

Data and AI Virtual Forum Korea

IBM Integrated Analytics System (IIAS)
(데이터 분석의 속도를 높여라!)

—

Darum

기존 모델의 장점은 더 강화

Simplicity (간편성)

- **인덱스 관리 불필요** : Storage Index / Zonemap / 자동압축 / No Admin
- **물리적 모델링 없이 바로 사용** : Load & Go / 성능개선용 물리모델설계 불필요

Performance (성능)

- **고성능 하드웨어** : 더 빠른 CPU, 더 많은 Cache, 더 큰 Memory
- **Micro latency Flash Storage** : I/O 지연 없는 데이터 전송
- **In-Memory 기술** : 컬럼기반의 인메모리 기술로 N3001대비 수배 성능향상
- **신뢰성 높은 스토리지** : 2-Dimensional flash RAID (99.999% 신뢰 있는 하드웨어 컴포넌트)

In-Place Expansion (확장)

- **모듈 방식 확장** : Server + Storage 증설
- **스토리지 확장 지원** : Storage 증설



최신 기술 적용

In-Memory (인메모리 기술)

- **집계 및 분석에 용이한 컬럼기반 데이터베이스 기술** : IBM의 BLU기술을 이용하여 성능 및 기술 안정성 확보
- **속도 향상** : Analytic Query PDA N3001대비 수배 성능향상
- **리소스 최적화** : I/O를 최소화 하고 CPU사용률을 절감

Analytic Tool (분석을 위한 툴)

- **데이터 분석가를 위한 환경** : 분석에 필요한 머신러닝 Lib, Jupyter Notebook, Rstudio등을 기본 탑재하여 Analytic Sandbox 분석환경 구축 지원
- **내장된 Spark 과의 연계**

Cloud Ready

- **SW Defined Analytic DW** : 동일한 구성으로 Public Cloud, Private Cloud, On-prem DW 구성가능
- **공통 SQL 엔진** : 공통 엔진 / 공통 SQL 사용으로 상호 이관이 빠르고 용이함



NO requirements for armies of administrators
(관리자 부대의 **불필요**)

NO compression administration
(압축에 대한 관리 **불필요**)

NO tuning, no configuration
(튜닝과 구성의 **불필요**)

NO storage administration
(스토리지 관리 **불필요**)

NO physical data modeling needed
(물리모델 **불필요**)

NO indexes and tuning necessary
(인덱스 튜닝 **불필요**)



Load data, **start querying within hours**, not days
(Load & Go)

Built-in data migration tools
(내장된 마이그레이션 툴 from PDA to IIAS)

Single management console
(H/W,S/W 통합 모니터링)

System is **performance optimized** for analytics
(분석 성능에 최적화되어 디자인)

In-place **expansion** capabilities
(용량 및 성능에 대한 In-place 확장)

Single support number
(H/W,S/W에 대한 일원화된 지원)

Power Architecture

- 더 적은 노드로 고성능
- 멀티 코어와 SIMD 병렬을 통한 CPU 가속
- 향상된 안정성 및 가용성



Flash Storage, standard

- 고속 데이터 전송 시 준 실시간의 지연을 통해 99.9%의 안정성과 운영 효율성

프로세서

분석 알고리즘의 유연하고 빠른 수행

4X

코어당 스레드 수 vs. x86
(시스템당 최대 1344 스레드)

메모리

비즈니스 통찰력을 극대화하기 위한 크고 빠른 작업 공간

4X

메모리 대역폭 vs. x86¹
(최대 32TB 메모리)

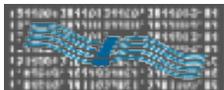
캐시

빠른 응답을 위해 중단 없는 데이터로드 보장

6X

캐시 vs. x86²
(코어당 19MB이상의 캐시)

