이터베이스 관리 시스템의 작동 방식에 대한 일반적인 개념을 잠시 잊어두시고, Netezza의 고객과 파트너가 해 온 것처럼 새로운 개념에 대한 개방된 마음을 갖고, 보다 좋은 결과를 얻기 위해 '더'가 아니라 '덜' 할 준비가 된 마음으로 본서를 읽어 주시기 바랍니다.

## ...보다 좋은 결과를 얻기 위해 '더'가 아니라 '덜' 할 준비가 되어 있으시길 바랍니다.

한 가지 주의점: Netezza는 Exadata 머신에 직접적인 액세스를 하지 않습니다. 운 좋게도 저희는 두 가지 기술 모두와 선별된 TwinFin을 평가한 많은 기업들로부터 구체적인 피드백을 받을 수 있었습니다. Exadata에 대한 Oracle의 회사규모와 관심도를 고려할 때 Exadata에 대한 공식적인 정보는 놀라울 정도로 부족합니다. Oracle이 내놓은 사용 사례는 논의를 하기에는 아주 적은 정보에 불과하며 이는 그 자체로 몇몇 업계 분석가들(예: Information Week)도 염려하고 있는 사항입니다.<sup>2</sup> 본 백서에서 공유된 정보는 공개적으로 사용할 수 있습니다. 부정확한 결과가 있다면, 이는 오도의 목적이 아니라 당사의 실수로 인한 것입니다.

<sup>1</sup> [http://oracle.com.edgesuite.net/ivt/4000/8104/9238/12652/lobby\\_external\\_flash\\_clean\\_480x360/default.htm](http://oracle.com.edgesuite.net/ivt/4000/8104/9238/12652/lobby_external_flash_clean_480x360/default.htm) 참조

<sup>2</sup> [http://www.informationweek.com/news/business\\_intelligence/warehouses/showArticle.jhtml?articleID=225702836&cid=RSSfeed\\_IWK\\_News](http://www.informationweek.com/news/business_intelligence/warehouses/showArticle.jhtml?articleID=225702836&cid=RSSfeed_IWK_News) 참조

## 온라인 트랜잭션 처리(OLTP) 및 데이터 웨어하우징

OLTP 시스템은 많은 짧은 트랜잭션들을 실행합니다. 각 트랜잭션의 범위는 작으며, 하나 또는 작은 수의 기록으로 제한되고, 빈번하게 데이터가 캐시메모리에 저장되는 것은 예사입니다. OLTP 시스템이 대용량의 데이터 쿼리를 처리하기는 하나, 현재 데이터 집합(Data Set)에 쓰는 작업(UPDATE, INSERT, DELETE)하는 것에 더욱 집중합니다. 이러한 시스템은 당좌계좌의 현재 잔액을 관리하는 작업과 같은 특정 비즈니스 프로세스나 기능에 한정적인 경우가 일반적입니다. 그러한 데이터는 보통 제3 정규형(3NF, third normal form)으로 구성되어 있습니다. OLTP 시스템의 트랜잭션 유형은 안정적이며 그 데이터 요건은 잘 이해되고 있어서 인덱스와 같은 2차 데이터 구조가 처리를 위해 메모리에 전송되기 전에 디스크의 기록을 유용하게 찾을 수 있습니다.

비교하면 데이터 웨어하우스 시스템은 현재 및 과거의 데이터 집합보다 훨씬 더 많은 데이터베이스 읽기(SELECT) 작업을 하는 것이 특징입니다. OLTP 작업은 작은 수의 기록에 액세스하는 반면, 데이터 웨어하우스 쿼리는 수많은 행의 표를 스캔하여 그 기록을 여러 개의 다른 표와 조인할 수 있을 것입니다. 또한 캐싱 가치 및 인덱싱 전략을 줄일 수 있으며, 비즈니스 사용자들이 웨어하우스의 데이터 중 어떤 쿼리를 요청할지 예측하는 것이 불가능하기 때문에 웨어하우스의 데이터 구성을 위한 선택은 3NF부터, Star 및 Snowflake 구조 등의 모델까지 있습니다.

일반적인 웨어하우스를 피딩하는 각 시스템 내의 데이터는 특정 비즈니스 프로세스의 필요사항을 반영하도록 구성되어 있습니다. 데이터가 웨어하우스에 로딩되기 전에 데이터는 정리, 중복제거 및 통합됩니다.

본 백서는 데이터 웨어하우스를 1세대 또는 2세대로 나누었습니다. 이러한 분류는 매우 심층적인 조사에는 적합하지 않더라도, 얼마나 많은 Netezza의 고객들이 데이터로 훌륭한 가치를 창출해 내는 혁신적인 방향에 대해 이야기하기 시작했는지 설명할 수 있을 것입니다.

1세대 데이터 웨어하우스는 보통 하룻밤 동안 로드됩니다. 1세대 데이터 웨어하우스는 천천히 전개되는 SQL 기반 리포트 및 대시보드의 안정적인 중심체를 통해 비즈니스에 정보를 제공합니다. 이와 같이 단순한 웨어하우스가 OLTP 시스템과 비슷한 면이 있으므로, 즉 워크로드와 데이터 요건이 이해되고 안정적이기 때문에, 기업들이 OLTP에 사용하는 것과 같은 데이터베이스 관리 제품을 채택하는 경우가 많습니다.

제품을 실행하게 되면 데이터베이스 관리자들이 각 리포트의 데이터 요건을 분석하고 데이터 검색을 가속화하기 위해 인덱스를 구축합니다. 웨어하우스의 데이터 볼륨이 트랜잭션 시스템에서 일반적으로 관리되는 양을 증가하기 전까지는 OLTP 기술 및 테크닉에 단점은 없해보입니다.

오늘날 기업들과 공공분야의 기관들은 연간 30-50%의 데이터 성장률을 보통으로 여깁니다. OLTP 세계에서 성공적인 기술과 관례는 데이터 웨어하우징에 점점 더 적용이 어려워지고 있음이 입증되고 있습니다. 데이터 검색을 지원하는 인덱스가 그 적절한 예가 될 수 있습니다. 데이터베이스 시스템은 데이터 로드를 위한 업무를 처리하는 동시에 여러 개의 인덱스를 업데이트하느라 바쁜 경우가 많습니다. 데이터 볼륨이 크기 때문에 이러한 프로세스는 매우 느리게 진행되며 로드 작업이 그 할당된 처리 창을 오버런하게 될 수 있습니다. 긴 작업 시간에도 불구하고 기술 팀은 업체와 협의된 서비스 수준을 놓치지 일쑤입니다. 사업부서가 리포트와 데이터가 이용 가능해질 때까지 기다리는 동안 생산성은 저하됩니다.

## OLTP 세계에서 성공적인 기술과 관례는 데이터 웨어하우징에 점점 더 적용이 어려워지는 것이 입증되고 있습니다...

많은 기업들은 필요한 데이터나 데이터의 활용 방법을 재정의하고 있습니다. 본 백서는 이러한 발전사항을 2세대 데이터 웨어하우스라고 부릅니다. 손쉽게 초대량의 데이터 집합을 관리하는 이러한 새로운 웨어하우스는 기업의 중추역할을 합니다. 검색을 하면 여러 해 전에 기록된 이벤트를 불러오고, 이와 같은 과거의 메모리는 예측적인 분석 어플리케이션의 정확성을 높여줍니다. 지속적인 트리클 피드는 밤샘 배치 작업을 대체하여 관련한 범위의 이벤트의 기록 및 그 분석과 다른 수많은 이벤트 간의 대기시간을 줄입니다. 웨어하우스는 리포트 및 대시보드를 채우는 데 사용되는 단순한 SQL을 뛰어넘어 선형 회귀, Naïve Bayes 및 기타 고급 분석의 수학적 알고리즘을 처리합니다. 단지 다섯개 매장에서 마진이 높은 제품 판매가 급등하는 사실을 파악함으로써 업체는 어떤 일이 일어났고 그 이유가 무엇인지 알 수 있으며 이러한 정보는 150개 위치에 있는 매장 모두에 유사한 판매 활동을 촉진하는 전략을 알려줍니다. 웨어하우스를 지원하는 컴퓨팅 시스템은 정규 리포트 및 대시보드에 영향을 주지 않고 이와 같은 수요급등을 관리할 수 있어야 합니다. 비즈니스 사용자들은 그들이 원하는 시간과 방법으로 데이터를 활용할 수 있는 자유를 요구합니다. 그들이 요구하는 속도는 관리자들과의 튜닝 작업에 의존성이 높은 특성을 가진 기술을 기다려주지 않습니다.

테라바이트-페타바이트 볼륨을 분석하려면 많은 컴퓨터 노드에 데이터를 퍼뜨릴 수 있는 컴퓨터가 필요합니다. 각 컴퓨터 노드는 다른 피어 노드가 모두 그리드에 있을 때 보다 작은 데이터 집합을 병렬로 분석합니다. 끊임없이 변하는 워크로드 조건 하에서 지속적으로 빠르고 일관적인 쿼리 성능을 구현 가능한 웨어하우스는 새로운 등급의 어플리케이션을 가능하게 하여 가치있는 인사이트를 창출하며 데이터 주도적인 의사결정을 채택할 수 있도록 합니다.

### Oracle Exadata의 쿼리 성능

Sun을 인수하면서, Oracle은 Netezza가 10년 전에 내린 것과 같은 결론을 도출했습니다. 모든 부품, 소프트웨어 및 하드웨어가 목표를 위해 최적화되었을 때 데이터 웨어하우스 시스템이 최고의 효율성을 가지게 된다는 것입니다. Exadata는 신속한 네트워크로 연결된 두 가지의 하위 시스템으로부터 만들어졌습니다. 해당 네트워크는 InfiniBand를 통해 리얼 어플리케이션 클러스터(RAC)가 있는 Oracle Database 11g V2와 커뮤니케이션을 하는 스마트 스토리지 시스템입니다. 단일 랙 시스템은 8개의 비대칭 멀티프로세싱 노드의 공유 디스크 클러스터로 실행되는 Oracle RAC 데이터베이스와 연결된 MPP(초고속 병렬 처리) 그리드에 있는 Exadata 셸이라는 14개 스토리지 서버의 스토리지 층을 포함합니다.

