

Technology and Business Issue Report

TaB Issue Report

www.cccr.or.kr

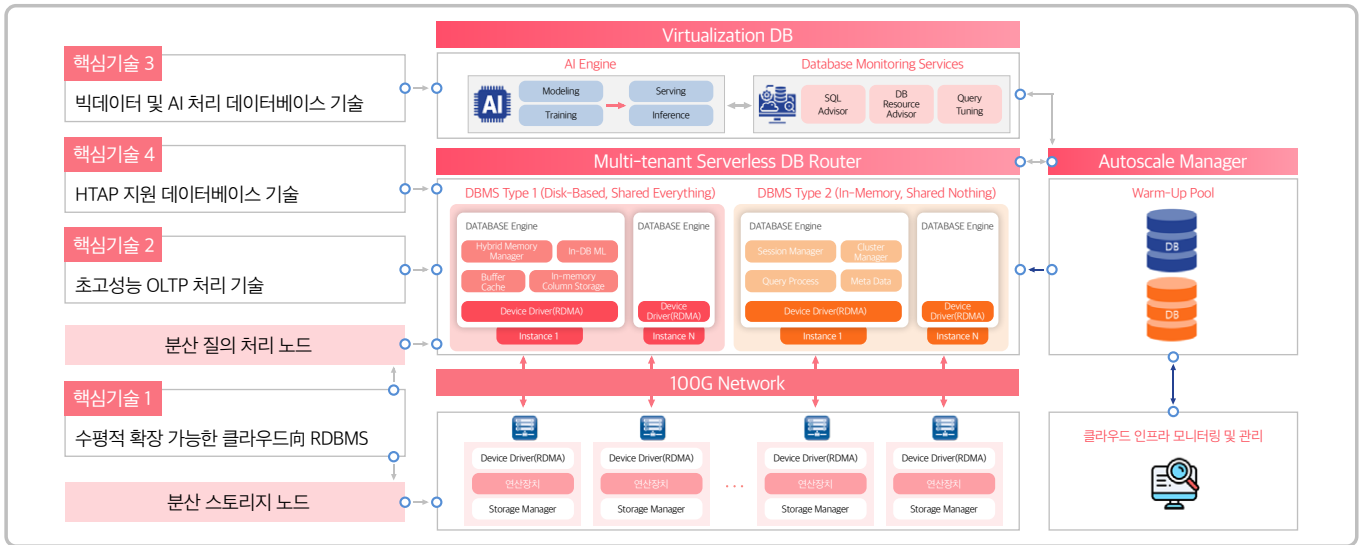


TaB Issue Report

빅데이터 분석 및 AI 처리를 위한 클라우드向 차세대 DBMS 기술 - 빅데이터 분석 및 AI 처리를 위한 클라우드向 차세대 DBMS 기술

기술 개요

- 고성능 HW 기반 OLTP, 빅데이터 및 AI 데이터를 효율적으로 저장 관리하는 차세대 클라우드向 (Cloud-Native) DBMS 기술



핵심 기술

핵심 기술 1 수평적 확장 가능한 클라우드向 RDBMS

스토리지 분리 환경에 적합한 샤딩 기반 분산 DBMS 기술, 데이터 접근 최적화 기술, 임시데이터 저장 및 관리 기술

핵심 기술 2 초고성능 OLTP 처리 기술

고성능 하드웨어를 이용한 분산 데이터베이스의 신뢰성, 가용성, 동시성 제어, 성능 향상, 복제 관리 기술과 이를 활용한 OLTP 데이터베이스 최적화 기술

핵심 기술 3 빅데이터 및 AI 처리 데이터베이스 기술

빅데이터 및 AI 처리를 위한 AI 엔진 연동 및 In-Database ML 기술과 분석질의처리 과정에서 필요한 대용량 임시 데이터 저장 및 관리 기술, 칼럼기반 저장과 압축 기술, auto-scaling 및 저지연 콜드 스타트 기술