중단 없는
데이터 로드



방대한 IO
대역폭



병렬 처리



극한의 성능을
위한 Flash사용

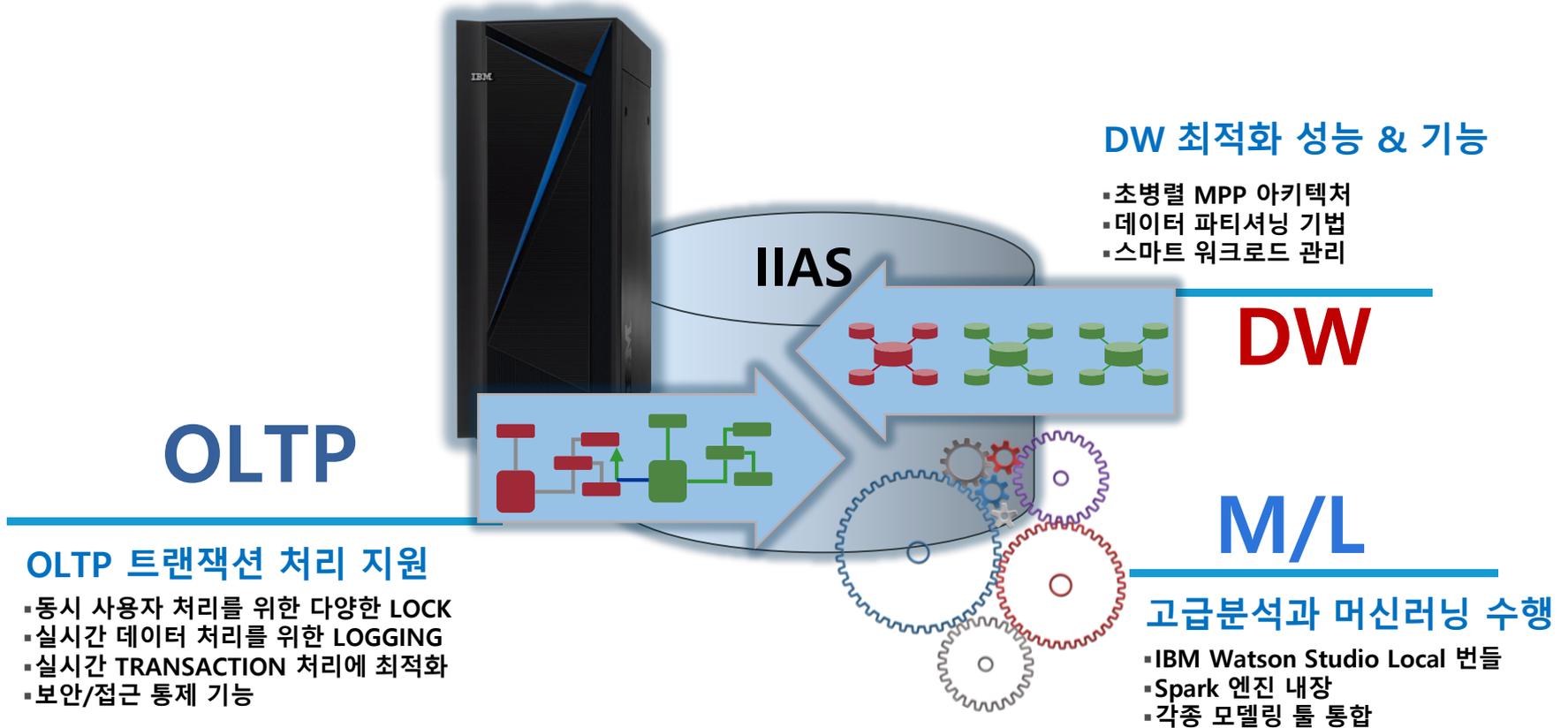


대규모
메모리 프로세싱



1. 특정 x86 및 POWER8 서버에 따라 최대 4 배까지 차이를 보임
2. Intel e7-8890 서버와 12 코어 POWER8 서버를 비교하여 최대 6 배 더 많은 캐시를 제공

MPP (massively parallel processing)

