

PureData System으로 시간 단축을 통해 가치 구현 가속화

워크로드에 맞게 최적화된 시스템으로 구축을
간소화하고 더 빠르게 결과 달성

통합된 전문성이 내재화되어 있는 PureData System

- 구축의 간소화
 - ▶ IT 라이프사이클의 *모든* 부분을 더 쉽게 수행
- 내재화된 전문성
 - ▶ 전문가가 수행하는 작업 캡처 및 자동화
- 설계 단계부터 고려된 통합
 - ▶ 하드웨어 및 소프트웨어 성능 튜닝과 긴밀한 통합을 통해 워크로드 최적화

다른 특성을 가지고 있는 다양한 워크로드

■ 비정형 분석

- ▶ 소수의 사용자 – 매우 숙련된 프로그래머
- ▶ 여러 시스템에 분산되어 있는 비정형 데이터
- ▶ 대규모 병렬화와 빠른 처리를 위해 데이터 레이어에서 프로세스 처리

■ 심층 분석

- ▶ 적은 수의 사용자와 많은 복합 보고서가 있는 데이터 마트
- ▶ 효율적인 보고 및 집계를 위해 비정규화된 데이터
- ▶ 쿼리 시간 단축을 중시

■ 운영 분석

- ▶ 많은 사용자와 다양한 복잡도의 보고서가 있는 데이터 웨어하우스
- ▶ 정규화된 데이터와 비정규화된 데이터의 혼합
- ▶ 처리량을 중시

■ 트랜잭션 처리

- ▶ 수많은 사용자의 동시 액세스가 필요한 운영 데이터
- ▶ 동시성을 최대화하기 위해 정규화된 데이터
- ▶ 비즈니스에 중요한 응답 시간

다른 특성을 가지고 있는 다양한 워크로드

■ 비정형 분석

- ▶ 소수의 사용자 - 매우 숙련된 프로그래머
- ▶ 여러 시스템에 분산되어 있는 비정형 데이터
- ▶ 대규모 병렬화와 빠른 처리를 위해 데이터 레이어에서 프로세스 처리

■ 심층 분석

- ▶ 적은 수의 사용자와 많은 복합 보고서가 있는 데이터 마트
- ▶ 효율적인 보고 및 집계를 위해 비정규화된 데이터
- ▶ 쿼리 시간 단축을 중시

■ 운영 분석

- ▶ 많은 사용자와 다양한 복잡도의 보고서가 있는 데이터 웨어하우스
- ▶ 정규화된 데이터와 비정규화된 데이터의 혼합
- ▶ 처리량을 중시

■ 트랜잭션 처리

- ▶ 수많은 사용자의 동시 액세스가 필요한 운영 데이터
- ▶ 동시성을 최대화하기 위해 정규화된 데이터
- ▶ 비즈니스에 중요한 응답 시간

PureData System For Hadoop으로 Hadoop 인프라 간소화

커스텀 빅데이터에서

구조화된 간소화로



설계 목적...

- Hadoop 클러스터 빌드, 구축 및 관리 간소화
- Hadoop 및 비정형 데이터에 대해 시간 단축을 통한 가치 구현 가속화
- 전체 분석 에코시스템의 역량 극대화
- 엔터프라이즈 보안 및 플랫폼 관리 제공



*IBM 내부 테스트 및 고객 피드백 기준. "커스텀 빌드 클러스터"는 전문적으로 사전 빌드, 사전 테스트 및 최적화되지 않은 클러스터입니다. 개별 결과는 달라질 수 있습니다.

