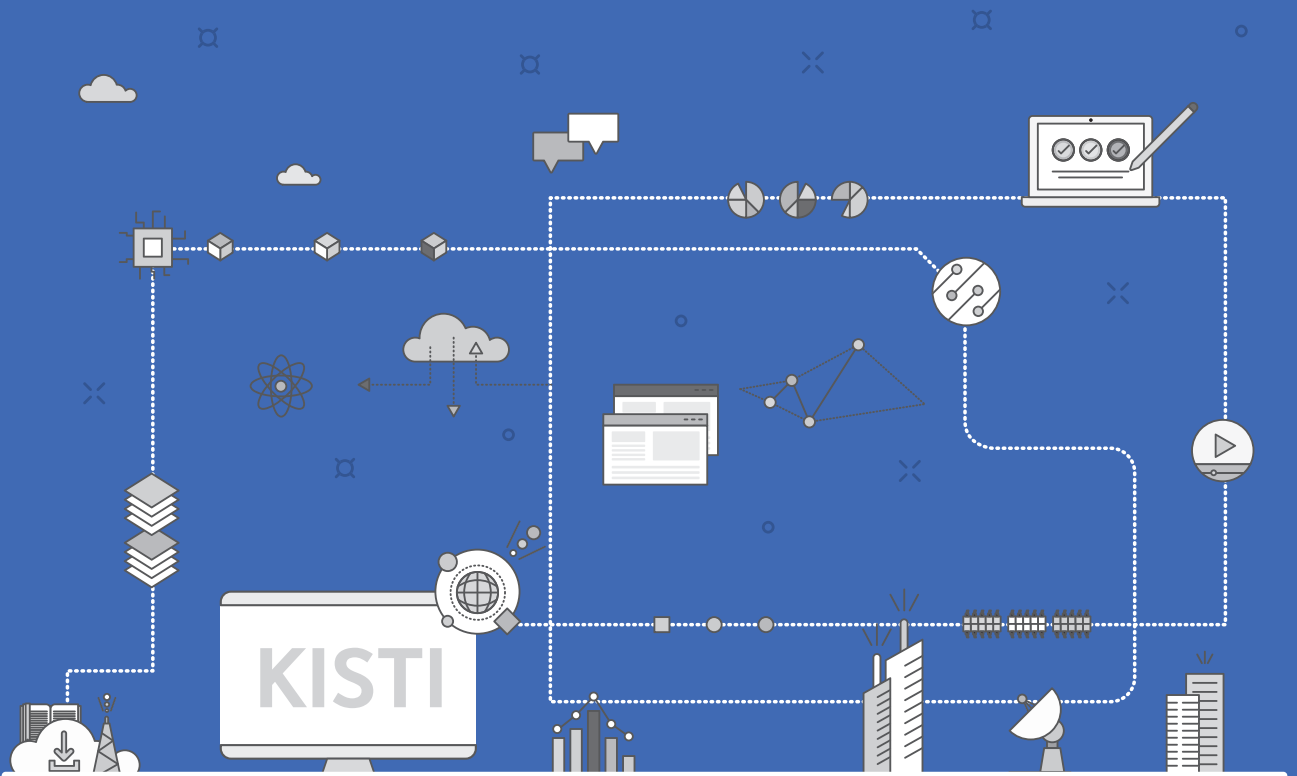


다차원 빅데이터 분석기술(TUPIX)





Keyword 빅데이터 분석, 다차원 빅데이터, TUPIX
연구책임자 박경석
기술 완성단계(TRL) 6단계(시작품 현장 테스트 단계)

기술개요

빅데이터 분석시스템인 ‘투픽스(TUPIX, Turning Pixels into Knowledge and Science)’에 대한 기술로, 클러스터와 같은 빅데이터 처리 환경에서 데이터 처리 및 분석에 소요되는 시간을 줄여 기존의 상용 빅데이터 분석 플랫폼보다 빠른 다차원 빅데이터 분석에 대한 기술임

