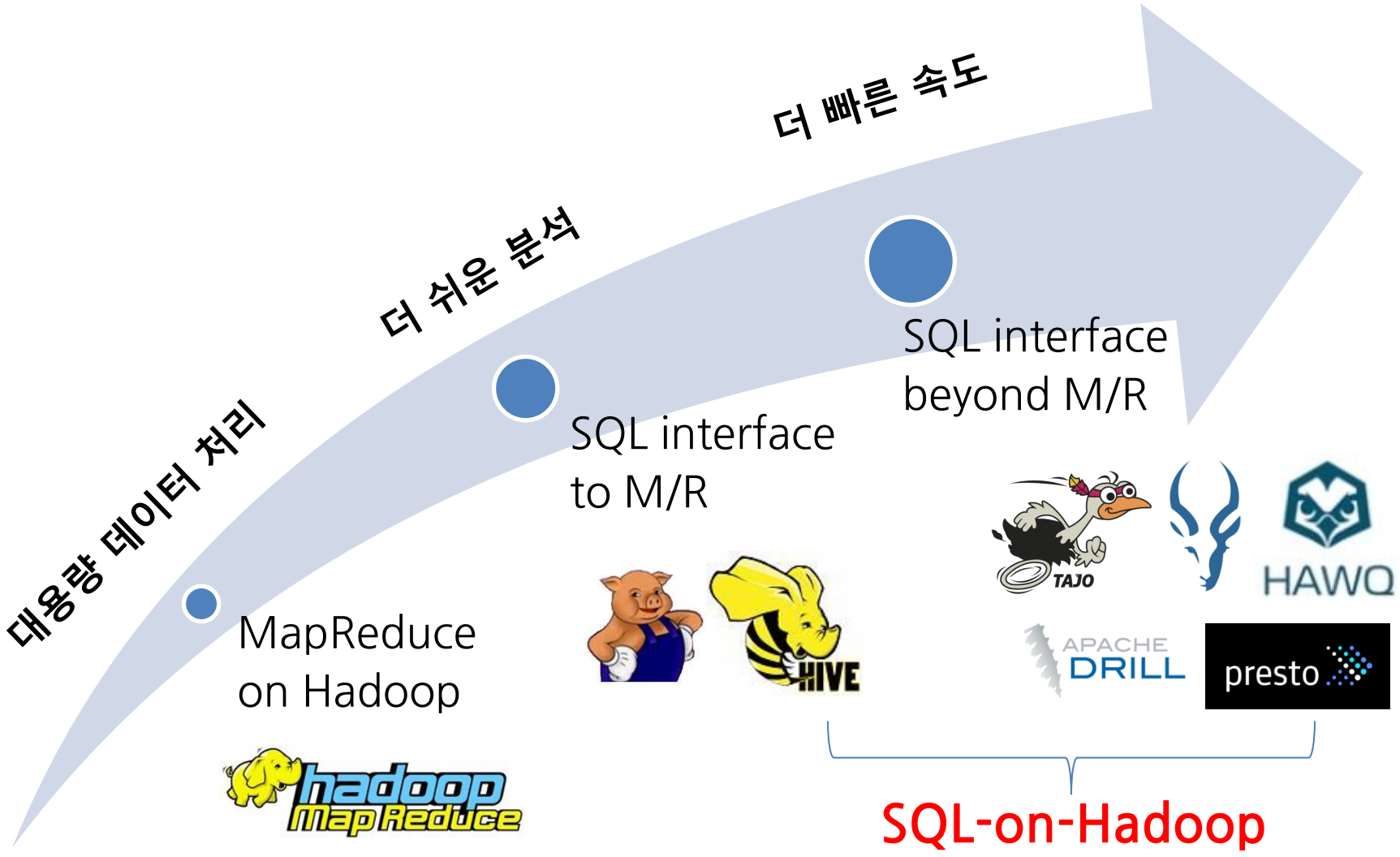
GRUTER
Enjoy Connecting

Apache Tajo를 활용한 빅데이터 분석 어플라이언스 사례

㈜그루터

빅데이터 분석 기술 트렌드



SQL-on-Hadoop 시스템

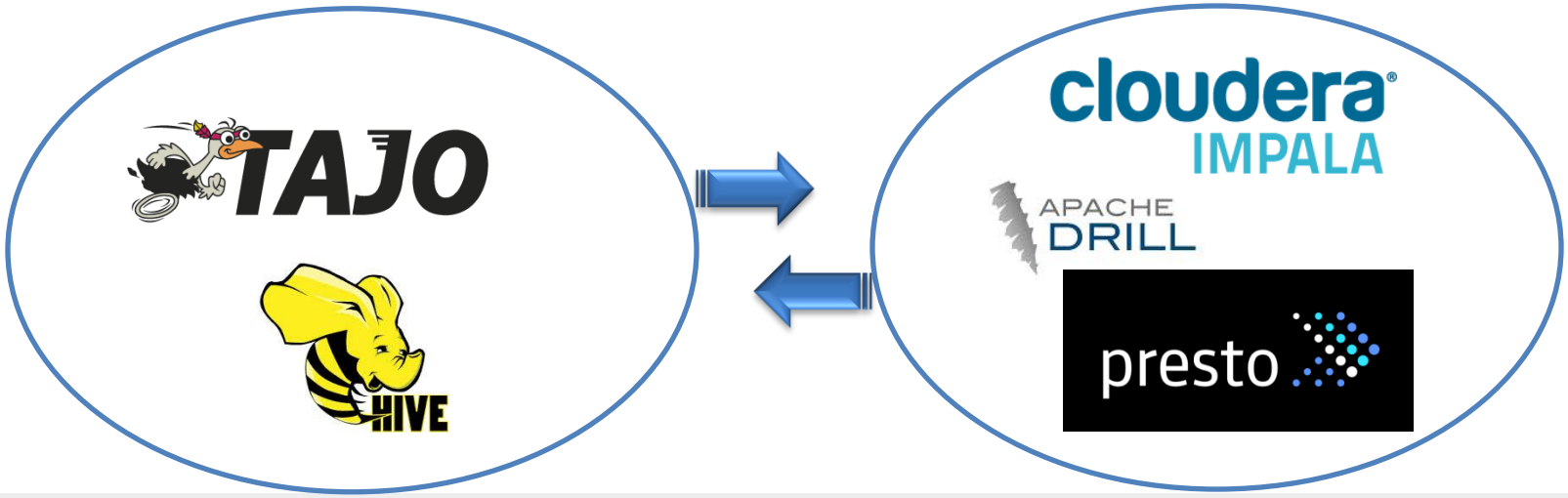
HDFS에 저장된 데이터에 대한 SQL 처리를 제공하는 시스템
MapReduce 를 대체하는 새로운 분산 처리 모델 & 프레임워크 기반
다양한 설계 목표 지향 (외형적으로 비슷해 보이지만 다름)

Data Warehouse System

안정적인 대용량 처리를 중시한 아키텍처
Fault-tolerance