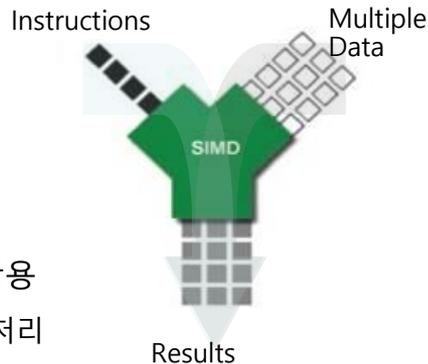


In-Memory Optimized

- 메모리에서 더 유용한 데이터 유지
 - 데이터는 압축 상태 유지
 - LRU Caching보다 진보된 Scan Friendly Caching
- 메모리로 유입되는 데이터량 최소화
 - Columnar access
 - Late materialization
 - Data skipping
- 최적화된 메모리 응답시간 이용
 - 스캔(Scans)
 - 조인(Joins)
 - 집계(Aggregation)

CPU Optimized

- CPU 가속화
 - SIMD 처리 for
 - 스캔
 - 조인
 - 그룹핑
 - 연산
- CPU 사용을 최대한 활용
 - Core 친화적인 병렬처리
- 적은 CPU 처리량
 - 압축된 데이터의 처리
 - Late materialization
 - Data skipping



I/O Optimized

- 읽기 횟수를 줄임
 - Columnar I/O
 - Data skipping
 - Late materialization
 - Scan Friendly Caching
- 효율적인 I/O
 - 특화된 columnar prefetching 알고리즘

Hoffman Encoding 압축

- Register-friendly encoding
- 데이터 처리를 위하여 적은 I/O , 적은 CPU cycle

LAST_NAME Encoding

Brown	
Johnson	
Johnson	
Johnson	
Johnson	
Brown	
Johnson	
Gilligan	
Wong	
Johnson	

Actionable Compression

- Predicates 조건, Group By, Aggregation, Join 등의 작업시 압축해제없이 Query 수행

```
SELECT COUNT(*) FROM T1
WHERE LAST_NAME = 'JOHNSON'
```



Encode JOHNSON

```
SELECT COUNT(*) FROM T1
WHERE LAST_NAME = ''
```

향상된 Storage Indexing

- *Netezza Zonemap*은 Integer, Date, Timestamp 컬럼의 min/max 를 관리하는 Storage indexing 기술로 이제 **Character 컬럼지원으로 향상됨**

예시

10TB 데이터 쿼리 수행 예시

- 데이터 크기 : 100개 컬럼을 가진 10TB 테이블 (10년 동안 축적된 데이터)
- 조회 : 2017년에 얼마나 많은 "SALES"가 일어났는가?
>> SELECT COUNT(*) FROM MYTABLE WHERE YEAR = '2017'

손쉬운 사용 (Ease of Use)

- 인덱스 불필요, 집계 테이블 최소화, No 튜닝
- Load and Go!

동적 인메모리(In-memory) columnar processing

- 동적 데이터 이동기술을 통해 성능 극대화 (Storage ↔ Memory)

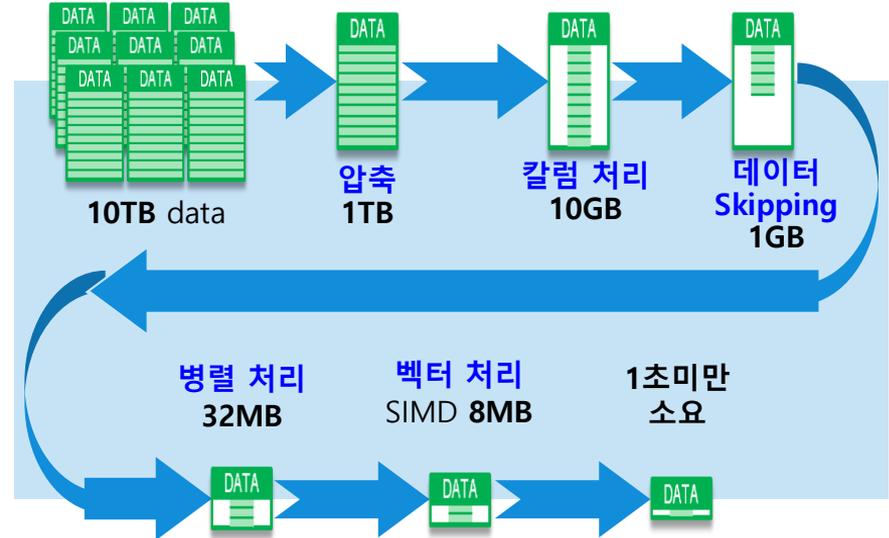
Data skipping (zone maps)