Hadoop 구현의 복잡함을 제거한 IBM PureData System For Hadoop

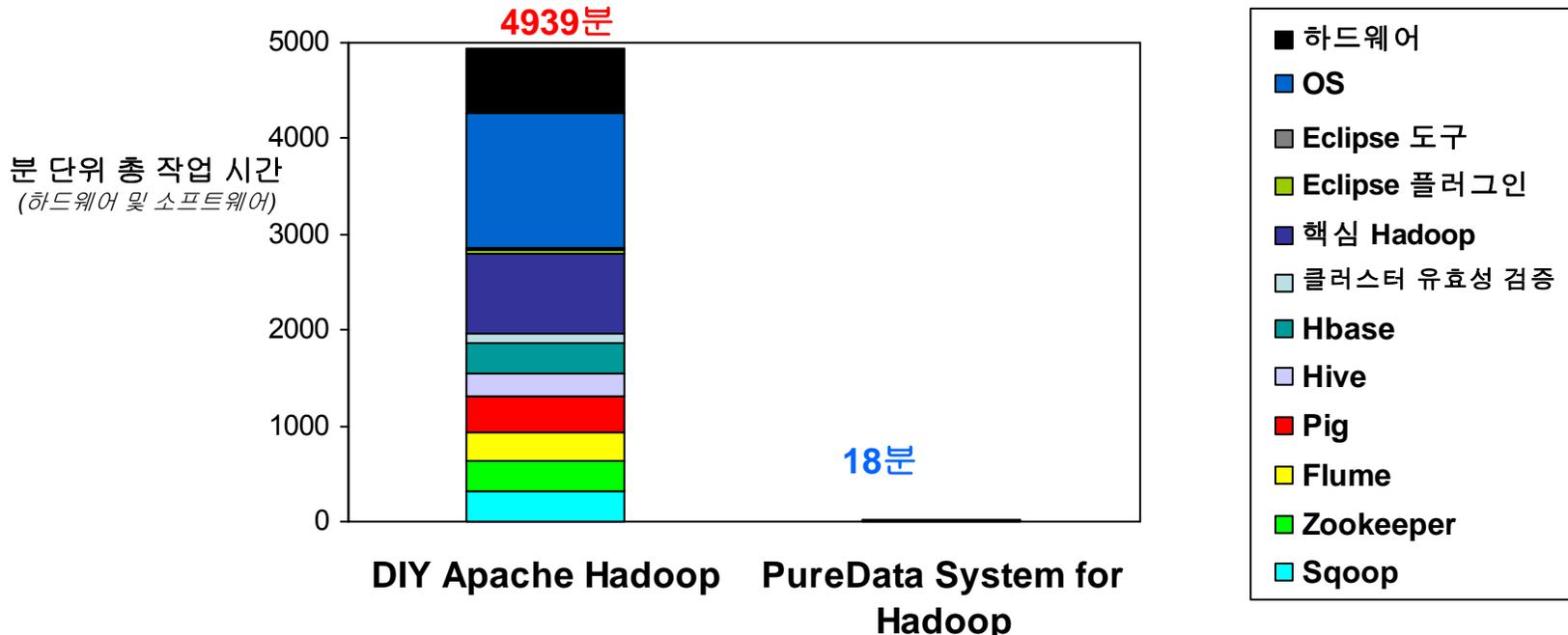
DIY Apache Hadoop의 문제점

- 올바른 버전의 Hadoop 하위 프로젝트를 모두 찾아야 함
- 모든 요소가 제대로 구성되었는지 확인해야 함
- 안정감 있는 환경을 제공하지 않음
- Hadoop 릴리스마다 핵심 Hadoop 및 모든 하위 프로젝트를 다시 설치해야 함
- 포럼에서만 지원 제공



Hadoop을 업그레이드할 때마다 문제가 계속 반복됨

82시간 vs 18분



다양한 워크로드의 다양한 특성

■ 비정형 분석

- ▶ 소수의 사용자 – 매우 숙련된 프로그래머
- ▶ 여러 시스템에 분산되어 있는 비정형 데이터
- ▶ 대규모 병렬화와 빠른 처리를 위해 데이터 레이어에서 프로세스 처리

■ 심층 분석

- ▶ 적은 수의 사용자와 많은 복합 보고서가 있는 데이터 마트
- ▶ 효율적인 보고 및 집계를 위해 비정규화된 데이터
- ▶ 쿼리 시간 단축을 중시

