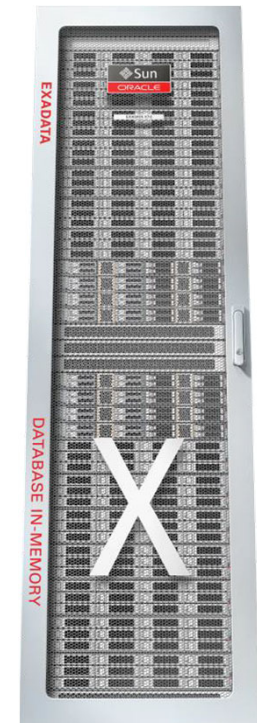




Exadata 기술 소개

- Exadata Hardware 기초




경북산업직업전문학교 27-JAN-2022

Exadata 개요



최고의 성능 / 무한의 확장성 / 준비된 엔터프라이즈 사용 환경

Runs Oracle Data Warehouses 10x Faster



Announcing
The World's Fastest Database Machine

- Hardware by HP
- Software by Oracle

ORACLE®

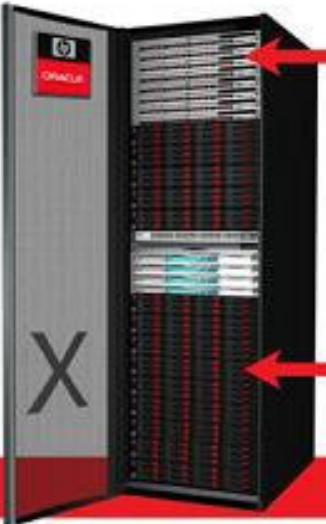
oracle.com/exadata
or call 1.800.ORACLE.1

The performance improvement based on customer reports comparing average performance of Oracle data warehouses on existing customer systems versus Oracle Database Machines. Actual results may vary.

Copyright © 2009, Oracle. All rights reserved. Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

HP Super Architecture 2100

Announcing The World's Fastest Database Machine



Oracle Database Server Grid

10x Faster Query Results

Exadata Storage Grid Fast Query Processing

ORACLE®

Hardware by HP Software by Oracle

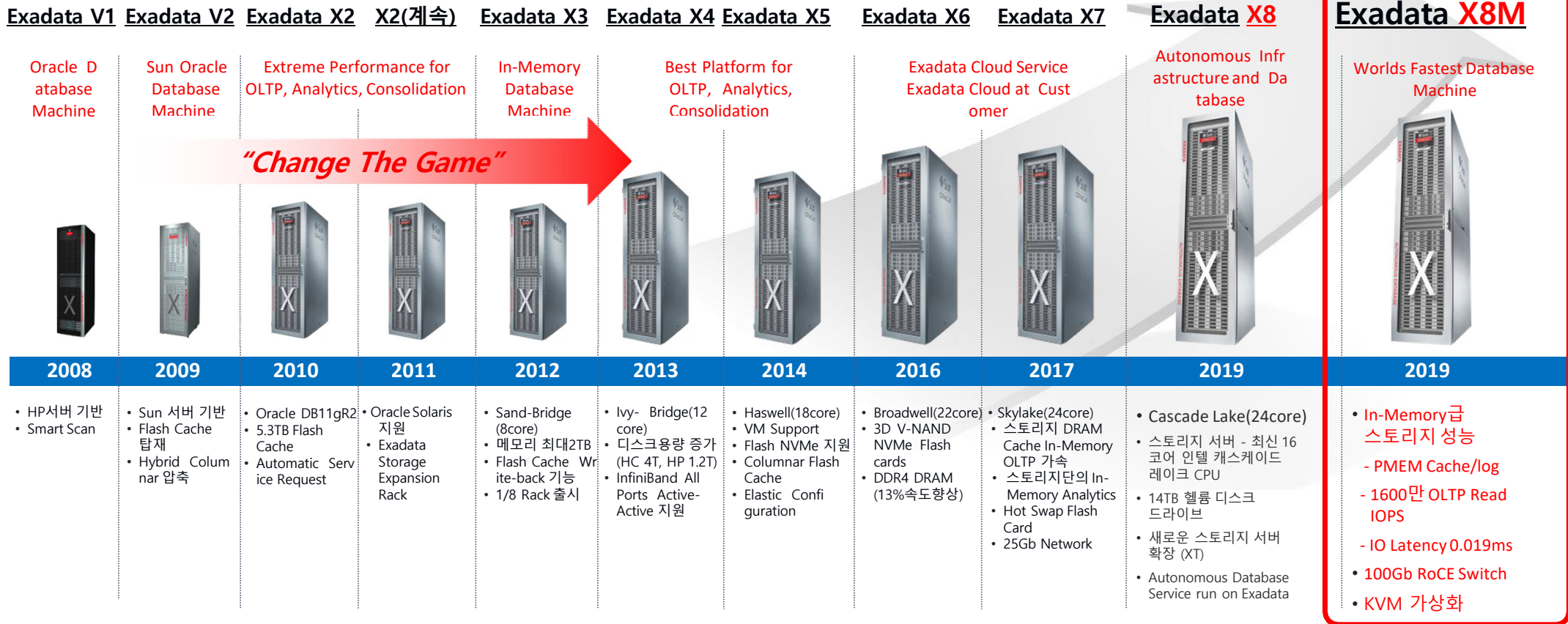
The performance improvement based on customer reports comparing average performance of Oracle data warehouses on existing customer systems versus Oracle Database Machines. Actual results may vary.

Copyright © 2009, Oracle. All rights reserved. Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

HP Super Architecture 2100

Exadata – The Game Changer

혁신에서 출발, 진화와 발전을 거듭한 혁신적인 기능 탑재



Exadata의 Loadmap

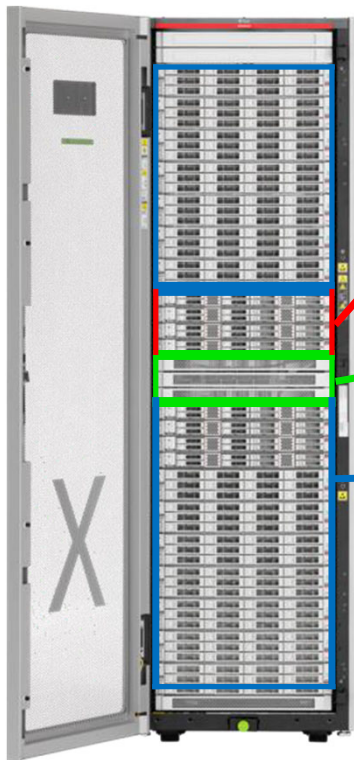
Exadata는 Intel의 최신 Xeon Chip을 사용하기 때문에 Intel의 CPU 출시와 Version을 같이함

V1/ V2	X2	X3	X4	X5	X6	X7	X8/X8M
네할렘 4 core	웨스트미어 6 core	샌드브릿지 8 core	아이비브릿지 12 core	하스웰 18 core	브로드웰 22 core	스카이레이크 24 core	케스케이드레이크 24 core



최근 PC는 커피 레이크, 코멧레이크 ?

Exadata Database Machine X8-2M 하드웨어 구성



- 확장 가능한 데이터베이스 서버



2소켓 Xeon 8260 프로세스 (2.4GHz)
서버 당 48 코어
384 GB ~ 1.5 TB DRAM

- 고속 내부 네트워크

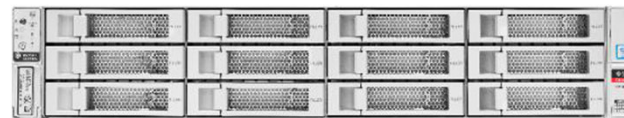
100 Gb/s RoCE
(RDMA over Converged Ethernet)
외부접속 25, 10, 1 GbE

- 확장 가능한 지능형 스토리지



High-Capacity Storage Server

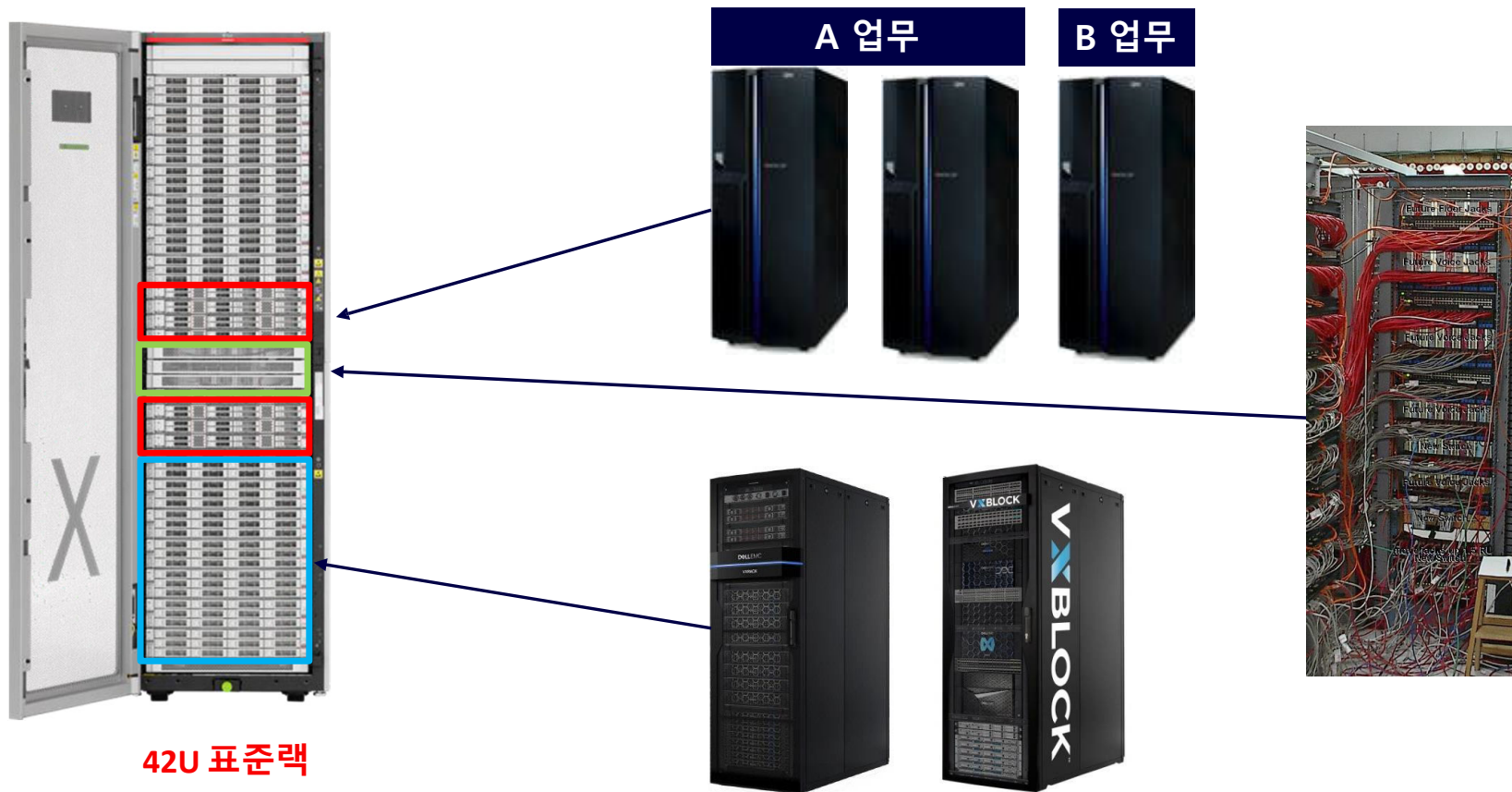
168 TB disk (14 TB 헬륨 디스크)
25.6 TB NVMe Flash
1.5 TB Persistent Memory
32 cores for SQL offload