핵심 기술 4 HTAP 지원 데이터베이스 기술

ML 기술적용을 통한 클라우드向 고성능 HTAP 지원을 위한 OLTP 및 OLAP를 통합한 HTAP 지원 기술 데이터베이스 최적화 기술

핵심 기술 5 Virtual Database 기술

Virtual Database Package를 이용한 통합 기술

핵심 기술 6 HTAP 환경에서 고속분석 질의처리를 위한 In-memory Column 기반 저장 및 압축 기술

HTAP 환경에서 Column 단위의 분석형 쿼리를 빠르게 처리하기 위한 Column 기반 데이터 저장, 처리 기술

핵심 기술 7 서버리스 환경에서 DB 인스턴스 풀 관리 및 워크로드 증감에 따른 Auto-scaling 및 저지연 콜드스타트 기술

Warm-up 된 DB 인스턴스 풀 관리 기술 및 Auto-scaling 기술과 증감에 따른 Load Balancing 기술

핵심 기술 8 In-Database ML 기능

SQL 확장 통한 DBMS 엔진에서의 ML 응용의 직접 지원 및 고성능 컴퓨팅 HW기반의 AI 기술

핵심 기술 1

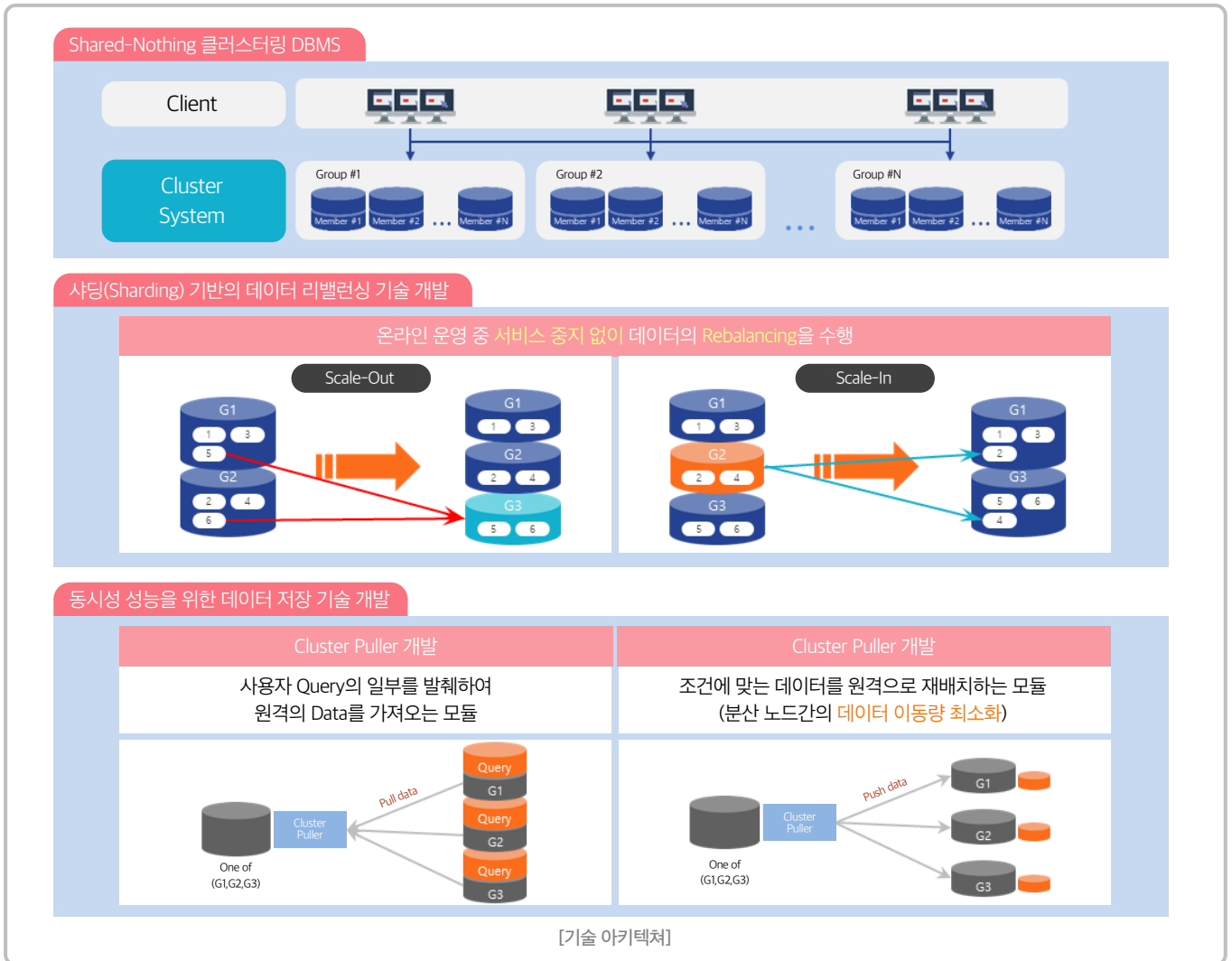
▶ 스토리지 분리 환경의 클라우드向 데이터베이스 지원

- 스토리지 분리 환경에서 클라우드向 데이터베이스 지원을 위한 핵심 기반 기술

1.1 | 스토리지 분리 (Disaggregated Storage) 환경에서 데이터 접근 최적화

1.2 | NoSQL의 수평적 확장성(Scale-Out) 장점과 기존 RDBMS SQL 및 ACID 장점을 모두 제공하기 위해, 샤딩(Sharding) 기반 분산 DBMS 기술 및 관련 요소 기술

1.3 | 임시 데이터(Ephemeral Data) 저장 및 관리 기술



핵심 기술 2

▶ 초고성능 OLTP 데이터베이스 최적화 기술

- 고성능 하드웨어 클라우드 환경에 최적화된 OLTP 데이터베이스 기술

2.1 | NVM, RDMA, NVMeOF, 100GB 네트워크 스위치 등 고성능 HW 기반 분산 데이터베이스 신뢰성, 가용성, 동시성 제어, 성능 향상 기술