■ 운영 분석

- ▶ 많은 사용자와 다양한 복잡도의 보고서가 있는 데이터 웨어하우스
- ▶ 정규화된 데이터와 비정규화된 데이터의 혼합
- ▶ 처리량을 중시

■ 트랜잭션 처리

- ▶ 수많은 사용자의 동시 액세스가 필요한 운영 데이터
- ▶ 동시성을 최대화하기 위해 정규화된 데이터
- ▶ 비즈니스에 중요한 응답 시간

IBM PureData System For Analytics

*내재화된 전문성과 효율적인 전용 어플라이언스를 통해
심층 분석 간소화*

- 스마트
 - ▶ 혁신적인 스트리밍 기술
- 간소화
 - ▶ 빠른 설치
 - ▶ DB 관리자 작업 불필요
 - ▶ 매우 쉬운 조작
- 속도
 - ▶ 용도에 맞게 최적화



혁신적인 기술을 통해 획기적인 간소화를 제공하는 PureData System For Analytics

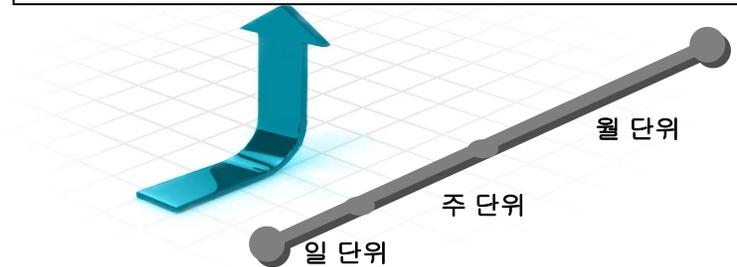
■ 소프트웨어 설치 없음

■ 인덱스 및 성능 조정 없음

■ 스토리지 관리 없음

- ▶ DB 공간/테이블스페이스 크기 조정 및 구성 없음
- ▶ Redo/물리/논리 로그 크기 조정 및 구성 없음
- ▶ 테이블 페이지/블록 크기 조정 및 구성 없음
- ▶ 테이블 Extent 크기 조정 및 구성 없음
- ▶ 임시 공간 할당 및 모니터링 없음
- ▶ DB 공간 RAID 레벨 결정 없음
- ▶ 파일의 논리 볼륨 작성 없음
- ▶ OS 커널 권장 통합 없음
- ▶ OS 권장 패치 레벨 유지보수 없음
- ▶ 호스트/네트워크/스토리지 구성 JAD 세션 없음

관리 시간은 줄이고
혁신을 위한 시간은 더 많이 확보



정상 가동:

PureData for Analytics

2.5일

(학습 기간 1일 포함)

VS

Exadata

2.5주

(숙련된 관리자 필요)

BI Day 워크로드 - 심층 분석 테스트를 통해 동시 사용자 보고서 성능 측정

4명의 사용자가
복합 보고서 작성



사용자 2명

사용자 2명



20명의 사용자가
중간 보고서 작성



사용자 6명

사용자 8명

사용자 6명



4개 연결

20개 연결

각 보고서에서
하나 이상의
쿼리 실행



- 동시 사용자 24명
- 복합 보고서와
중간 보고서의 혼합
- 처리량을 중시

참고: Forrester Research, Business Intelligence 분석
사용자 프로파일링, 2004에 기반을 둔 복합, 중간 및 단순
워크로드 분산

경쟁제품을 능가하는 Netezza의 뛰어난 분석 쿼리 실행 성능

동일한 쿼리, 동일한 클라이언트, 동일한 데이터...
다른 결과

The image displays two side-by-side screenshots of BI report dashboards, comparing the performance of Netezza and a competitor. Both dashboards show the same data table, but the Netezza version is significantly faster.

