

데이터웨어하우스 및 분석 현대화를 위한 IBM NPS 어플라이언스 소개

최석재 실장
Data & AI
IBM Cloud and Cognitive Software

2020-09



IBM

DW Appliance의 탄생 & PDA의 등장

전통 데이터 웨어하우스 이슈 1 – 관리이슈

지나치게 복잡합니다.

전통적인 데이터 웨어하우스는 트랜잭션 처리 작업에 최적화된 데이터베이스를 기반으로 하기 때문에 대량의 데이터를 토대로 한 고급 분석 작업의 요구사항을 충족하지 못합니다.

- 너무 복잡한 인프라
- 너무 복잡한 배치
- 너무 많은 튜닝 필요
- 너무 비효율적인 분석 작업
- 유지에 너무 많은 인력 필요
- 운영에 너무 많은 비용 소모

필요한 정보를 얻는 데 너무 많은 시간 소요



전통 데이터 웨어하우스 이슈 2 – 성능이슈

하나의 쿼리 처리에 수일 소요
끊임없는 튜닝

“ 거의 70%의 데이터 웨어하우스는 성능을 제약하는 다양한 형태의문제로 고민을 하고 있습니다. ”

- Gartner Magic Quadrant

특수 리소스 필요
배치에 수개월 소요



IBM PDA

(PureData System for Analytics)

- **빠르고 간단하게** 복잡한 분석 업무를 수행할 수 있도록 고안된 시스템
- 데이터베이스, 서버, 스토리지가 결합된 어플라이언스 형태의 DW 전용 솔루션
- **Netezza**, 2003년 세계 최초로 DW형태의 어플라이언스
- IBM이 2010년 인수

“ 2010년 IBM이 Netezza인수 후 2018년 기준 69개 고객사에서 92대의 PDA 도입 ”

What is a Modern Data Warehouse?

다양한 팀에서 활용

Cloud based
(클라우드 기반)

Business ready data fast access
(빠른 데이터 접근, 즉시 업무 적용)

Business leaders

Scalable (확장가능)

High Performances
(높은 성능)

Data Scientists

Cost efficient
(비용 효율)

Trusted (신뢰)

Data Analysts

Data Engineers

Optimized
(최적화)

Integration ready
(데이터 통합)

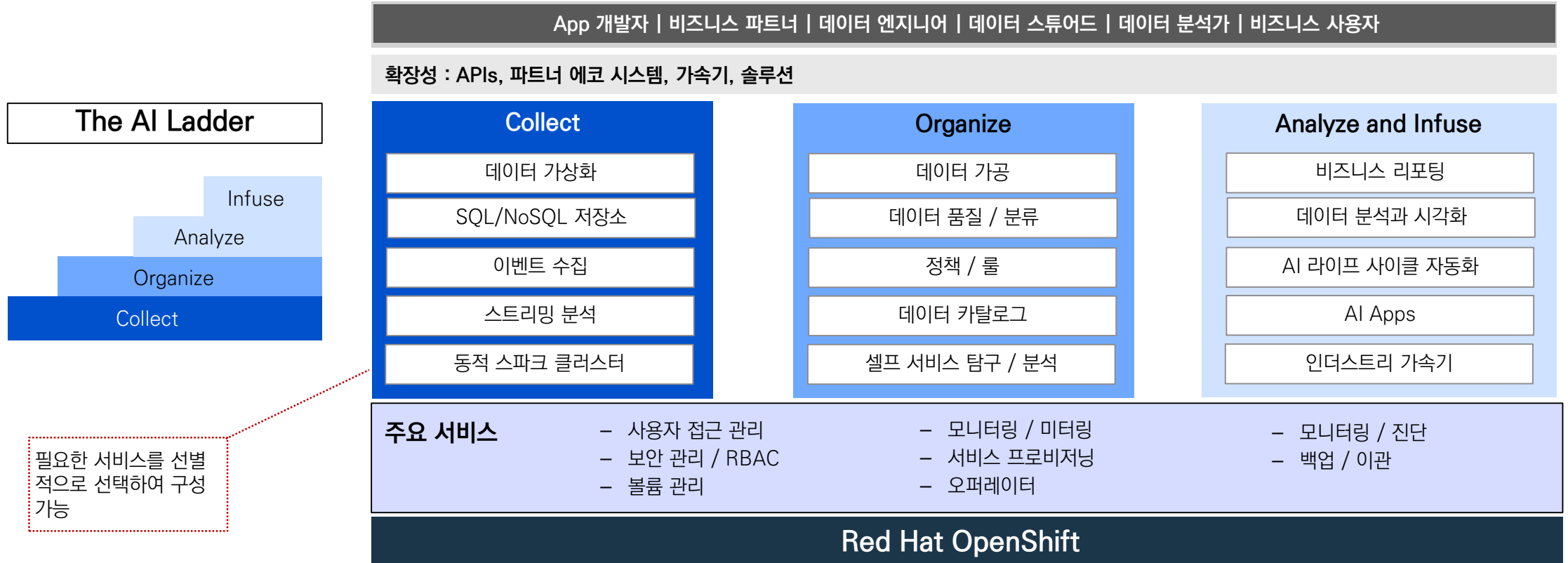
IT

Provides business value

**Netezza Performance Server (NPS)
in
IBM Cloud Pak for Data System**

IBM Cloud Pak for Data

Cloud Pak for Data는 AI 플랫폼 구현에 필요한 여러 구성 요소들을 MSA 기반으로 구현하여 필요한 서비스만 선별적으로 선택하여 구성하는 것을 지원합니다.



NPS의 기반 플랫폼 : Cloud Pak for Data System

SOFTWARE

Cloud Pak for Data

+

HARDWARE

HCI (Hyper Converged System)

- **Data & AI 플랫폼**
사전 테스트 된 빌딩 블록으로 신속하게 조립
- **거버넌스가 내장된 플랫폼**
데이터 관리, 통제 및 분석 방법을 단순화하고 통합합니다.
- **손쉽게 확장 가능**
add-on 카탈로그에서 제공되는 서비스
- **하이퍼컨버지드 시스템 (Hyper Converged System)**
스토리지, 컴퓨팅, 네트워크 및 소프트웨어를 단일 시스템으로 결합하여 복잡성을 줄이고 확장성을 높입니다.
- **확장 가능한 시스템**
레고 블럭 처럼 구성할 수 있는 데이터 센터



Netezza Performance Server 아키텍처 개요

NPS의 기반 플랫폼은 IBM Cloud Pak for Data이며 Red Hat OPEN SHIFT에 기반을 두고 데이터와 관련된 각종 서비스들을 사용할 수 있는 데이터 분석 전용 플랫폼으로 멀티클라우드를 기반으로 설계되었습니다.

1. 서비스 에코 시스템

단순한 클릭으로 45개 이상의 IBM 또는 3rd Party 분석 서비스 추가

2. 데이터 가상화 (Data Virtualization)

데이터 이동없이 쉽고 빠르게 다수의 데이터 소스에서 데이터 액세스

3. 플랫폼 인터페이스 **Cloud Pak for Data**

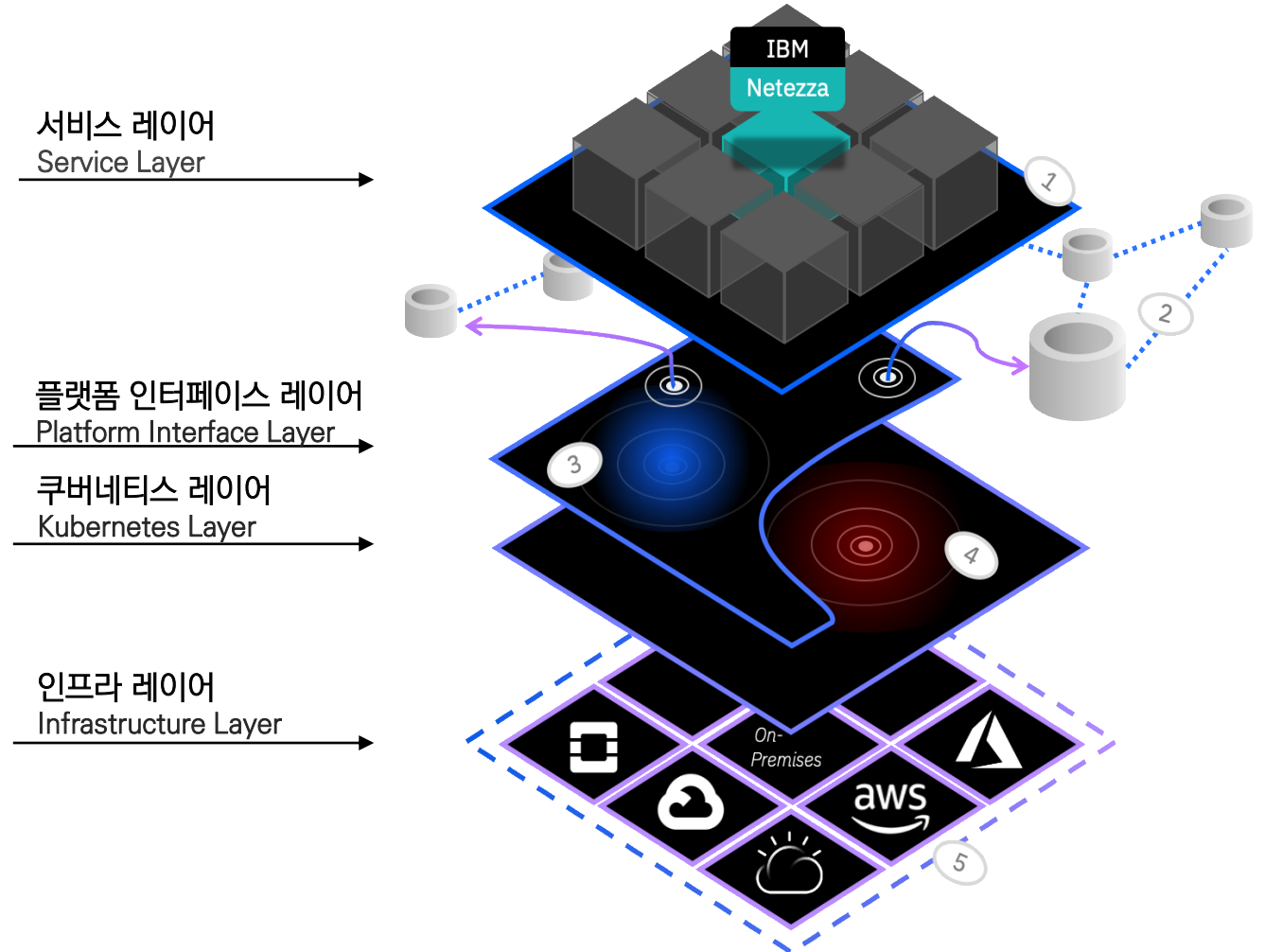
데이터 관리, 거버넌스, 분석등의 작업을 통합된 UI 및 인터페이스로 이용