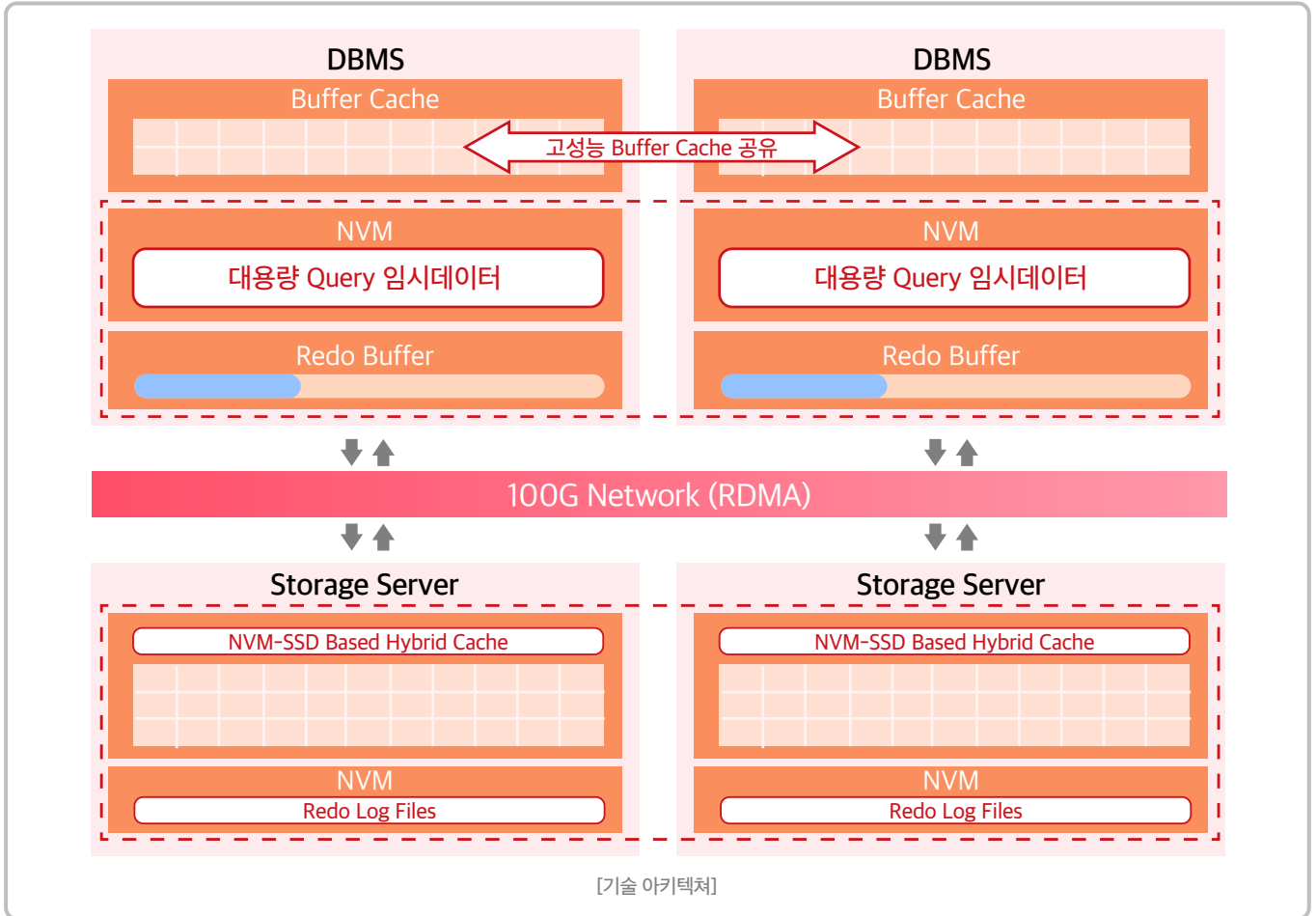
2.2 | 네트워크 쓰기 트래픽 및 지연 최소화를 통한 차세대 복제 (Replica) 관리 기술

2.3 | 비휘발성 메모리 도입에 따른 딥-메모리 계층구조에서 최적 데이터 Placement 및 캐싱 기술

2.4 | OLTP 데이터베이스 AI 기반 최적화 기술 (예: 인덱싱, 버퍼관리, 질의 최적화, 실시간 튜닝 등)

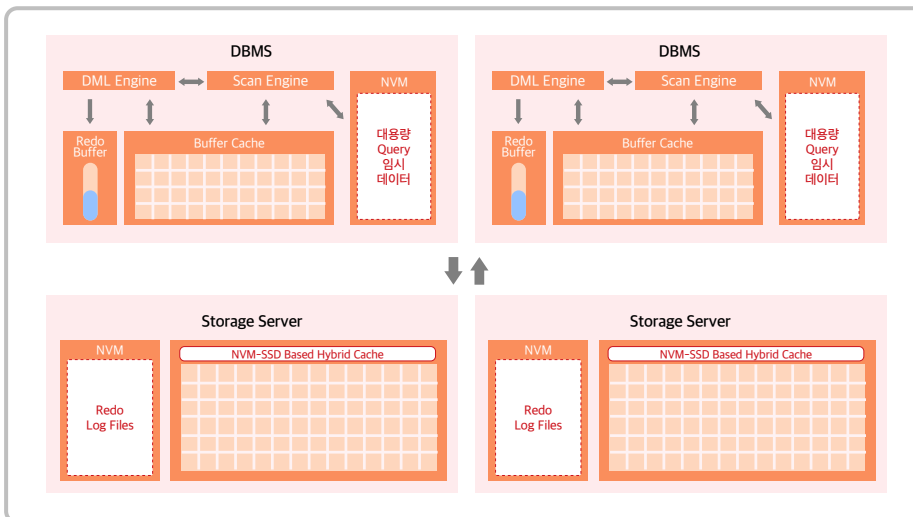
핵심 기술 2

▶ 초고성능 OLTP 데이터베이스 최적화 기술



주요 기술 | 고성능 스토리지를 이용한 데이터베이스 최적화 기술

- DBMS - 대용량 임시 데이터를 Local NVM에 저장하는 기술
- Storage Server - NVM/SSD Based Hybrid Cache + NVM Based RDMA Redo Log File 을 구성하는 기술