Netezza - BI Day Demo

Report 1 - Netezza

Report 1 Timer: 2 : 10
Min Sec
Actual Report Time

Start Clear Results

Netezza Results

2	Mountaineering Equip	1099
1	Camping Equipment	1099
1	Camping Equipment	1099
2	Mountaineering Equip	1099
2	Mountaineering Equip	1099
2	Mountaineering Equip	1099
1	Camping Equipment	4199
1	Camping Equipment	5199
1	Camping Equipment	4199

Number of rows returned = 5895
Number of rows shown = 10

Competitor - BI Day Demo

Report 1 - Competitor

Report 1 Timer: 18 : 24
Min Sec
Actual Report Time

Start Clear Results

Competitor Results

Product_linekey	Product_line0	Retailer_territor
3	Personal Accessories	5199
2	Mountaineering Equip	6199
5	Golf Equipment	6199
2	Mountaineering Equip	1099
4	Outdoor Protection	5199
5	Golf Equipment	1099
1	Camping Equipment	7199
4	Outdoor Protection	4199
3	Personal Accessories	6199
5	Golf Equipment	5199



netezza vs competitor

Netezza vs. Competitor

YouTube에서 자세한 내용 확인하기:

http://www.youtube.com/watch?v=T3O6yJ_hdUU

다양한 워크로드의 다양한 특성

■ 비정형 분석

- ▶ 소수의 사용자 – 매우 숙련된 프로그래머
- ▶ 여러 시스템에 분산되어 있는 비정형 데이터
- ▶ 대규모 병렬화와 빠른 처리를 위해 데이터 레이어에서 프로세스 처리

■ 심층 분석

- ▶ 적은 수의 사용자와 많은 복합 보고서가 있는 데이터 마트
- ▶ 효율적인 보고 및 집계를 위해 비정규화된 데이터
- ▶ 쿼리 시간 단축을 중시

■ 운영 분석

- ▶ 많은 사용자와 다양한 복잡도의 보고서가 있는 데이터 웨어하우스
- ▶ 정규화된 데이터와 비정규화된 데이터의 혼합
- ▶ 처리량을 중시

■ 트랜잭션 처리

- ▶ 수많은 사용자의 동시 액세스가 필요한 운영 데이터
- ▶ 동시성을 최대화하기 위해 정규화된 데이터
- ▶ 비즈니스에 중요한 응답 시간

분석 쿼리와 운영 트랜잭션이 혼합된 운영 분석

■ 운영 분석 워크로드의 특성

▶ 상세 분석 쿼리와 짧은 운영 트랜잭션의 혼합

- 상세 분석 쿼리에서는 순차 테이블 스캔을 주로 사용
- 운영 트랜잭션에서는 작은 무작위 테이블 액세스를 더 많이 사용

▶ 지속적인 업데이트

- 다수의 수집 스트림

▶ 동시성

- 수천 명의 사용자
- 정규화된 스키마 및 비정규화된 스키마 모두 통합 가능

■ 다양한 쿼리 유형 간에 밸런스를 유지하는 것이 매우 중요

DB2 z/OS와 Netezza 기술의 결합

Netezza 기술을 기반으로 한 IBM DB2 Analytics Accelerator

- 현재 전 세계 운영 데이터의 대부분은 z/OS에 있음
 - ▶ 심층 분석을 위해 DB2 Analytics Accelerator에 추가
- 애플리케이션에 투명함
 - ▶ DB2 for z/OS와 통합됨
 - ▶ DB2 쿼리 옵티마이저 및 z/OS 워크로드 관리자 통합
 - ▶ DB2 for z/OS에 있는 데이터에서 액셀러레이터 로드
- 응답 시간을 크게 단축
- 비즈니스 분석 비용 절감 촉진
 - ▶ 쿼리를 통해 하드웨어 사용을 자동으로 최적화
 - ▶ 액셀러레이터 내에 테이블 또는 파티션의 사본을 유지하고 증분 업데이트 수행(고성능 스토리지 서버)



BI Day 워크로드 - 전형적인 분석 보고 워크로드

BI Day 실행 테스트: 총 보고서 수 = 161,166

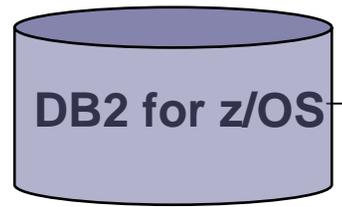


4개 연결

20개 연결

56개 연결

- DB2 for z/OS에 80개의 동시 연결 설정
- 모든 단기 실행 SQL 쿼리는 DB2 for z/OS에서 처리됨
- 장기 실행 SQL 쿼리는 액셀러레이터로 자동으로 전송되어 처리됨