4. Red Hat **OPENS**HIFT

컨테이너 기반의 하이브리드 클라우드 환경 제공

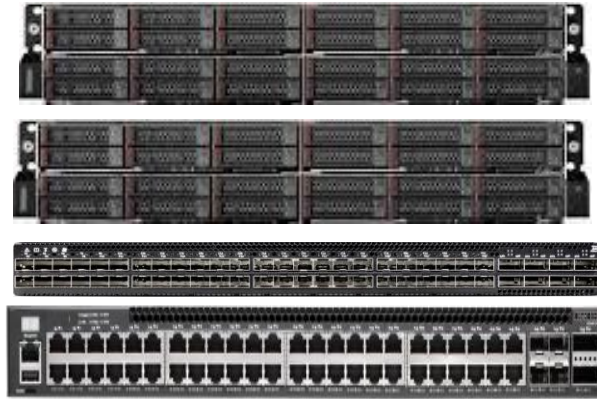
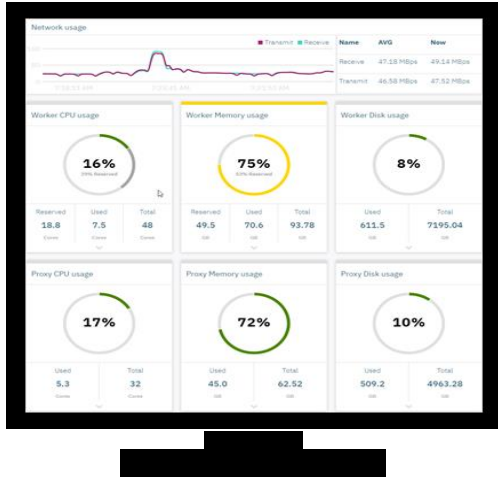
5. Any Cloud

벤더에 종속되지 않은 멀티클라우드 환경 구축



Netezza Performance Server 구성 단위

Base (Starter 유닛) + Extension (4개 S-Blade단위)



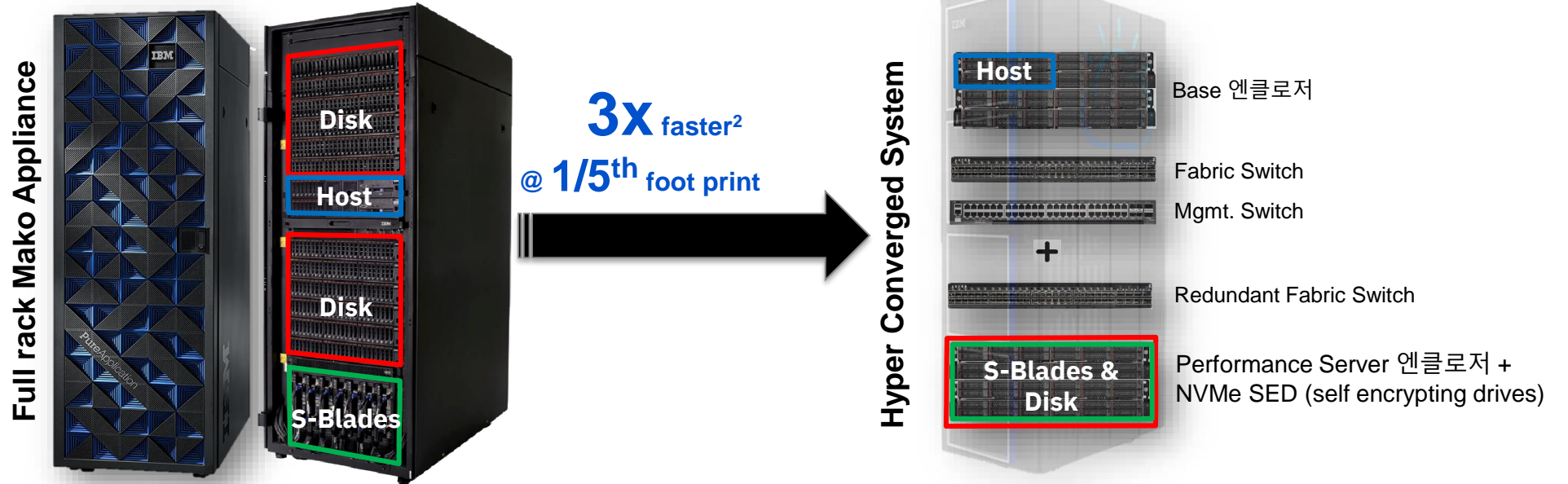
Netezza Performance Server – Base (2 Enclosure)

- OpenShift (master + worker)
- Cloud Pak for Data
- NPS Host 서버 컨테이너

Netezza Performance Server – Extension (n Enclosures)

- Netezza Performance Server 엔진
- Enclosure당 4개의 S-Blade
- Enclosure단위 확장가능

Netezza Performance Server vs N3001 Full Rack과의 비교



System	IBM PureData System for Analytics N3001-010	Netezza Performance Server Base + 2
Blade 서버 Core	140 cores (+ Host Server core 추가)	128 Cores / VPCs (+ Cloud Pak for Data Server core 추가됨)
Memory	896 GB	1.5 TB
Storage – RAW	148 TB (SAS Disk)	128 TB (NVMe)
Storage – 스캔속도	119.5 TB/h	288TB/h (엔클로저별 40 GB/s)
Available User Space	48 TB ¹	34 TB ¹
Performance	1X	3X+ (Twin Fin의 경우 5-8x) ²
Footprint (공간)	Full rack (42U Full)	랙 공간의 70%이상 절약 (12U)

¹ 통상 4배 압축을 지원하므로 수용하는 데이터 용량은 기재된 수치의 4배를 셈함

² TPC-DS 10, 20, 30 TB 테스트 및 고객사 환경과 유사하게 테스트 진행하였을 때의 성능 비교

상세 HW 비교

각 파트의 성능을 비교 하였을 때, 성능에 가장 중요한 영향을 미치는 SPU영역의 Disk의 Scan 성능이 약 3배정도 빨라졌고, 네트워크 대역폭도 2.5배 증가되어 HW수치로 본 성능은 약 3배 정도 향상되었음을 알 수 있습니다.

구분	파트	측정 기준	PDA N3001-010		NPS Base + 2		수치 비교
			모델	수치	모델	수치	
Host Server	CPU	총 Core수	Intel Xeon E5-4640 v2 10C 2.2GHz	20	Intel Xeon Silver 4110 8C 2.1GHz	16	20% 적음
		성능 ¹⁾		19,209		16,855	10% 느림
	Memory	총 용량 (GB)	-	64	192GB	192	3배 많음
S-Blade	CPU	총 Core수	Intel Xeon E5-2660 v2 10C 2.2GHz	140	Intel Xeon Silver 4110 8C 2.1GHz	128	10% 적음
		성능 ¹⁾		125,076		134,840	유사
	Memory	총 용량 (GB)	-	896	192GB	1,536	1.7 배 많음
	FPGA	처리단위 ²⁾	Xilinx Virtex-6 XC6VLX550T	1,202,880	Xilinx Kintex Ultrascale XCKU15P	4,030,464	3.4 배
Storage	Disk	총 Disk 수	Seagate ST600MM0026 600GB 10K 6Gbps SED	288	ThinkSystem U.2 CM5-R 3.84TB NVMe PCIe 3.0 Hot Swap SSD SED (4XB7A14060)	32	90% 작음
		IOPS Read		129,024		24,000,000	186 배 빠름
		IOPS Write		106,272		2,240,000	21 배 빠름
		Scan Rate (GB/s) ³⁾		33.6		96	2.9 배 빠름
Network (Fabric)	Switch	총 개수	Lenovo RackSwitch G8264 10Gb	2	Mellanox Switch SN2410	2	동일
		Bandth Width (GB)		10		25	2.5 배 많음

1) PassMark Software (<https://www.passmark.com/>)로 PassMark AVG CPU Mark 수치로 비교

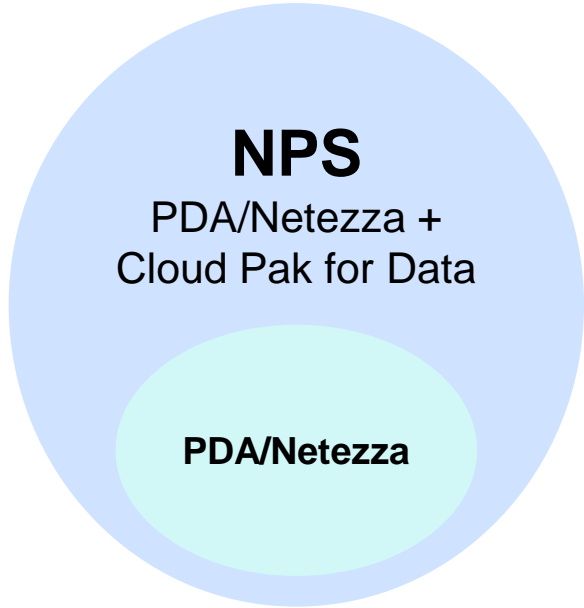
2) Xilinx 사의 칩셋별 DSP Slice (병렬 처리 수)로 성능을 비교함

3) Disk Spec에서 Sequential Read 값의 총합

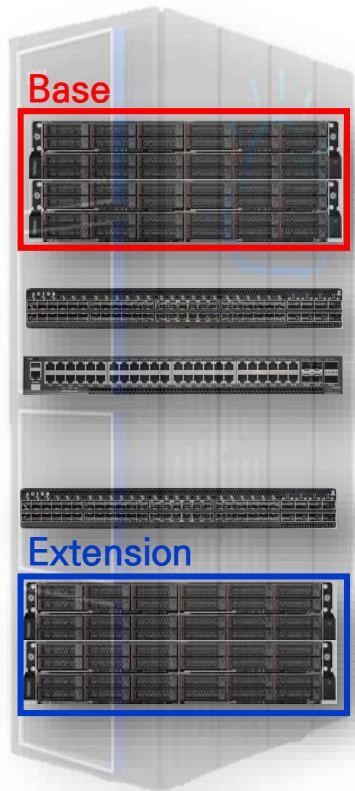
주요 변경 사항 (아키텍처 / 기능)

Netezza Performance서버는 PDA 및 Netezza의 기능을 모두 가지고 Cloud Pak for Data의 기능까지 추가된 제품입니다. DBMS 처리 아키텍처는 동일하며, H/W향상으로 성능은 높아졌습니다.

