

Data and AI Virtual Forum Korea

데이터웨어하우스사용 고객을
위한 최신의 데이터 통합
플랫폼

—

Data and AI Summit

최석재 실장

Client Technical Professional

Data & AI 플랫폼 최신 동향

AI 통합 플랫폼 구현을 목표로 하고 있는 고객들은 **데이터와 AI가 상호 유기적으로 통합되어 있는 환경을 요구**



플랫폼 현대화

클라우드 기반 쿠버네티스

- 디지털 혁신을 위한 **빠른 프로비저닝과 유연한 확장**이 가능한 플랫폼
- 환경 변화 대응을 위한 **하이브리드 멀티 클라우드** 지원
- 다양한 요구 사항을 수용할 수 있는 **컨테이너 기반 플랫폼**



데이터 관리

컴퓨팅과 스토리지 분리

- 분석에 필요한 데이터의 Hub 역할을 위한 **정형/비정형을 포함한 데이터 저장소** 구축
- **컴퓨팅과 스토리지 영역을 분리**하여 업무 증가 및 워크로드에 따른 유연한 확장성 제공



데이터거버넌스

데이터 카탈로그 셀프서비스

- 데이터 생산자와 사용자간 협업을 위한 **DataOps** 지원
 - * **DataOps** : 신뢰할 수 있는 데이터를 비즈니스에 즉시 제공하기 위한 프로세스 및 기술
- 신뢰 있는 데이터 제공을 위한 **거버넌스** 체계
- 셀프 서비스 분석을 위한 **데이터 탐색** 환경