Query Engine

빠른 응답을 중시한 아키텍처
In-memory



Apache Tajo – 차세대 Big DW System

Tajo는

하둡 기반의 대용량 데이터웨어 하우스 시스템 (SQL-on-Hadoop)
Apache 재단의 탑 레벨 프로젝트로 선정된 글로벌 오픈소스
최신 0.9 릴리즈 (2014/10)

Standard, Speed, Scalable

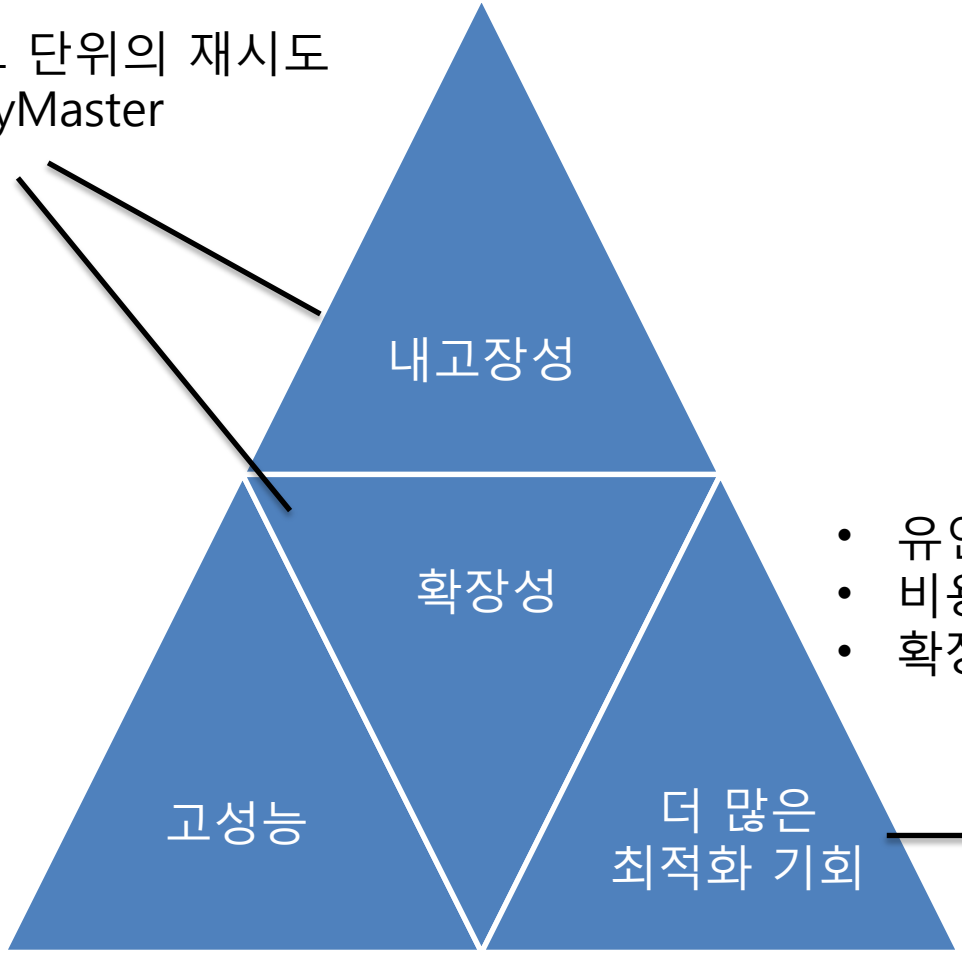
SQL 표준 지원
느린 MapReduce 대신 자체 고성능 분산처리 엔진을 사용하여 평균 3~5배 빠름
In-memory 방식의 솔루션들과 달리, 메모리 크기를 초과하는 데이터도 안정적 처리