구분	파트	PDA N3001	Netezza Performance Server	비고
HW 아키텍처	Host 서버	두대의 서버가 DRBD (디스크 미러링)로 구성	세대의 서버가 GPFS (디스크 공유)로 구성	기존 미러링 유지 관리 작업 불필요
	Disk	SAS2	NVMe	낮은 장애비율, 빠른 성능
		SPU와 분리된 스토리지 Enclosure	노드 일체형	Local 파티션의 데이터 전송 속도 향상
	네트워크	10G	25G	노드간 데이터 전송 속도 향상
	확장	불가능 (상위모델로 교체)	가능 (Enclosure단위)	
SW 아키텍처	파티션 구성	디스크별 하나의 Data Slice Data파티션 별 크기 200GB	디스크별 6~7개의 Data Slice Data파티션 별 크기 200GB	NPS의 경우 고성능, 고용량 디스크로 여러 개의 데이터 슬라이스로 분할, 파티션 크기 동일
	SW	NPS (Netezza Performance Server)	NPS (Netezza Performance Server)	동일한 엔진 사용
	Tool	Fluid Query, INZA, Netezza Replication, SQL Extension	Fluid Query, NPSA, NPS Replication, SQL Extension	100% 동일 (기능 이름만 변경)
	Bitness	Host : 32bit S-Blade : 32bit	Host : 32bit S-Blade : 64bit	S-Blade영역이 64비트로 변경
확장	연계 솔루션			동일
	가상화	Fluid Query	Fluid Query, Data Virtualization	
	Machine Learning	없음	Watson Machine Learning, Auto AI	
	거버넌스	없음	Watson Knowledge Catalog	
	분석 솔루션	없음	Watson Studio	



Netezza Performance Server 구성 및 확장



	Base	Base + 1	Base + 2	Base + 4	Base + 6
Performance Server Blades	4	4	8	16	24
Cores	64	64	128	256	384
Memory	0.76 TB	0.76 TB	1.5 TB	3 TB	4.5 TB
Available User Space^{1,2,3}	17 TB	17 TB	34 TB	68 TB	102 TB

	Base + 8	Base + 10	Base + 12	Base + XX	Base + 28...
Performance Server Blades	32	40	48	...	112
Cores	512	640	768	...	1,792
Memory	6 TB	7.5 TB	9 TB	...	21.2 TB
Available User Space	136 TB	170 TB	204 TB	...	476 TB

Base +10 and larger Offering Notes

- Solutions requires additional fabric switches
- Solutions require additional management switches
- Solutions require the two (2) new spine switches
- Solution requires additional racks & PDU's
- Assume custom build times

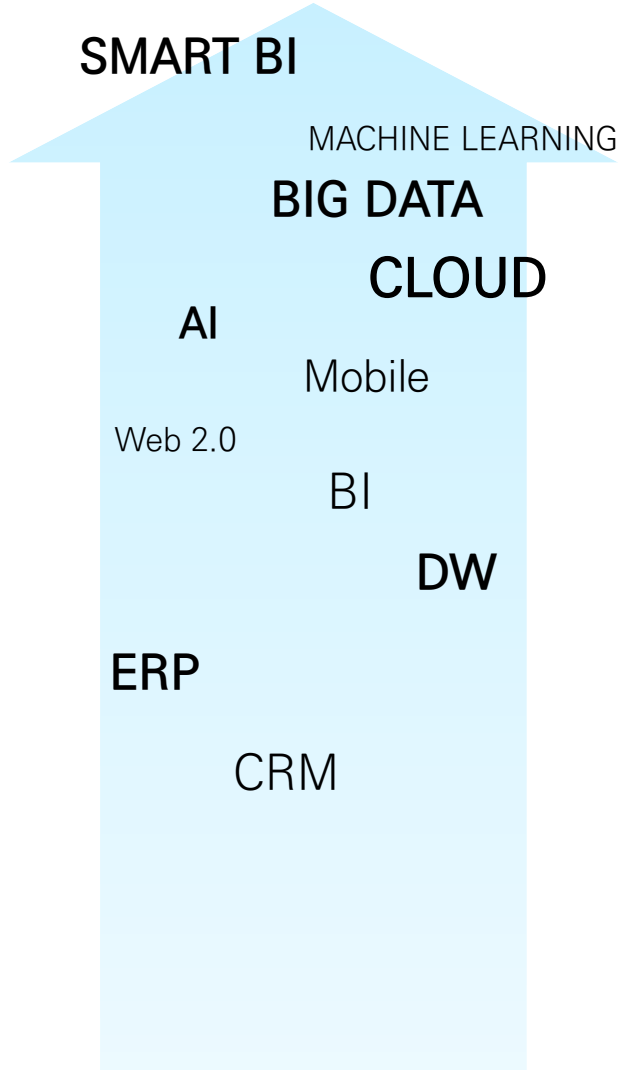
Netezza Performance Server Base+28 서버 구성 예시

Rack U	Rack 1	Rack 2	Rack 3	Rack 4	Rack 5	Rack 6
42						
41						
40						
39						
38						
37						
36	13U Blank		10U Blank			
35						
34						
33						
32			Hub Switch			
31			Spine Switch B			
30			Spine Switch A			
29	Management Switch 2A		1U Blank			
28	Management Switch 1A	27U Blank	Management Switch 3A	Management Switch 4A	Management Switch 5A	Management Switch 6A
27	1U Blank		1U Blank		1U Blank	
26	Fabric Switch 1B		Fabric Switch 2B		Fabric Switch 3B	
25	Fabric Switch 1A		Fabric Switch 2A		Fabric Switch 3A	
24						
23						
22						
21						
20	9U Blank		9U Blank	12U Blank	9U Blank	12U Blank
19						
18						
17						
16						
15	Enclosure 5 (Expansion)	Enclosure 10 (Expansion)	Enclosure 15 (Expansion)	Enclosure 20 (Expansion)	Enclosure 25 (Expansion)	Enclosure 30 (Expansion)
14	1U Blank	1U Blank	1U Blank	1U Blank	1U Blank	1U Blank
13	Enclosure 4 (Expansion)	Enclosure 9 (Expansion)	Enclosure 14 (Expansion)	Enclosure 19 (Expansion)	Enclosure 24 (Expansion)	Enclosure 29 (Expansion)
12						
11	1U Blank	1U Blank	1U Blank	1U Blank	1U Blank	1U Blank
10						
9	Enclosure 3 (Expansion)	Enclosure 8 (Expansion)	Enclosure 13 (Expansion)	Enclosure 18 (Expansion)	Enclosure 23 (Expansion)	Enclosure 28 (Expansion)
8						
7	1U Blank	1U Blank	1U Blank	1U Blank	1U Blank	1U Blank
6	Enclosure 2	Enclosure 7 (Expansion)	Enclosure 12 (Expansion)	Enclosure 17 (Expansion)	Enclosure 22 (Expansion)	Enclosure 27 (Expansion)
5						
4	1U Blank	1U Blank	1U Blank	1U Blank	1U Blank	1U Blank
3	Enclosure 1	Enclosure 6 (Expansion)	Enclosure 11 (Expansion)	Enclosure 16 (Expansion)	Enclosure 21 (Expansion)	Enclosure 26 (Expansion)
2						
1	1U Blank	1U Blank	1U Blank	1U Blank	1U Blank	1U Blank

- Racks : 6
- Enclosures : 30
(Base : 2, Expansion : 28)
- Management Switches : 6
- Fabric Switches : 6
- Spine Switches : 2
- Hub Switches : 1

Netezza Performance Server 소개

Netezza/PDA의 차세대 모델 출시



3 세대 (2019~)

- ✓ OpenShift 기반의 Cloud Pak for Data
- ✓ Hybrid Cloud 기반의 Netezza Performance Server 출시
- ✓ NVMe 기술을 통한 디스크 I/O속도의 비약적 발전
- ✓ 4 S-Blades/Enclosure



Netezza Performance Server

2 세대 (2011~2018)

- ✓ PDA 어플라이언스
- ✓ Scales to 4 racks(768 TB)
- ✓ 7 S-Blades/rack



PureData System for Analytics N2001/N2002



PureData System for Analytics N3001

1 세대 (2003~2010)

- ✓ 세계 최초의 DW 어플라이언스 상용 버전 출시
- ✓ IBM Netezza 어플라이언스
- ✓ 12 S-Blades/rack



NPS® 8000 Series



NPS® 10000 Series



TwinFin™



TwinFin (N1000)



Cloud Pak for Data System 기반

Netezza Performance Server

Cloud in-a-box

Cloud Pak for Data System위에 올라가 Hyper-Converged 의 장점을 활용

Blazing-fast

오픈 소스를 이용해 만들어졌고 내장 하드웨어 가속(FPGA)을 통해 고성능 분석에 최적화

Simple

Load and go, 최소화된 관리와 튜닝

Scalable & elastic

multi-tenant, start small, 필요한 만큼 스토리지와 컴퓨팅 파워를 늘릴 수 있음.