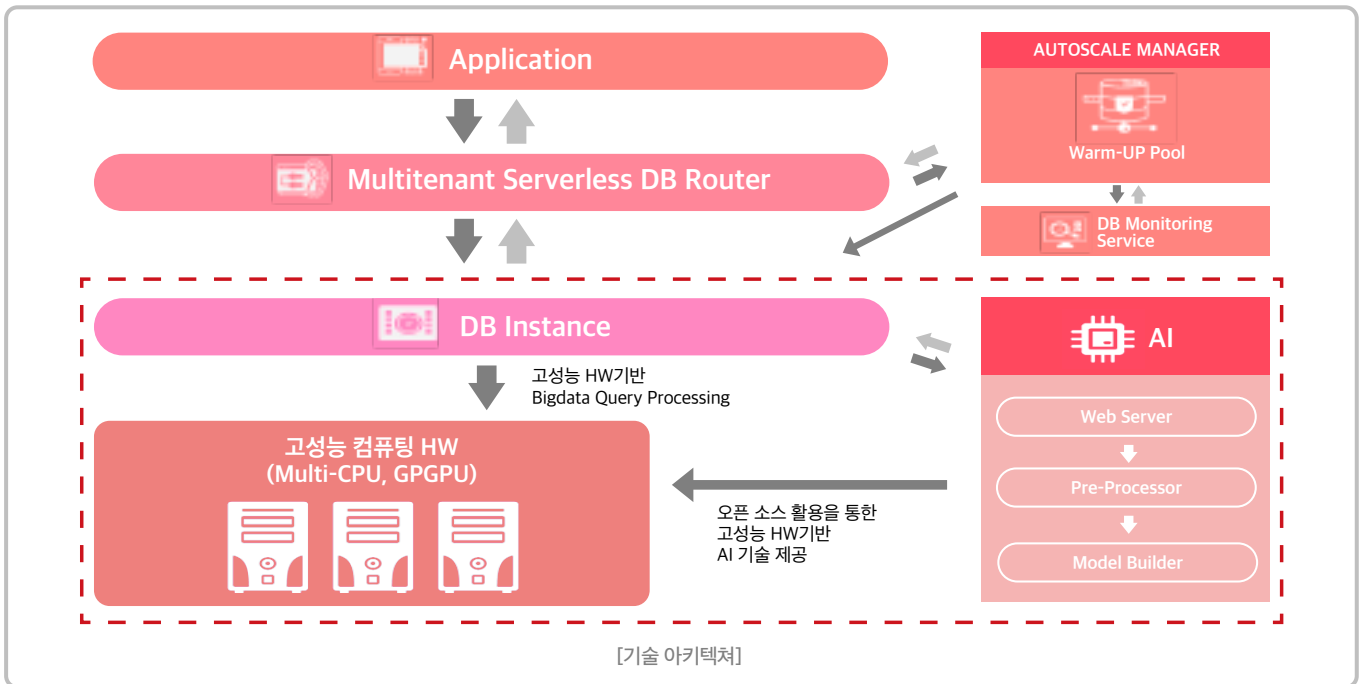
기대효과

- ✓ DBMS의 대용량 Query 데이터를 로컬 NVM에 저장함으로써 Query 수행 성능 향상
- ✓ Storage Server의 NVM/SSD를 Hybrid Cache로 이용함으로써 I/O 성능 향상
- ✓ NVM 기반의 Redo Log File에 RDMA로 접근함으로써 DML 성능 향상

핵심 기술 3

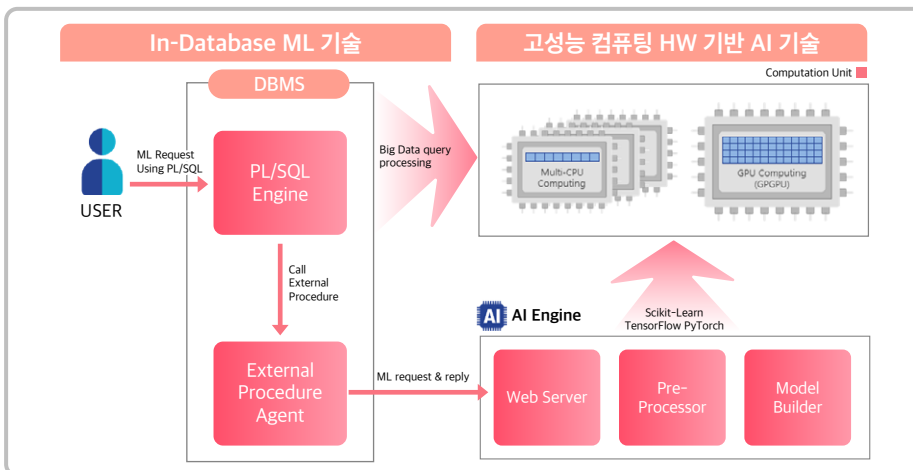
▶ 차세대 컴퓨팅 자원 기반 빅데이터 및 AI처리 DBMS 기술

- 차세대 컴퓨팅 자원 기반의 빅데이터 및 AI 처리 데이터베이스 기술
 - 3.1 | 멀티코어 CPU, GPGPU 융합을 통한 빅데이터 및 AI 처리 가속화 기술
 - 3.2 | 서버리스(Serverless) 환경에서 Auto-Scaling 및 저지연 콜드 스타트(Cold Start) 기술
 - 3.3 | 대용량 임시 데이터에 대한, 고성능 HW기반의 효과적인 분산 저장, 관리 기술
 - 3.4 | 분석질의처리 가속을 위한 칼럼기반 저장과 압축 기술
 - 3.5 | DB 엔진에서 ML 응용의 직접 지원 (즉, In-Database ML)을 위해 SQL 확장 및 처리 최적화 기술