Extreme Flash Storage Server

51.2 TB NVMe Flash
1.5 TB Persistent Memory
32 cores for SQL offload

Exadata와 타사 Solution H/W 비교

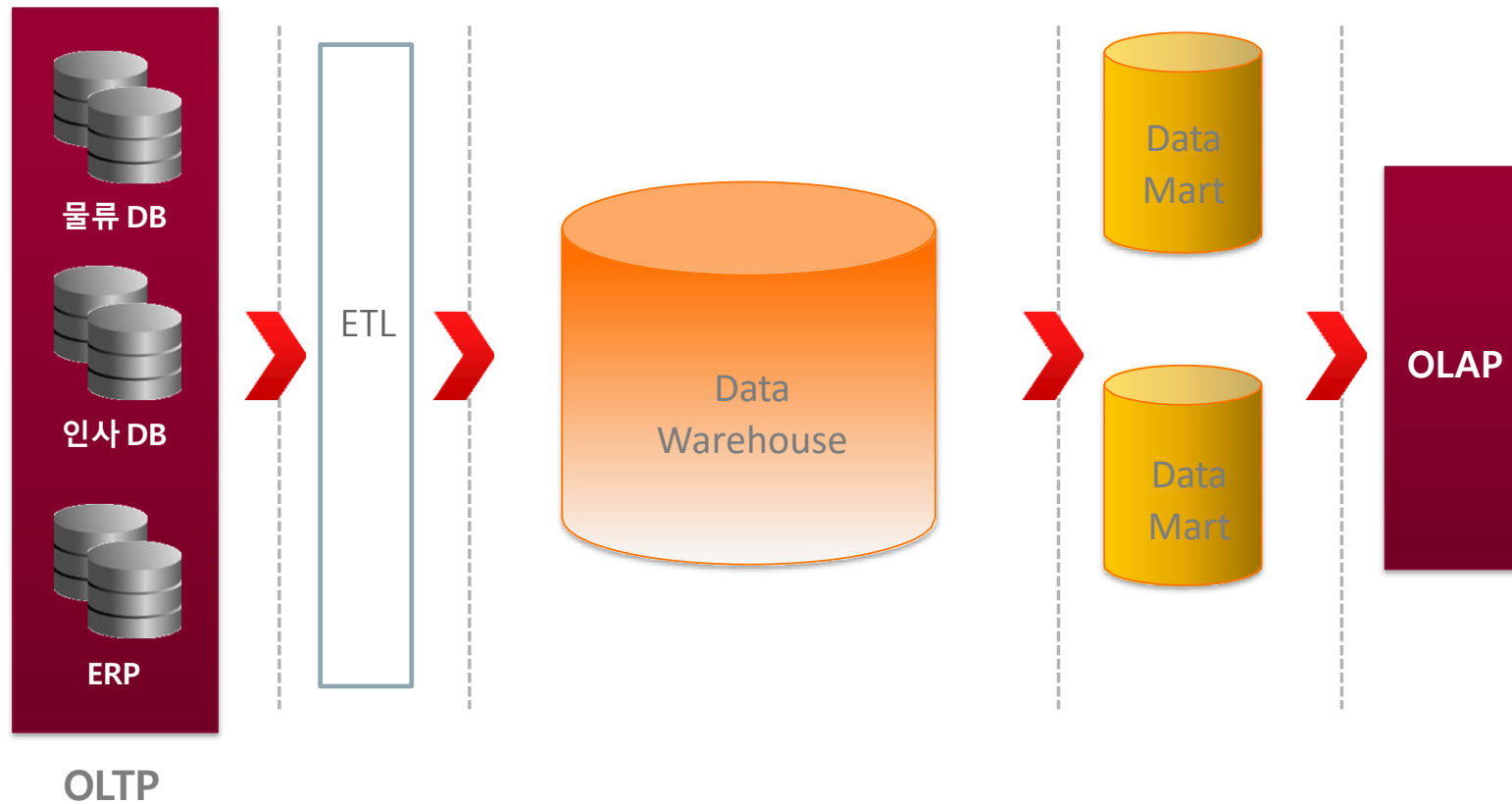


Exadata 출현 배경



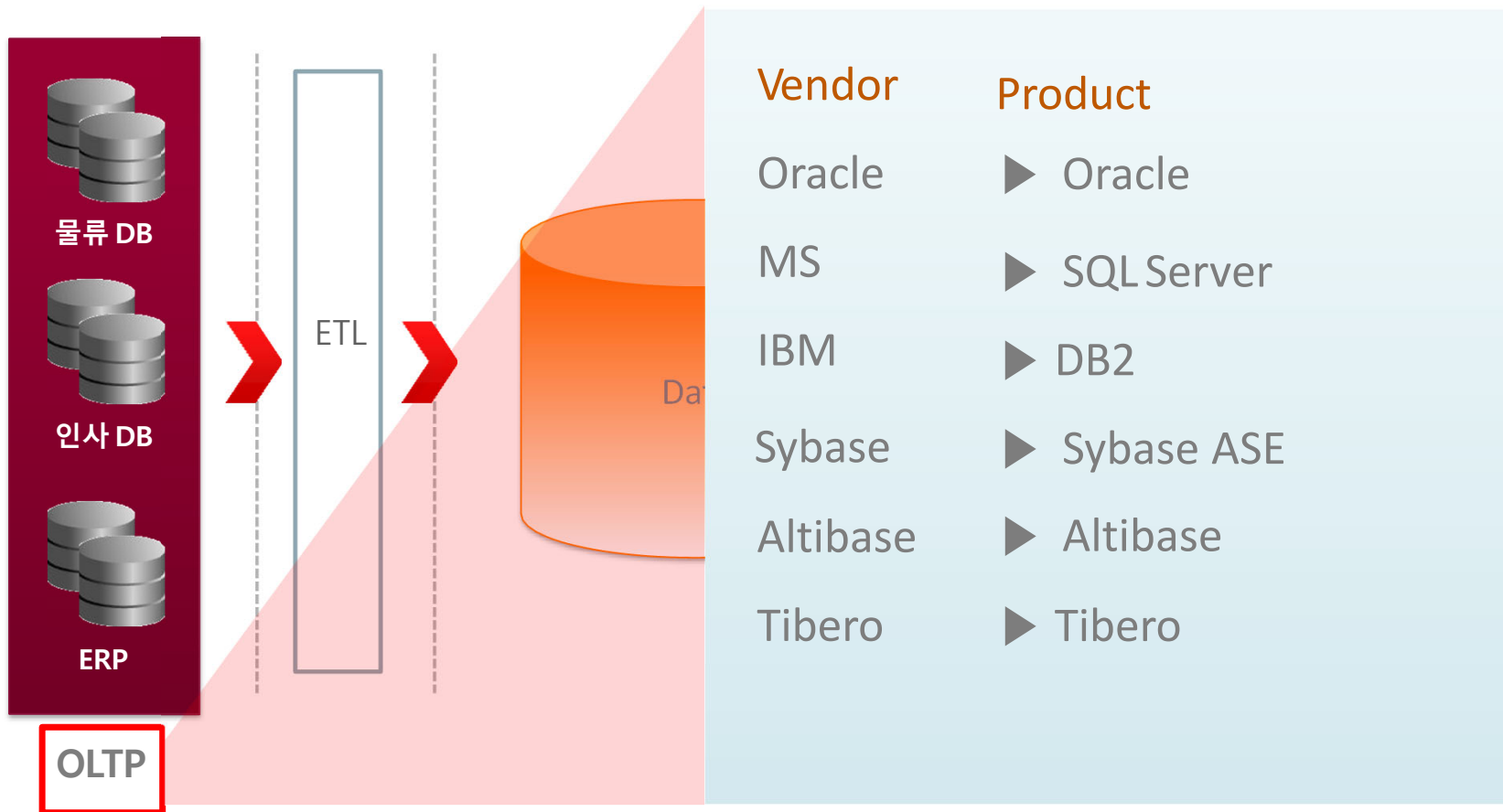
데이터베이스 용도별 DBMS