Availability (Cloud Pak for Data System 기능)

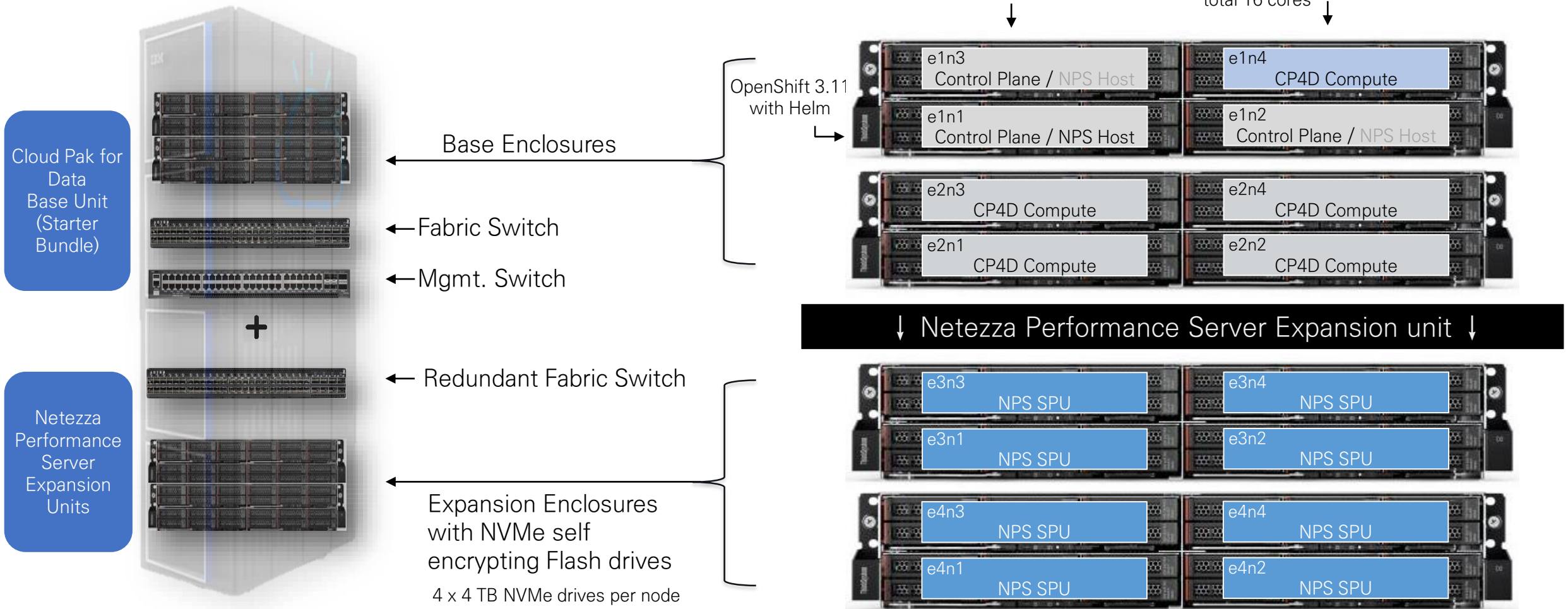
Red Hat Open Shift (Kubernetes) 컨테이너 기반 관리, highly available

Support & usage

IBM이 모두 지원

Netezza / Pure Data System과 100% 완벽하게 호환되는 어플라이언스 제품, 모든 nz * 유틸리티 호환, 데이터 통합 및 BI 도구와 Synergy, 매우 쉬운 기존 Netezza 시스템 이관 "fork-lift", 'nz_migrate'

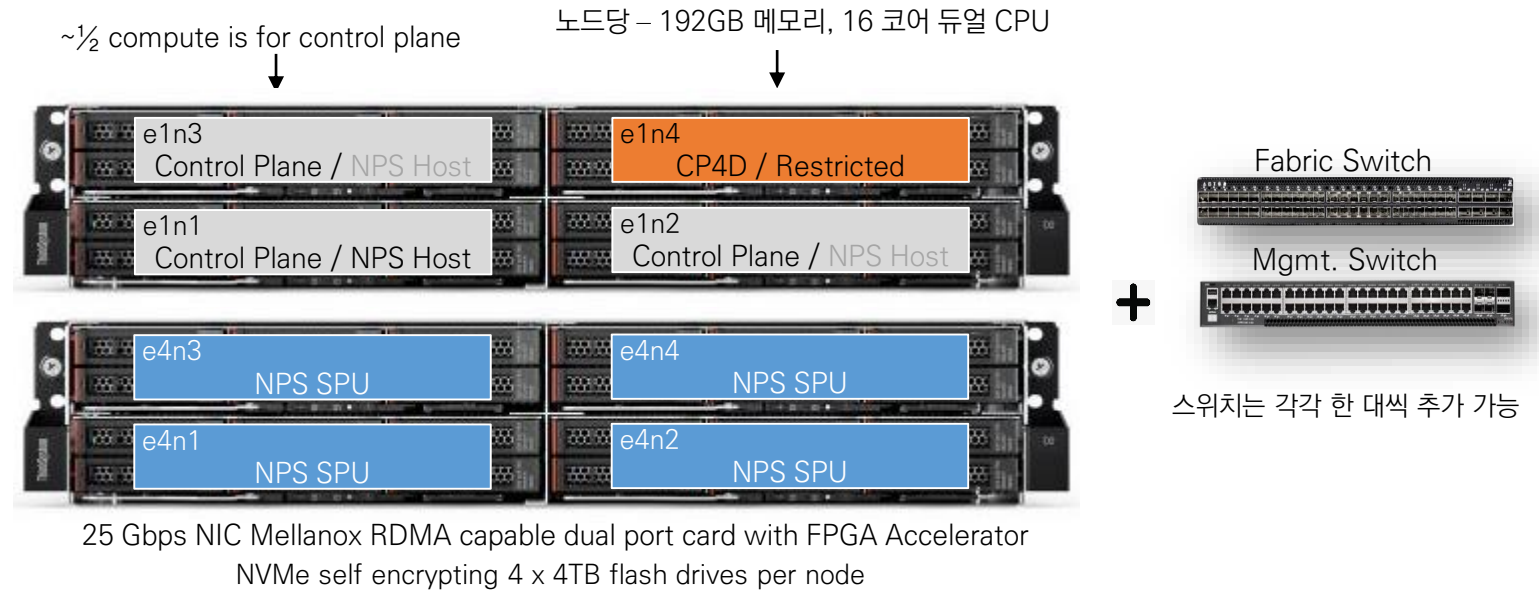
Netezza Performance Server System 아키텍처



25 Gbps NIC Mellanox RDMA capable dual port card with FPGA Accelerator

NPS Base Offering 기술 스펙




- NPS Base는 두대의 Base Enclosure중 한대에 NPS가 설치된 형태
- NPS Base와 NPS Base+1의 NPS노드 스펙은 동일
- 더 큰 모델로 확장 가능



	NPS Base
NPS 노드 개수	4
NPS노드 – 총 코어수 (VPCs)	64
NPS 노드 – 총 Memory	0.76 TB
NPS 노드 – 가용 User Space ¹	17 TB

¹4배 압축 가정, 사용자 데이터 = 4 x 17TB = 68TB

PDA vs NPS HA

PureData System for Analytics	Netezza Performance Server
2 front-end host nodes <ul style="list-style-type: none"> - Active / Passive - ha1 / ha2 	3 “front-end host” nodes <ul style="list-style-type: none"> - Active / Passive / Passive (NPS standpoint) - e1n1 / e1n2 / e1n3
NPS software installed on both nodes	NPS Docker container deployed on each node
DRBD volumes <ul style="list-style-type: none"> - /nz - /export/home 	 IBM Spectrum Scale / GPFS cluster <ul style="list-style-type: none"> - /opt/ibm/appliance/storage/NPS - /NPSHost1 - /export_home - /nz - /nzscratch
Linux-HA 	Platform Manager (Magneto) 
Virtual / Wall IP	Virtual / Wall IP

Less Admin, More Analytics



NO 수많은 관리자

데이터 로딩 후, 바로 쿼리 (Load & Go)