- 물리적인 인덱스 없이 필요한 데이터만 읽을 수 있도록 자동 처리하는 기술

전세계 일등 압축 기술

Actionable compression

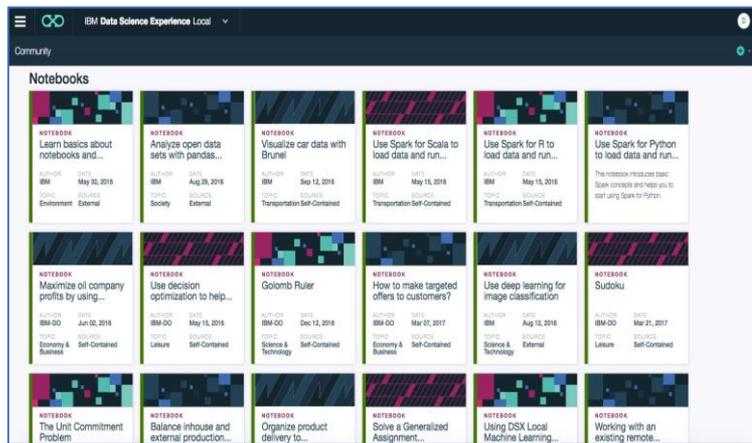
- 압축 해제없이 데이터를 조회할 수 있는 특허 받은 압축 처리 기술



데이터 적재만으로 Query 수행 1초 미만 소요

Watson Studio Local (WSL)

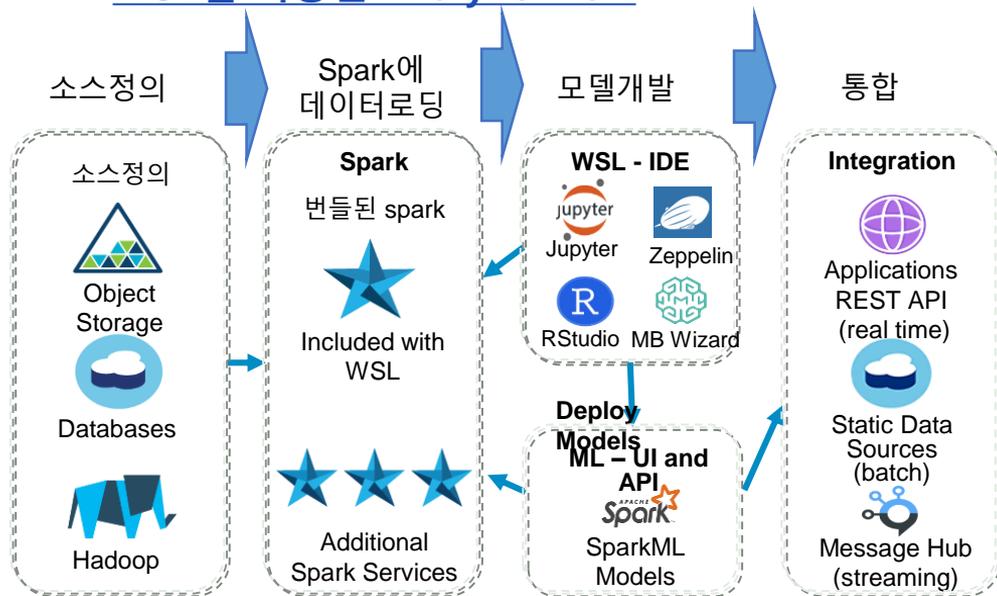
- 분석 애플리케이션을 공동으로 개발 및 배포 하기 위한 플랫폼
- 분석 응용 프로그램을 빌드하고 배포하는 데 필요한 모든 것을 포함
- 확장 가능 : remote Spark, 다중 데이터 소스, 라이브러리 추가,LDAP