- DW 기본 Architecture



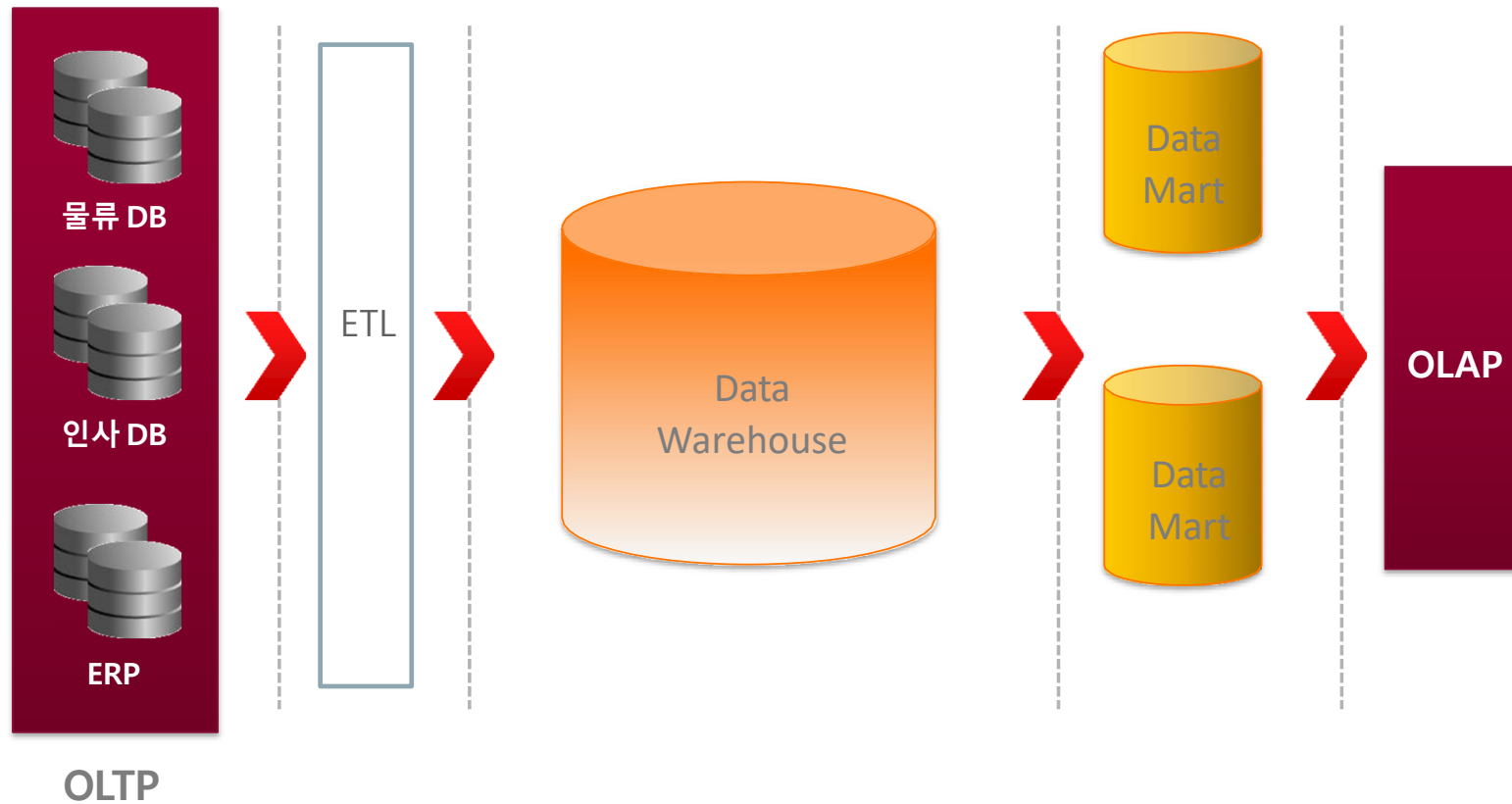
데이터베이스 용도별 DBMS

- DW 기본 Architecture



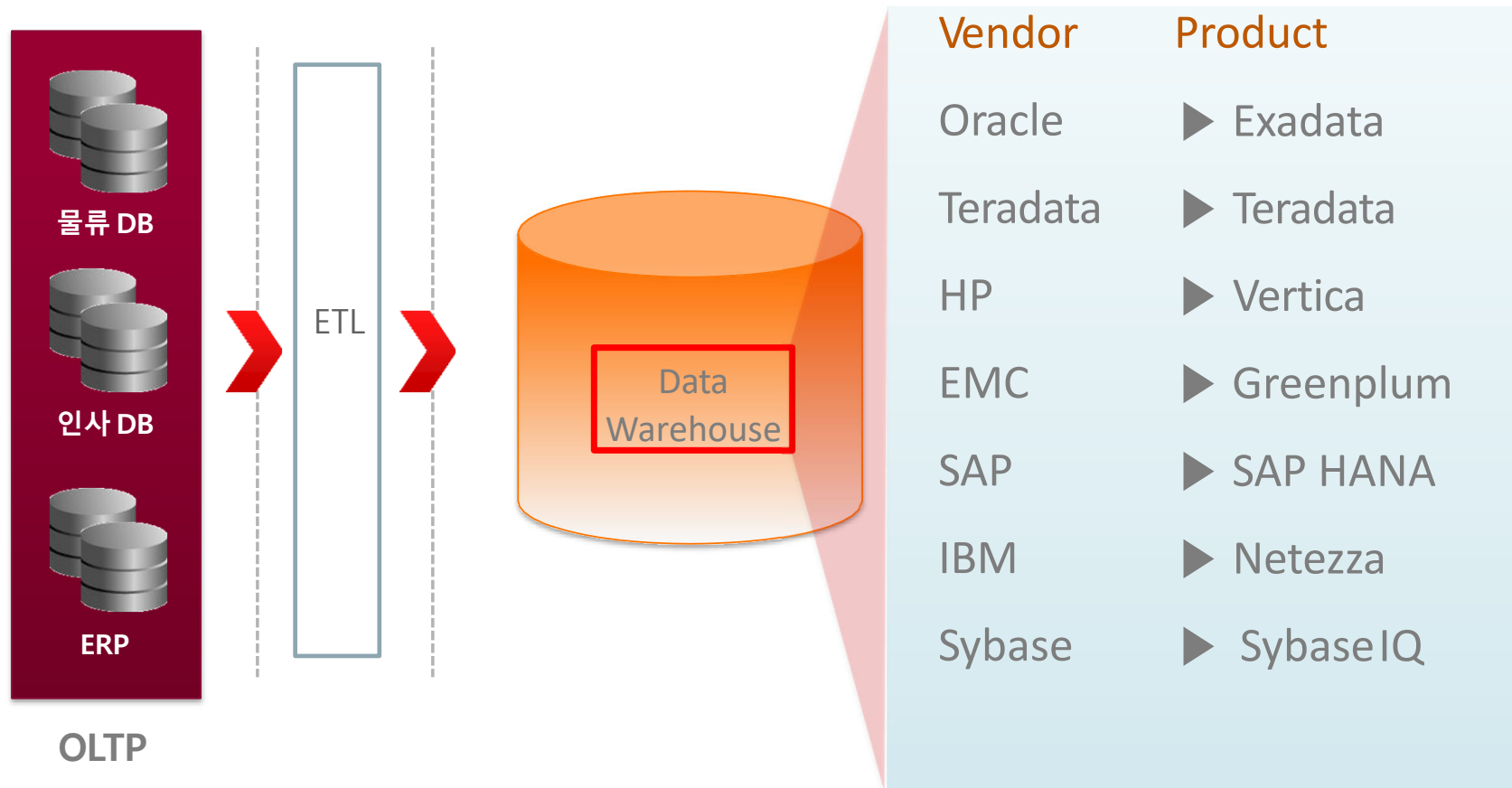
데이터베이스 용도별 DBMS

- DW 기본 Architecture

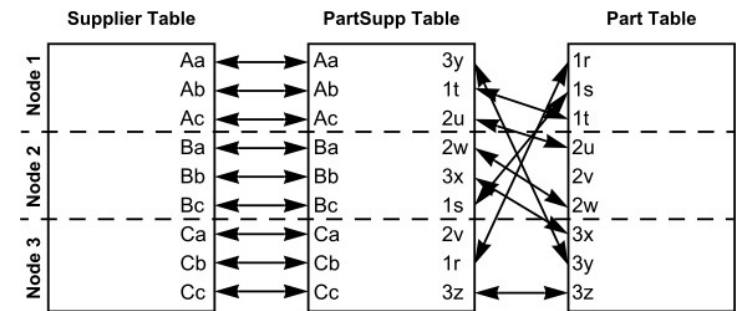
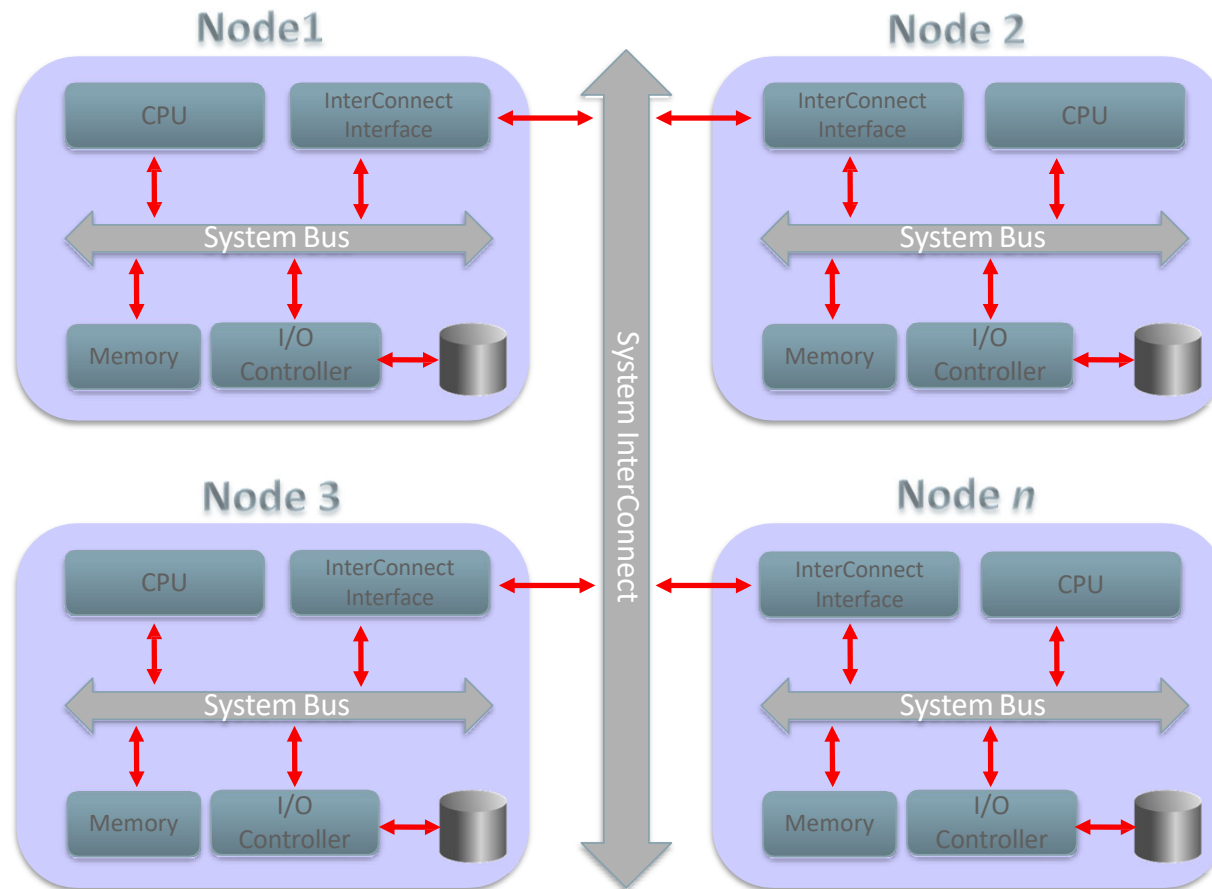