NO 데이터의 과도한 가공작업

내장된 데이터 마이그레이션 툴

NO 압축 관리

단일 관리 콘솔

NO 스토리지 관리

분석에 최적화된 성능

NO 물리모델링

하이퍼 컨버지드 확장 시스템

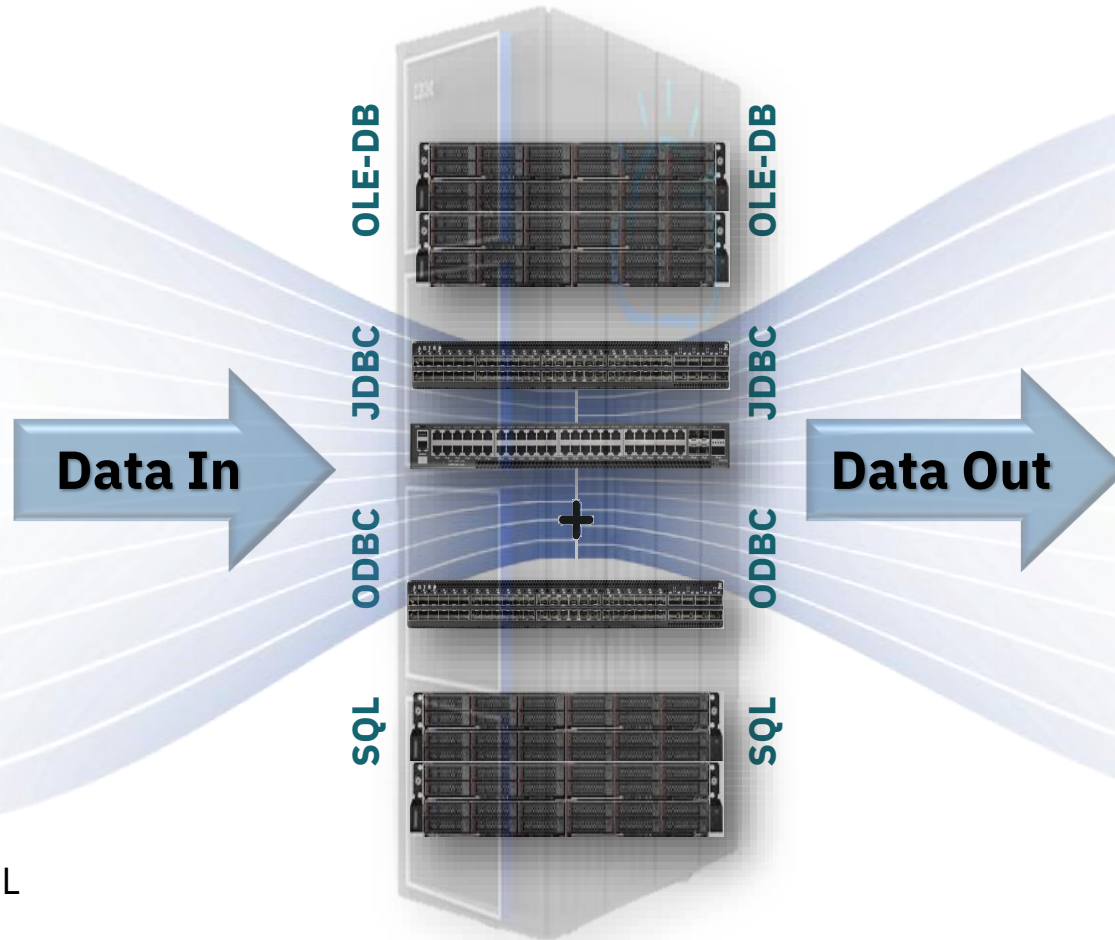
NO 인덱스 및 과도한 튜닝

H/W, S/W 단일 지원 조직

Less Admin, More Analytics

데이터 통합¹

- IBM
 - Information Server
 - InfoSphere Streams
- Ab Initio
- Hadoop
- Informatica
- Microsoft
- Oracle
- SAP
- SAS
- Others using standard ODBC/JDBC/OLE-DB/SQL



리포팅 및 분석¹

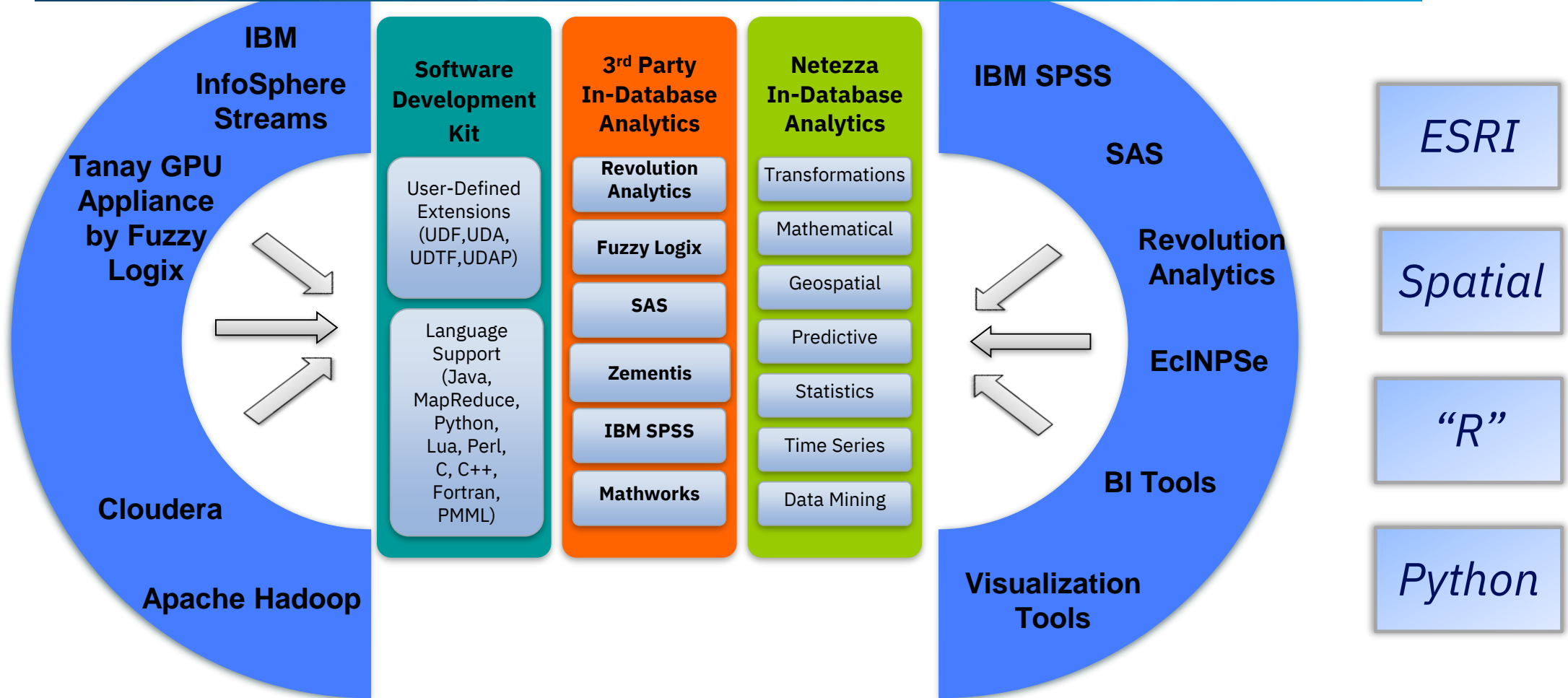
- IBM
 - Cognos
 - SPSS
 - Campaign
- Hadoop
- Information Builders
- Microsoft
- MicroStrategy
- Oracle
- SAP
- SAS
- Tableau
- Others using standard ODBC/JDBC/OLE-DB/SQL

¹Netezza 및 PDA와 100% 동일

Embedded IBM Netezza Analytics¹

In database 분석 – in place 머신 러닝을 위한 핵심

ESRI, Geo Spatial and Revolution



¹Netezza 및 PDA와 100% 동일

내장된 Data Science 와 머신러닝

Powered by state of the art Data Science Hardware Performance Accelerator

데이터의 이동 없는
Machine Learning
+
High Speed
데이터 가속기
(FPGA)

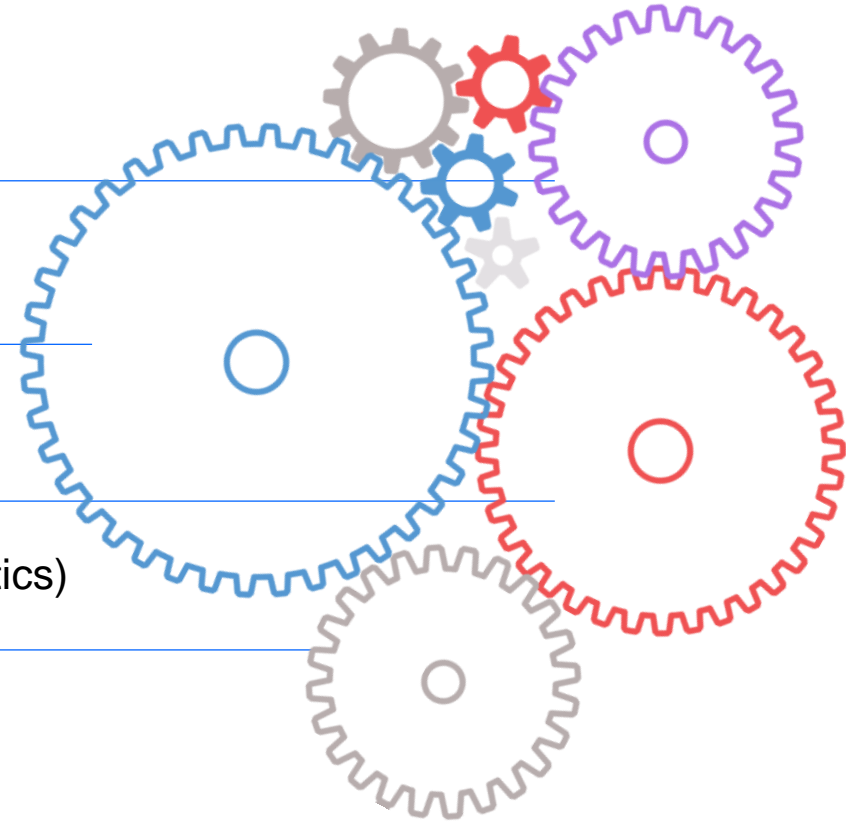


Watson Studio내장

모델 개발 도구와 통합

INZA (In-database analytics)