**Sun을 인수하면서, Oracle은 Netezza가 10년 전에 내린 것과 같은 결론을 도출했습니다. 모든 부품, 소프트웨어 및 하드웨어가 목표를 위해 최적화될 때 데이터 웨어하우스 시스템이 최고의 효율성을 가지게 된다는 것입니다.**

Oracle은 Exadata의 스토리지 층을 스마트로 라벨링하는데 이는 SQL 프로젝션, 제한 및 단순한 조건 필터링을 처리한 후에 <sup>3</sup> Oracle RAC가 처리한 다운스트림 네트워크에 결과 데이터 집합을 배치하기 때문입니다. 이 테크닉을 스마트 스캔이라고 합니다. 그러나 데이터가 Exadata의 MPP 층에서 처리되는지 여부는 데이터 저장을 위해 사용된 표의 유형, 웨어하우스에 데이터가 도착한 시기, 그 외 아래에서 논하게 되는 요소들을 포함하여 많은 요소들에 따라 다릅니다.

<sup>3</sup> Sun Oracle Exadata 스토리지 서버 및 데이터베이스 머신의 기술적 개요- Oracle 백서, 2009년 10월.



IOT(Index Organized Table, 인덱스 조직 표)라고 부르는 Oracle 구조 내에서 관리되는 데이터는 스마트 스캔으로 처리되지 않습니다. 2010년 2월 Oracle® 데이터베이스 개념 11g 릴리스 2 (11.2)의 3-20 페이지에서 설명서는 "인덱스 조직 표는 연관된 데이터들을 함께 저장해야 하거나 데이터를 물리적으로 특정 순서에 따라 저장해야 하는 경우 유용합니다. 이 유형의 표는 정보 검색, 공간 및 OLAP 어플리케이션에 자주 사용됩니다"라고 권장하고 있습니다. 2010년 2월 Oracle® 데이터베이스 관리 자용 설명서 11g 릴리스 2 (11.2)의 19-53페이지는 "인덱스 조직 표는 어플리케이션 인덱스 구조를 모델링하는 데 적합합니다. 예를 들어, 텍스트, 이미지 및 오디오 데이터를 포함하고 있는 콘텐츠 기반 정보 검색 어플리케이션은 인덱스 조직 표를 이용하여 효과적으로 모델링될 수 있는 새롭게 구성된 인덱스를 필요로 합니다"라고 명시합니다. Netezza 고객들은 TwinFin에서 이와 같은 2세대 데이터 웨어하우스 어플리케이션을 실행합니다. 예를 들어, Levenshtein 편집 디스턴스 알고리즘은 마케팅 캠페인을 위한 이름 및 주소 데이터 정렬에서부터 국가 기밀 기관에서 입력 데이터 포트에 나열된 관심 대상의 이름을 조사할 때 그들이 맡게 되는 복잡한 텍스트 분석에 이르기까지 어플리케이션에 있어서 매우 중요합니다. Oracle의 조언에 따라서 IOT에서 텍스트, 이미지, 오디오 및 공간 데이터를 관리하는 기업들은 Exadata가 그들 데이터에 초고속 병렬 처리력을 활용할 수 없으므로 어플리케이션이 최적으로 수행력을 발휘하지 못하는 것을 확인하게 됩니다.

SMP 서버의 OLTP 데이터베이스로 일반적으로 실행하는 1세대 웨어하우스는 한마디로 업체가 필요로 하는 모든 쿼리를 실행할 처리력이 없었습니다. ODS(Operational Data Store)는 이러한 단점을 보완하기 위해 구축자가 채택하는 하나의 구조를 말합니다. 여러 개의 운용 시스템으로 규칙적인 업데이트가 되는 ODS는 데이터를 통합하고 어플리케이션에 트랜잭션 및 분석 서비스를 모두 제공합니다. MPP 아키텍처를 실행하는 2세대 웨어하우스는 기업에 작업 데이터 저장소를 웨어하우스에 통합할 수 있는 기회를 제공합니다. Exadata는 단일 활성 트랜잭션(insert, update, delete)을 포함하고 있는 블록에 스마트 스캔을 활용할 수 있는 능력을 가지고 있지 않으며 Exadata 블록은 많은 기록을 포함할 수 있습니다. Oracle의 이러한 변동성은 Exadata를 ODS로 고려하는 고객들에게 복잡성의 문제를 초래합니다.

Oracle의 최근 데이터베이스 릴리스는 511가지 기능을 제공하지만, 그 중 319가지 기능만이 Exadata의 MPP 스토리지 층에서 처리 가능합니다. 스마트 스캔의 조건성은 기능의 조건자에까지 확대되는데, 예를 들어 Exadata는 날짜/시간 기능을 조건자 months\_between(d,sysdate) = 0, 또는 months\_between(d,current\_date) = 0, 또는 months\_between(d,to\_date('01-01-2010','DD-MM-YYYY')) = 0을 사용하여 활용하지 않을 것입니다.<sup>4</sup>

<sup>4</sup> <http://antognini.ch/2010/08/exadata-storage-server-and-the-query-optimizer-%E2%80%93-part-4/>

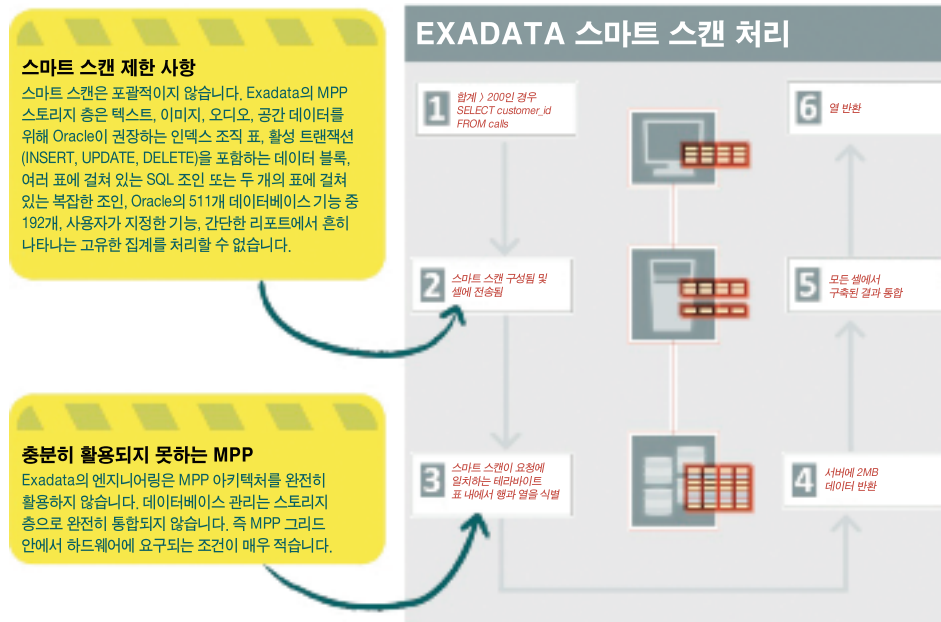
Exadata의 스토리지 층은 Bloom 필터를 제공하여 하나의 큰 표와 다른 하나의 작은 표 간의 단순한 조인을 실행하도록 합니다. 그보다 더 복잡한 것은 MPP에서 처리될 수 없습니다. 분석 쿼리는 보통 Exadata가 지원하는 것보다 더 복잡한 조인을 필요로 합니다. 다국적 소매기업이 영국에 위치한 자사 매장에서의 달러 가치에 대한 인사이트를 필요로 하는 경우를 생각해 봅시다. 이 단순한 SQL 쿼리에는 세 가지 표-판매, 통화 및 매장 간의 조인이 필요합니다.

```
select sum(sales_value * exchange_rate) us_dollar_sales
from sales, currency, stores
where sales.day = currency.day
and stores.country = 'UK'
and currency.country = 'USA'.
```

Exadata는 MPP 층에서 이 세 표의 조인을 처리할 수 없으며 계산에 필요한 모든 데이터를 네트워크에 걸쳐 Oracle RAC로 비효율적으로 옮겨야 합니다.

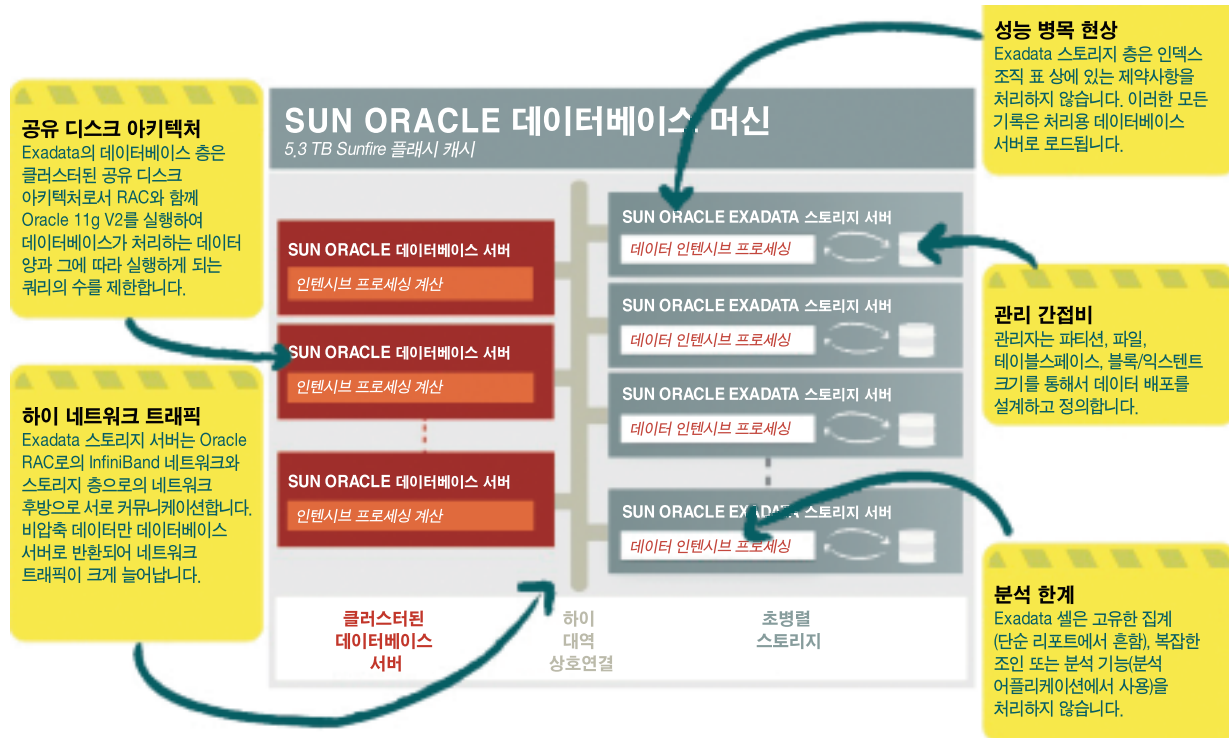
UDF(사용자 정의 함수)는 2세대 웨어하우스에서 중심적인 역할을 합니다. UDF는 쿼리 기능을 확장하여 SQL을 능가하며 임베디드 분석 처리가 가능합니다. Exadata는 MPP에서 UDF 처리가 불가능하며 큰 데이터 집합을 분석하도록 설계되지도 않았으며 그런 능력도 없는 OLTP 데이터베이스인 Oracle RAC에 그 처리 임무를 넘겨야 합니다.