Enterprise Big DW

표준 SQL + 확장성 + 안정성 + 고성능으로 빅데이터 처리 시간과 비용을 대폭 절감
수 시간의 대규모 ETL 작업과 수 초 내의 인터랙티브 분석을 동시에 지원
엔터프라이즈 환경의 다양한 데이터 분석 요건을 하나의 솔루션으로 해결

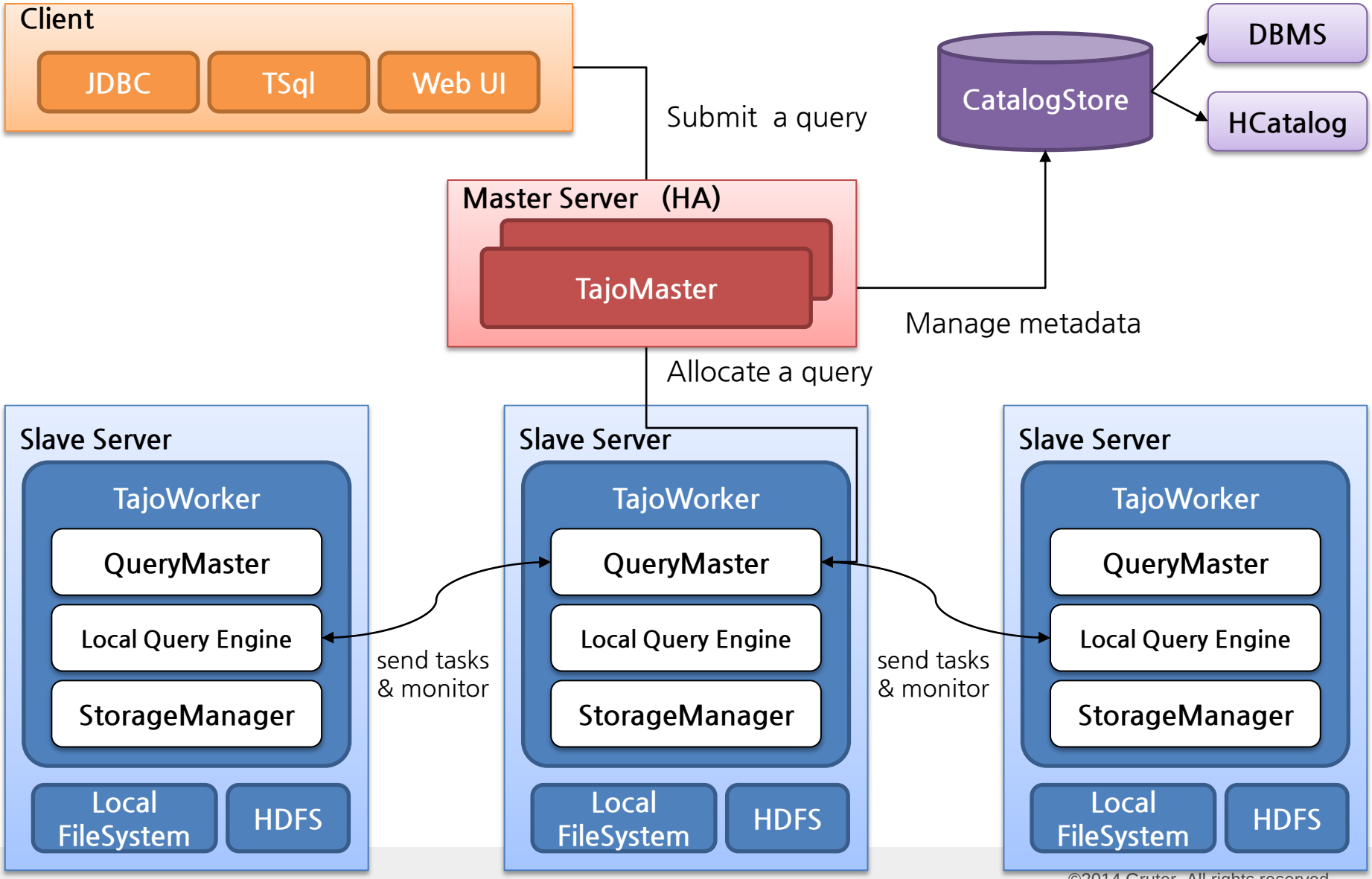
Tajo의 주요 설계 원칙

- 실패한 테스크 단위의 재시도
- 질의 별 QueryMaster

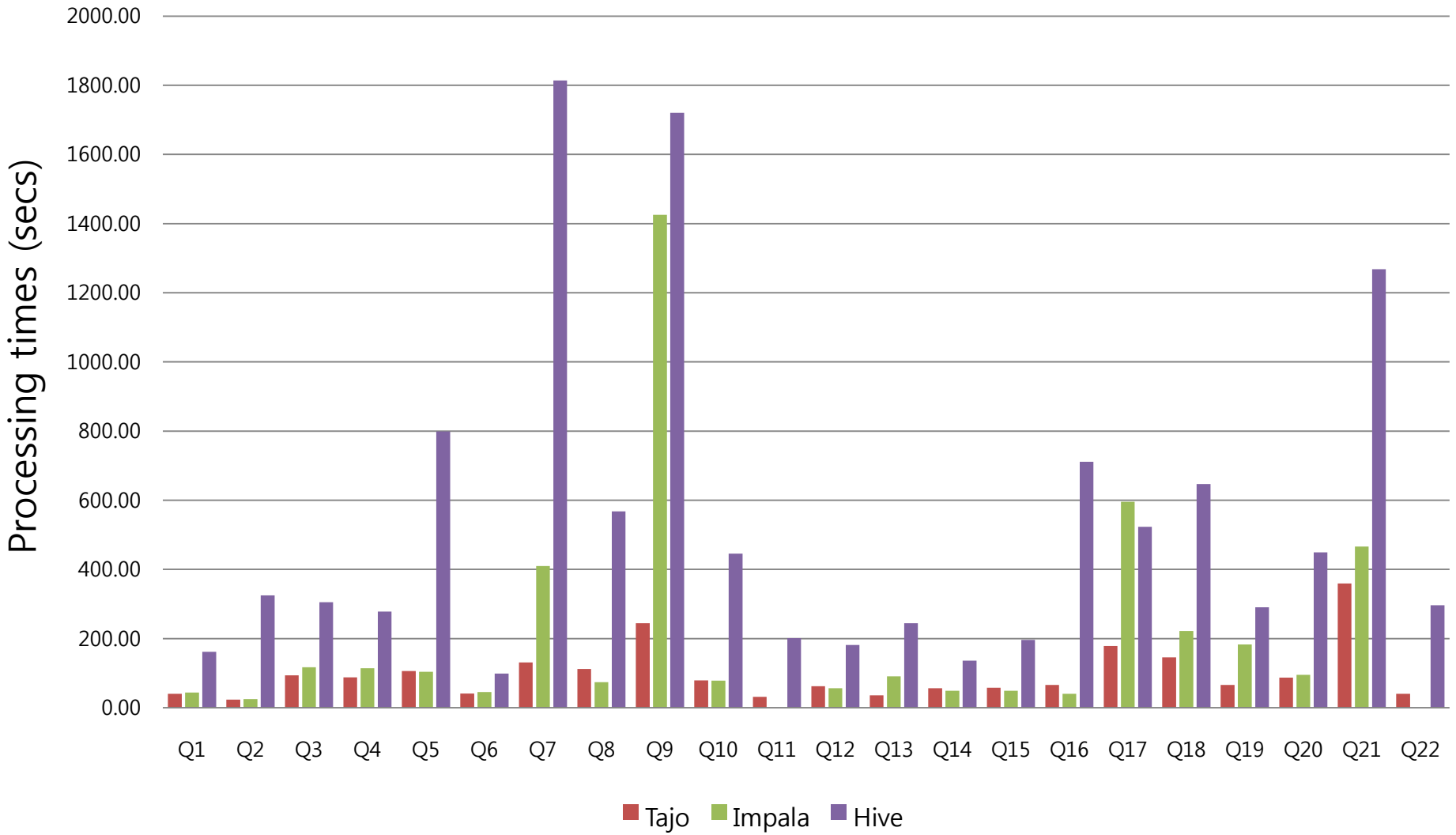