IIAS와 WSL 의 통합 시너지

- 일반 DB사용자 뿐만 아니라 Data Scientist 까지 사용자 계층 다양화하고 Analytic Sandbox 로 활용

WSL를 이용한 Analytic Flow



다양한 DBMS 가 지원되는 **Transparent Access 기능**을 통해 이기종 소스에 대한 가상 뷰를 사용할 수 있음

Hadoop Data 이동

하둡을 데이터 웨어하우스의 아카이빙 및 복구에 활용



- 실 고객 활용 (예) : 하둡에 수집된 대량의 Level 0 (raw) 데이터를 IIAS 로 이동시켜 배치작업으로 Cleansing/Enrich 함.
- Enrich 된 Level 1 데이터는 다시 하둡으로 이동하여 저장하고 활용
- 또한 Hadoop 은 data warehouse 의 archiving 영역 또는 DR 로 활용함



이기종 RDBMS 연계

다양한 RDBMS 지원



- 기업내부에서 사용하는 다양한 RDBMS 에 저장된 데이터를 IIAS에서 원격 연계조회

TERADATA ORACLE SAP

Microsoft
Microsoft Azure

mongoDB



AMAZON
REDSHIFT

HAWQ



HIVE



HOME

MONITOR

Dashboard

System

Workloads

Overview

Spark

Database Partitions

Statement

Query Execution

Package Cache

Stored Procedures

Locking

Applications

Throughput

Buffer Pools

Table Performance

Notifications

ADMINISTER

RUN SQL

LOAD