주요 기술 | 고성능 HW 기반의 In-Database ML 기능

- 시스템 패키지를 통한 DBMS 내에서의 ML 기능 제공
- 오픈소스 활용을 통한 고성능 컴퓨팅 HW기반의 AI 기술 제공



기대효과

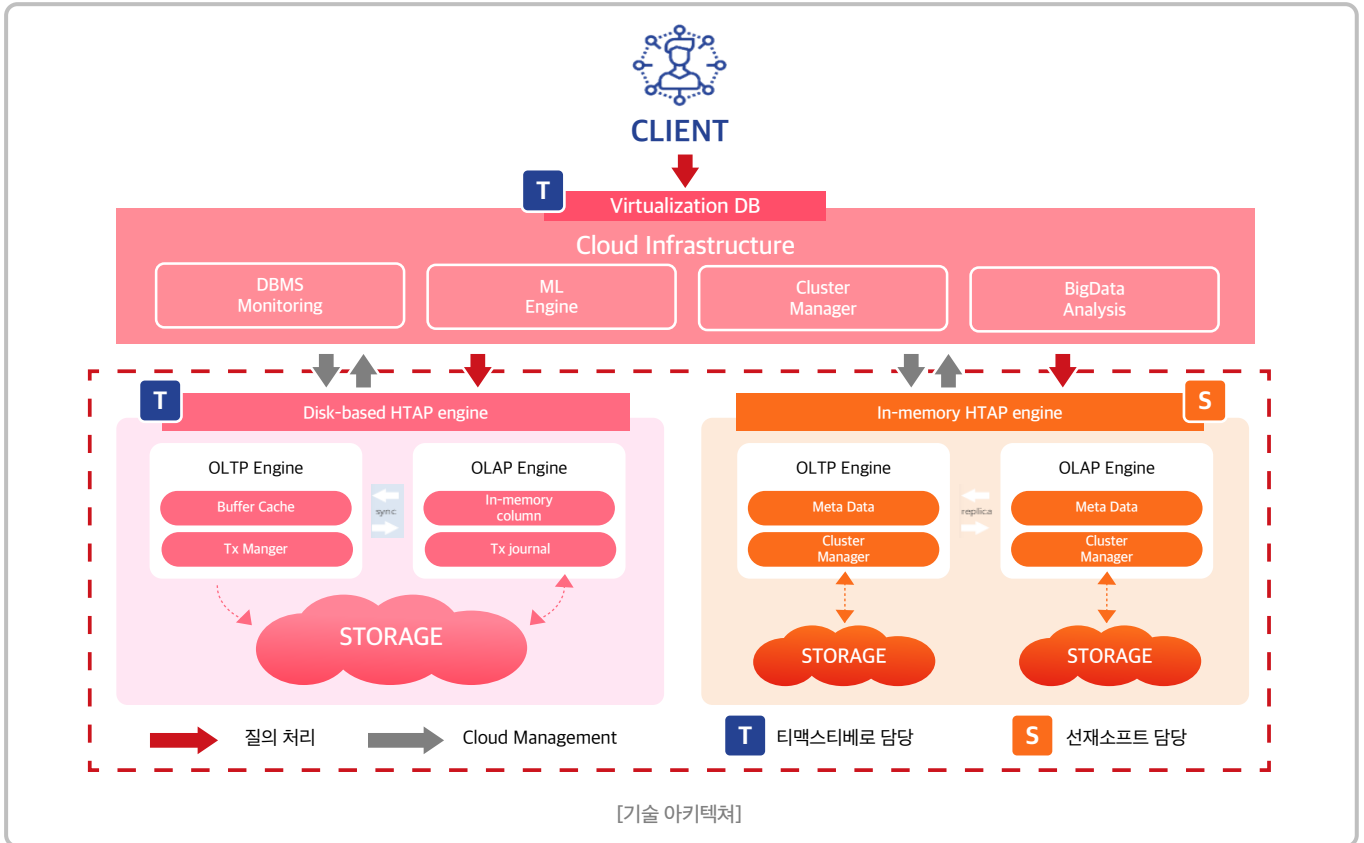
- ✓ 고성능 CPU 및 GPGPU를 통한 최신 AI 기술 적용
- ✓ In-Database ML을 패키지 형태로 편리하게 사용

핵심 기술 4

▶ HTAP 지원 데이터베이스 기술

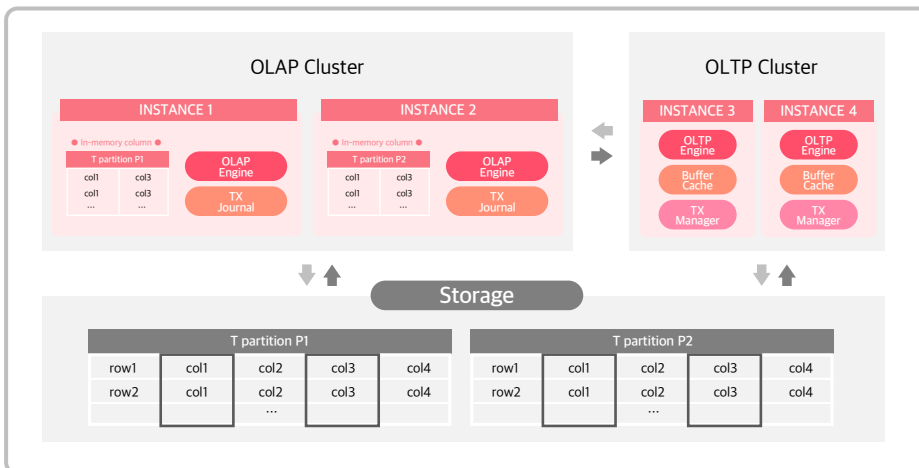
• HTAP 지원 데이터베이스 기술

- 4.1 | OLTP 및 OLAP를 통합한 HTAP 지원 기술
- 4.2 | 클라우드 向 고성능 HTAP 지원을 위한 데이터베이스 최적화 기술
- 4.3 | ML 기술 적용을 통한 HTAP 데이터베이스 최적화 기술