- 유연한 분산 처리 모델
- 비용 기반 최적화
- 확장 가능한 rewrite rule

Tajo Architecture



Tajo 성능 벤치마크



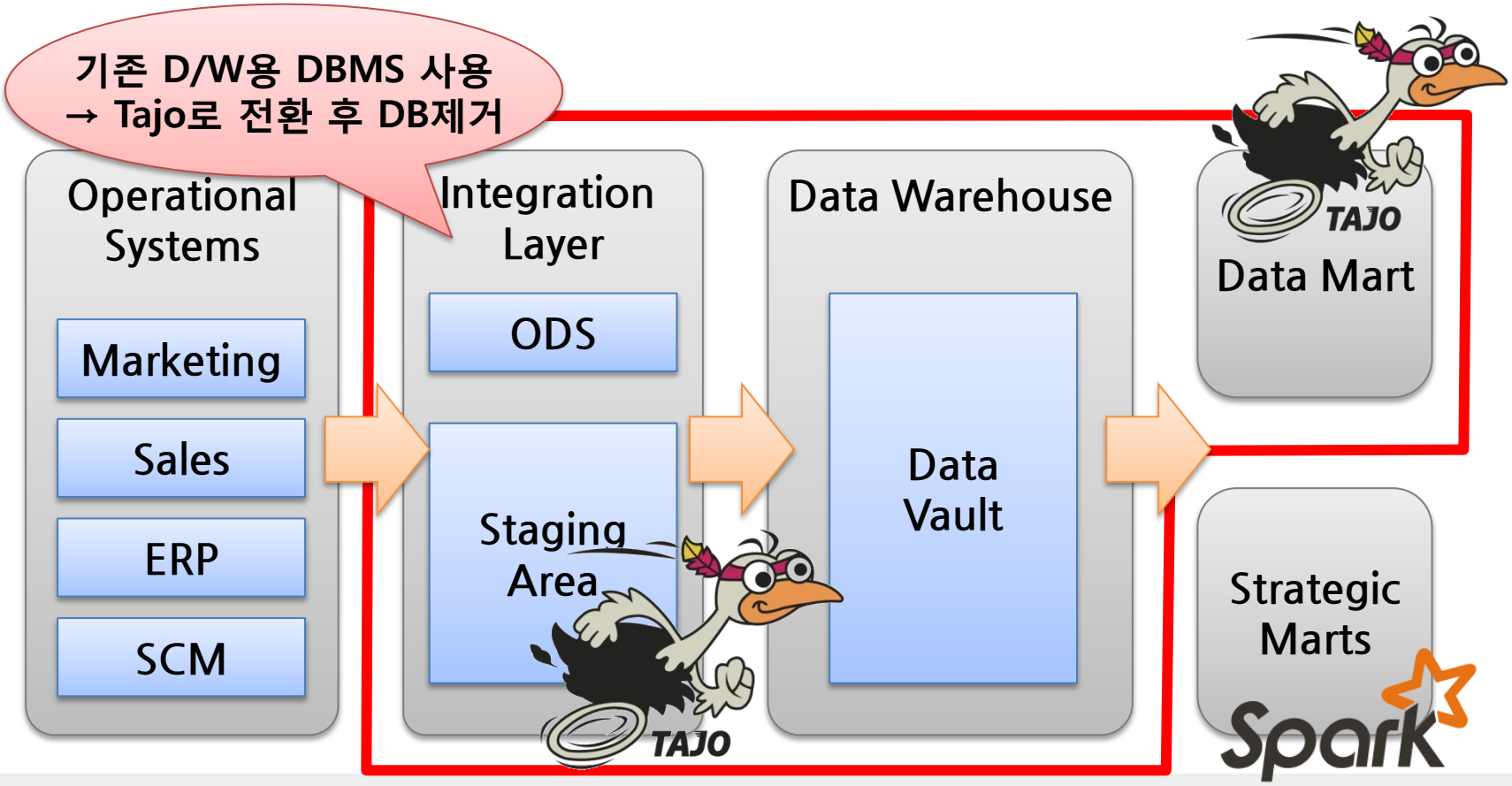
Benchmark Data: TPC-H 100GB
Tajo 0.9.0-SNAPSHOT, Impala: 2.0.0-CDH 5.2.0, Hive: 1.2.0-CDH 5.0.0, HDFS: 2.3.0-CDH 5.1.3