데이터베이스 용도별 DBMS

- DW 기본 Architecture



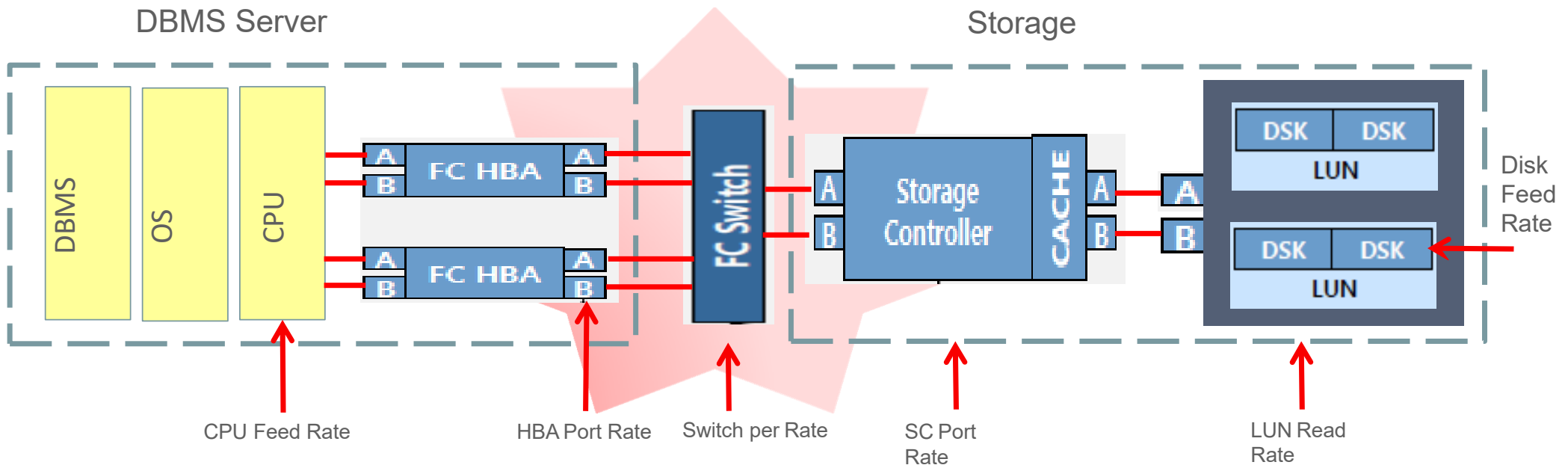
MPP란 무엇인가?



MPP: costly inter-node communication required

Exadata 스토리지의 출현 배경은 **SMP**의 한계

- 일반 스토리지 구성 시 스토리지 데이터 대역폭(Bandwidth)한계로 병목 현상 발생



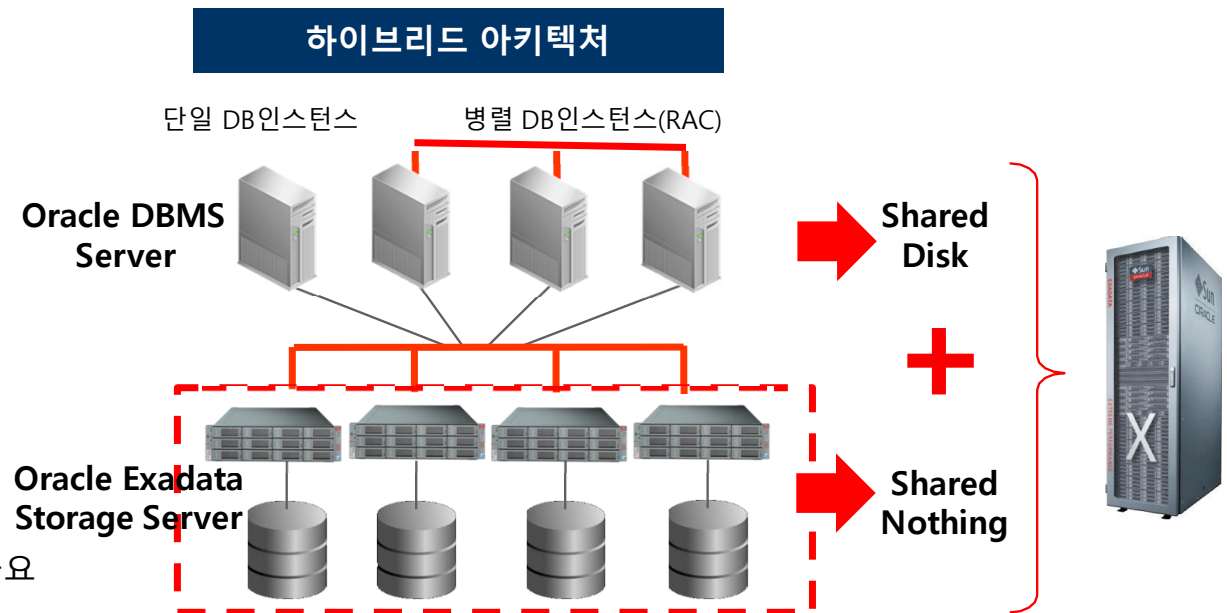
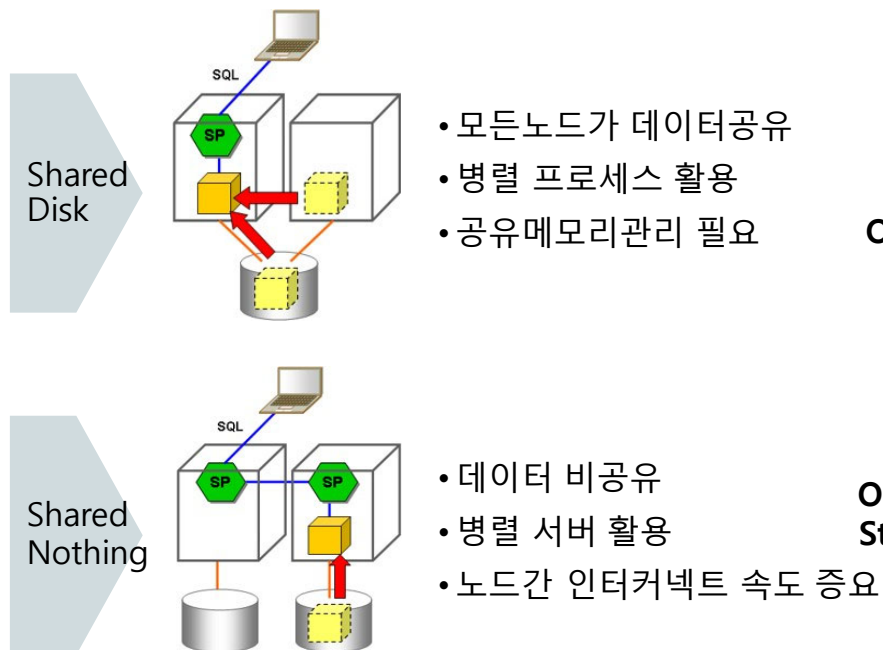
- 데이터베이스 성능은 Storage에 의해 제한되는 상황
 - 데이터는 기하급수적으로 늘어나고 있는 상황에서 기존의 SAN(Storage Array Network) 기반의 구조에서는 Database의 성능이 제한되어 대용량 데이터 처리에 한계가 존재

Exadata 아키텍처

Exadata는 SMP의 장점과 MPP의 장점을 수용한 하이브리드 아키텍처 방식

- SMP와 MPP는 수십년 동안 논란이 이어져 왔음
 - SMP(Symmetric Multi Processing : Shared Disk) → OLTP
 - MPP(Massive Parallel Processing : Shared Nothing) → DW

Exadata는 이러한 논쟁에 있어서 새로운 관점을 제공



Exadata 스토리지의 출현 배경

- 대역폭의 한계를 해결하는 방안



고속도로 정체가
심한 경우
해결 방안은?

해결 방안 1

통행 차량을 줄인다

해결 방안 2

고속도로의 차선을 넓힌다

해결 방안 3

고속도로를 추가로 건설한다

Exadata 스토리지의 출현 배경

- 대역폭의 한계를 해결하는 방안

해결 방안 1

통행 차량을 줄인다



Cell-offload 을 통해 Storage에서
필요한 데이터만 추출 / 전송

해결 방안 2

고속도로의 차선을 넓힌다



Network로 기존 대역폭을 6배 향상
(8/16Gbps -> 100Gbps)