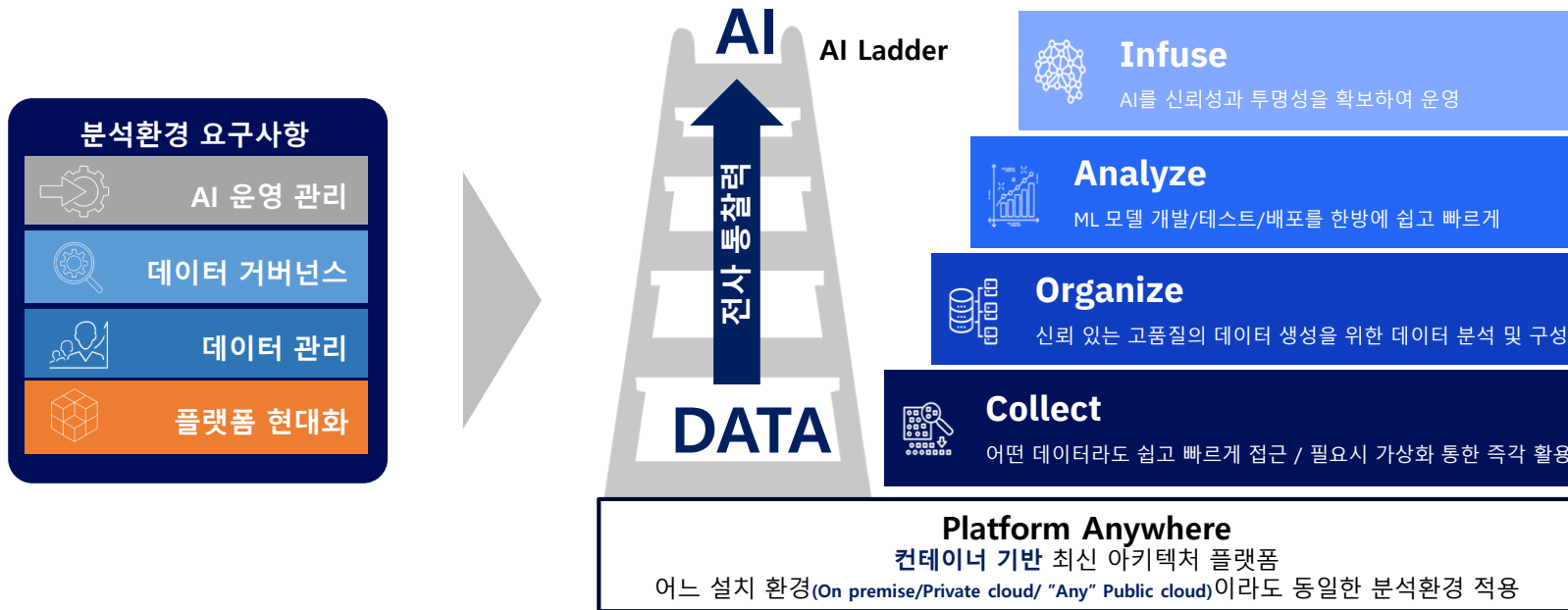
AI 운영관리

데이터사이언스 및 AI 운영관리

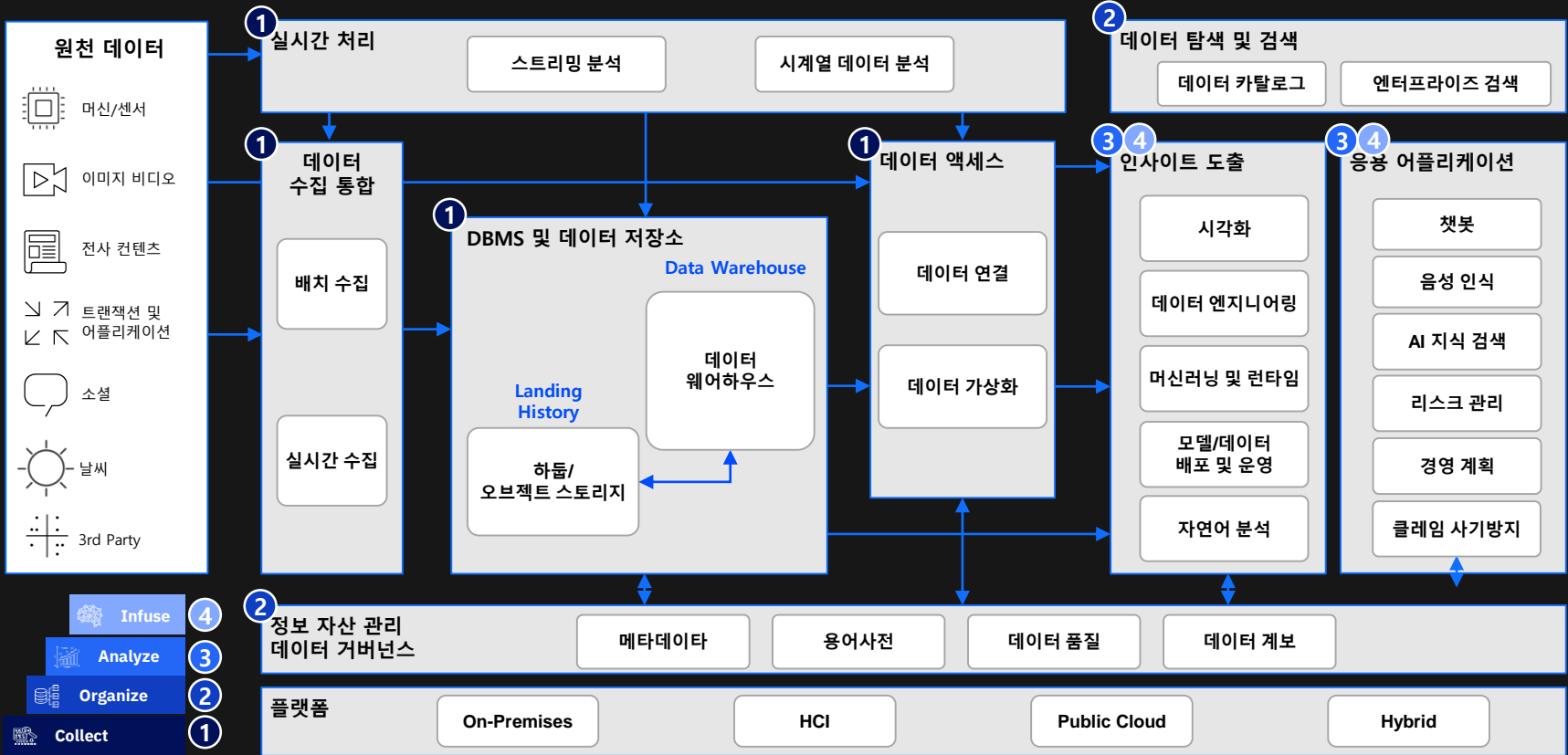
- 분석 모델 개발, 테스트, 배포, 모니터링을 포함한 **AI관리 기술 도입**
- 개방성과 최신 트렌드 반영을 위해 **오픈소스 기반 AI 기술 반영**
- 분석 모델 개발의 효율성을 높이기 위한 **Auto AI 기술 활용**