Exadata의 스토리지 층은 단순한 리포트에서도 고유 집계를 처리할 수 없는 경우가 많습니다. 이는 분석 쿼리에서 매우 많이 사용되는 사항입니다.



Exadata 스토리지 서버는 다른 서버와 커뮤니케이션을 할 수 없습니다. 모든 커뮤니케이션은 InfiniBand 네트워크를 통해 Oracle RAC로 강제로 보내진 뒤 다시 네트워크를 통해 스토리지 층으로 이루어집니다. 이 아키텍처는 하나 또는 몇 개의 기록을 가진 각 트랜잭션을 이용해 작은 데이터 집합을 스토리지에서 데이터베이스로 옮기게 되는 정도의 온라인 트랜잭션 처리에 유용합니다. 매우 큰 데이터 집합의 분석은 Exadata가 그 스토리지와 데이터베이스 층 사이의 데이터 대용량을 옮기도록 하여 쿼리 성능에 부정적인 결과를 초래합니다.

Oracle은 TwinFin보다 강력한 40 Gb/sec 스위치 InfiniBand를 사용합니다. 실제로 Exadata는 시스템의 불균형과 비효율로 인해 이와 같이 고가의 네트워크를 필요로 합니다. Exadata 스토리지 서버는 많은 업무를 처리하지 못해서, 필요한 것보다 더 많은 데이터를 네트워크에 배치하여 과도한 업무 요청을 받게 되는 Oracle RAC가 처리하게 합니다.



Exadata 스토리지 층의 모든 디스크는 Oracle RAC를 실행하는 그리드의 모든 노드로 공유됩니다. 이러한 공유 스토리지는 한 노드가 페이지를 읽는 동안 다른 노드가 페이지를 업데이트하게 되는 위험을 초래합니다. 이를 관리하기 위해 Oracle은 노드 간의 조정을 강제합니다. 각 노드는 충돌을 막기 위해 피어의 디스크 활동을 확인합니다. Oracle의 기술자들은 이러한 활동을 블록 핑잉(Block Pining)이라고 합니다. 각 노드가 동료의 디스크 활동을 확인할 때 소비되거나, 한 노드가 작업 완료를 위해 다른 노드를 기다리는 시간만큼 컴퓨팅 시간은 낭비됩니다. 데이터 웨어하우징을 위해 특별히 설계된 데이터베이스에서는 이런 시간이 낭비되지 않고 사용되어 쿼리를 처리하고, OLTP에는 유용하며, 데이터를 마이닝하며 분석을 실행합니다.

주식 시장 분야에서의 상관성 확인을 위한 주식거래 데이터 분석에 관심이 있는 기업의 경우를 생각해 봅시다. 해당 기업의 알고리즘은 두 가지 변수 간의 관계가 얼마나 잘 설명될 수 있는지를 평가하고 두 변수 간의 통계적 상관관계를 측정하여 Spearman의 순위 상관계수(Spearman의 rho)를 계산합니다. 이러한 분석은 특정 주식이 같은 시장 분야의 다른 주식 행위에 영향을 미치는지에 대해 1-10분 내에 의미있는 인사이트를 창출해 냅니다. 데이터 집합은 큼니다. 시장의 움직임에 따라 분석될 과거 기록된 시장 데이터와 일일 거래의 긴 이력이기 때문입니다. 분석을 위해서는 분야의 모든 주식 간에 Cartesian 조인이 필요하며, 동시에 조사 중인 주식에 대한 거래량 가중 평균 가격 및 이전 마감으로부터의 수익 가치를 계산해야 합니다. 결과는 Spearman의 순위 상관계수 기능으로 넘겨져 해당 기간에 대한 모든 주식 결합의 모공분 및 표준편차를 계산합니다. 테스트를 시작한 지 몇 시간 후에도 고객은 여전히 Oracle 10g로부터의 결과를 기다리고 있었습니다. 이러한 낮은 성능으로 인해 Oracle 10g는 최대 10분만 주어지는 비즈니스 사례의 기술로서는 적합하지 못하게 됩니다.

Exadata Oracle 11gR2를 MPP 스토리지 시스템 및 신속한 네트워크에 연결 시킨다 해도 스마트 스캔은 이 분석 과정의 첫 단계인 Cartesian 조인을 처리할 수가 없으며 이 계산에 0이란 값을 내놓습니다. Exadata는 네트워크에 모든 기록을 배치하여 Oracle RAC로 전송할 것입니다. 이제 'Oracle RAC가 Oracle 10g보다 얼마나 더 효율적인가?'에 대한 질문을 하게 됩니다. 알고리즘 처리에서 Oracle은 임시 데이터 집합을 만들고 관리해야 하며 스토리지를 위해 이를 메모리에서 써야 합니다. Exadata의 플래시 캐시가 여기서 몇 가지 역할을 할 수도 있지만 데이터 집합의 크기와 알고리즘의 복잡성으로 인해 데이터베이스 처리는 디스크에 써야만 하게 될 것입니다. Oracle RAC의 이러한 흐름은 네트워크로 돌아가 MPP 스토리지 층 데이터에서 오는 것으로 여전히 막혀 있고, 완전 작업 중인 Oracle RAC의 주목을 받기 위해 차대를 기다리지만 처리되지 못합니다. Exadata의 서버 및 공유 디스크 간의 네트워크 연결성은 병목현상을 보입니다. OLTP 데이터베이스로 설계된 Oracle RAC는 병목현상을 보입니다.

# 모든 단순 쿼리에 대해 Exadata는 큰 데이터 집합을 스토리지 층에서 데이터베이스 층으로 옮기는데 이는 현대 데이터 웨어하우스를 위한 플랫폼으로서의 적절성에 의문을 제기합니다.

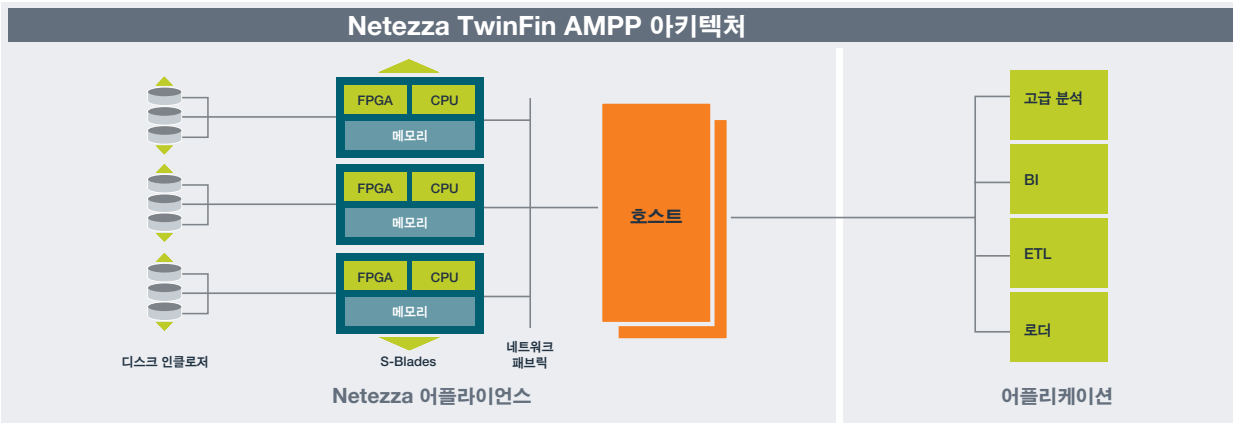
데이터 웨어하우징에 대해, Exadata의 MPP 스토리지 층은 우선적으로 조건자 수준의 필터링(SQL 설명의 "where" 절을 기반으로 하는 행 수준 제한, "select" 절을 기반으로 하는 열 수준 프로젝션) 수행과, 데이터가 스캔되면서 디스크에 저장된 데이터 압축풀기에 대한 CPU 처리력을 제공하는 것으로 제한됩니다. 또한 Exadata의 MPP 층 처리는 조건적이어서 관리자들이 머신의 복잡성을 깊게 연구해야만 합니다. 나머지 모든 복잡한 조인, 집계 및 필터링 논리, 웨어하우징 및 데이터 분석의 헤비 리프팅은 그 다음 Oracle 11g RAC를 실행하는 클러스터된 SMP 헤드 노드의 모든 Exadata 노드의 결과에 따라 실행됩니다. Exadata는 여러 개의 OLTP 시스템을 단일 플랫폼으로 통합하고자 하는 CIO에 흥미로운 기회를 제공합니다. 모든 단순 쿼리에 대해 Exadata는 큰 데이터 집합을 스토리지 층에서 데이터베이스 층으로 옮기는데 이는 현대 데이터 웨어하우스를 위한 플랫폼으로서의 적절성에 의문을 제기합니다.

## IBM Netezza TwinFin의 쿼리 성능

IBM Netezza TwinFin은 데이터 웨어하우징 플랫폼으로서 설계되었습니다. Netezza는 AMPP(비대칭 초병렬 처리) 아키텍처를 사용합니다. 비대칭 Multiprocessing 호스트<sup>5</sup>는 초병렬 처리 노드의 그리드를 향하고 있습니다. TwinFin은 이 MPP 그리드를 활용하여 웨어하우징 및 분석 데이터의 헤비 리프팅을 처리합니다.