Tajo 적용 사례: 기존 상용DW 대체 (SKT)

수백대의 대규모 Tajo 클러스터에서 페타바이트급 빅데이터 처리에 사용 중

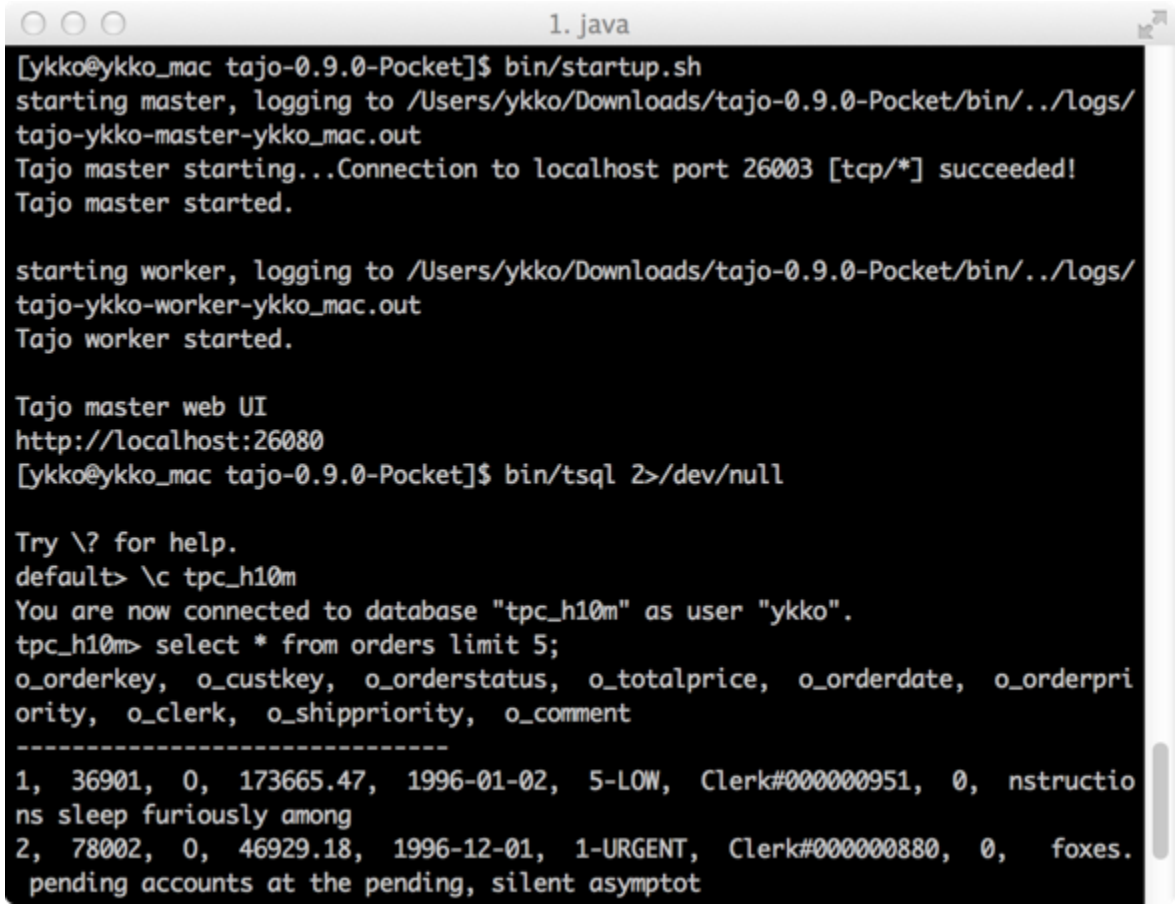
대규모 Batch ETL 처리: Hive를 Tajo로 대체 (120+ queries, ~4TB read/day)