Data & AI 플랫폼 최신 동향

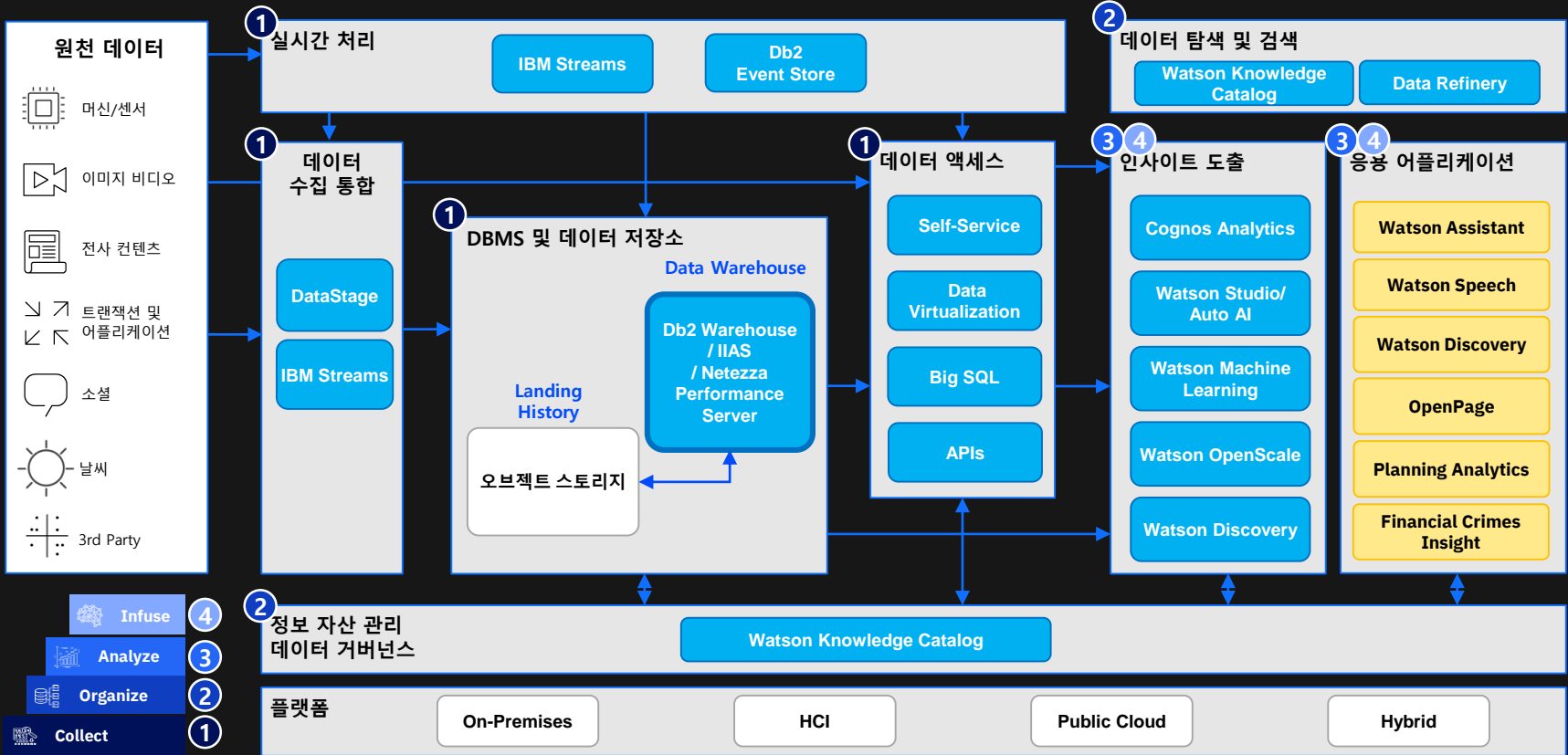
“정보 아키텍처(IA) 없이는 AI를 성공적으로 수행할 수 없다”
“There’s no AI without IA(Information Architecture)”