TwinFin의 그리드에 있는 노드를 멀티코어 CPU(중앙처리장치)를 포함하고 있는 독립 서버인 S-Blade(Snippet-Blade)라고 합니다. 각 CPU는 멀티엔진 FPGA 및 RAM의 기가바이트와 연결됩니다. CPU는 자체 메모리를 가지고 있으므로 데이터 분석에만 집중할 수 있으며 Oracle RAC의 블록 핑잉에서 발생하듯 다른 노드의 활동 추적으로 방해 받는 일이 없습니다.

<sup>5</sup> TwinFin은 중복성을 위해 두 개의 SMP 호스트를 가지지만 한 번에 하나만 활성화할 수 있습니다.



FPGA는 거의 모든 논리적 기능을 실행할 수 있도록 프로그램 가능한 수 많은 내부 게이트가 장착된 반도체 칩이며, 특히 스트리밍 처리 작업을 관리할 때 효과적입니다. Netezza 외부에서 FPGA는 디지털 신호 처리, 의료 영상 및 음성인식과 같은 어플리케이션에 사용됩니다. Netezza의 엔지니어들은 당사 어플리케이션의 FPGA 내에 소프트웨어 머신을 구축하여 FPGA가 CPU에 연결되기 전에 데이터 처리를 가속화하였습니다. 각 Oracle은 각 Exadata 랙 내에서 14개의 8방향 스토리지 서버를 사용하여 블레이드 서버 내에 포함된 48 FPGA를 사용한 Netezza보다 낮은 성과를 보입니다. 각 FPGA(실리콘 1"x1" 정사각형)은 현저한 효율성을 자랑하며 더 적은 전력을 사용하고 열 발생도 적습니다.

**Netezza는 알려진 단점이 있는 구식 시스템을  
채택하지 않았으며 새로운 스토리지 층과  
결합시키지도 않았습니다. TwinFin은 데이터  
웨어하우징을 위한 최적의 플랫폼으로 설계되었습니다.**

**Netezza MPP** 그리드의 노드 간 커뮤니케이션은 네트워크 패브릭에서 발생하며 사용자 지정 IP 기반 프로토콜을 실행하고 총 단면 대역을 완전히 활용하여 지속적이며 매우 활동적인 네트워크 트래픽 하에서도 정체를 제거해 줍니다. 네트워크는 1000개 이상의 노드를 조정하도록 최적화되어 있으며 각 노드가 모든 다른 노드에 동시에 큰 데이터 전송을 시작할 수 있도록 합니다. 이러한 전송은 데이터 웨어하우징과 고급 분석의 처리 작업 전형에 엄청난 효율성을 허용합니다. SQL문이 **TwinFin**의 **MPP** 아키텍처 내 처리를 유용하게 사용하는 것과 마찬가지로 고급 분석 중심의 복잡한 계산 알고리즘 역시 그렇습니다. 이전 세대의 기술은 물리적으로 어플리케이션 처리를 데이터베이스 처리와 분리하여 큰 데이터 집합이 웨어하우스에서 분석 처리 플랫폼으로 옮겨질 때와 그 반대의 경우에 비효율성과 제약이 있었습니다. **Netezza**는 그 **MPP** 그리드에 고급 분석 계산법을 도입하여 데이터에 물리적으로 가까운 각 CPU에서 알고리즘을 실행하고 데이터 이동을 충분히 하여 성능을 강화합니다. 알고리즘은 확장이 어렵게 클러스터된 시스템에 부과된 제약에서 자유로운 **Netezza MPP** 그리드의 많은 노드에서 실행하는 혜택을 받습니다.

**TwinFin**은 True **MPP** 아키텍처가 실제 데이터집합의 실제 워크로드인 **Spearman**의 순위 상관계수 산출에서 어떻게 뛰어난지를 보여줍니다. 고객은 초용량의 거래 데이터를 **TwinFin**에 로드하여 라이브 시장에서 웨어하우스로 지속적으로 트리클 피드를 합니다. **TwinFin**은 모든 하드웨어와 소프트웨어 리소스를 활용하여 병렬로 모든 쿼리 단계를 실행합니다. **Netezza**의 지능적인 스토리지는 해당 분야에서 필요로 하는 행만 선택하며 평가에 필요한 열만 프로젝팅합니다. 조인 결과는 **TwinFin**이 그 **MPP** 그리드의 모든 프로세서에 다운로드하는 통계적 분석을 실행하는 코드에 직접 스트림되어 복잡한 계산을 병렬로 실행합니다. **MPP** 그리드의 각 노드로부터의 결과는 최종 어셈블리를 위해 네트워크에서 호스트로 반환되며 요청 어플리케이션으로 다시 렌더링됩니다. **TwinFin**은 몇 분 후(기업의 10분 내에)에 분석을 완료하고 기업에서 역동적인 오픈된 시장의 여러 주식 행위를 평가하고자 하는 경우 규칙적으로 분석을 실행합니다. 금융 서비스 업계가 아니더라도 **Spearman**의 순위 상관계수는 계량경제학, 품질 관리, 생물정보학 및 마케팅 분야에서 일하는 통계가들에게 중요한 정보입니다.

**Netezza**는 알려진 단점이 있는 구식 시스템을 채택하지 않았으며 새로운 스토리지 층과 결합하지도 않았습니다. **TwinFin**은 데이터 웨어하우징을 위한 최적의 플랫폼으로 설계되었습니다. **TwinFin**은 초병렬 그리드에서 모든 쿼리를 실행하며 아낌없이 성능을 발휘하여 프로그래머, 관리자, 사용자에게 편리함을 전해 줍니다.

IBM Netezza의 한 금융 서비스 업계 고객은 Oracle 데이터 웨어하우스 관리에 필요한 리소스 경비를 분석하기 위해 반드시 필요하지 않은 일체의 모든 과정을 제거하는 린 접근법(Lean Approach)을 사용했습니다. 그리고 인덱스, 집계, 구체화된 뷰 및 데이터 마트 등을 구축하고 유지하는 과정의 자사 IT 업무 90% 이상이 낭비였거나 비 부가가치적인 처리였음을 알게 되었습니다.

#### Oracle Exadata의 작업 간편성

Oracle은 쿼리에 성공적인 답을 보장할 수 있는 데이터베이스 설계를 위해서 경험 많고 교육을 많이 받아 기술력이 높은 데이터베이스 관리자들에게 의존해야 합니다. 한 설계 테크닉은 데이터의 물리적 파티셔닝으로 곧 쿼리가 관련 데이터를 포함한다고 알려진 파티션 내로 제한될 수 있다는 것을 의미합니다. 파티션은 Oracle OLTP에서 기원합니다. 2세대 웨어하우스 관리자들은 내일 또는 다음 주에 어떠한 쿼리와 분석이 실행될지 알 수 없습니다. Exadata는 모든 이용 가능한 디스크 파티션의 스트라이핑을 자동화하는 Oracle의 자동 스토리지 관리(Automatic Storage Management)를 포함합니다. 그러나 관리 팀은 여전히 파티션을 만들고, 인스턴스에 걸친 공유 스토리지를 위해 디스크 그룹을 구성 및 관리하며, 2-way 미러링 또는 3-way 미러링을 선택 및 실행하고, 할당 단위 크기를 구성해야 합니다. 또한 Exadata 구성을 위해 관리자들은 테이블스페이스, 인덱스 공간, 임시 공간, 로그 및 익스텐트를 생성하여 관리해야 합니다. 업계 분석가인 Curt Monash는 "우수한 성능을 얻기 위해 복잡한 파티셔닝에 지나치게 의존해야 한다면 더 좋은 시스템이라고 할 수 없다"라고 평했습니다.<sup>6</sup>

<sup>6</sup> Curt Monash [www.dbms2.com/2009/09/21/notes-on-the-oracle-database-11g-release-2-white-paper/](http://www.dbms2.com/2009/09/21/notes-on-the-oracle-database-11g-release-2-white-paper/)



이전 섹션에서 설명된 바와 같이 Exadata는 그 MPP 스토리지 층에서 조건적으로 데이터를 처리할 수 있는지 여부와 관계없이, 이 복잡한 제품 사용을 위해서는 이전 Oracle 버전 사용 경험이 있으며 수 많은 교육 수업에 참여한 관리자들이 필요합니다. Oracle은 "DBA가 Exadata 지식과는 별도로 최신 (11gR2) 지식을 보유해야 합니다."라고 권장합니다.<sup>7</sup> 최신 지식 습득을 위해서는 각 관리자가 총 18일에 걸쳐 최소 네 가지의 코스를 이수해야 합니다. Oracle RAC를 처음 사용하는 관리자들은 관리상의 복잡함을 배우기 위해 5일을 더 공부해야 합니다.

IBM Netezza의 한 금융 서비스 업계 고객은<sup>8</sup> Oracle 데이터 웨어하우스 관리에 필요한 리소스 경비를 분석하기 위해 반드시 필요하지 않은 일체의 모든 과정을 제거하는 린 접근법(Lean Approach)을 사용했습니다. 그리고 인덱스, 집계, 구체화된 뷰 및 데이터 마트 등을 구축하고 유지하는 과정의 자사 IT 업무 90% 이상이 낭비였거나 비부가가치적인 처리였음을 알게 되었습니다. 이와 같은 낭비 비용으로는 불필요한 하드웨어 및 소프트웨어 라이선스 비용, 낭비된 스토리지 테라바이트, 연장된 개발 및 데이터 로드 기간, 장기간의 데이터 사용 불가능성, 부실 데이터, 저성능의 로드 및 쿼리, 과도한 관리 비용 등이 있습니다.