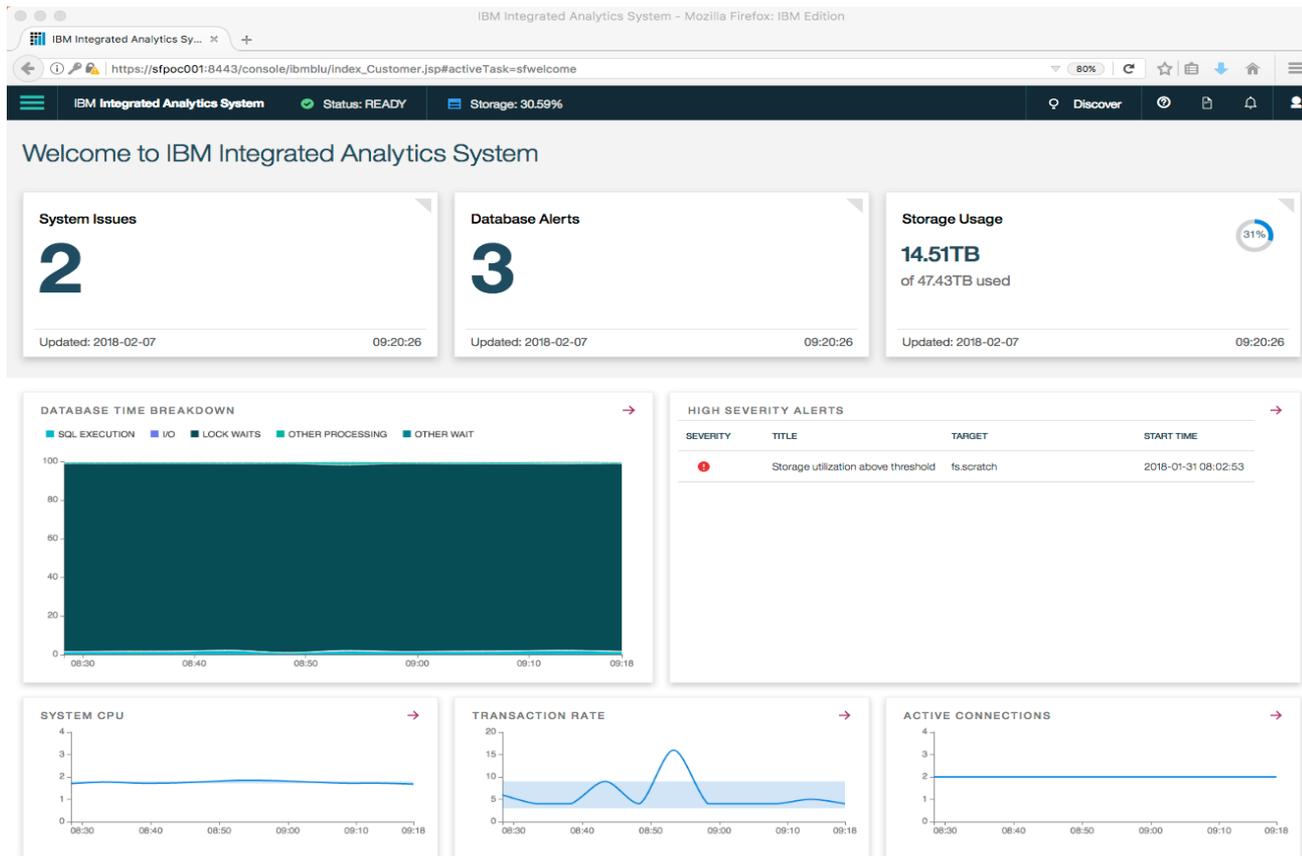
MOVE TO CLOUD

DEVELOP ANALYTICS

SETTINGS

CONNECT

HELP



Built-in Console – Monitor (Hardware)



HOME

MONITOR

Dashboard

System

Workloads

Overview

Spark

Database Partitions

Statement

Query Execution

Package Cache

Stored Procedures

Locking

Applications

Throughput

Buffer Pools

Table Performance

Notifications

ADMINISTER

RUN SQL

LOAD

MOVE TO CLOUD

DEVELOP ANALYTICS

SETTINGS

CONNECT

HELP

MTM: 3452-GB1 Serial: 7840158

rack1

Legend: OK (green), STARTING (blue), WARNING / ATTENTION (orange), FAILED (red), UNREACHABLE (grey)

- Nodes (7): No issue
- Management Switches (2): No issue
- Fabric Switches (2): 1 Warning
- Fiber Channel Switches (2): No issue
- FlashSystem (3): No issue
- RPCs (4): No issue

Last updated at 2018-02-07 09:47:25 (UTC-6)

Built-in Console – Monitor (Resource)