DATA&AI 참조 아키텍처



DATA&AI 참조 아키텍처



SOFTWARE

Cloud Pak for Data

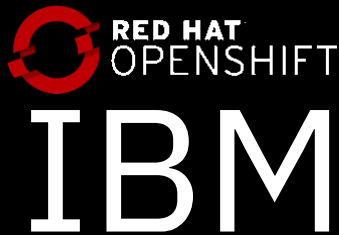
+

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 블록 처럼 구성할 수 있는 데이터 센터





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 아키텍처 개요

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**

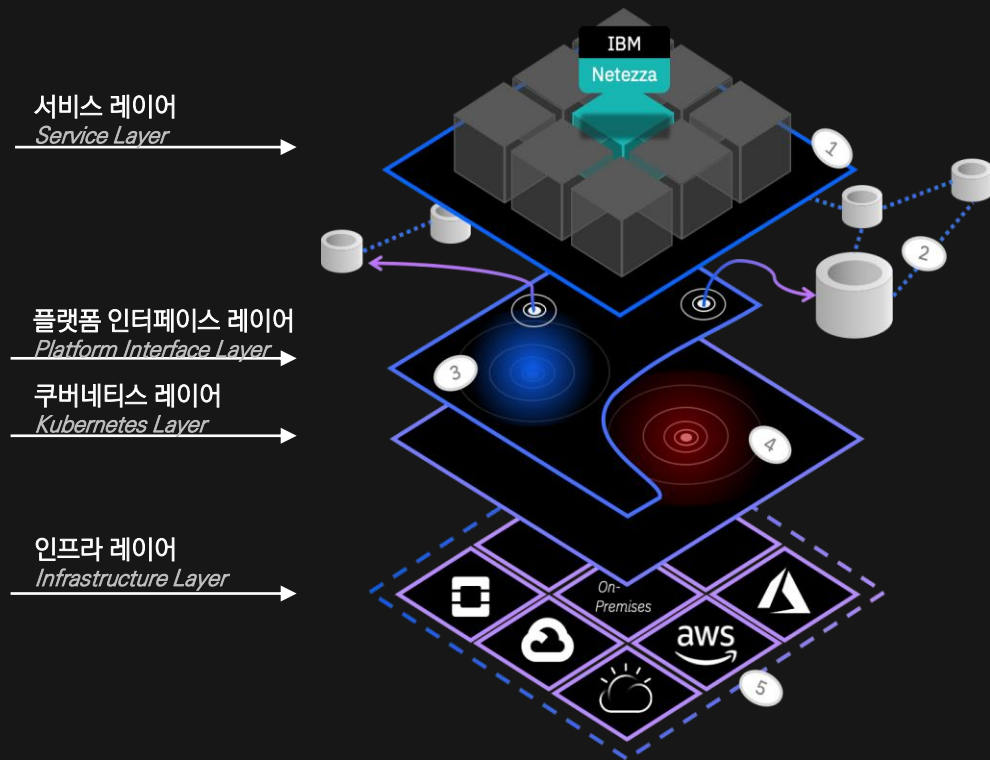
데이터 관리, 거버넌스, 분석등의 작업을 통합된 UI 및 인터페이스로 이용

4. Red Hat **OPENShift**

컨테이너 기반의 하이브리드 클라우드 환경 제공

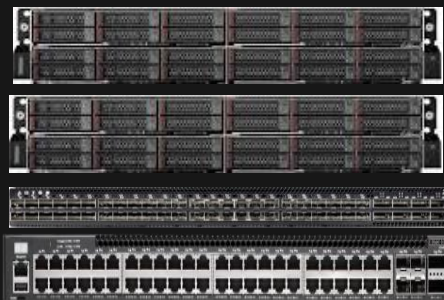
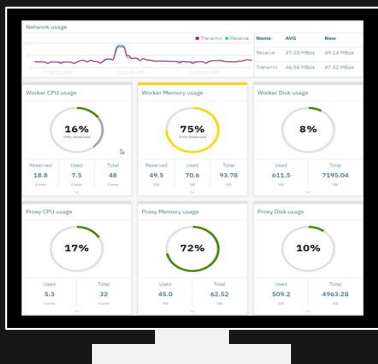
5. Any.Cloud

벤더에 종속되지 않은 멀티클라우드 환경 구축



Netezza Performance Server 구성 단위

Base (Starter 유닛) + Extension (4개 S-Blade 단위)



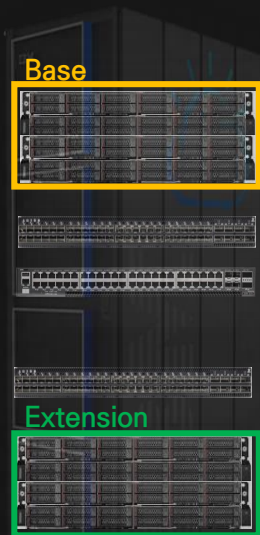
Netezza Performance Server – Base (2 Enclosure)