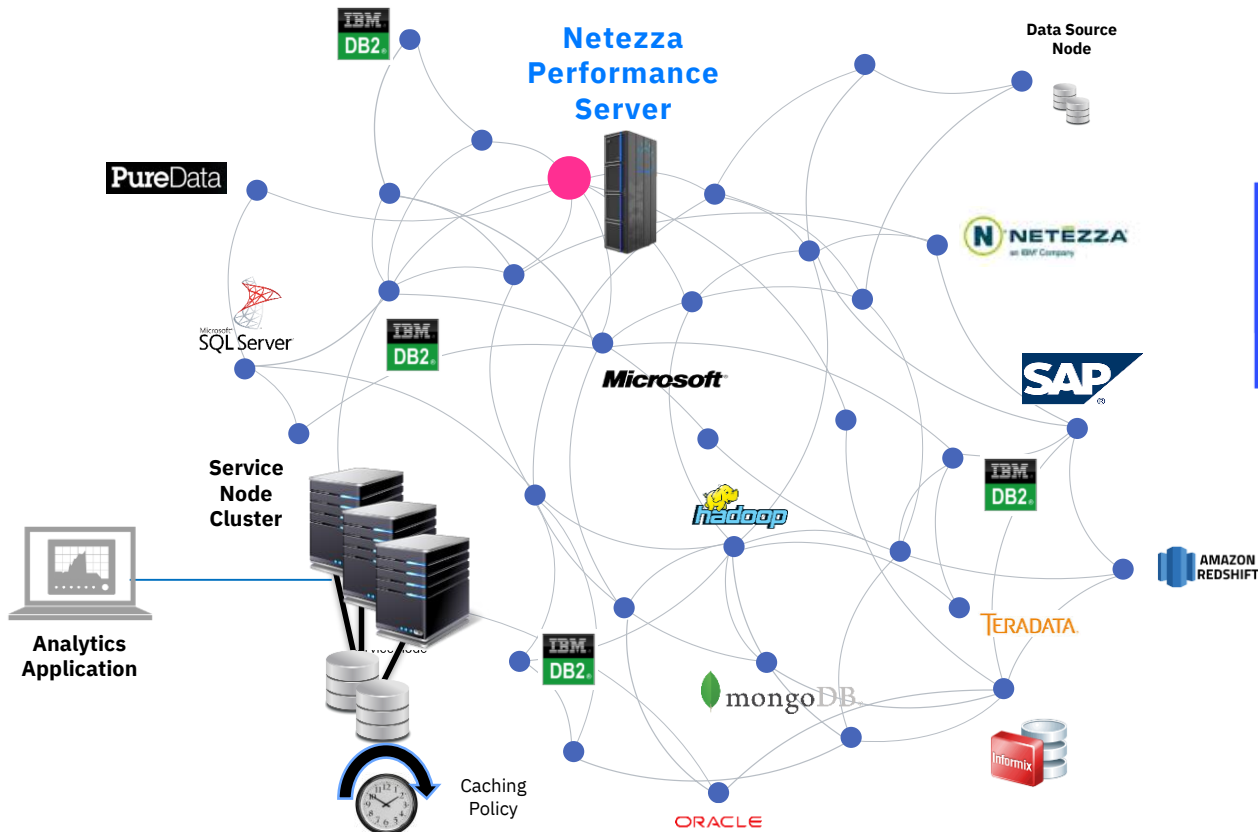
BI 및 시각화 도구와 통합



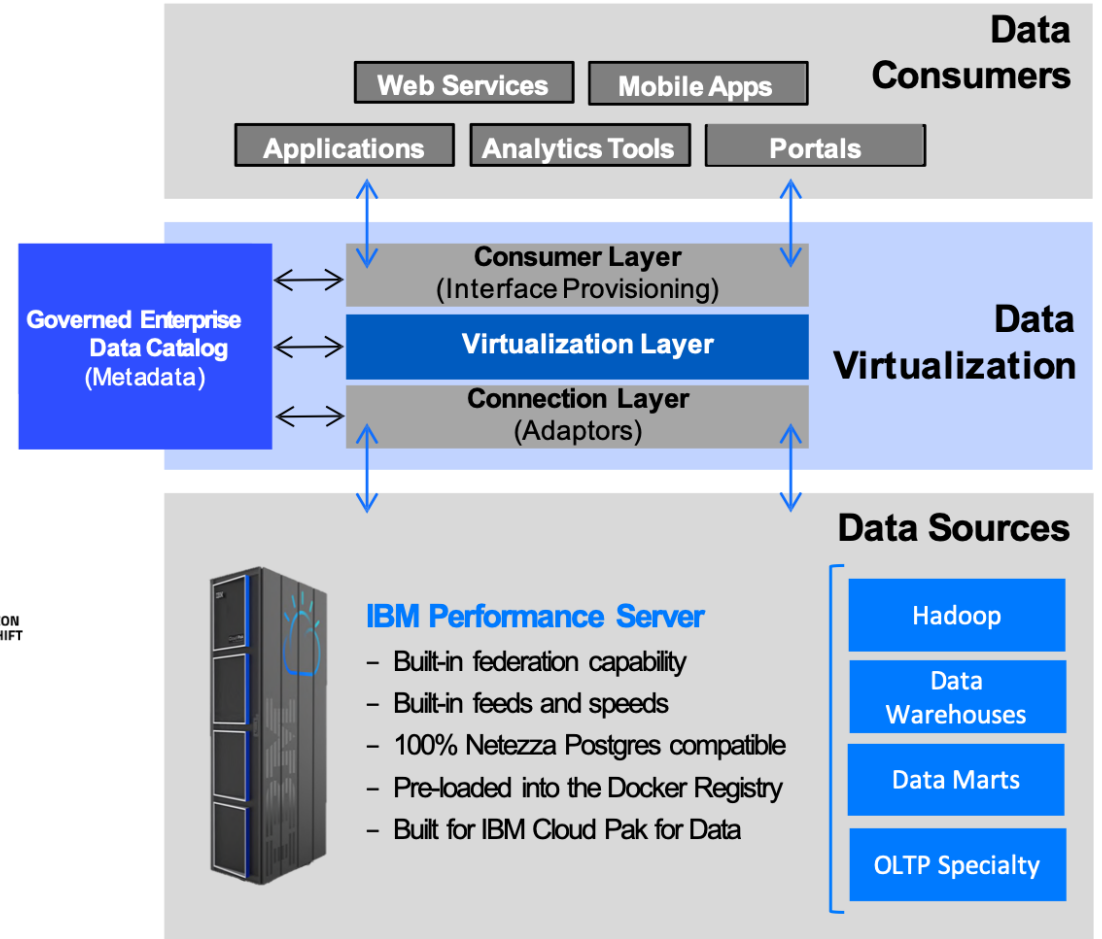
데이터베이스에서 분석작업을 실행할 수 있어 더 많은 가치 창출
툴과 고급 알고리즘을 사용하는 머신 러닝 모델이 내장
데이터 과학자 및 분석 전문가가 선호

데이터 가상화

- 고성능 / 엔터프라이즈 급
- 이기종 데이터 연계



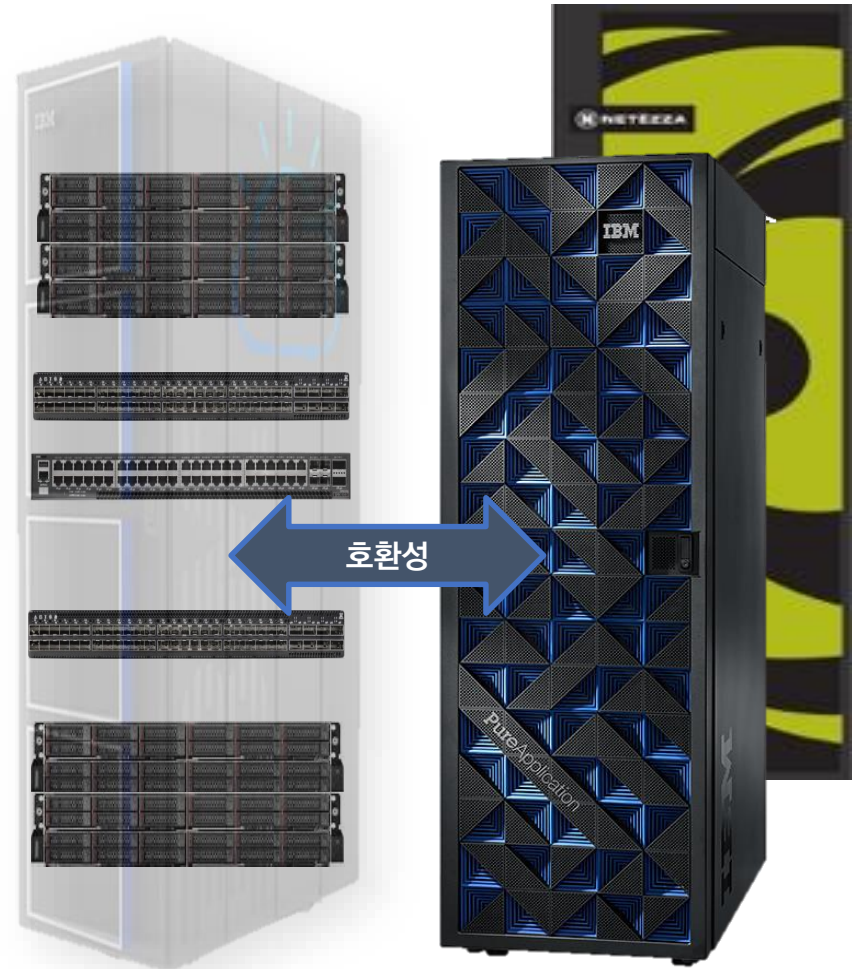
- 셀프서비스를 위한 단일 데이터 소스 구성
- 기존 Fluid Query와 연계 가능



**Zero-Cost
Migration
from PDA/Netezza to NPS**

Netezza Performance Server – Native Netezza 호환성

- **단순성 (Simple)**
 - Load and Go, 관리 및 튜닝 최소화
- **호환성 (Compatible)**
 - Netezza (Mustang/Skimmer/TwinFin/Striper/Mako) 호환성
 - "Fork-lift" existing Netezza/PDA systems
 - 'nz_migrate'를 이용한 손쉬운 마이그레이션
 - 기존 UDFs / UDXs 동작 (재 컴파일 필요)
 - 양방향 D/R 복제, 과거 Netezza/PDA와 BnR(Backup & Restore) 가능
- **내장된 FPGA DB 가속기를 통한 활용**
 - Data Stream processing
 - Machine Learning
 - In-place Advanced Analytics & Geo Spatial
- 유사한 인터페이스 및 툴



마이그레이션의 간편성

기존 UDF/UDX
그대로 사용 가능

nz_backup 유틸리티를 통해
백업/복구 방식으로 전환 가능

기존 Zone maps 유지를 통한
성능 보장

데이터 변환 등
다른 절차 불필요

nz_migrate 유틸리티 제공

이관과 동시에
1차 검증(cksum) 가능

이행
절차의
단순화

전환
유틸리티
제공

무결성 및
성능
보장

기존 PDA와 100% 호환

초고속 데이터 이관 사례

“ 기존 IBM PDA로 수행하던 고객/상품 분석 업무 사용량이 증가하여 시스템 증설을 고민하다가 그동안 IBM PDA가 보여준 활용성과 성능 때문에 Netezza Performance Server를 선택했습니다.”