HOME

MONITOR

Dashboard
System

Workloads

Overview

Spark

Database Partitions

Statement

Query Execution

Package Cache

Stored Procedures

Locking

Applications

Throughput

Buffer Pools

Table Performance

Notifications

ADMINISTER

RUN SQL

LOAD

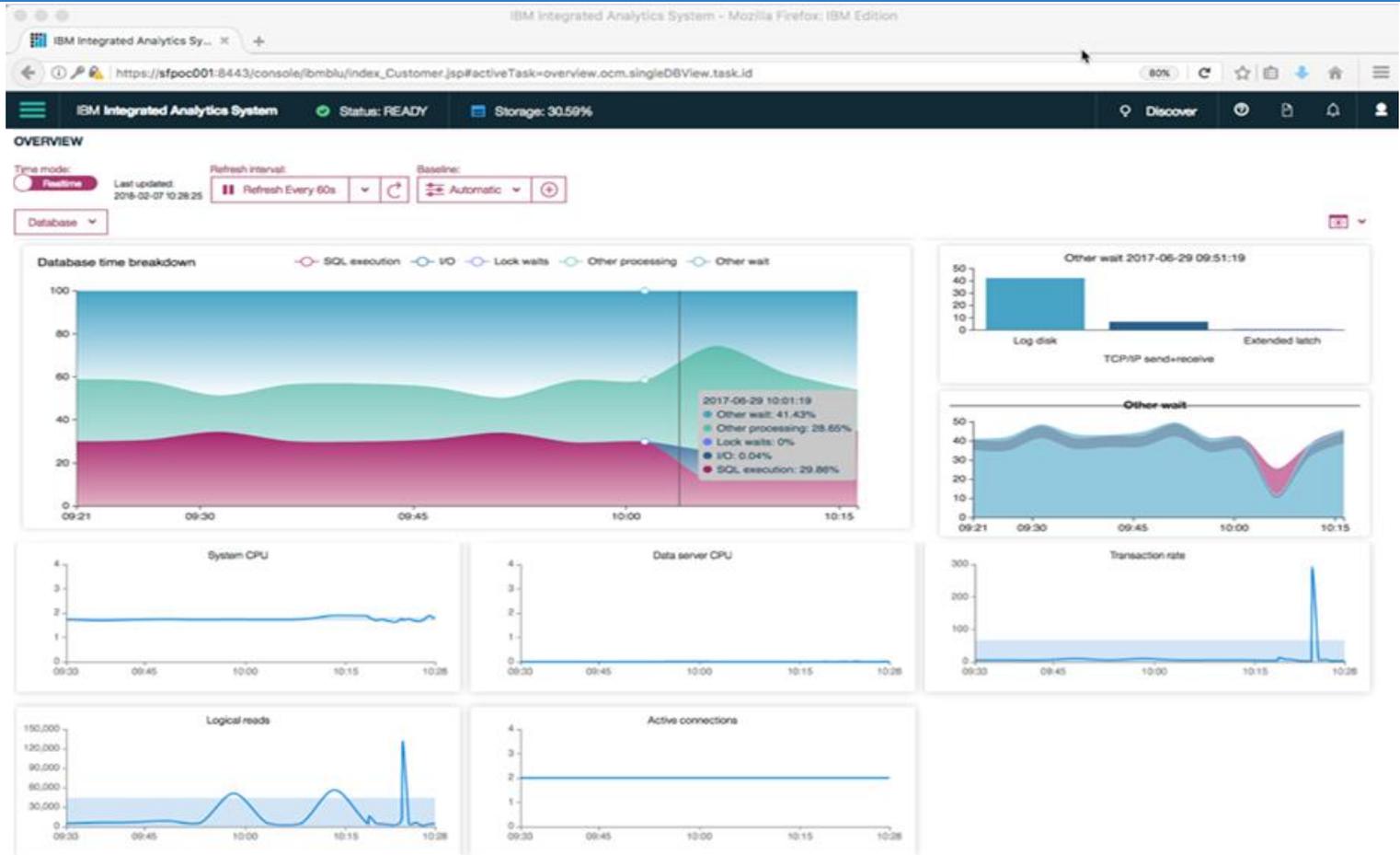
MOVE TO CLOUD

DEVELOP ANALYTICS

SETTINGS

CONNECT

HELP



Built-in Console – Administer

IBM Integrated Analytics System | Status: READY | Storage: 30.59%

Privileges Explorer | Privileges of Group BLUADMIN

Object Privileges | Membership

Tables | Add Objects

WITH GRANT | GRANT | REVOKE | ... None

Tables	SCHEMA	GRANTEE	GRANTEE TYPE	CONTROL	SELECT	INSERT	UPDATE	DELETE	ALTER	REFERE
Views	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
MQTs	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
Schemas	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
Sequences	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
Procedures	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
Functions	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
Servers	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
Remote Tables	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSUSERAUTH	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSEVENTABLES	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSSECURITYPO...	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSAUDITPOLICIES	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSSECURITYLA...	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSSMT	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSTABAUTH	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSPLANAUTH	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSSECURITYPO...	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSSEQUENCEA...	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSTSPACEAUTH	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSAUDITUSE	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSSCHEMATA	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSXSPROJECTA...	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSPLAN	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSROLES	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--
<input type="checkbox"/> SYSSECURITYLA...	SYSIBM	BLUADMIN	Group	--	✓	--	--	--	--	--