인터랙티브 분석: BI 툴과 Tajo 직접 연결, 상용DW 대체 (500+ OLAP queries)



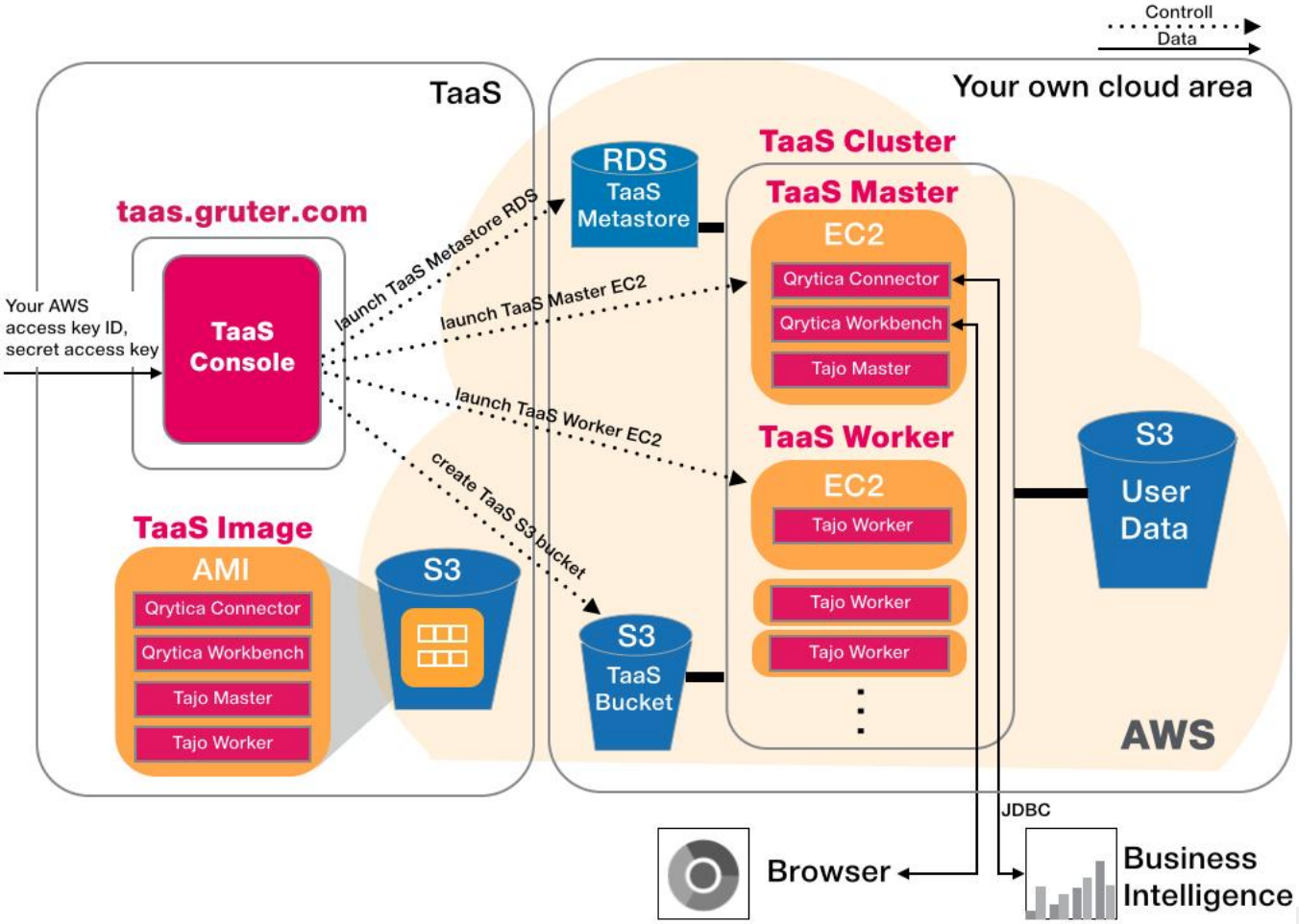
Tajo on Desktop – Tajo Desktop Package

데스크탑에 바로 설치하는 싱글노드 Tajo (http://gruter.com/download.html)
엑셀로 다루기 힘든 큰 데이터를 변환/Load 없이 바로 SQL로 분석