Exadata는 Oracle 데이터 웨어하우스 관리를 간소화하는 데 큰 영향을 미치지 못합니다. 관리자들은 여러 개의 서버 층을 관리해야 하며, 각 층은 운영 시스템 이미지, 펌웨어, 파일 시스템 및 소프트웨어를 유지해야 합니다. Oracle은 DBA가 Exadata의 데이터베이스 버전인 11g를 관리하는데<sup>9</sup> 이전 버전인 10g 구축에서 사용한 것과 비교하여 26%의 시간을 감축할 수 있다고 제안합니다. 실제로 이것이 확인된다면 Exadata는 고객들이 관리에 쓸데없이 낭비하는 시간의 1/4을 줄이게 되는 것이며 Oracle은 올바른 방향으로 나아가고 있는 것이 됩니다. Netezza의 어플라이언스는 고객의 시간을 낭비하지 않도록 설계되었습니다. "DBA 팀은 환경을 지원하고 어플라이언스에 대한 높은 수준의 보안 모델을 관리하기만 하면 됩니다. 그 이외에는 할 것이 없습니다(예를 들어, 인덱싱 개념은 Netezza에서는 완전히 새로운 개념입니다)."<sup>10</sup>

<sup>7</sup> [http://www.oracle.com/global/hu/events/20100126/exadata\\_migration\\_budapest\\_robert\\_pastijn.pdf](http://www.oracle.com/global/hu/events/20100126/exadata_migration_budapest_robert_pastijn.pdf)

<sup>8</sup> 제조에 근원을 두는 "린(Lean)"이란 식스시그마(Six Sigma)의 도구와 테크닉을 사용하는 사례로서, 낭비적인 리소스 경비와 제품 또는 서비스에 가치를 추가하지 않는 대상 활동을 분석하여 제거하는 것입니다.

<sup>9</sup> <http://www.dbms2.com/2009/09/21/notes-on-the-oracle-database-11g-release-2-white-paper/>

<sup>10</sup> *Linked-In Exadata Vs Netezza 포럼에서 OLTP를 위해 Oracle을, 데이터 웨어하우스를 위해 Netezza를 사용하는 고객의 코멘트 인용*  
[http://www.linkedin.com/groupAnswers?viewQuestionAndAnswers=&gid=2602952&discussionID=11385070&sik=1275353329699&trk=ug\\_qa\\_q&goback=.ana\\_2602952\\_1275353329699\\_3\\_](http://www.linkedin.com/groupAnswers?viewQuestionAndAnswers=&gid=2602952&discussionID=11385070&sik=1275353329699&trk=ug_qa_q&goback=.ana_2602952_1275353329699_3_)

비즈니스 사용자들은 쿼리가 신속히 완료되기를 바라는 것과 동시에 일관성 있는 성능도 기대합니다. 어제는 5초만에 완료된 리포트가 오늘은 3분이 걸렸다면 IT 헬프데스크 직원에게 불만을 제기할 가능성은 높아질 것입니다. 웨어하우스는 다양하고 역동적인 워크로드의 수요에 불가피하게 대처해야 하기 마련입니다. 일괄 작업 또는 트리클 피드를 통해 OLTP 시스템에서 도착하는 데이터는 비즈니스 뷰에 나타나지 않고 Background에서 실행되는 백업, 복원 및 정리와 같은 관리 작업이며 대시보드는 끊임없이 업데이트됩니다. 동시에, 계산 집중적인 어플리케이션(어떤 청구 또는 거래가 사기이거나 불규칙한 가능성이 있는지 예측하는 어플리케이션)은 웨어하우스 인프라에 갑작스럽게 많은 로드를 생성합니다. 업체에 일관성 있는 성능을 제공하기 위해서 웨어하우스에 필요한 두 가지 요건이 있습니다. 일관성 있는 쿼리 성능과 효과적인 워크로드 관리가 그것이며 워크로드 관리는 서비스(보통 업체와 합의한 우선순위를 기반)가 필요한 모든 작업에 대해 이용 가능한 컴퓨팅 능력할당을 간소화하는 것을 말합니다.

Oracle에서 워크로드를 관리하기 위해 관리자들은 여러 튜닝 매개변수(다수가 다른 매개변수에 대한 의존도가 높음)에 대해 배워야 합니다. 특정 매개변수를 그리드의 모든 프로세서에 대해 같은 값으로 설정해야 하는 Exadata의 경우는 상황이 더 복잡합니다. 관리자들은 탑-다운 시스템 컨트롤 대신 실험적으로 개별 매개변수 설정을 변경하며 프로덕션 환경에서의 효과를 관찰하면서 바텀-업으로 작업을 해야 합니다. Oracle의 워크로드 관리 철학은 온라인 트랜잭션 처리에는 성공적입니다. 온라인 트랜잭션 처리에서, 각 어플리케이션이 데이터베이스와 서버에 배치하는 로드를 개발 환경에서 프로파일하고, 매개변수 설정의 효과를 확인 및 이해하며, 증분 조정을 할 수 있습니다. 안정적인 시스템이 이루어지면 매개변수 설정을 프로덕션에 배치할 수 있습니다.

2세대 웨어하우스에는 "정상적인 로드"라는 게 없습니다. 비즈니스 사용자는 외부 세계에서 이벤트가 발생하는 것에 따라 대량의 데이터 집합에 대해 계산 요구가 많은 분석을 요청할 것입니다. 분석가들은 다른 로드가 안정될 때까지의 유휴시간을 기다려 주지 않을 것입니다. 웨어하우스의 역동적인 특성으로 인해 워크로드 관리 시 Oracle 튜닝 접근법의 허약함이 노출됩니다. 상승/하강하는 로드를 통해 다양한 어플리케이션 및 데이터 요건을 가진 많은 사용자 커뮤니티를 위한 일관성 있는 성능을 이루어 유지하는 것은 복잡한 업무입니다. 많은 교육을 받고 경험이 풍부한 Oracle 사용 경험이 있는 관리자 팀조차 규칙적인 리포트 및 로드 업무에 대한 일관성 있는 성능을 제공하고, 예상치 못했지만 중요한 분석 활동을 위한 여유를 남겨둘 수 있는 안정적인 웨어하우스를 유지하기란 쉽지 않은 일입니다.

"우리가 Netezza와 PoC(개념 증명)를 한 방법은 Netezza가 우리에게 상자를 배송하고 우리는 데이터 센터에 그것을 배치하여 우리 네트워크에 연결한 것이 다입니다. 24시간 이내에 설치를 완료하여 실행할 수 있었습니다. 제가 과장하는 것이 아닙니다. 정말 그렇게 쉬웠습니다."

Oracle RAC는 복잡한 기술이며 그 튜닝 매개변수에 대한 정보는 다 알려져 있지 않은 상태입니다. Exadata의 복잡성은 Oracle이 Oracle의 이전 데이터베이스 버전에 익숙한 관리자들에게조차 수일의 교육을 권장하는 것만으로도 확인할 수 있습니다.

#### IBM Netezza TwinFin의 작업 간편성

Netezza의 고객들은 당사 어플라이언스의 설치와 사용이 얼마나 간단한지에 대해 자신있게 말합니다. "우리가 Netezza와 PoC(개념 증명)를 한 방법은 Netezza가 우리에게 상자를 배송하고 우리는 데이터 센터에 그것을 배치하여 우리 네트워크에 연결한 것이 다입니다"라고 그는 말합니다. "24시간 이내에 설치를 완료하여 실행할 수 있었습니다. 제가 과장하는 것이 아닙니다. 정말 그렇게 쉬웠습니다." 이 인용문은 이미 Oracle의 데이터베이스와 RAC 소프트웨어를 사용 중인 기업 eHarmony, Inc.의 부사장인 Joseph Essas의 의견입니다.

시간 단축을 통해 생산성을 향상시킨다는 것은 좋은 시작입니다. Netezza의 철학은 데이터 웨어하우스의 모든 단계에 단순성과 간편성을 부여하는 것입니다. 고객이 만나게 되는 첫 작업은 데이터 로드입니다. TwinFin은 데이터 배포를 자동화합니다. PoC(개념 증명) 프로젝트에서 고객들은 자동 배포를 이용해 Netezza에 데이터를 로드하며, 쿼리를 실행하고, Oracle에 맞추어 높게 조절된 환경에서 나온 결과와 비교하는 것을 경험하게 됩니다. 모든 그러나 가장 단순한 쿼리에 대한 TwinFin의 자동 배포는 Oracle을 능가할 만큼 우수합니다. 고객들은 추후에 쿼리를 분석하여 다른 키에 데이터를 재배포함으로써 가속화할 수 있는 쿼리를 식별할 수 있습니다. TwinFin은 작업을 단순화합니다.

<sup>11</sup> <http://www.itworld.com/software/61575/eharmony-finds-data-warehouse-match-netezza>

## 단순한 접근법은 다음을 의미합니다.

- 클러스터 상호연결(GES & GCS) 모니터링/튜닝 필요 없음
- RAC용 지식/튜닝 필요 없음 (RAC 경험이 있는 DBA일 필요가 없음)
- 데이터베이스 공간/테이블스페이스 크기 조절 및 구성 필요 없음
- 다시 실행/실제 로그 크기 조절 및 구성 필요 없음
- 저널링/논리적 로그 크기 조절 및 구성 필요 없음
- 표를 위한 페이지/블록 크기 조절 및 구성 필요 없음
- 표를 위한 익스텐트 크기 조절 및 구성 필요 없음
- 임시 공간 할당 및 모니터링 필요 없음
- OS 커널 권장 사항 통합 필요 없음
- OS 권장 패치 수준 유지 필요 없음
- 호스트/네트워크/스토리지 구성을 위한 JAD 세션 필요 없음
- 쿼리(예: `first_rows`) 및 옵티마이저(예: `optimizer_index_cost_adj`) 힌트 필요 없음
- 스탯팩(토계, 캐시 히트, 대기 이벤트 모니터링) 필요 없음
- 메모리 튜닝(SGA, 블록 버퍼 등) 필요 없음
- 인덱스 계획/생성/유지 필요 없음
- 간단한 파티셔닝 전략: 해시 또는 라운드 로빈

TwinFin에 제출되는 모든 쿼리는 데이터베이스 관리자들의 참여 없이 초병렬 그리드에서 자동으로 처리됩니다. 쿼리와 분석은 호스트 머신을 통해 TwinFin에 들어갑니다. 호스트 머신에서는 옵티마이저, 컴파일러, 스케줄러가 수 많은 다양한 피스와 스니펫으로 쿼리와 분석의 구성을 해제하고, 노드 또는 S-Blade를 처리하는 MPP 그리드에 이러한 지침을 배포하며 이들 모두는 로컬로 관리되는 데이터 슬라이스에 대해 동시에 워크로드를 처리합니다.