주요 기술 | OLTP 및 OLAP을 통합한 HTAP 지원 기술

- Disk-Based 데이터베이스에서 OLAP 질의에 필요한 칼럼을 In-Memory로 클러스터에 나눠서 저장하는 기술



기대효과

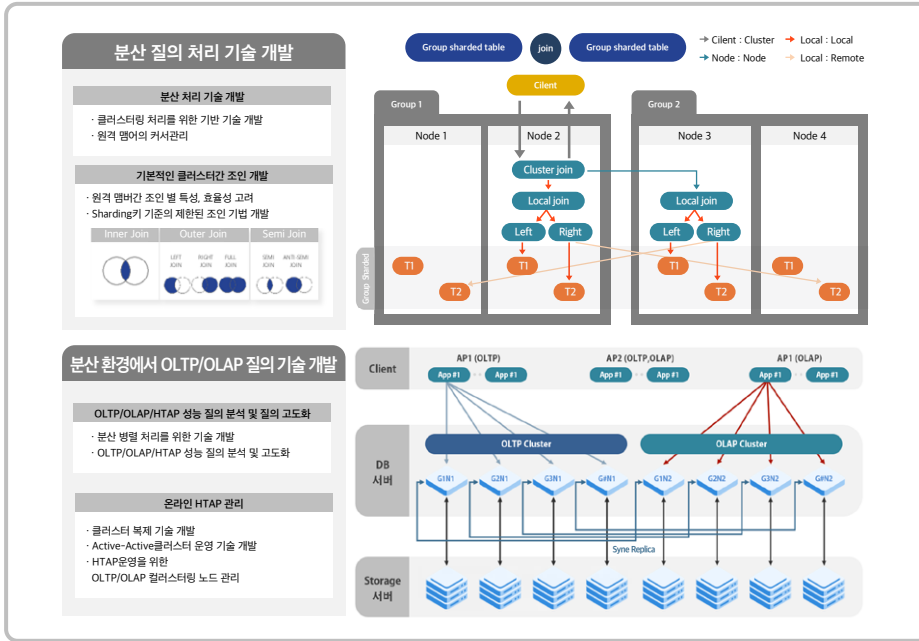
- ✓ OLAP 질의는 In-Memory 칼럼 대상으로 수행하여 OLTP와 독립적으로 수행
- ✓ 인스턴스 간 In-Memory 칼럼 분산 저장 및 분산 질의를 통한 리소스 효율 향상

핵심 기술 4

▶ HTAP 지원 데이터베이스 기술

주요 기술 | OLTP 및 OLAP을 통합한 HTAP 지원 기술

- 온라인 HTAP 지원을 위한 In-Memory 데이터베이스 기술



기대효과

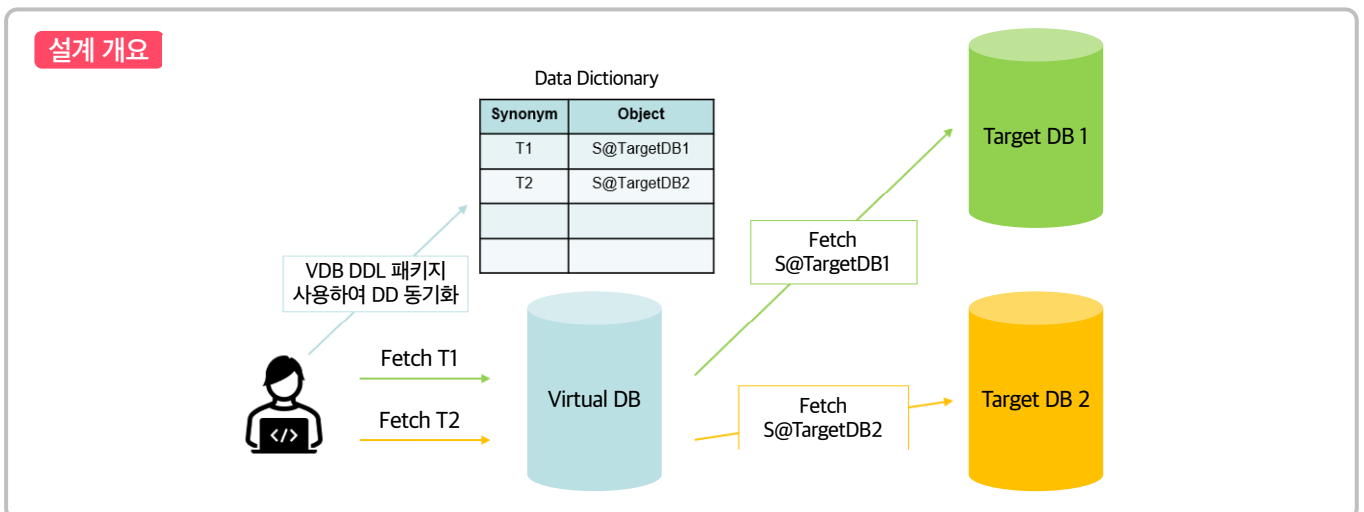
- ✓ OLTP/OLAP 대용량 분산 병렬 질의 처리 성능 향상
- ✓ OLAP, OLTP 클러스터 분리 운영으로 데이터 처리 성능 및 효율성 향상
- ✓ DB 부하에 따른 OLAP/OLTP 클러스터 노드 Auto-Scaling

핵심 기술 5

▶ Virtual Database 기술

주요 기술 | Virtual Database Package를 이용한 통합 기술

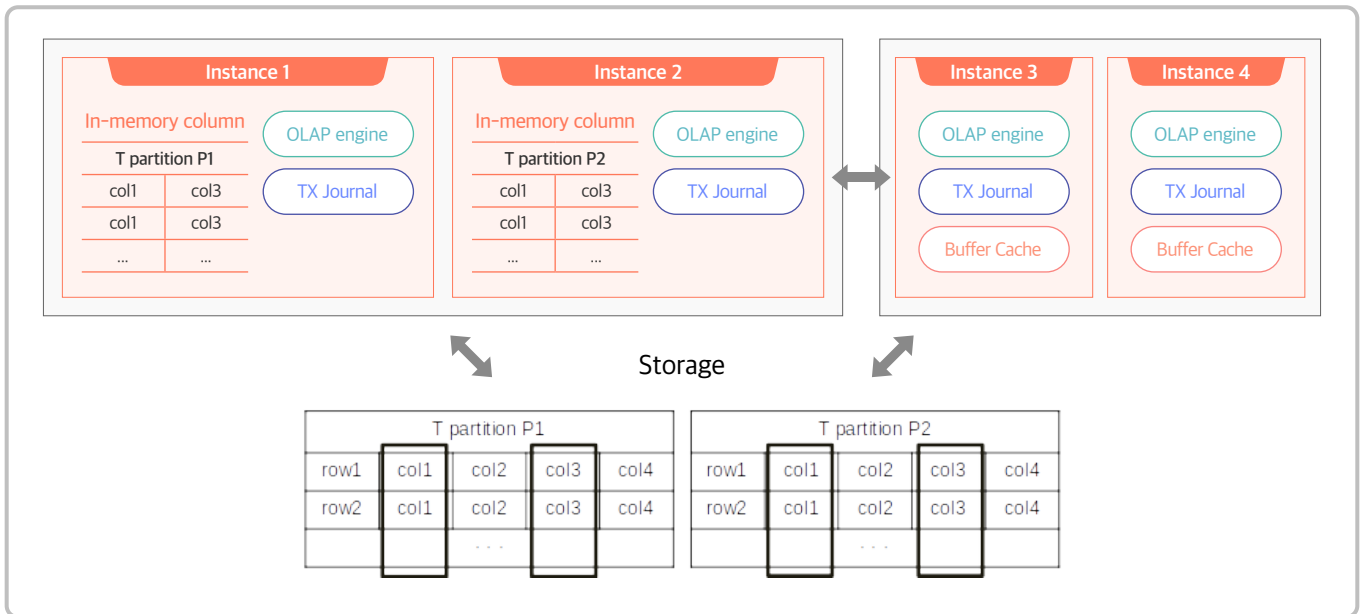
- Remote DBMS에 Data를 생성, 조회, 삭제, 변경 할 수 있도록 해주는 패키지 제공
- DDL을 Remote DBMS에 수행하고 Virtual DBMS의 Data Dictionary에 저장하는 기능 제공