HOME

MONITOR

ADMINISTER

Tables

Views

MQTs

Remote Tables

Aliases

Schemas

Application Objects

Privileges

Workloads

RUN SQL

LOAD

MOVE TO CLOUD

DEVELOP ANALYTICS

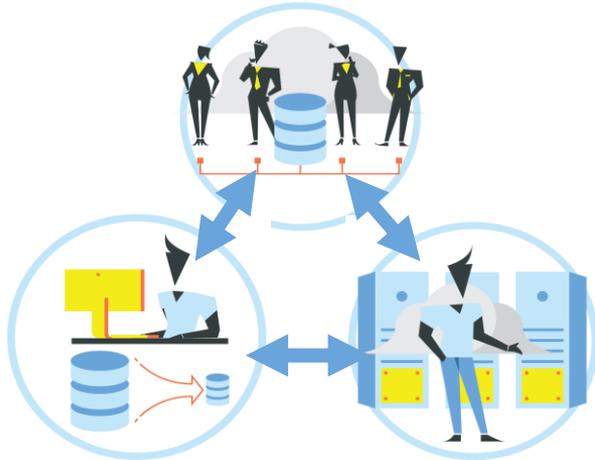
SETTINGS

CONNECT

HELP

퍼블릭 클라우드 데이터 웨어하우스

(new data sources, deep analytics)



온프레미스 데이터 웨어하우스

(structured, defined set of
data, traditional BI)

프라이빗 클라우드 데이터 웨어하우스

(new data sources,
deep analytics)

Common SQL Engine

On-Premises, Private 클라우드 및 Public
클라우드에 있는 모든 IBM데이터 관리 제품군에
손쉽게 배포

운영 환경 통합

모든 데이터 플랫폼을 위한 단일 운영
인터페이스

클라우드로의 빠른 데이터 마이그레이션

클라우드로의 빠르고 안전한 데이터 이동 지원
(IBM Lift)

모든 형태의 데이터 지원

데이터 가상화를 사용하여 Hadoop을 비롯하여
더 많은 유형의 데이터를 쿼리에 포함



서버 & Storage

IBM S822L 24 core server 3.02GHz
(RedHat Linux)
IBM FlashSystem 900

확장 방식

In-place Expansion, Tiered storage

네트워크

Mellanox 10G Ethernet switches
Brocade SAN switches

확장(증설)지원

- 선형 성능 확장 : 상위 model 로 증설 가능
예, 1/3 Rack → 2/3 Rack
예, 1/3 Rack → Full Rack
- 스토리지 확장 : Tiered Storage 추가
예, 1/3 Rack → 1/3 Rack with TS,
예, 1/3 Rack → Full Rack with TS,

Model	1/3 Rack	1/3 Rack Tiered Storage	2/3 Rack	2/3 Rack Tiered Storage	Full Rack	Full Rack Tiered Storage
Server	3	3	5	5	7	7
CPU cores	72	72	120	120	168	168
Memory	1.5TB	1.5TB	2.5TB	2.5TB	3.5TB	3.5TB
¹⁾ Usable User Disk	Flash 27TB	Flash 59TB HDD 166TB	Flash 54TB	Flash 86TB HDD 299TB	Flash 81TB	Flash 113TB HDD 432TB

1) DBMS Log 저장 등이 제외된 순수 User Data 영역

Thank you for working with



(주)다름테크는 IBM 소프트웨어를 판매하고 기술지원을 책임지는 한국IBM에 PARTNER 입니다.