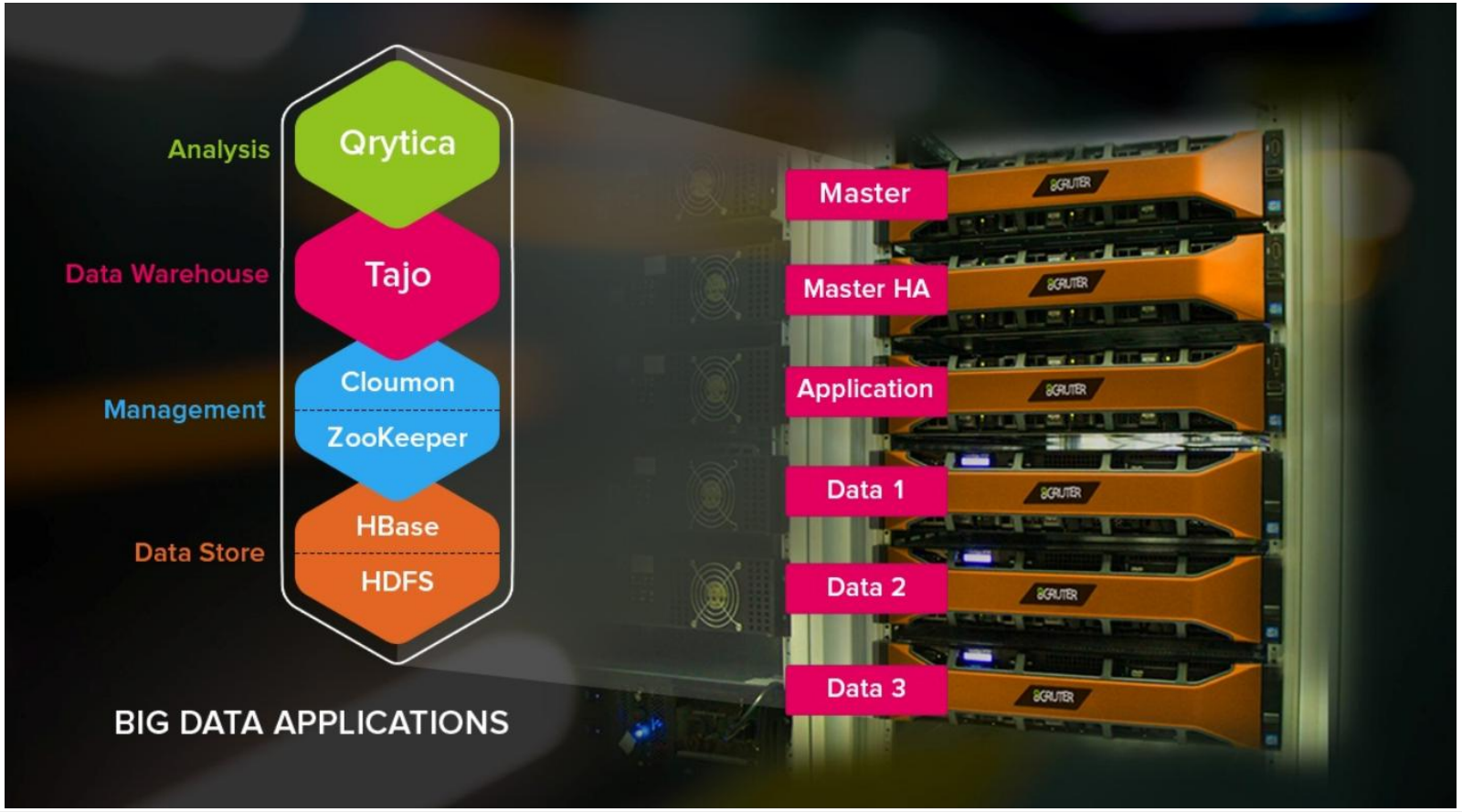
Tajo on Cloud – G-TaaS

클라우드 환경에 최적화된 Tajo를 PaaS 방식으로 제공 (http://taas.gruter.com)
아마존 AWS 베타 테스트 중, MS Azure, Google Cloud 버전도 개발 중



Tajo on Premises - G-DPU

빅데이터 분석 S/W 스택을 미리 탑재한 설치형 빅데이터 플랫폼 장비
최적화된 H/W, 미리 튜닝된 설정으로 플랫폼 구축 시간 단축, 손쉬운 운영 가능
빅데이터 전문인력의 on-site 기술 지원, 실무 교육을 통해 기술 내재화 지원



G-DPU 특징

- **고성능 빅데이터 분석 플랫폼**
  - 빠른 빅데이터 처리를 지원하는 차세대 빅데이터 DW 시스템 Tajo 탑재
  - 수시간 걸리는 대용량 처리와 수초 이내의 인터랙티브 분석을 동시에 지원
- **수평 확장성**
  - 데이터 노드 추가를 통해 선형적인 용량 및 성능 확장 가능
  - 대용량 분산 처리에 최적화된 S/W 스택
- **고가용성**
  - 마스터 노드 이중화로 HA (High Availability) 구성
  - 하둡의 3 복제본 저장 정책으로 일부 노드 장애시에도 데이터 가용성 보장
- **비용 효율성**
  - X86 아키텍처 기반 H/W 구성
  - 오픈소스 기반으로 상용 벤더 제품 대비 라이선스 비용 절감
- **쉬운 설치와 운영**
  - 미리 튜닝/테스트된 설정으로 구축 시간 크게 단축
  - 통합관리도구 Cloumon Enterprise® 기본 탑재
- **기술 내재화 지원**
  - 도입시 on-site 기술 지원 및 아키텍처 컨설팅 제공
  - 빅데이터 실무 교육 프로그램으로 기술 내재화 지원

S/W 스택

Hadoop, Tajo, 분석도구, 관리도구까지 End-to-End 빅데이터 솔루션 제공

S/W Stack	S/W	Version
통합 관리도구	Cloumon	Cloumon Enterprise 2.0
빅데이터 분석 도구	Qrytica	Qryitica 1.0 (탑재 예정)
빅데이터 DW 시스템	Tajo	Tajo-0.9.1
Hadoop 기반 응용S/W	분산 코디네이터 NoSQL *고객 요청 S/W 추가 가능	Zookeeper-3.4.6 Hbase-0.98.4-hadoop2
Hadoop	Apache Hadoop 2	Apache Hadoop 2.4.1
OS	리눅스	CentOS 6.5