- 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 구성 및 확장



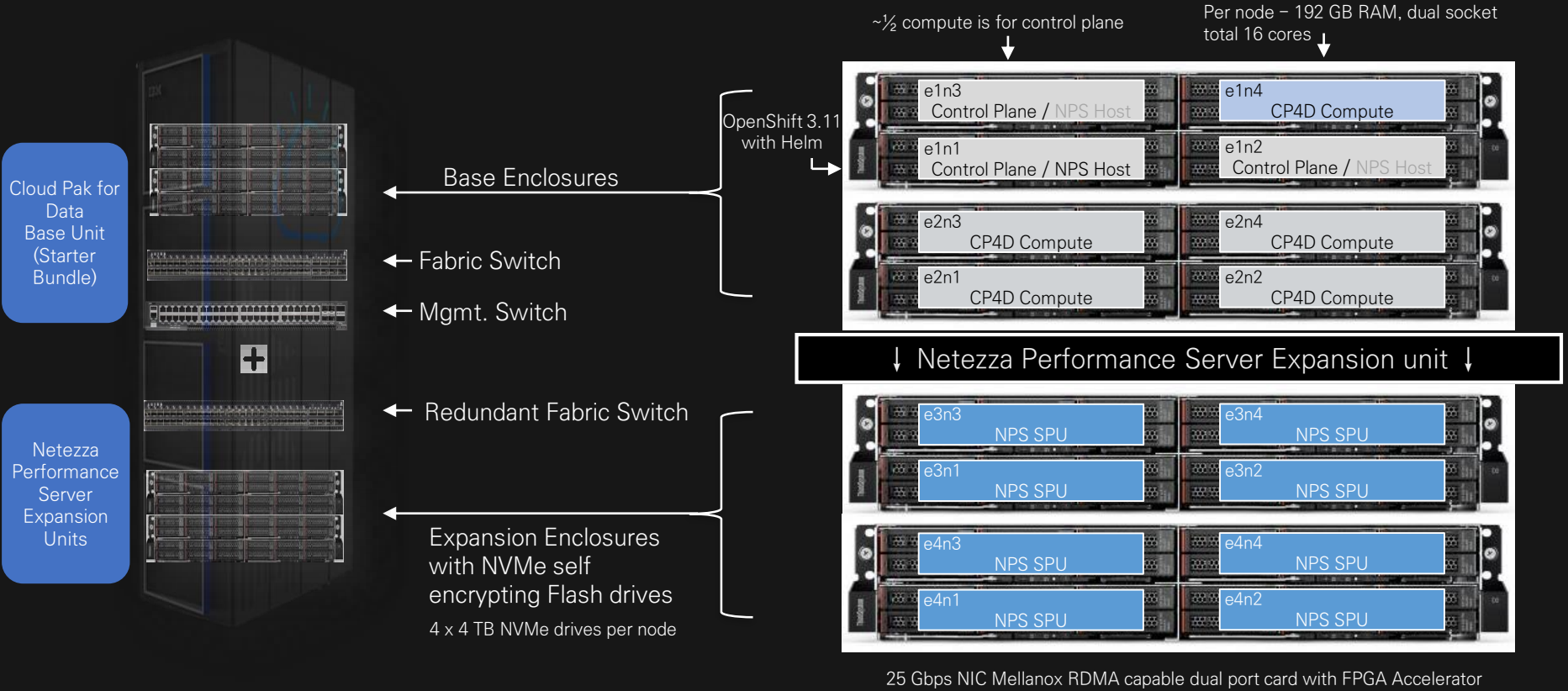
	Base	Base + 1	Base + 2	Base + 4	Base + 6
Performance Server Blades	4	4	8	16	24
Cores	64	64	128	256	384
Memory	0.76 TB	0.76 TB	1.5 TB	3 TB	4.5 TB
Available User Space^{1,2,3}	17 TB	17 TB	34 TB	68 TB	102 TB

	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 System 아키텍처



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

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

“ 새로운 Netezza Performance Server로 업그레이드하였습니다.
마이그레이션은 간단하고 문제가 전혀 없었으며,
시작하고 실행하는데 단 7주가 걸렸으므로 인사이드 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
올렸지만 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 성과 별류가 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



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

¹⁾ 4배 압축 기준

IBM