각 보고서에서
하나 이상의
SQL 쿼리 실행

분산 플랫폼을 위한 IBM PureData System For Operational Analytics

가치 실현 시간을 단축하는 기능 제공

구축의 간소화

- 단일 통합 콘솔

- IBM PureApplication System과 동일한 사용자 인터페이스를 사용하여 시스템 간 일관성 제공

- 내장된 분석 기능

내재화된 전문성

- 병렬 분산에 맞게 최적화된 분석

- ▶ 내장된 임시 분석 포함

- 어댑티브 압축 기술

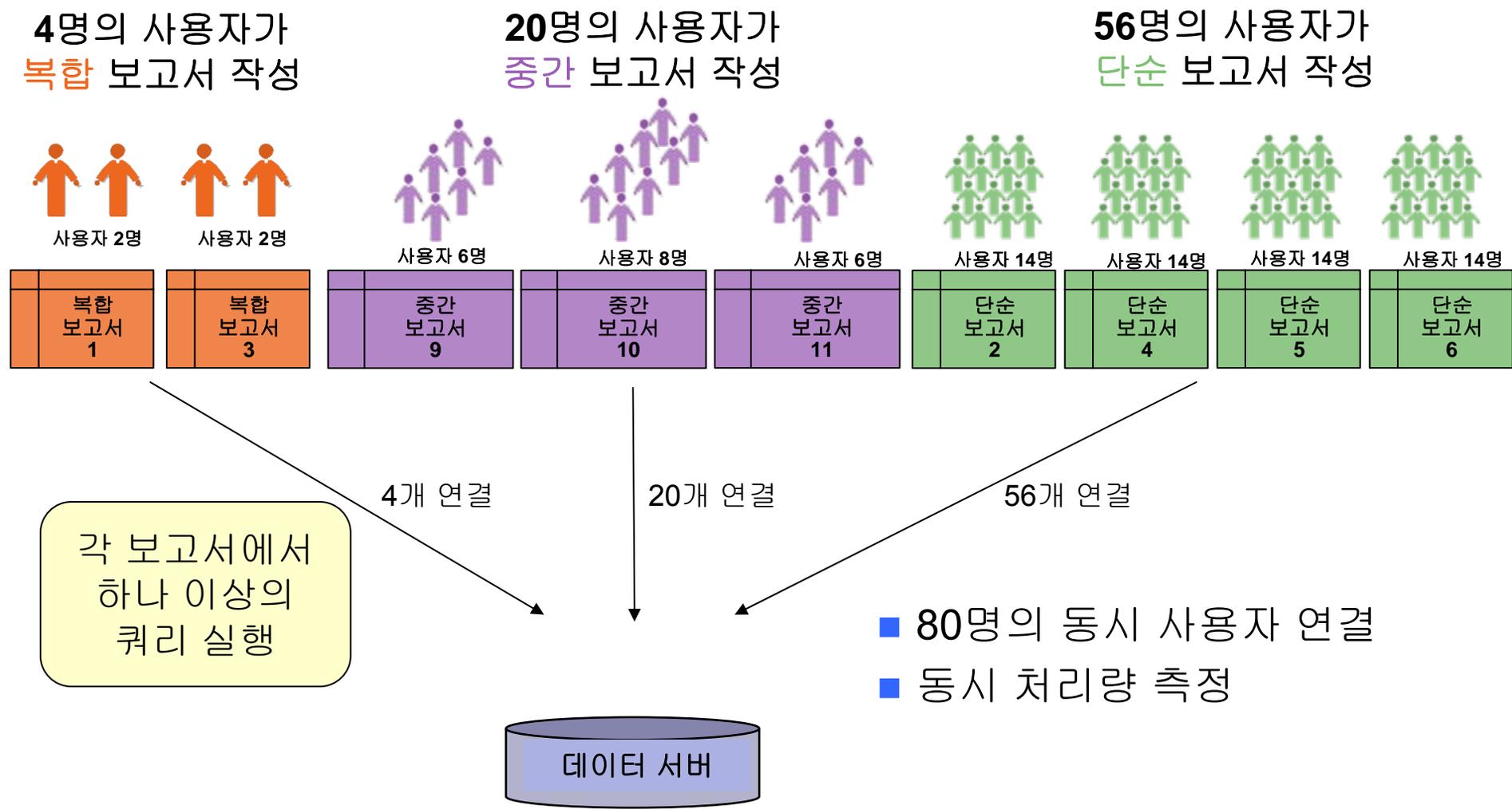
설계 단계부터 고려된 통합

- 서버, 스토리지, 데이터베이스를 하나의 사용하기 쉬운 패키지로 구성

- 광범위한 엔터프라이즈 모니터링 도구 및 프로세스와 손쉽게 통합



BI Day 워크로드 - 고수준의 동시 실행 워크로드를 측정