TwinFin의 각 S-Blade에 도착하는 스니펫은 디스크에서 메모리로 압축된 데이터를 읽기 시작합니다. FPGA가 그 다음 메모리 버퍼로부터 데이터를 읽고 Compress 엔진을 활용하여 데이터의 압축을 풀며, 디스크의 각 블록을 FPGA 내에 있는 4-8 데이터 블록에 바로 전송합니다. Netezza의 엔지니어링은 모든 데이터 웨어하우스의 가장 느린 구성요소(디스크)도 가속화할 수 있습니다. 그 다음 FPGA 내에서 데이터는 처리 중인 SQL 쿼리의 SELECT 절에 지정된 매개변수를 기반으로 하는 열을 필터링하는 Project 엔진으로 스트리밍합니다. SELECT 절을 충족하는 기록만 WHERE 절에 지정된 제한을 기반으로 Restrict 엔진(쿼리를 처리하는 데 필요하지 않은 행은 게이트 통과에서 차단)으로 다운스트림 쪽으로 통과됩니다. Visibility 엔진은 스트리밍 속도로 ACID(원자성, 일관성, 독립성, 지속성)를 유지합니다. 지속적으로 불필요한 열과 행을 잘라내는 이 모든 작업은 겨우 사방 3cm가 안되는 에너지 효율적인 FPGA 칩에서 이루어집니다. TwinFin은 데이터를 옮길 필요가 없는 경우 옮기지 않습니다.

FPGA의 사전 처리가 완료되면 크게 줄어든 기록의 결과 집합만 S-블레이드 메모리(CPU가 정렬, 조인, 집계 등의 높은 수준의 데이터베이스 작업을 수행, MPP 그리드 내에서 다른 모든 CPU와 병렬로 이 작업을 진행)로 스트리밍합니다. CPU는 또한 고급 분석 처리를 위해 스니펫 코드에 포함된 복잡한 알고리즘을 적용할 수도 있습니다. CPU는 최종적으로 전체 데이터 스트림에서 모든 중간 결과를 어셈블하며 스니펫 코드의 지시에 따라 네트워크 패브릭을 통해 다른 S-Blades 또는 호스트에 전송된 스니펫에 대한 결과를 생성합니다. JOIN이 필요로 하는 데이터가 노드에서 연결되지 않은 경우, TwinFin의 노드 간 네트워크 패브릭은 데이터베이스가 제한사항과 프로젝션을 완료한 뒤 처리 기간의 후반에 효율적으로 간단히 재배포합니다. 일부 매우 복잡한 알고리즘은 그 해답을 산출하기 위해 노드 간의 커뮤니케이션을 필요로 합니다. TwinFin은 중간 결과를 커뮤니케이션하고 최종 결과를 생성하기 위해 인터페이스를 통과하는 메시지를 활용합니다.

또한 압축된 오리지널 데이터 블록이 여전히 메모리에 있으므로 TwinFin의 테이블 캐시를 통해 유사한 데이터를 필요로 하는 차후 쿼리에 자동으로 재사용할 수 있습니다(DBA 교육 또는 참여를 필요로 하지 않는 자동화 메커니즘).

# Netezza로 전환한지 겨우 3개월 뒤, 한 고객은 그의 팀이 과거 3년 동안 Oracle에서 했던 것보다 더 많은 분석 어플리케이션을 제공할 수 있었다고 말합니다.

TwinFin은 모든 작업에 대해 병렬 처리를 적용하므로 그 워크로드 관리 시스템은 어플라이언스의 계산 리소스를 각각 그리고 모든 업무에 이용할 수 있도록 컨트롤하는 데 중요한 역할을 합니다. Netezza의 어플라이언스 아키텍처에서 하나의 소프트웨어 구성요소가 모든 시스템 리소스, 즉 프로세서, 디스크, 메모리, 네트워크를 컨트롤합니다. 이러한 간결함이 TwinFin 워크로드 관리 시스템의 기초입니다. TwinFin의 워크로드 관리 시스템은 관리자들이 업체와 합의한 우선순위를 기반으로 사용자 및 그룹에 계산 리소스를 할당하고 여러 커뮤니티를 위한 일관성 있는 응답 시간을 유지하기 쉽도록 도와줍니다.

TwinFin은 데이터 튜닝 작업 낭비를 제거해 줍니다. 지능적인 결정을 할 수 있도록 장착된 Netezza의 어플라이언스는 튜닝 작업이 필요 없으며 시스템 관리도 거의 필요 없습니다. 역동적으로 변하는 워크로드의 일관성 있는 성능 유지를 위해 필요한 몇 가지 관리 업무는 Linux 및 SQL 경험이 있는 분이라면 누구라도 쉽게 할 수 있습니다. 관리자들이 해야 할 부분이 있다면 그것은 사용자 커뮤니티 내의 그룹에 TwinFin의 리소스를 할당하고 그를 워크로드 관리 시스템에 맡기는 것입니다. 지속적인 데이터베이스 관리 기간에서 자유로워진 기술 직원들은 데이터를 활용하는 새롭고 가치 창출적인 방법을 연구할 수 있습니다. Netezza로 전환한지 겨우 3개월 뒤, 한 고객은 그의 팀이 과거 3년 동안 Oracle에서 했던 것보다 더 많은 분석 어플리케이션을 제공할 수 있었다고 말합니다. 데이터가 관리되는 곳에 가까운 분석 어플리케이션 처리, SQL 처리에서 사용되는 것과 같은 MPP 플랫폼 활용을 통해 기업은 데이터에서 얻는 가치를 극대화할 수 있는 실제적인 기회를 가질 수 있습니다.

### Exadata의 가치

엄청난 볼륨의 데이터, 복잡하고 계산 집중적인 경우가 많은 워크로드, 데이터 웨어하우스 실행을 위해 필요한 하드웨어 및 소프트웨어는 곧 그만큼의 많은 경비를 의미합니다. 1세대 웨어하우스 구축을 위해 OLTP 데이터베이스를 사용하는 많은 기업의 경험은 전문 직원 교육 및 고용 비용이 자산 비용을 초과한다는 사실 뿐만 아니라, 해마다 반복되는 비용을 초래한다는 사실도 깨닫게 하였습니다. 이러한 비용은 데이터 웨어하우스 프로젝트에 부담을 주며 기업의 가치 창출을 방해합니다. Netezza와 Oracle을 모두 사용하는 한 금융 서비스 고객이 시행한 낭비 분석에서 볼 수 있는 것처럼 데이터 웨어하우징을 위해 Oracle을 사용하는 것은 노동 집중적입니다. Netezza 고객들은 수준도 낮고 기술적으로도 힘든 관리 업무는 불필요한 일이라고 증언합니다. 이러한 측면에서 Oracle 데이터베이스 작업을 위해서 관리자들은 그들 시간의 대부분을 기본 기술의 관리와 피딩에 사용해야 하는 반면 Netezza 고객들은 데이터를 활용하여 가치를 창조하는 데 보낼 수 있다는 것은 명확한 사실입니다.

Exadata의 새로운 스토리지 층은 관리자들에게 튜닝과 관리에 또 다른 복잡성을 더해줄 뿐입니다. Exadata는 매우 새로운 것이기 때문에 해당 기술을 이용하는 소수의 데이터 웨어하우스만이 프로덕션 중에 있으며 그 TCO에 대한 예측이 시기 상조적인 면이 있습니다. 그러나 고객들은 Exadata를 통해 일관성 있는 고성능을 달성하기 위해서는 교육, 데이터베이스 설계, 관리에 엄청난 비용이 발생할 것을 예상해야 합니다.

새로운 스토리지 층의 추가가 Oracle의 데이터베이스에 대한 디스크 작업량 병목현상을 조건적으로 줄여주는 하겠지만, Exadata의 엔지니어링은 아키텍처 전체의 활용보다는 초병렬 처리에 더 맞추어져 있습니다. 데이터 관리를 Exadata의 스토리지 층에 완전히 통합하지 못하는 Oracle은 그 MPP 그리드의 하드웨어 성능이 그리 좋지 못함을 말해줍니다. 이는 Exadata 구매 비용을 높이며 고객들은 소프트웨어를 통해 100% 활용될 수 없는 하드웨어에 대해 비용을 치루어야 합니다. 이러한 비용은 웨어하우스의 수명에 걸쳐 점점 늘어납니다. 고객들은 보다 효율적인 컴퓨터 시스템을 사용했다면 훨씬 더 높은 가치를 반환했을 데이터 센터의 활용되지 않는 공간에 대해서도 비용을 지불하게 됩니다.

시스템 평가		Netezza	Oracle
항목		TwinFin 12	Exadata v2 (SAS)
성능 및 아키텍처	MPP	<ul style="list-style-type: none"> <li>· True MPP</li> <li>· 데이터 웨어하우징 및 분석에 최적화됨</li> </ul>	<ul style="list-style-type: none"> <li>· 하이브리드 – 병렬 스토리지 노드 &amp; SMP 클러스터된 헤드 노드</li> <li>· 트랜잭션 처리(OLTP)에 최적화됨</li> </ul>
	하드웨어 아키텍처	<ul style="list-style-type: none"> <li>· 풀 프로세싱 S-Blades (CPU 코어 1개 + FPGA 코어 1개 / 디스크 드라이브 1개)</li> <li>· 사용자/어플리케이션 인터페이스에 주로 사용되는 SMP 호스트 노드</li> <li>· 독립 블레이드-블레이드 재배포</li> </ul>	<ul style="list-style-type: none"> <li>· 지능적인 스토리지 (CPU 코어 1개 / 디스크 드라이브 1.5개)</li> <li>· Oracle 11g RAC 실행 SMP 클러스터 노드</li> <li>· InfiniBand (Exadata 노드에서 SMP 클러스터)</li> <li>· 모든 데이터 재배포에서 헤드 노드 참여</li> </ul>
	데이터 스트리밍	<ul style="list-style-type: none"> <li>· S-Blade에 대한 FPGA 성능 지원 – 압축 풀기, 조건자 필터링, 행-수준 보안 적용</li> <li>· S-Blades에서 작업의 &gt;95% 수행</li> </ul>	<ul style="list-style-type: none"> <li>· 압축 풀기 및 조건자 필터링에 주로 사용되는 Exadata 노드.</li> <li>· 대부분의 DW &amp; 분석 작업은 SMP 헤드 노드에서 이루어짐.</li> </ul>
	In-DB 분석	<ul style="list-style-type: none"> <li>· 분석을 위해 완전히 활용되는 MPP 플랫폼</li> <li>· 사용자 지정 기능, 집계 및 표</li> <li>· 언어 지원: C/C++, Java, Python, R, Fortran</li> <li>· 패러다임 지원: SQL, Matrix, Grid, Hadoop</li> <li>· &gt;50개의 핵심 분석 내장 집합 (완전 병렬처리)</li> <li>· 오픈 소스: GNU Scientific &amp; CRAN 라이브러리</li> <li>· 통합 개발 환경: 마법사가 있는 Eclipse &amp; R GUI</li> </ul>	<ul style="list-style-type: none"> <li>· 헤드 노드 클러스터로만 제한된 분석 처리</li> <li>· 사용자 지정 기능 및 집계</li> <li>· 언어 지원: C/C++, Java</li> <li>· 패러다임 지원: SQL, Matrix (마이너)</li> <li>· 기본 분석 기능</li> </ul>
	확장성	<ul style="list-style-type: none"> <li>· 선형 성능 및 데이터 크기 확장 가능성</li> <li>· 전기능의 엔터프라이즈급 워크로드 관리 및 기타 기능</li> </ul>	<ul style="list-style-type: none"> <li>· 비선형 성능 및 데이터 크기 확장 - 헤드노드 클러스터의/로의 성능 및 i/o 병목현상</li> </ul>
단순성	어플라이언스 시스템 관리 및 통합	<ul style="list-style-type: none"> <li>· 튜닝, 인덱싱, 파티션 필요 없음</li> <li>· 최고의 가격 성능 제공을 위해 개발한 균형화된 시스템</li> </ul>	<ul style="list-style-type: none"> <li>· 높게 튜닝된 성능 의존도</li> <li>· 파티션 및 인덱스의 설계와 관리에 의존하는 성능</li> </ul>