H/W 구성

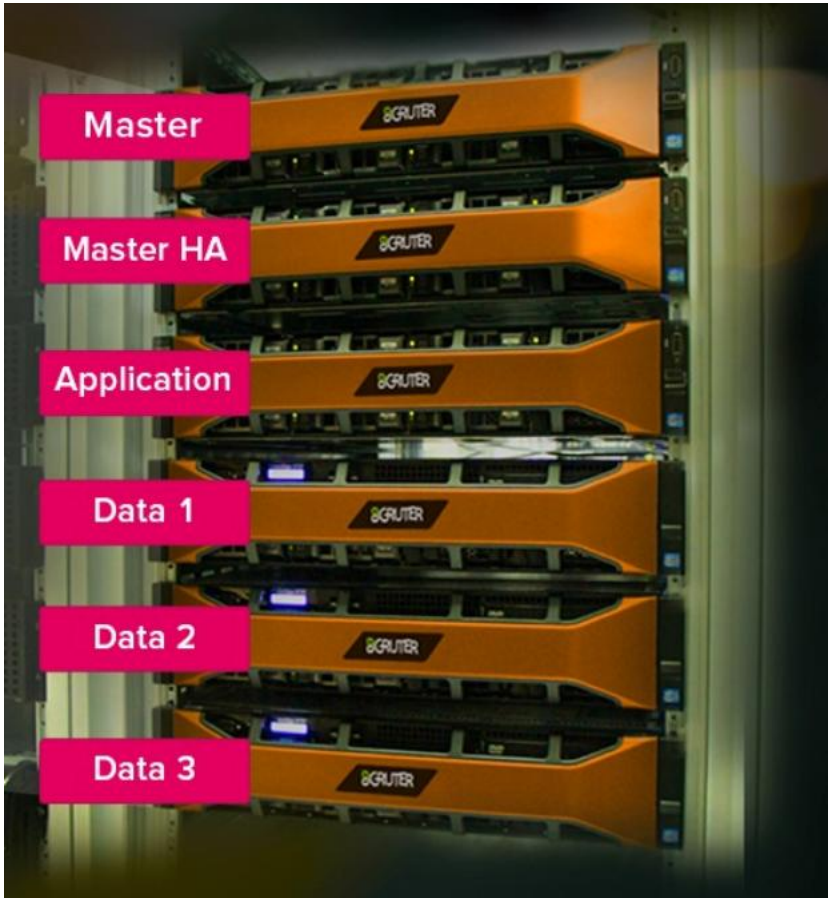
기본 구성 6 노드 1 Set, 데이터 노드 추가로 고객 요구에 따른 용량 ·성능 확장

2U / 2 CPU 지원 범용 x86서버 (3대)
. Master, Master HA, Application 노드용 각 1대
. CPU: Intel Xeon 6 cores, 12 threads
. Memory: 32 GB
. Disk : 4.8TB

2U / 2 CPU 지원 범용 x86서버 (3대)
. 데이터 노드용 3대
. CPU: Intel(R) Xeon(R) 6 cores, 12 threads
. Memory: 48 GB
. Disk: OS용 600GB, 데이터용 30TB
. 데이터 노드는 고객 요청에 따라 추가 가능

Network Switch
. Dell Networking N2024

* 서버 제조사 선택은 고객 편이에 따라 조정 가능



기술 지원 체계

- **Professional Service**
  . 2명 engineer 기본(커미터 급) / 2주 간 10 business day 기준
  . 아키텍쳐 Workshop(5명 내외) : 16시간 지원
  . 하둡 / Tajo 운영 관련 기술 교육 (10명 내외): 40시간 지원
  . 아키텍쳐 수립 및 디자인 적용
  . 하둡 클러스터 최적화 / 안정화 / 현장 튜닝

- **기술 지원 세부**
  . 24X7 / 전담 기술 인력의 장애 접수 및 현장 방문

  . H/W : 최초 장애 접수 후 2시간 이내 전자 우편 / 전화 회신
  최초 장애 접수 후 4시간 이내 현장 방문*
  [OS 및 Part 장애 지원 처리 / HW 공급사가 명기하는 기술 지원 기반 지원]

  . S/W : 최초 장애 접수 후 2시간 이내 전자 우편 / 전화 회신
  최초 장애 접수 후 4시간 이내 현장 방문
  [Hadoop 및 Hadoop Eco 시스템 장애 지원 처리]
  [Tajo 엔진상의 쿼리 장애 지원 처리]

  . 구매등록 고객 전용 Hot line ID 및 Call bridge 제공

* 4시간 이내 현장 방문 해당 지역 : 서울 경기 및 충남, 강원 영서 권
경남 / 호남 / 제주 권역은 +2시간

Gruter – the company behind Tajo

오픈소스 Apache Tajo 의 메인 스폰서로서, 핵심 개발 인력 및 기술력 보유
엔터프라이즈 Tajo 솔루션 개발, 구축, 기술지원 및 프로페셔널 서비스 제공
Tajo를 중심으로 데이터 수집, 저장, 분석, 관리에 이르는 End-to-End 빅데이터 시스템 개발

Scalable SaaS
"Tajo Cloud"

Optimized Appliances
"G-DPU"

Professional Services
"GruterTech"

GRUTER: YOUR PARTNER IN THE BIG DATA REVOLUTION

Phone +82-70-8129-2950
Fax +82-70-8129-2952

Phone +1-415-841-3345

E-mail contact@gruter.com
Web www.gruter.com