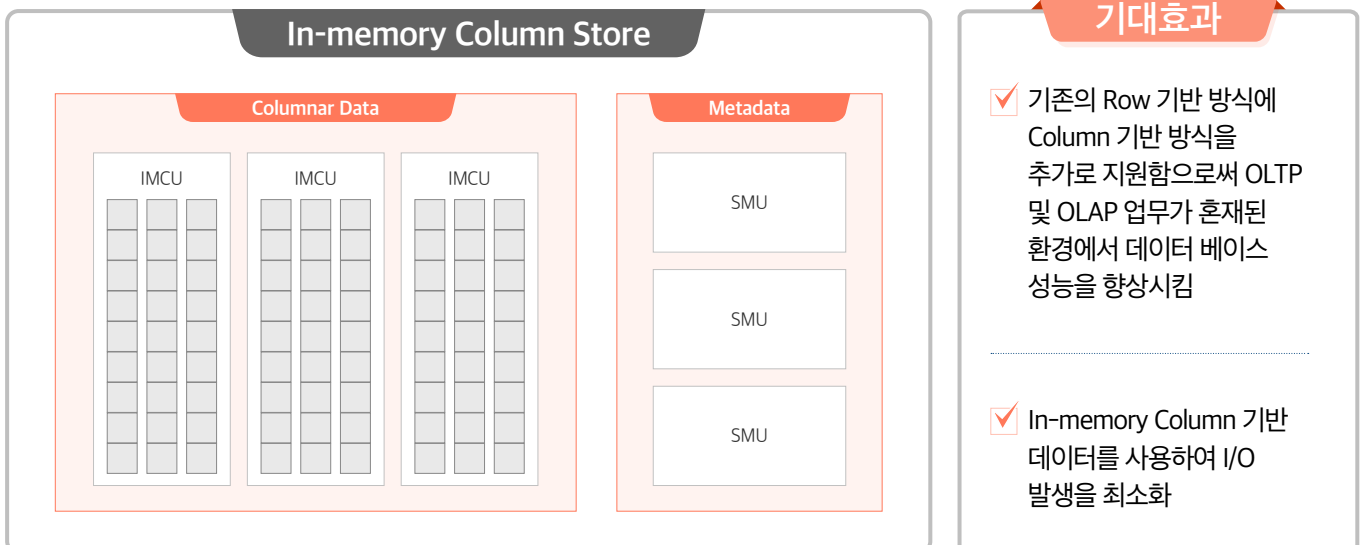
핵심 기술 6

▶ HTAP 환경에서 고속분석 질의처리를 위한 In-memory Column 기반 저장 및 압축 기술

- HTAP 환경에서 Column 단위의 분석형 쿼리를 빠르게 처리하기 위한 Column 기반 데이터 저장, 처리 기술
 - 6.1 | Row 기반 데이터를 Column 기반 데이터 포맷으로 만들어 In-memory로 저장하는 기술
 - 6.2 | 메모리 효율을 높이기 위한, Column 기반 데이터를 압축하는 기술
 - 6.3 | Row 기반 데이터와의 정합성 유지 및 데이터 최신화를 위한 Transaction Journal 기술
 - 6.4 | Scan 효율을 높이기 위한, 불필요한 데이터를 Pruning 하는 기술과 SIMD Vector 연산을 통한 Filter 처리 기술