분석 데이터베이스로서 Oracle RAC의 저성을 보상하기 위해 Exadata는 많은 하드웨어를 포함시킵니다. 머신은 구매 자체도 비싸지만 데이터 센터에서 운영하는 비용도 만만치 않습니다. Exadata 단일 랙은 전력을 켜고 생각하는 데 단일 TwinFin 랙보다 약 60% 전력을 더 소모합니다.

비용이 가치를 파괴한다는 사실보다 기초적인 질문은 Exadata가 고객의 가치 창출을 돕느냐 하는 것입니다. 1세대 데이터 웨어하우스는 기업에 최근의 정보를 지속적으로 알려주는 중요한 역할을 하지만 데이터는 백서의 앞부분에서 언급된 2세대 웨어하우스의 다른 능력과 고급 분석을 통해 보다 훌륭한 잠재력을 발휘합니다. 기존 스토리지와 결합한 Oracle RAC도 현재까지 이 역할에 대한 성공을 입증하지 못하고 있습니다. Exadata의 스토리지 층은 복잡한 조인, 고유한 집계 및 분석 기능을 처리할 수 없습니다. 매우 큰 데이터 집합을 고성능으로 심층 분석하기에는 어렵게 장착된 두 가지 기술이 빠른 네트워크와 연결되고 같은 랙에 하우징됐을 경우에 이러한 기능을 달성할 수 있을지 예측하기는 어렵습니다.

### IBM Netezza TwinFin의 가치

Netezza의 엔지니어들은 초병렬의 무공유 그리드 내에서 데이터 관리와 심층 분석을 통합합니다. 이러한 혁신의 한 가지 결과는 고객을 위한 단순성입니다. 이러한 단순성은 Oracle 제품과 같이 OLTP 지향 데이터베이스 제품에서 가능한 것보다 TCO 및 데이터 웨어하우스 작업 비용을 크게 줄여줍니다.

데이터 웨어하우스에 대한 수요는 간단한 SQL 처리 이상이 되었습니다. 데이터를 완전히 활용하기 위해서는 웨어하우스가 공간 분석, 예측 모델, 조사 그래프 및 기타 분석 어플리케이션을 실행할 수 있어야 합니다. 예를 들어, 보험에 가입한 재산의 대형 포트폴리오에 대해서 산불 및 홍수와 같은 재해로 생길 수 있는 위험을 평가하고자 하는 기업이 위험성 높은 지역의 보험 가입 자산을 식별하기 위해 Oracle Spatial과 Oracle 10g를 사용하여 어플리케이션을 구축했다고 가정해 보겠습니다. 400여 곳의 현장을 분석하기 위해서는 공간 조인과 계산 집중적인 공간 알고리즘 처리가 필요합니다. 산불의 위험을 평가하려면 10,000 다각형에 대해 430억 건의 계산을 해야 하며, 홍수의 경우 7,000 다각형에 대해 300억 건의 계산이 필요합니다. Oracle은 각 평가를 마치는데 1시간 이상이 걸렸습니다. 바람은 갑자기 방향과 강도를 변하게 하여, 홍수의 피해도가 높아질 수 있습니다. 자연 재해의 역동적인 특성은 위험 프로파일이 1시간 내에도 크게 변할 수 있다는 것을 의미합니다. 그러므로 이 회사는 사업에 필요하다고 판단될 때면 언제든 상황을 분석할 수 있도록 보다 우수한 민첩성이 필요합니다. Oracle 10g 또는 Exadata와는 달리, Netezza는 초병렬 처리 그리드에서 공간 조인과 계산 집중적인 공간 분석 알고리즘을 실행합니다. Netezza는 1.6초만에 산불 위험 평가를 마치고 4.4초만에 홍수 위험 평가를 마쳐서 업체에 데이터의 전체 가치를 얻어낼 수 있는 민첩성을 부여할 수 있습니다.

TwinFin은 토대부터 SQL 및 고급 분석 어플리케이션 모두의 초병렬 처리를 위해 설계되었습니다. Netezza 덕분에 기업들은 소유언어로부터 자유로워졌습니다. 파트너 및 고객들은 기존의 어플리케이션을 TwinFin으로 이동하거나 직접 선택한 C++, Java, Python, R, Fortran 등을 포함한 다양한 언어로 새로운 분석 어플리케이션을 개발하는 것을 선택할 수 있습니다. C 사용 기업들은 GNU Scientific Library로부터의 무료 소프트웨어로 이용 가능한 1,000여 개의 분석 기능을 활용할 수 있습니다.<sup>12</sup> R 사용 기업들 또한 CRAN(Comprehensive R Archive Network)<sup>13</sup>에서 공식적으로 이용 가능한 2,000여 패키지를 사용할 수가 있습니다. 또한 Netezza의 파트너 및 고객들은 그것들을 가지고 있는 높은 확장성을 지닌 통합 메커니즘으로서 MapReduce / Hadoop 등을 이용하여 일반인들이 사용하는 웹 어플리케이션이나 웹 로그에서 만들어진 엄청난 데이터 집합을 사전 처리하기 위해 이용한 후 TwinFin에 로드할 수도 있습니다.

<sup>12</sup> [www.gnu.org/software/gsl/](http://www.gnu.org/software/gsl/)

<sup>13</sup> <http://cran.r-project.org/>

# 6

## TwinFin을 통한 Exadata 성능 향상

섹션 2에서는 온라인 트랜잭션 처리와 데이터 하우징의 차이에 대해 논의했습니다. OLTP의 최적화, 데이터 캐싱 전략 및 성능 튜닝은 데이터 웨어하우징과 관련성이 적습니다. OLTP용으로 설계된 데이터베이스로, Exadata는 CIO에게 여러 개의 서로 다른 트랜잭션 시스템(자체 하드웨어와 Oracle 데이터베이스를 각자 포함)을 단일 관리 가능 환경으로 통합할 수 있는 기회를 제공합니다.

이러한 통합은 비용을 절감하고 소스 시스템에서 웨어하우스로의 피드에 사용되는 ETL(extract-transform-load) 단계의 성능을 개선할 수 있는 기회를 제공할 가능성이 높습니다. 1세대 웨어하우스에 더 많은 소스를 추가할 때 일부 기업들은 무의식중에 복잡한 ETL 웹을 만들게 됩니다. 많은 Oracle 고객들은 PL/SQL 및 기타 구문이 있는 Oracle의 데이터베이스를 비롯해 해당 데이터베이스 시스템 전용의 인터페이스를 이용해 자체 ETL을 구축합니다. PL/SQL의 열 기반 처리는 1세대 웨어하우스 초기의 상대적으로 낮은 데이터 볼륨 전형을 변환하기에는 충분한 속도였을 수 있으나 2세대 웨어하우스에 일반적인 큰 데이터 집합에는 집합 기반 처리에서 이용 가능한 고성능이 필요합니다. Oracle의 GoldenGate 소프트웨어나 ETL 전문 공급 업체의 기타 제품은 이러한 기능을 제공하고 있습니다. 이 백서의 주제를 벗어나기는 하나 Netezza Migrator™ 및 Netezza Data Virtualizer™ 역시 프로젝트 재설계 및 쿼리 성능 가속화를 통해 ETL의 성공에 이바지할 수 있습니다.

여러 개의 Oracle 어플리케이션을 실행하는 Exadata가 분석을 위해 TwinFin에 데이터를 넘깁니다



IBM Netezza는 데이터 웨어하우징을 위해 Oracle에 대한 근본적인 대안으로 출시되었습니다. Oracle에서 Netezza로 데이터 웨어하우스 및 마트를 옮기는 것을 통해 위험이 아닌 새로운 기회를 창출할 수 있습니다. 대부분의 Netezza 고객들은 이 과정을 이미 거쳤으며 그들 중 다수가 성공적인 이전에 대한 강력한 이력을 가지고 있는 시스템 통합 업체들과의 제휴를 통해 이러한 이전 작업을 시행했습니다.

Exadata는 Oracle OLTP 플랫폼의 혁신입니다. Oracle의 데이터베이스 관리 시스템은 데이터 웨어하우스에 비해 데이터 볼륨이 상대적으로 낮은 OLTP를 위해 설계되었습니다. OLTP 시스템의 데이터베이스 활동은 프로덕션에 배치되기 전에 평가될 수 있습니다. 관리자들은 각 트랜잭션의 데이터 검색을 설계, 시험 및 최적화할 수 있는 시간을 가지게 됩니다. 데이터 웨어하우스는 그들의 데이터에 대해 비즈니스에 필요한 어떠한 쿼리 요청이 온다 해도 즉각적으로 처리할 수 있어야 합니다. 관리자 중재를 필요로 하는 기술은 이러한 작업에 맞지 않습니다. 트랜잭션 처리 이외의 이런 특성을 가진 기술에 대한 관리 역할을 더 하는 것은 데이터 웨어하우스를 관리하고 작동해야 하는 사람들과 프로세스에 엄청난 스트레스가 됩니다.