참고: Forrester Research, Business Intelligence 분석 사용자 프로파일링, 2004에 기반을 둔 복합, 중간 및 단순 워크로드 분산

다양한 워크로드의 다양한 특성

■ 비정형 분석

- ▶ 소수의 사용자 – 매우 숙련된 프로그래머
- ▶ 여러 시스템에 분산되어 있는 비정형 데이터
- ▶ 대규모 병렬화와 빠른 처리를 위해 데이터 레이어에서 프로세스 처리

■ 심층 분석

- ▶ 적은 수의 사용자와 많은 복합 보고서가 있는 데이터 마트
- ▶ 효율적인 보고 및 집계를 위해 비정규화된 데이터
- ▶ 쿼리 시간 단축을 중시

■ 운영 분석

- ▶ 많은 사용자와 다양한 복잡도의 보고서가 있는 데이터 웨어하우스
- ▶ 정규화된 데이터와 비정규화된 데이터의 혼합
- ▶ 처리량을 중시

■ 트랜잭션 처리

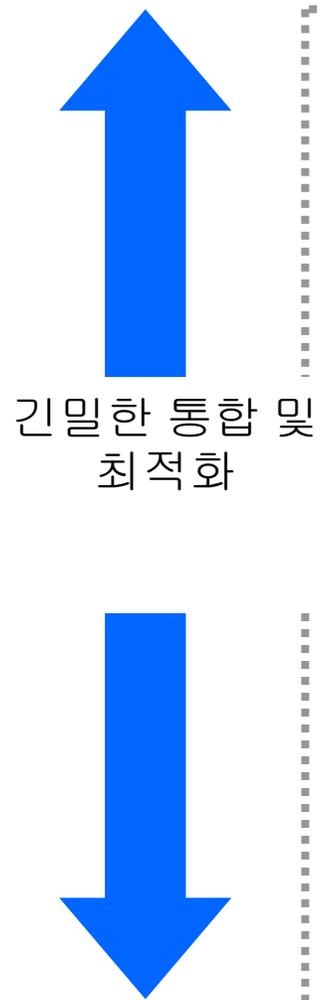
- ▶ 수많은 사용자의 동시 액세스가 필요한 운영 데이터
- ▶ 동시성을 최대화하기 위해 정규화된 데이터
- ▶ 비즈니스에 중요한 응답 시간