해결 방안 3

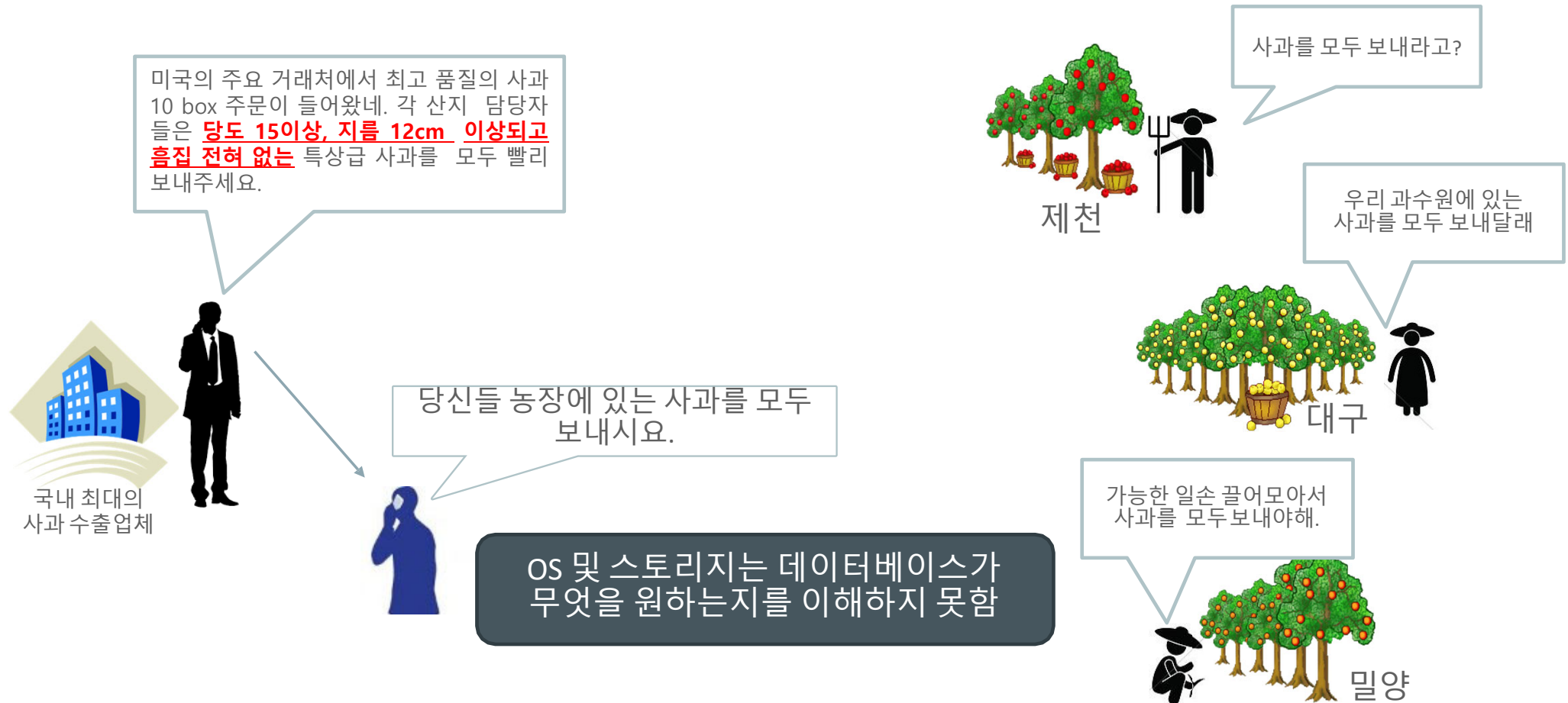
고속도로를 추가로 건설



스토리지 단위별로 Channel을
추가하는 새로운 Architecture 도입

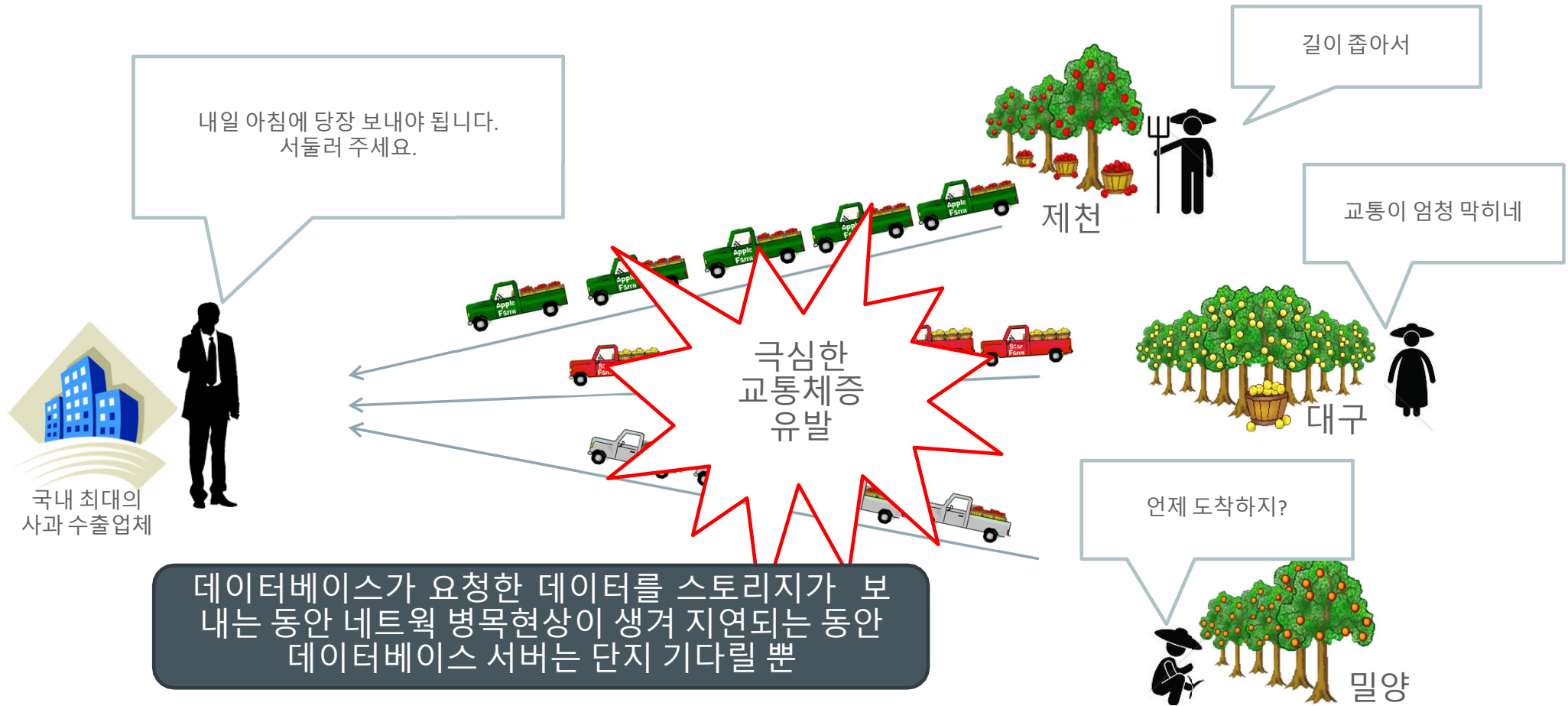
전통적인 Database Server Architecture

Non-Exadata Platform – 무슨일이 일어나고 있나?



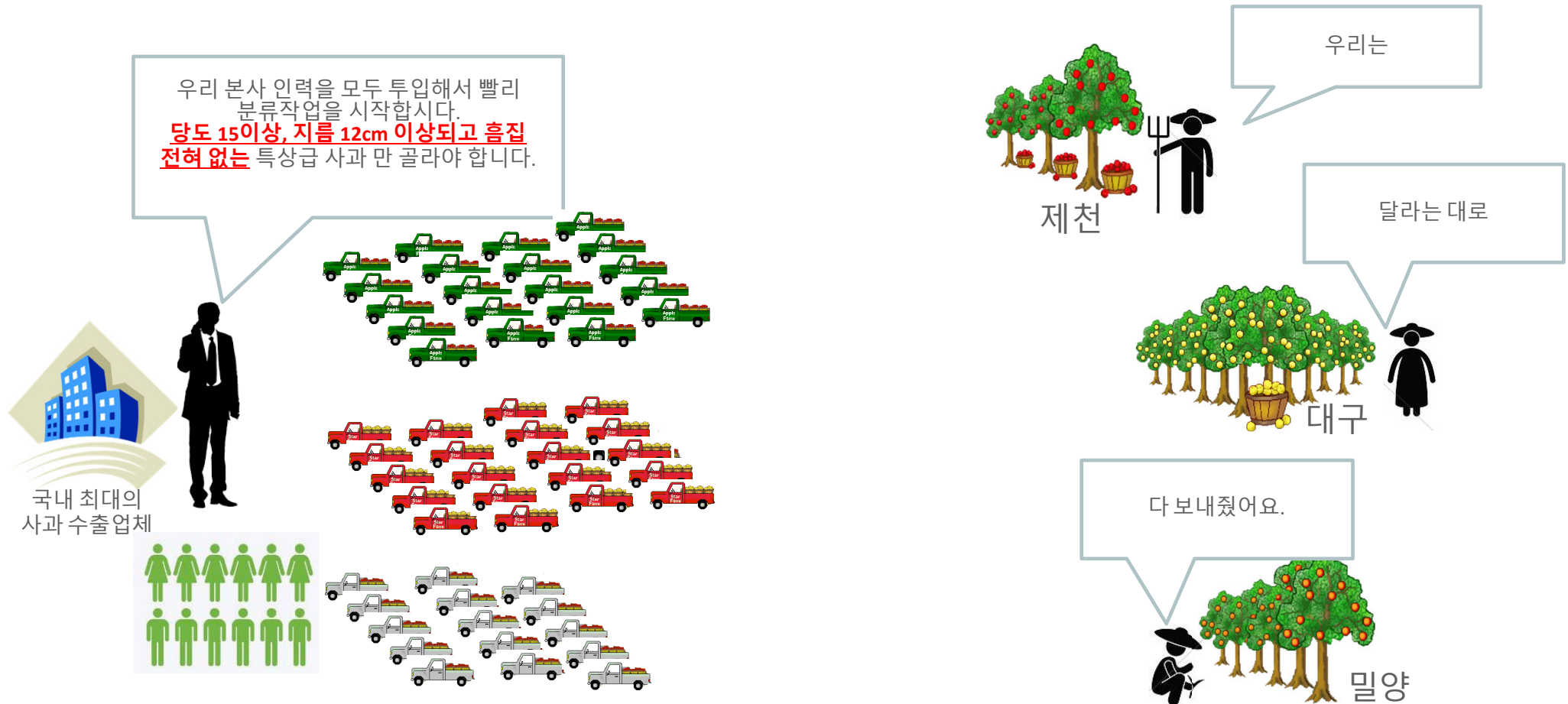
전통적인 Database Server Architecture

Non-Exadata Platform – 무슨일이 일어나고 있나?



전통적인 Database Server Architecture

Non-Exadata Platform – 무슨일이 일어나고 있나?



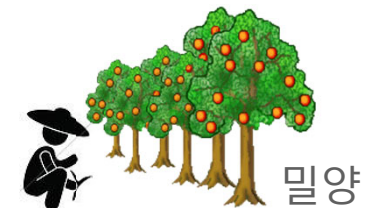
전통적인 Database Server Architecture

Non-Exadata Platform – 무슨일이 일어나고 있나?

믿기 힘들시겠지만, 지금 이 시간에도 귀사의 오라클 데이터베이스를 구동하는 Non-Exadata 데이터베이스 서버에서 일어나고 있는 일입니다



밤샘 작업으로 분류 완료했습니다.
기준에 맞는 사과는 모두 100 box 입니다.
10 box 는 출고하고 나머지 90 box는
폐기처분 합니다.



Exadata Platform – 데이터처리의 혁신

미국의 주요 거래처에서 최고 품질의 사과 10 box 주문이 들어왔네. 각 산지 담당자들은 당도 15이상, 지름 12cm 이상되고 흠집 전혀 없는 특등급 사과를 모두 빨리 보내주세요.