**"Netezza는 우리가 가야 할 비즈니스 방향과 정확히 일치하는 장기적 관점의 제품 로드맵을 가진 최초의 데이터베이스 제품입니다. 우리는 이것을 주문형 데이터베이스라고 합니다."**

**- 뉴욕 증권거래소, Chief Data Officer, Steve Hirsch**

Oracle은 Exadata가 아키텍처 측면에서 Netezza와 유사하지만 TwinFin이 모든 데이터 유형 또는 SQL 기준을 지원하지 않으며 데이터 마이닝이나 높은 동시성을 지원하지 않으므로 Netezza보다 더 좋다고 합니다. Netezza의 고객들은 이에 반대합니다. NYSE Euronext (뉴욕 증권거래소)의 Chief Data Officer, Steve Hirsch는 "Netezza는 우리가 가야 할 비즈니스 방향과 정확히 일치하는 장기적 관점의 제품 로드맵을 가진 최초의 데이터베이스 제품입니다. 우리는 이것을 주문형 데이터베이스라고 합니다"<sup>14</sup>

<sup>14</sup> [www.netezza.com/customers/nyse-uronext-video.aspx](http://www.netezza.com/customers/nyse-uronext-video.aspx)

다양한 워크로드 특성을 고려할 때 지속적인 튜닝과 최적화 작업을 해야 하기 때문에 OLTP와 데이터 웨어하우스 시스템을 실행하려는 고객은 거의 없습니다. 기술자들은 OLTP 및 데이터 웨어하우스 모두에 대해 저성을 수락하느냐, 다양한 워크로드에 대한 충돌 수요를 만족시키기 위해 끊임 없이 데이터베이스를 재구성하는 쓸데없는 시도를 하느냐를 결정해야 하는 어려운 상황에 놓이게 되었습니다. 기업들은 각각의 워크로드 필요에 맞춘 구성을 한 OLTP와 웨어하우스를 다른 플랫폼에서 계속해서 실행하려고 할 것입니다. OLTP를 위해 Oracle Exadata 사용을 계획하고 있는 기업들은<sup>15</sup> 데이터 웨어하우스를 위한 Netezza TwinFin과 연결한다면 최고의 결과를 얻을 수 있습니다.

정말로 중요한 유일한 데이터 웨어하우스는 여러분의 데이터 웨어하우스입니다. 즉, 여러분의 데이터 센터에서 여러분의 데이터를 실행하는 여러분의 어플리케이션인 것입니다. PoC(on-site proof-of-concept)는 IT 부서가 기술을 심층 연구할 수 있는 기회를 주며 TwinFin을 사용해 비즈니스 부서가 데이터에서 보다 훌륭한 가치를 추출할 수 있도록 돕는 방법을 배우게 합니다. 이 기회를 가장 잘 활용하기 위해서 PoC는 다른 프로젝트에 적용되는 것과 같은 규율을 가지고 관리되어야 합니다. IBM Netezza는 웨어하우스가 어떻게 수행하는지 알고자 하는 기업들을 위해 비용, 위험 부담 없이 Netezza를 시험 사용할 수 있는 기회를 제공합니다.

## IBM Netezza TwinFin: 사용하는 것이 바로 즐기는 것이다

Netezza는 여러분이 Netezza의 솔루션을 마음에 들어 하실 것을 확신합니다.

<sup>15</sup> 몇 개의 큰 데이터베이스 관리 업무 실행이 아닌 기업의 많은 Oracle 데이터베이스 통합을 위한 플랫폼으로서 Exadata 같은 기술의 역할에 대한 Curt Monash의 논의를 보려면 그의 DBMS2 블로그 참조 [www.dbms2.com/2010/01/22/oracle-database-hardware-strategy/](http://www.dbms2.com/2010/01/22/oracle-database-hardware-strategy/).

## 공유

이 eBook을 친구들이나 동료들과 공유하세요.

- 보내기
- 링크 게시: [www.netezza.co.kr/eBooks/Exadata-TwinFin-Compared.pdf](http://www.netezza.co.kr/eBooks/Exadata-TwinFin-Compared.pdf)

## 피드백

이 eBook의 아이디어와 주장에 대해 어떻게 생각하셨습니까? 마음에 드셨는지, 그렇지 않은지 더 논하고 싶으신지 알려주십시오.

- Netezza 커뮤니티와의 채팅: [www.enzeecomunity.com/groups/twinfin-talk](http://www.enzeecomunity.com/groups/twinfin-talk)

## 연락처

당사의 블로그 방문: [www.enzeecomunity.com/blogs/nzblog](http://www.enzeecomunity.com/blogs/nzblog)

Netezza 웹 사이트 방문: [www.netezza.co.kr](http://www.netezza.co.kr)

Netezza 커뮤니티 웹 사이트 방문: [www.enzeecommmunity.com](http://www.enzeecommmunity.com)

여러분의 코멘트를 보내주세요: [www.netezza.com/company/contact\\_form.aspx](http://www.netezza.com/company/contact_form.aspx)

## 저자 소개



### Phil Francisco(Netezza의 제품 관리 및 제품 마케팅 부사장)

Phil Francisco는 기술 개발 및 글로벌 기술 마케팅 분야에서 20여 년의 경력을 갖고 있습니다. Netezza의 제품 관리 및 제품 마케팅 부사장으로서, Phil Francisco는 새로운 사업 및 제품 전략을 수립하고 제품 포트폴리오를 감독하며 제품 마케팅 프로그램을 주도합니다. Netezza에 입사하기 전에 Francisco는 코어 텔레커뮤니케이션 네트워크 제공업체를 위한 40 Gb/s 광전송 시스템의 핵심 개발업체인 PhotonEx에서 마케팅 부사장으로 활약했습니다. Francisco는 PhotonEx 근무 전에 Lucent Technologies의 Optical Networking Group 제품 마케팅 부사장직에 있었으며, 광 네트워크 솔루션 계획 및 실행을 위해 세계 최대의 텔레커뮤니케이션 업체들과 협력한 바 있습니다. Francisco씨는 고급 광네트워크 아키텍처 특허를 보유하고 있습니다. 그는 University of Pennsylvania의 Moore School of Electrical Engineering에서 2등 성적으로 전기공학 및 전산학 학사 학위를 취득했습니다. Stanford University에서 전기공학 석사 학위를 받고, Duke University의 Fuqua School of Business에서 고급 경영 프로그램을 이수했습니다. Phil의 블로그 읽기: [www.enzeecomunity.com/blogs/nzblog](http://www.enzeecomunity.com/blogs/nzblog).

## Netezza 소개

IBM Netezza는 데이터 웨어하우스 어플라이언스의 발명으로, 데이터 처리에 어려움을 겪으며 그에 대한 분석 및 이해를 위한 처리 속도와 능력을 찾기 위해 노력 중인 업체들을 위해 분석업무를 혁신화하고 간소화했습니다.

현재 IBM Netezza는 스토리지, 처리, 데이터베이스 및 분석을 다른 접근법의 1/3 비용에 10-100배의 성능을 제공하는 단일 시스템을 보유한 데이터 웨어하우스 어플라이언스의 선도기업입니다. 당사는 가격/성능의 기준을 다시 한 번 재정립하고 있습니다. Netezza를 경제적인 가격과 효율성을 가진 페라리 수준의 고급차로 생각해 보십시오.

Nationwide, Neiman Marcus, Orange UK, The Sherwin-Williams Company, Virgin Media 등과 전 세계의 사무소들을 포함해 수많은 고객들을 보유한 Netezza는 비용 증가와 데이터 웨어하우징 및 분석의 복잡성에 대한 입증된 솔루션입니다. 당사의 다양한 글로벌 파트너와, 전 세계의 폭넓은 시스템 통합업체, 리셀러 및 개발자 파트너 목록을 통해 당사의 고객들은 그들이 가지고 있는 어떠한 기존의 인프라 내에서도 Netezza의 편안한 사용성에 대해 안심할 수 있습니다. Netezza는 여러분의 든든한 파트너로서 저비용 고효율의 혁신적인 분석 솔루션을 고객 여러분께 제공할 것입니다.

IBM Netezza는 세계 최초의 데이터 웨어하우스 어플라이언스를 도입하여 기존의 틀을 깨는 데 도전했습니다. 또한 새로운 시장 분야를 개척하여 전체 업계에 대한 지칭을 수립하였습니다. 이제 Netezza 고객의 주도하에 저희는 글로벌 연합, 데이터 통합, 레거시 시스템 통합, 비즈니스 연속성, 고급 분석 및 준수를 비롯한 보다 더 크고 복잡한 전사적인 난제들을 해결하기 위해 솔루션을 개발 중입니다.

IBM Netezza의 기술력은 기업이 확보한 방대한 양의 데이터를 매우 빠른 속도로 처리할 수 있도록 지원하여, 이 기술력을 활용한 기업은 디지털 미디어, 에너지, 금융, 정부기관, 의료 및 생명과학, 유통 및 텔레콤 등 오늘날의 정보 중심 산업 환경에서 경쟁적 우위를 선점하고 운영 이점을 누릴 수 있습니다.



한국아이비엠주식회사 (135-270) 서울시 강남구 도곡동 467-12 군인공제회관빌딩  
TEL : 02-3781-7800 한국IBM 마케팅 총괄본부

[www.netezza.co.kr](http://www.netezza.co.kr) [www.ibm.com/kr](http://www.ibm.com/kr)

IBM Netezza는 확장된 기업 환경에서 고성능 분석을 크게 간소화할 수 있는 데이터 웨어하우스 및 분석 어플라이언스 부문의 글로벌 선두 기업입니다. IBM Netezza의 기술력은 기업이 확보한 방대한 양의 데이터를 매우 빠른 속도로 처리할 수 있도록 지원하여, 이 기술력을 활용한 기업은 디지털 미디어, 에너지, 금융, 정부기관, 의료 및 생명과학, 유통 및 텔레콤 등 오늘날의 정보 중심 산업 환경에서 경쟁적 우위를 선점하고 운영 이점을 누릴 수 있습니다. IBM Netezza에 대한 보다 자세한 정보는 [www.netezza.co.kr](http://www.netezza.co.kr) 또는 [www.ibm.com/kr](http://www.ibm.com/kr)을 방문하십시오.