온라인 상거래 통합 DW

- 회사 내부의 분산된 정보계 시스템을 통합
- 영업, 마케팅, 상품 개발 등 모든 부서가 사용하는 분석 시스템
- 전사레벨의 영업 실적과 고객서비스 같은 주요 지표 제공

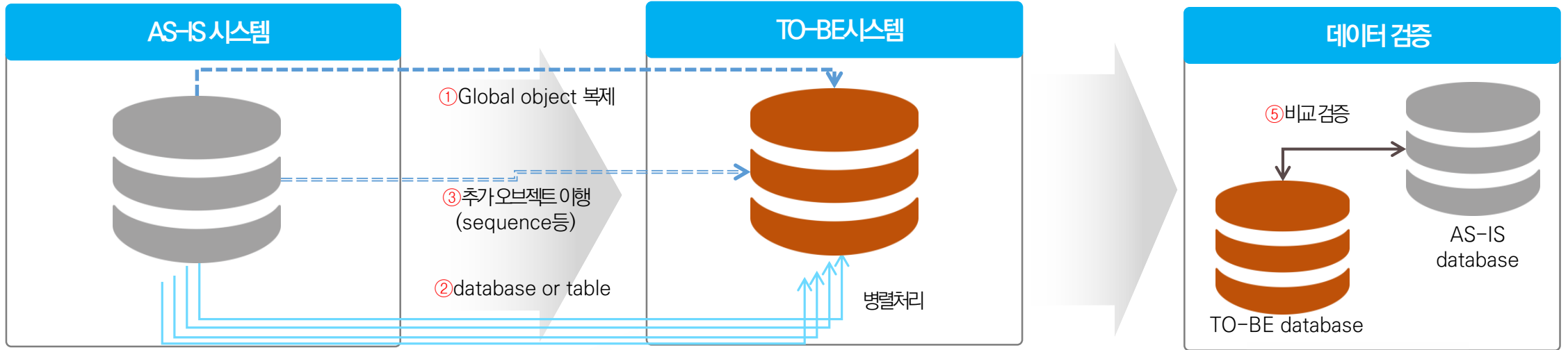
Netezza Performance Server Base+2

- 기존 PDA (N3001-002) 이관
- 데이터 사이즈 : 6TB
- 연계 솔루션
 - ✓ InfoSphere DataStage
 - ✓ Cognos Analytics
 - ✓ Aginity Netezza Workbench

초 스피드 이관

- 2020년 4월 (5일, 실제 이관 작업은 1일)
- 전용 이관 툴(nz_migrate)
- 검증은 건수 검증만 수행
- 연계 솔루션은 IP만 변경
- 평균 10배 속도 향상 (기존 PDA 대비)
- 기존에 성능 이슈로 인해 사용하지 않던 분석 장표 사용
- 사용자 업무 만족도 증가

이행 방식 1 (전용 유틸리티 사용)



운영시스템 중지 & 이행 시작

④이행 결과자동 검증

⑥데이터슬라이스 확인

1차데이터검증

- 마이그레이션 수행시 cksum 옵션으로 자동화된 데이터 검증 수행
- 마이그레이션 완료 후 로그 파일 분석 후 에러 도출

```

ERROR: nz_cksum's differ!

After each table is migrated, an (optional) cksum is calculated against
both the source + target tables, and then compared. For some reason
the checksum's don't match. The tables should be analyzed + compared
further to try to determine why this occurred.

ERROR: Reload column count mismatch.
# i.e., the total number of columns does not match
ERROR: Reload column type mismatch.
# i.e., the column data types do not match
# or the null/not null constraints do not match
ERROR: Reload distribution algorithm mismatch.
# i.e., the DISTRIBUTE ON clause does not match

These errors occur when you are doing a "-format binary" migration
between the two systems, and the definition of the target table does
not match that of the source table. They must match.

Note: The "nz_ddl_diff" script can be used to compare two table's DDL.
    
```

```

[nz@cs-spubox6 ~]$ nz_zonemap TPCDS40B INVENTORY INV_DATE_SK

Database: TPCDS40B
Object Name: INVENTORY
Object Type: TABLE
Object ID : 560485
Data Slice: 1
Datatype 1: INV_DATE_SK (INTEGER)

ZoneMaps are enabled on this Table/MView

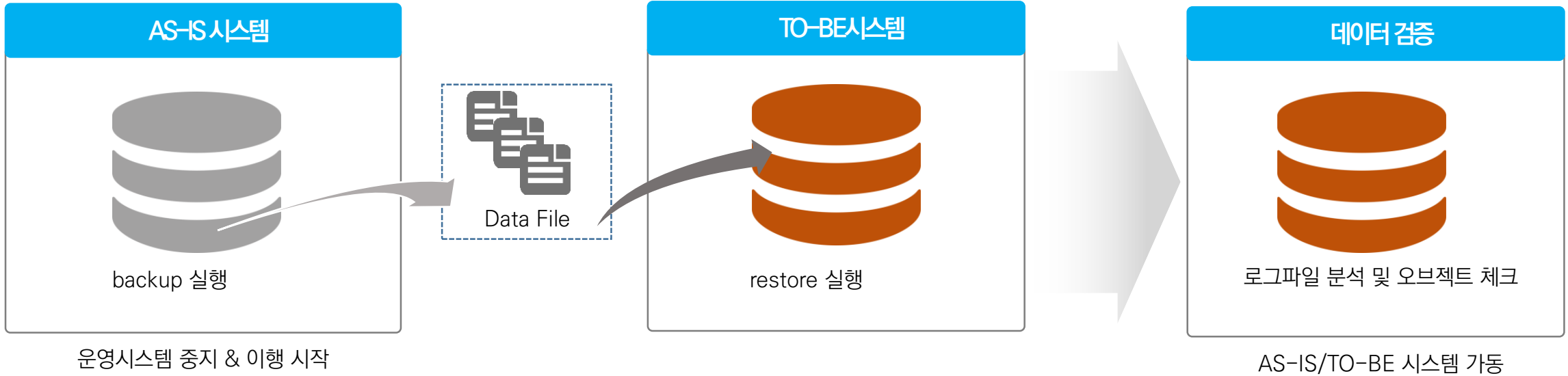
Extent # | INV_DATE_SK (Min) | INV_DATE_SK (Max) | ORDER BY
-----|-----|-----|-----
0 | 2450749 | 2450749 |
1 | 2450749 | 2450751 | TRUE
2 | 2450751 | 2450753 | TRUE
3 | 2450753 | 2450753 | TRUE
4 | 2450753 | 2450756 | TRUE
5 | 2450756 | 2450758 | TRUE
6 | 2450758 | 2450760 | TRUE
7 | 2450760 | 2450760 | TRUE
8 | 2450760 | 2450763 | TRUE
9 | 2450763 | 2450765 | TRUE
    
```

데이터 최적화

- groom 작업으로 테이블 재구성
- nz_zonemap 유틸리티를 통해서 AS-IS 와 TO-BE 테이블의 데이터 슬라이스 스캔
- 스캔 결과 데이터 정렬 후 데이터 재 적재
- 스캔 자동화 shell 생성 제공

AS-IS/TO-BE 시스템 가동

이행 방식 2 (백업 복구)



데이터 오브젝트 한번에 이행 **백업 시 저장 공간 확보 필요** **데이터 검증의 최소화**

**백업/복원
유틸리티**

nzhostbackup/ nzhostrestore	Netezza 호스트 데이터 (카탈로그 메타 데이터)
nzbackup / nzrestore	<ul style="list-style-type: none"> • 사용자 데이터베이스 • Netezza 사용자, 그룹 및 권한
nz_backup / nz_restore	<ul style="list-style-type: none"> • Table별 백업/복원 • multi-thread, 압축/비 압축 방식 지원

UDX이행

외부 라이브러리 변경 없이 활용 가능

	AS-IS	TO-BE
C 소스 -> C++ 변환	✓	
공유 라이브러리 생성	✓	✓
UDX 컴파일	✓	✓
라이브러리 배포	✓	✓
컴파일&배포 스크립트 제공	✓	✓

외부 라이브러리 변환으로 인한 불안감 해소

```

1 #include <stdio.h>
2 char compare(char *str1, char *str2);
3 int main() {
4     char str1[20];
5     char str2[20];
6     scanf("%s", str1);
7     scanf("%s", str2);
8
9     if (compare(str1, str2)) {
10        printf("s1 = %s > s2 %s\n", str1, str2);
11    } else {
12        printf("s1 = %s < s2 %s\n", str1, str2);
13    }
14    return 0;
15 }
16
17 char compare(char *str1, char *str2) {
18     while (*str1) {
19         if (*str1 != *str2) {
20             return 0;
21         }
22         str1++;
23         str2++;
24     }
25     if (*str2 == '\0') return 1;
26     return 0;
27 }
28
29
30
31
                
```

```

// extends UDF base class
class My_UDX : public Udf {
public:
    Rconnection * rc;

    My_UDX(UdxInit *pInit) : Udf(pInit) {
        initsocks();
        // IP address in place of ...
        rc = new Rconnection("...", 6311);
    };

    ~My_UDX() {
        delete rc;
    }

    static Udf* instantiate(UdxInit *pInit);

    virtual ReturnValue evaluate();
};

Udf* My_UDX::instantiate(UdxInit *pInit) {
    return new My_UDX(pInit);
}
                
```

import

```

gcc -m64 -fPIC -c xxxx.c -o
/nz/lib/udxclib_spu.o
                
```

C++ Compile

```


zudxcomple [options] srcfile
                
```

UDX Compile

```

nzsqli -d db -c
"create or replace library ...
                
```