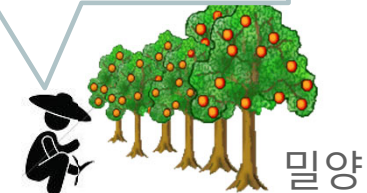


이번에는 특상품 주문이구나!

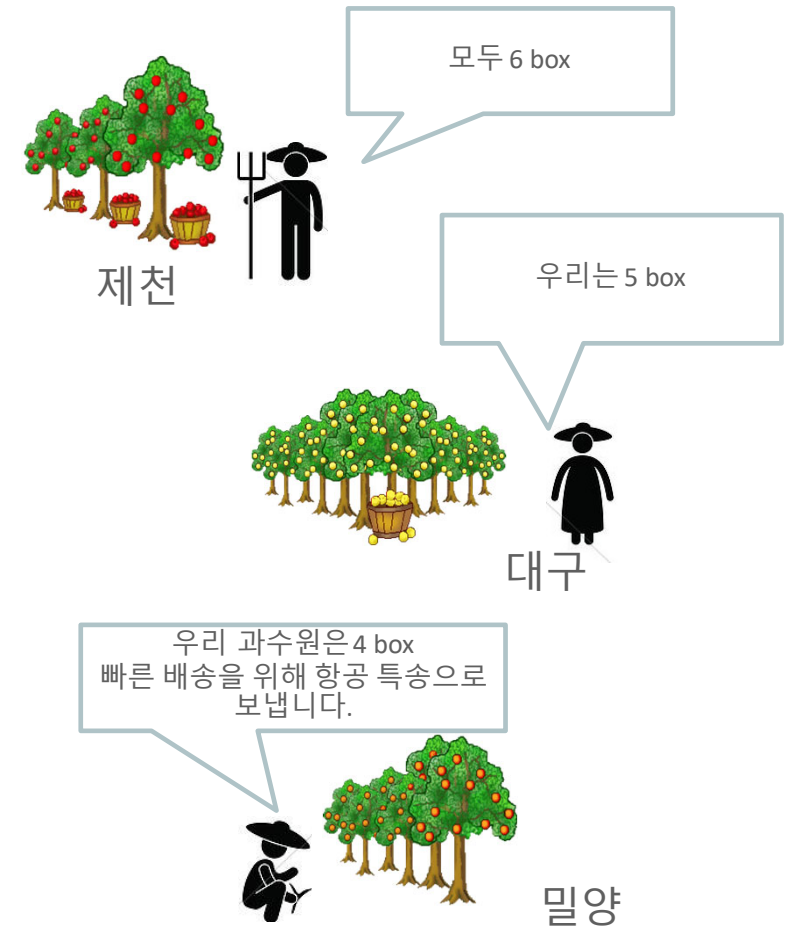


올해는 우리 과수원 사과 품질이 최고지!

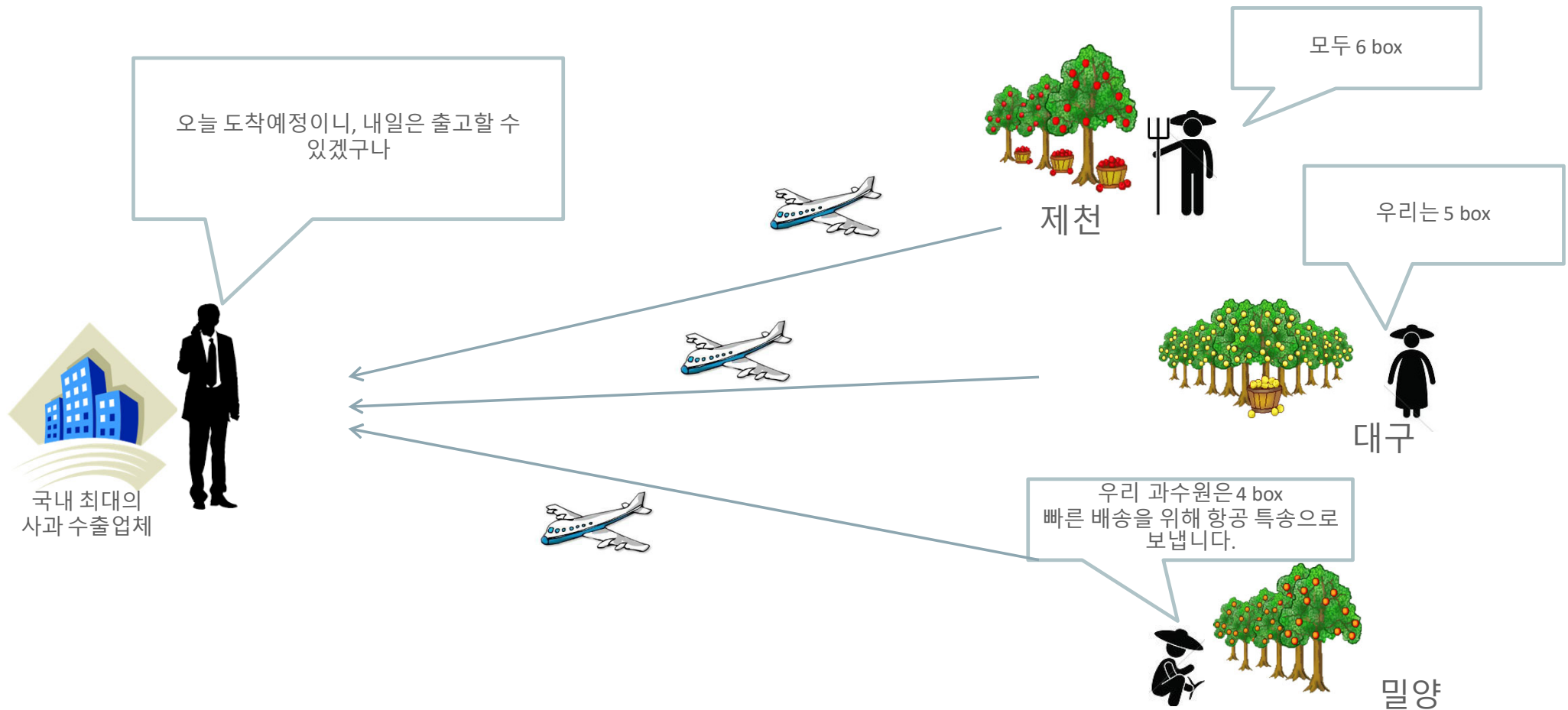
당도는 괜찮은데, 지름 12cm 이상은 많이 없을텐데..



Exadata Platform – 데이터처리의 혁신



Exadata Platform – 데이터처리의 혁신



Exadata Platform – 데이터처리의 혁신

Exadata의 스토리지는 오라클 데이터베이스 엔진을 이해하며, 필요없이 낭비되는 작업을 원천적으로 배제합니다.

산지에서 상품이 도착했습니다. 확인 후에 출고 부탁드립니다.

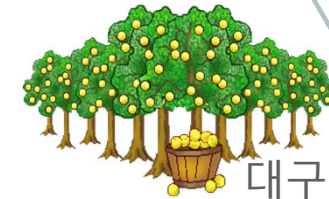


10 box 확인해서 출고완료, 나머지 5box는 창고에 저장합니다.

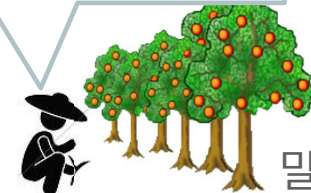


비슷한 주문이 들어올 수 있으니

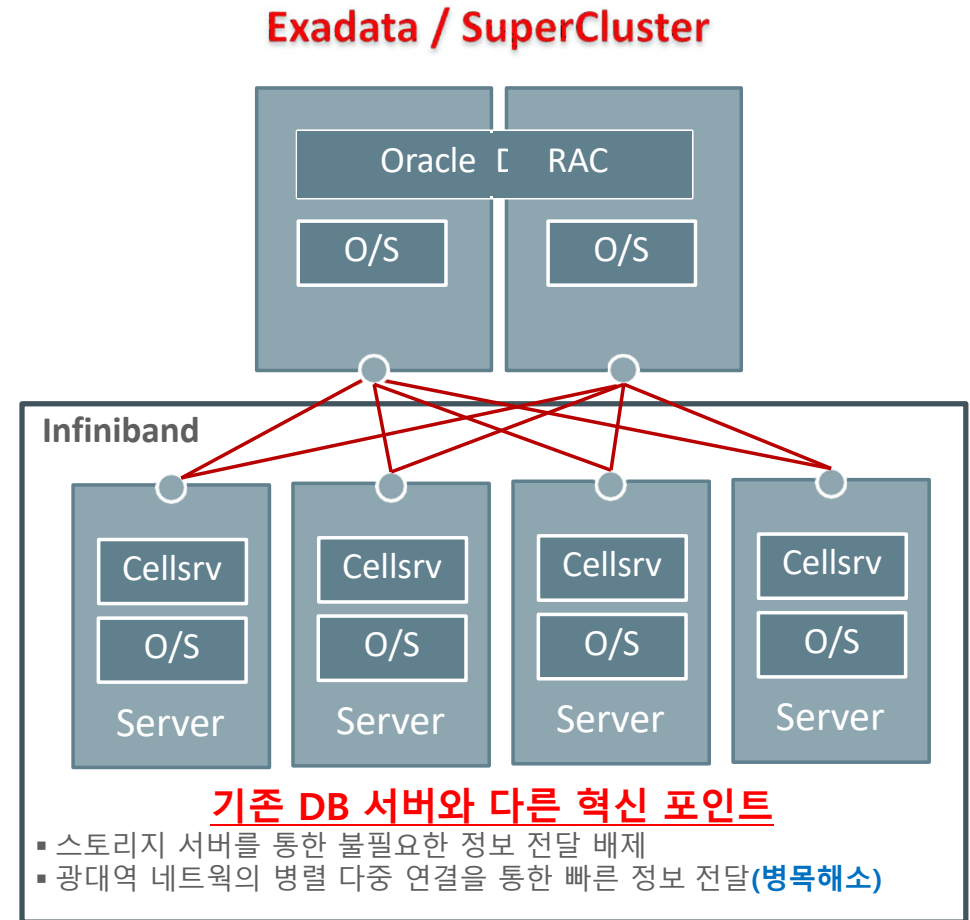
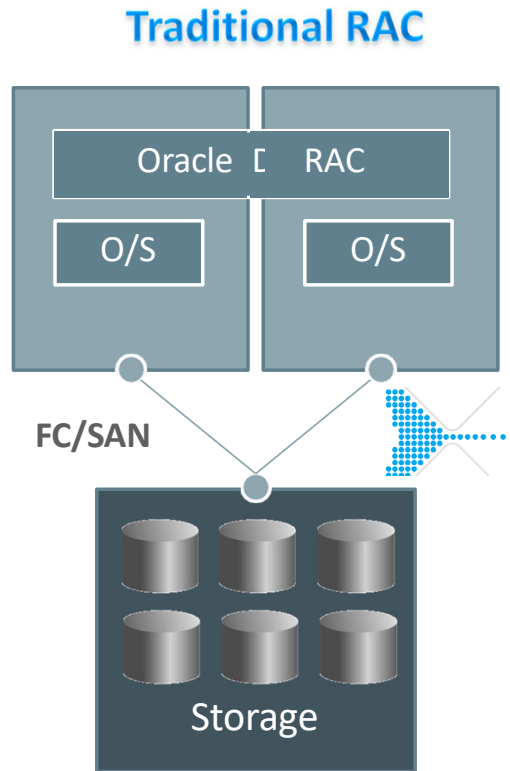
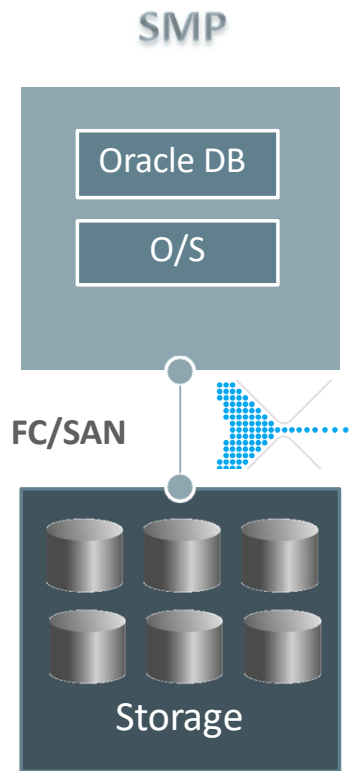
수확한 사과를 잘 분류해서