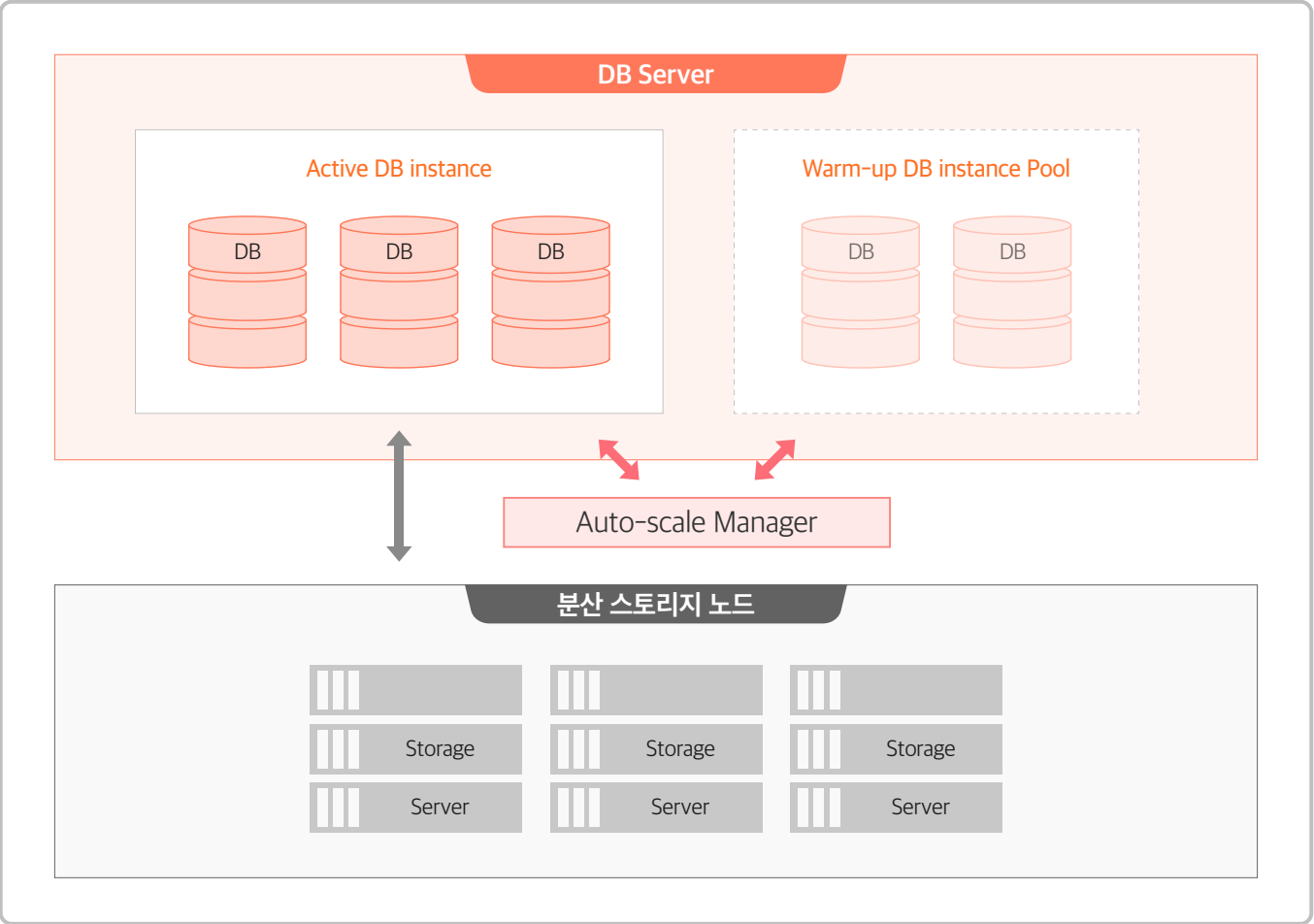
주요 기술 | Row 기반 데이터를 Column 기반 데이터 포맷으로 만들어 In-memory로 저장하는 기술



핵심 기술 7

▶ 서버리스 환경에서 DB 인스턴스 풀 관리 및 워크로드 증감에 따른 Auto-scaling 및 저지연 콜드스타트 기술

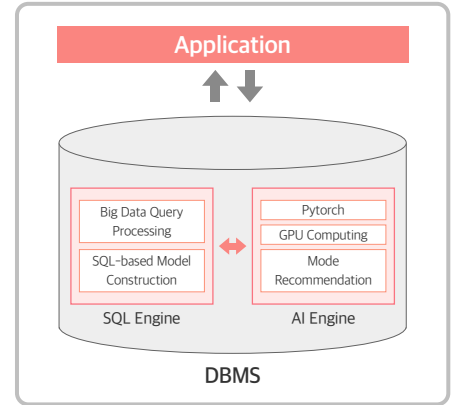
- 워크로드 증감에 따른 데이터베이스 Auto-scaling 기능 제공
 - 현재 워크로드에 따라 Auto-scale Manager의 판단 하에 데이터베이스 인스턴스의 Auto-scale In/Out 제공
- 저지연 콜드스타트(Cold-start) 데이터베이스 기능 제공
 - 워업(Warm-up) 데이터베이스 인스턴스 풀 관리를 통한 신속한 Auto-scale In/Out 제공
- 워크로드 모니터링을 통한 데이터베이스간 로드밸런싱 기능 제공
 - 워크로드를 모니터링하여 데이터베이스 인스턴스 간의 적절한 세션 중계를 통한 Load Balancing 기능 제공



핵심 기술 8

▶ In-Database ML 기능

- SQL문을 통한 DBMS 내에서의 ML 기능 제공
 - 데이터 과학자 없이 DBA 인력만으로 모델의 학습/추론과 같은 고성능 머신러닝 기능의 사용
- 데이터베이스 내 고성능 모델 제공
 - 주어진 과제에 맞는 기생성 / 학습된 모델을 통해 즉시 학습과 추론할 수 있도록 제공
- 티베로 데이터베이스 내의 데이터를 대상으로 한 대규모 학습 기능 제공
 - DBMS 내에서 대규모 데이터베이스를 대상으로 데이터 탐색, 변환, 분석 및 추론 기능 제공



기술 적용 분야

Row 기반 데이터를 Column 기반 데이터 포맷으로 만들어 In-memory로 저장하는 기술



공개 SW
커뮤니티



TmaxTibero <https://github.com/tibero-oss>

기술개발 기관

TmaxTibero

SUNJE SOFT
(주)선재소프트

기술 이전 및
협력 문의



한국클라우드컴퓨팅연구조합 연구개발팀
전화 : 02-2052-1263 이메일 : rnd@cccr.or.kr



Technology and Business Issue Report

TaB Issue Report

본 보고서의 저작권은 한국클라우드컴퓨팅연구조합에 있으며,
한국클라우드컴퓨팅연구조합과 사전 협의 없이 문서의 전체 또는
일부를 상업적 목적으로 복제 또는 배포해서는 안됩니다.

보고서 문의처

한국클라우드컴퓨팅연구조합 연구개발팀

전화 : 02-2052-1263 | 이메일 : rnd@cccr.or.kr