UDX 배포



NPS on Cloud

IBM Netezza on Cloud



Easy to get started

프로비저닝, 확장,
유지 관리를 위한
최소한의 작업



Runs Netezza

Netezza
업그레이드를 위한
확실한 선택



Scalable

독립적으로 컴퓨팅 및
스토리지 확장



Highly available

Red Hat
OpenShift를 사용한
관리형 인프라



Secure

데이터 암호화를
사용한 개인 배포



Reliable

가용영역(AZ)간 백업
및 복제

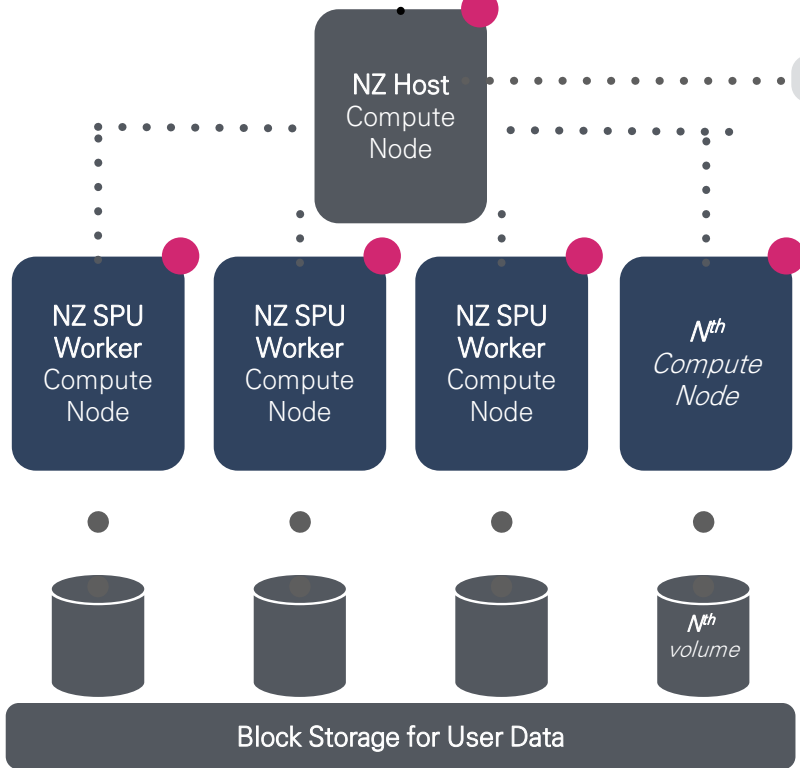
Netezza Performance Server
데이터웨어 하우스 DB엔진, 웹 콘솔

IBM Cloud Pak for Data
카탈로그 및 제어 센터가 있는 통합 데이터 및 AI 플랫폼

Red Hat OpenShift
엔터프라이즈 쿠버네티스 애플리케이션 플랫폼

3
OpenShift
Compute
Nodes (HA)

3
Cloud Pak
Compute
Nodes (HA)



Netezza Performance Server on Cloud Architecture

Cloud Native

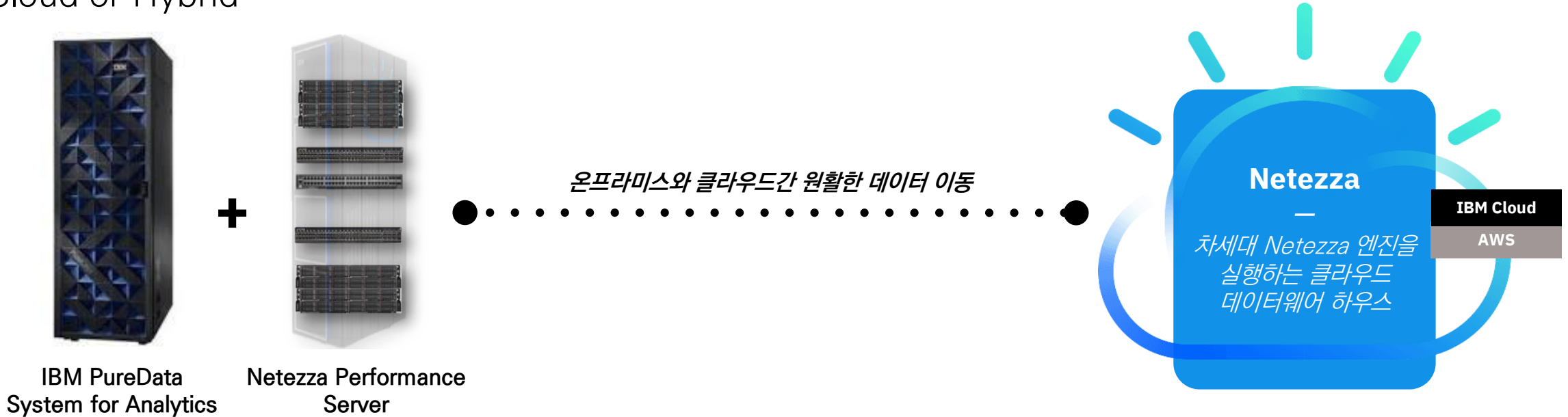
어느 환경에도 손쉽게 이식 및 이관

HW와 무관한
어플리케이션 중심 컨테이너 환경

Managed by Red Hat OpenShift

Netezza Performance Server on Cloud

All Cloud or Hybrid



- **모든 클라우드 지원** : nz_migrate 또는 nzbackup / restore를 이용해 어느 환경이라도 시프트 가능
- **개발/테스트 환경** : nzbackup으로 Object Storage에 데이터를 생성해 클라우드로 쉽게 이전
- **클라우드 데이터 마트** : 특정 현업 부서를 위한 데이터 세트 현지화
- **프라이빗 클라우드 구축을 위한 재해 복구** : 오브젝트 스토리지에 백업데이터를 업로드하고 클라우드 데이터웨어 하우스로 복원하고, 재해가 발생한 경우 전환

Netezza Performance Server

고객의 소리 및 Value

고객의 소리

“ IBM Netezza Performance Server를 사용하여 주요보고 작업을 **45 시간에서 12 초로 단축**했습니다. ” – 북미 은행고객

“ 연방 CCAR [종합 자본 분석 및 검토] 스트레스 테스트의 일환으로 매우 무거운 분석 작업을 수행하는데, 이전에는 다양한 도구를 사용하여 분석 서버로 데이터를 가져 왔습니다. 이제 Netezza Performance Server를 사용하여 **데이터 이동 없이 처리 시간을 며칠에서 몇 분으로 단축**했습니다. ”
– 북미 은행고객

“ 새로운 Netezza Performance Server로 업그레이드하였습니다. 마이그레이션은 간단하고 문제가 전혀 없었으며, **시작하고 실행하는데 단 7주**가 걸렸으므로 인사이트 및 의사 결정에 이르는 시간을 단축 할 수 있었습니다. ” – 북미 대형 자동차 회사

“ 다양한 클라우드 옵션을 살펴 보았지만 비용을 예측할 수 없다는 것이 주요 문제였습니다. Netezza Performance Server를 사용하면 **적절한 예산을 책정** 할 수 있으며 여러 항목이 청구되거나 갑작스러운 청구에 대해 걱정할 필요가 없습니다. ”



“ Netezza on Cloud는 데이터 센터에 새로운 공간 없이도 말도 안되게 빠른 속도를 제공했습니다. **ELT 및 프런트 엔드 SQL을 변경없이 그대로 옮겼지만 10배 더 빠릅니다.** Netezza를 사용하면 큰 비용이 들지 않습니다. ” – 서비스 공급자

“ 기존 환경에서 몇 시간이 걸렸던 보고서는 Netezza Performance Server에서 단 몇 분 만에 실행되어 **2~3배 정도 더 빨라졌습니다.** Netezza Performance Server로의 업그레이드는 전혀 걱정이 없었습니다. ”

“ Netezza Performance Server를 사용하여 엄청난 양의 데이터를 집계하는 통합 데이터 보고서를 모니터링하여 품질 문제와 마진 누출을 실시간으로 파악할 수 있습니다. **이전에는 실행하는 데 몇 시간이 걸렸던 보고서를 이제 단 몇 초** 만에 실행할 수 있습니다. ”

“ Netezza Performance Server에서 본 성과와 밸류가 정말 마음에 듭니다. nz_migrate라는 간단한 명령만으로 기존 테이블, 데이터 및 저장 프로시저를 Netezza를 새로운 Netezza Performance Server에 추가하였는데, 이 **모든 작업을 저희가 직접 수행했습니다.** 간단합니다. ”

Netezza Performance Server의 Value



¹⁾ 4배 압축 기준

Cloud Ready

- 컨테이너 기반 RedHat OpenShift 아키텍처를 사용하여 관리가 용이하고, 유연한 서비스 추가 변경이 가능함
- 퍼블릭 클라우드의 Netezza On Cloud 로 손쉽게 이전 가능하며, 향후 BYOL (Bring Your Own License)로 비용부담 없이 이관 가능

정보 통합 솔루션

- AI를 위한 IA (Information Architecture, 정보의 구조화)를 위한 Cloud Pak for Data 탑재
- 데이터 가상화, 데이터 거버넌스 등의 기능 활용 가능

성능

- 기존 PDA대비 3배 이상 빠른 성능
- 최근 국내 PoC에서 타사 대비 2배 이상 빠른 성능 보여줌

확장 및 지원

- In-Place 확장
- HW/SW 일체형으로 IBM이 모든 파트 및 이슈에 대해 일괄 지원하는 일원화된 지원 체계

호환성

- 100% 기존 PDA 쿼리와 호환
- 모든 유틸리티 (nz_*) 동일 및 상호 연결 가능
- 연계 조회 (Fluid Query)
- 기존 PDA를 개발 서버로 두고 NPS를 운영서버로 두는 형태의 아키텍처 가능

마이그레이션

- Zero 마이그레이션 비용
- 검증이 필요 없는 마이그레이션 (DDL, 데이터 및 정렬 순서까지 동일하게 이관)
- 기존 PDA로 부터 시간당 최대 16TB ¹⁾의 속도로 완벽한 데이터 이관
- 위험 부담 0

IBM