창고에 저장해 놓아야겠다.



스토리지 별로 Channel을 추가하여 무한 Channel 확장

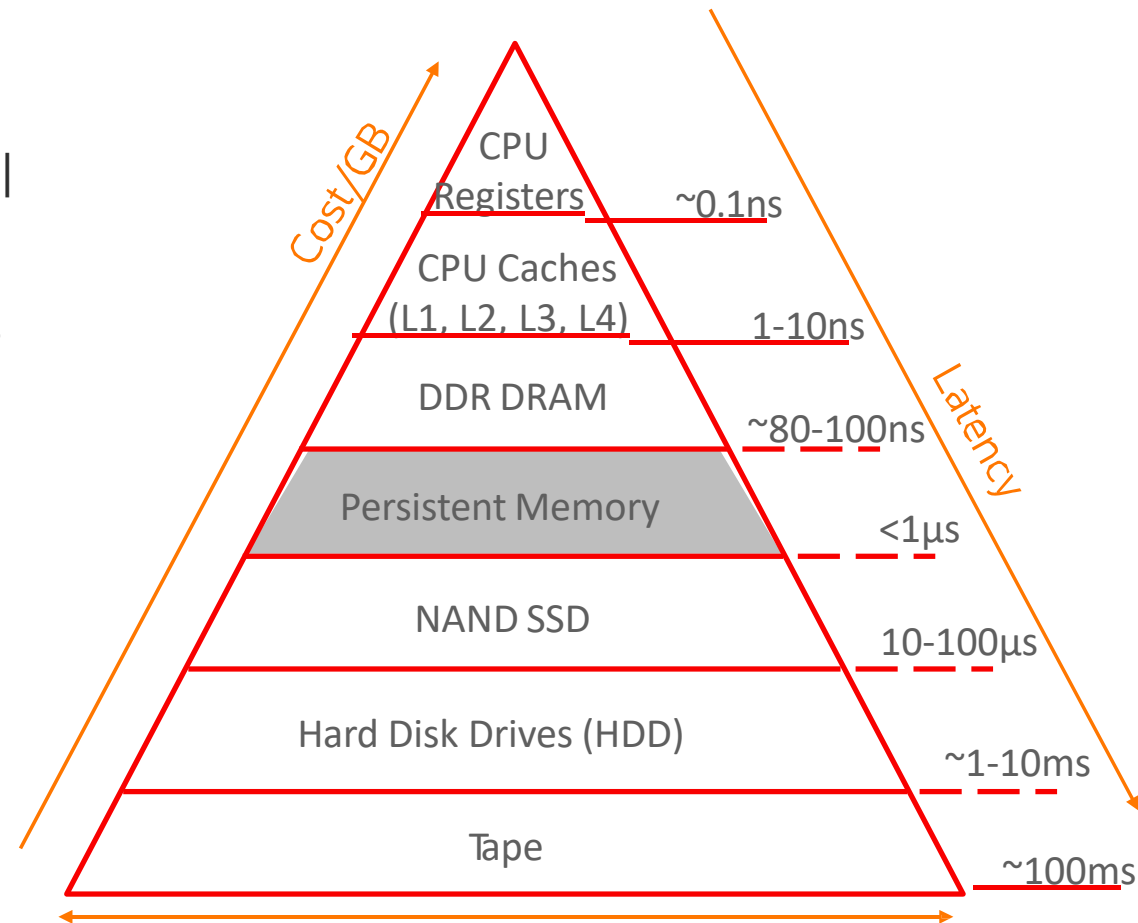
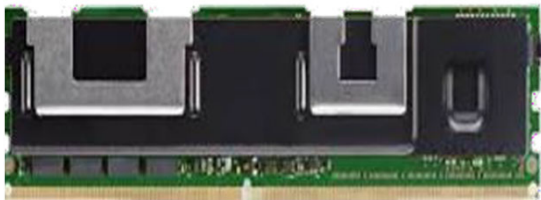


Exadata OLTP로의 도약 X8M의 새로운 기능



새로운 Persistent Memory의 Exadata 적용

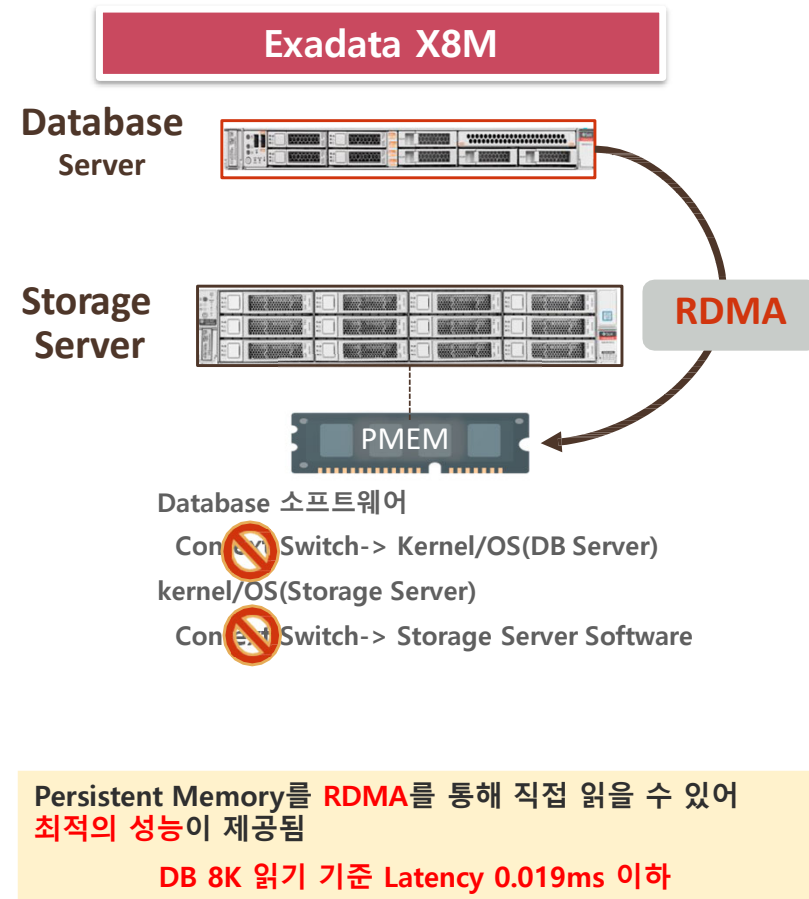
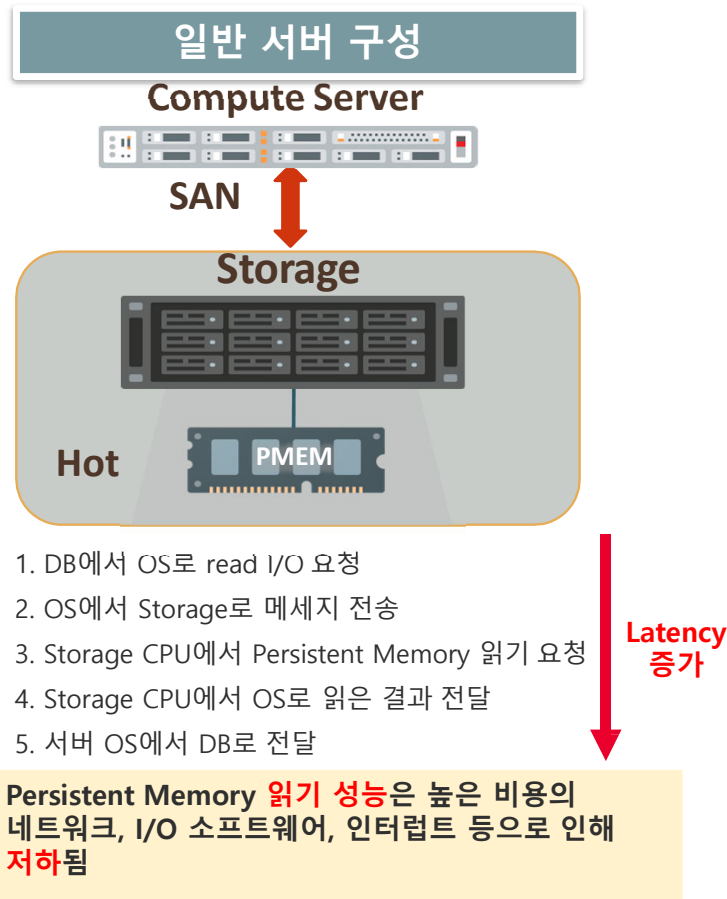
- Persistent memory는 새로운 실리콘 기술
 - 용량, 성능 및 가격은 DRAM과 플래시 사이
- Intel® Optane™ DC Persistent Memory:
 - 메모리 속도 읽기 - 플래시보다 훨씬 빠름
 - DRAM과 달리 정전시에도 쓰기 보존
- Exadata는 PMEM을 Read는 Cache로 사용하여 사용량을 최대로 유지하고 Re-log write에 PMEM을 사용함.