IBM PureData System For Transactions

고가용성, 속도, 신뢰성, 확장성 및 간소화를 위해 최적화됨



IBM PureData System



긴밀한 통합 및 최적화



내재화된 운영 전문성

+

데이터베이스 최적화 및 클러스터링

+

IBM Flex System 구성 요소

패턴 기반 배치, 작업 시간을 절약하는 자체 조정, 밸런싱 및 모니터링

- 클러스터링과 압축 소프트웨어를 포함하여 CPU와 스토리지 하드웨어를 균형있게 설계하여 최대의 성능 발휘

- DB2 pureScale을 통해 확장성과 가용성을 최대화하도록 설계됨

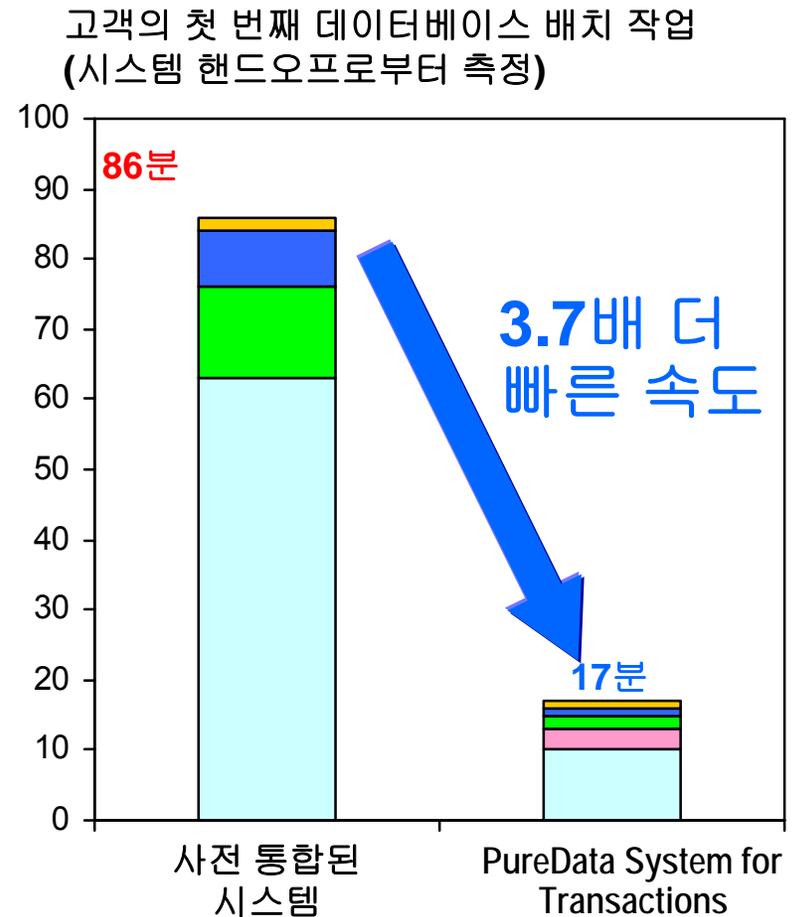
컴퓨팅 + 스토리지 + 네트워크 + 하드웨어 관리 + HA를 위한 가용성

PureData System For Transactions로 더욱 빠르게 데이터베이스 구축

활동	사전 통합된 시스템 X3	PureData System for Transactions
고객 설정/구성	63	10
클러스터 구축	0	3
데이터베이스 사용자 정의	13	2
데이터베이스 배치	8	1
스토리지 추가	2	1
합계	86	17

구성된 시스템을 공급업체가 전송하면
고객은 첫 번째 데이터베이스 배치 작업을 수행해야 함

사전 통합된 시스템에서는 **PureData System for Transactions**에 비견되는 기능을 제공할 별도의 관리자 시스템을 설치하는 데 상당한 시간이 필요함



복잡성과 비용을 감소시키는 IBM PureData System

트랜잭션, 분석, Hadoop 등
모든 빅데이터 워크로드에 사용

- 더 빠른 솔루션 구현
- 구축의 간소화
- 더 빠른 성능, 더 저렴한 비용
- 어떤 예산 규모에도 적합한 광범위한 시스템



Service Oriented Finance CTO

보고서: Contrasting IBM's Approach to Workload Optimization to Oracle's Exadata Single Server Approach (2012년 3월, 영문) http://www.clabbyanalytics.com/Free_Reports_Critiques.html
IBM's New Moves in Big Data: Acceleration, Optimization, and an Open Source Alternative(2013년 4월, 영문) <http://www.clabbyanalytics.com/uploads/BigDataAlmaden.pdf>