X8M 특징점 - 고성능

RoCE 네트워크 - Remote Direct Memory Access

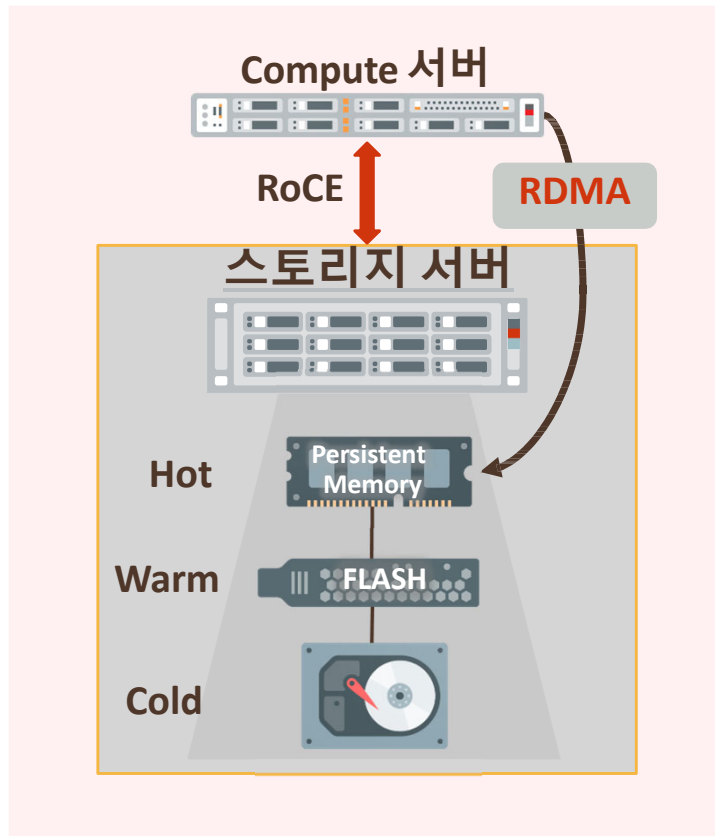
Exadata RDMA 기술은 RoCE 스위치를 통해 100Gbps의 속도로 Ethernet fabric을 통해 안정적으로 제공됨



X8M 특징점 - 고성능

PMEM(Persistent Memory) – PMEM Cache 데이터 가속

Exadata X8M 장비는 데이터베이스에 최적화된 세계 최초, 유일한 공유 방식의 Persistent Memory가 탑재 되어 IO 성능을 극대화 시킴



- Exadata 스토리지 서버는 플래시 메모리 앞에 Persistent Memory Accelerator를 투명하게 추가
 - ✓ 기존 X8 대비 2.5X 높은 IOPS – 1600 만 IOPS(FR)
- 데이터베이스는 IO 대신 RDMA를 사용하여 원격 PMEM 읽음
 - ✓ 네트워크 및 IO 소프트웨어, 인터럽트, 컨텍스트 스위치 우회
 - ✓ 기존 X8대비 10X 빠른 latency - <19 μ sec (8K database read)
- PMEM이 자동으로 계층화 되고 DB간 모두 공유
 - ✓ 가장 많이 사용되는 데이터를 캐시 하여 유효 용량 10X 증가
- Fault-tolerance를 위해 스토리지 서버에서 자동으로 Persistent Memory 미러링

Exadata System Software 19.3 및 Database Software 19c 에서 가능

X8M 특징점 - 고성능

PMEM(Persistent Memory) – PMEM Cache 데이터 가속 적용 전후 성능 비교

TPC-C 수행테스트에서 PMEM Cache를 적용한 결과 single block physical IO wait가 약 10배 개선됨

Top 10 Foreground Events by Total Wait Time

Event	Waits	Total Wait Time (sec)	Avg Wait	% DB time	Wait Class
cell single block physical read	47,113,315	14K	296.96us	50.8	User I/O
DB CPU		12.9K		46.8	
cell list of blocks physical read	5,837,281	4809.2	823.88us	17.5	User I/O
log file sync	1,744,262	657.4	376.92us	2.4	Commit
read by other session	7,753	10.5	1.35ms	.0	User I/O
SQL*Net message to client	14,735,787	7.4	499.81ns	.0	Network
control file sequential read	5,379	1	192.45us	.0	System I/O
latch free	1,835	.6	302.91us	.0	Other
PGA memory operation	16,983	.4	24.07us	.0	Other
Sync ASM rebalance	11				

single block physical
IO wait가
약 10배 개선됨

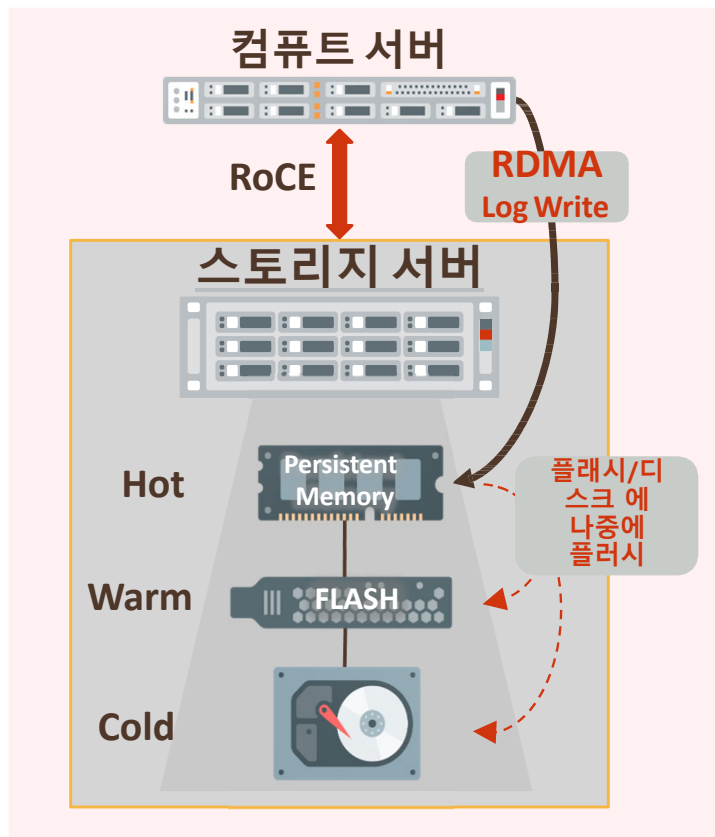
Top 10 Foreground Events by Total Wait Time

Event	Waits	Total Wait Time (sec)	Avg Wait	% DB time	Wait Class
DB CPU		25.5K		95.4	
cell list of blocks physical read	11,577,225	7091.6	612.54us	26.5	User I/O
cell single block physical read	94,805,407	2540.6	26.80us	9.5	User I/O
log file sync	3,080,055	1557.6	505.72us	5.8	Commit
read by other session	17,666	17.7	1.00ms	.1	User I/O
SQL*Net message to client	25,995,939	10.9	418.07ns	.0	Network
PGA memory operation	54,053	1.4	26.24us	.0	Other
latch free	6,413	.6	95.56us	.0	Other
latch: cache buffers chains	33,852	.5	14.95us	.0	Concurrency
Sync ASM rebalance	11	.4	38.93ms	.0	Other

X8M 특징점 - 고성능

PMEM(Persistent Memory) – PMEM Log Commit 가속

Exadata X8M 장비는 DB에 최적화된 세계 최초의 유일한 공유 방식의 Persistent Memory가 탑재 되어 Log write 성능을 극대화 시킴



■ Log Write latency는 OLTP 성능에 매우 중요

- ✓ 빠른 Redo Log write IO는 빠른 트랜잭션 Commit 시간을 의미함
- ✓ Log write 속도 저하는 전체 데이터베이스 성능에 영향을 줌

■ 자동 Commit 가속 기능

- ✓ 데이터베이스는 여러 스토리지 서버의 PMEM에 단방향으로 RDMA 쓰기를 수행함
- ✓ 네트워크 및 IO 소프트웨어, 인터럽트, 컨텍스트 스위치 등을 건너뛴
- ✓ 기존 X8대비 최대 8x 빠른 Log Writes

Exadata System Software 19.3 및 Database Software 19c 에서 가능

X8M 특징점 - 고성능

PMEM(Persistent Memory) – PMEM Log Commit 가속 적용 전후 비교

TPC-C 수행테스트에서 PMEM Cache를 적용한 결과 log file sync wait가 약 8배 개선됨

Top 10 Foreground Events by Total Wait Time

Event	Waits	Total Wait Time (sec)	Avg Wait	% DB time	Wait Class
log file sync	19,100,192	7920.8	414.70us	61.2	Commit
DB CPU		4867		37.6	
cell multiblock physical read	108,691	298.3	2.74ms	2.3	User I/O
cell single block physical read	1,688,422	293.2	173.67us	2.3	User I/O
cell list of blocks physical read	107,128	37	345.70us	.3	User I/O
SQL*Net message to client	22,150,135	9.2	413.09ns	.1	Network
Sync ASM rebalance	9	.3	31.27ms	.0	Other
control file sequential read	3,056	.2	78.43us	.0	System I/O
Disk file Mirror Read	565	.1	215.40us	.0	User I/O
cell statistics gather	108	.1	516.45us	.0	User I/O

log file sync wait 가
약 8배 개선됨

Top 10 Foreground Events by Total Wait Time

Event	Waits	Total Wait Time (sec)	Avg Wait	% DB time	Wait Class
DB CPU		10.9K		94.3	
log file sync	36,577,091	1940.5	53.05us	16.7	Commit
cell single block physical read	1,222,367	128.6	105.25us	1.1	User I/O
cell multiblock physical read	38,394	62.9	1.64ms	.5	User I/O
SQL*Net message to client	47,866,078	17.2	359.13ns	.1	Network
cell list of blocks physical read	26,104	4.4	166.94us	.0	User I/O
Sync ASM rebalance	9	.3	32.05ms	.0	Other
control file sequential read	3,056	.2	68.19us	.0	System I/O
read by other session	4	.2	40.72ms	.0	User I/O
Disk file Mirror Read	565	.1	201.23us	.0	User I/O