기존 기술의 문제점

기존 데이터베이스 관리 시스템이나 빅데이터 처리 프레임워크는 다차원 데이터 처리에 있어 비효율적

- 대용량 다차원 데이터를 효율적으로 관리하고 수치 모델 및 기계학습 모델 등을 빠르게 분석할 수 있는 기술 필요

기존의 대용량 다차원 데이터 분석은 데이터 적재 시간이 많이 걸림

- 전체 분석 시간이 크게 지연됨에 따라 이를 혁신적으로 개선할 수 있는 기술 필요

대용량 다차원 데이터 분석을 위한 시간과 관리 비용이 많이 소요됨

- 제한된 컴퓨팅 자원을 효율적으로 활용하여 대용량 다차원 데이터 분석이 가능한 클러스터 기반 플랫폼 필요

기술 내용 및 차별성

빅데이터 처리 환경에서 빠르게 데이터 처리/분석하는
 빅데이터 분석 기술 개발

기술 내용

- 컴퓨팅 노드를 추가하는 scale-out 방식의 확장
- 통계모형, 수리모형, 기계학습 등 분석 알고리즘의 병렬 처리
- 분석 결과를 고해상도로 시각화하고 다양한 방식으로 표현
- 데이터 획득, 분석, 시각화 등 전 과정을 단일 UI로 수행

차별성

- 클러스터 기반으로 다차원 빅데이터의 효율적 분산 병렬 처리 가능
- 다양한 계산 블록을 활용하여 신규 분석 알고리즘의 신속한 개발 가능
- 시각화에 요구되는 데이터만 선택 처리하여 질의 처리 성능 향상
- 대용량 데이터의 체계적 관리와 사용자 간 분석 과정 협력 지원



주요기술 구성 및 구현방법

| 주요기술 구성 |

데이터베이스의 인스턴스가 위치한 노드의 개수에 따라 다차원 데이터 병합

병합된 다차원 데이터를 데이터베이스 내의 배열 구조로 배치

배치된 다차원 데이터를 데이터베이스의 원래 데이터 적재 방식에 따라 재배치

재배치된 다차원 데이터를 대상으로 질의를 처리

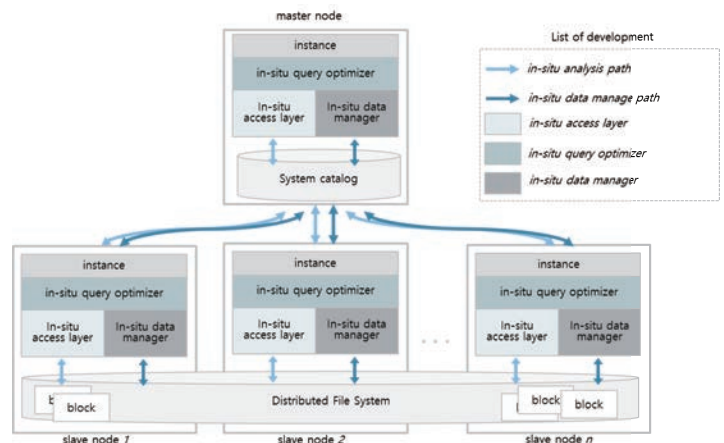
| 구현 방법 |

• 다차원 빅데이터 병합 및 배치 기술 개발

- 분산된 노드에서 Raw 데이터 파일들을 접근하기 위해서는 다수의 파일에 대한 접근이 필요하며 데이터 I/O에 대한 오버헤드(특정한 처리를 위해 소요되는 간접적인 처리 시간 및 메모리 등을 의미)가 큼
- Raw 데이터 파일들을 병합(merge)하여 분산 노드에 배치함으로써 데이터 처리 성능을 향상

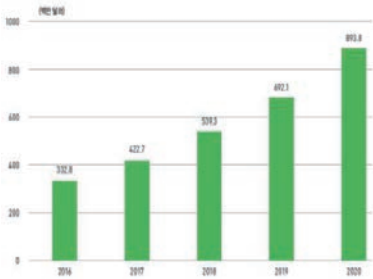
• Raw 데이터를 직접 분석 가능한 in-situ 분석 엔진 개발

- 원본 데이터를 분석 시스템에서 처리 가능한 포맷으로 변환하여 적재할 때까지 분석을 할 수 없는 기존 in-database 기반기술의 한계를 극복
- 데이터의 원형을 보존하고 저장소 간 대규모 데이터 전송을 하지 않고 원본 데이터에 대한 직접 고속 질의, 수확모델 및 고급 분석 알고리즘에 대한 병렬연산이 가능한 기술 개발
- in-situ 분석을 위한 질의 계획 변경 및 in-situ scan operator 적용 기술 개발

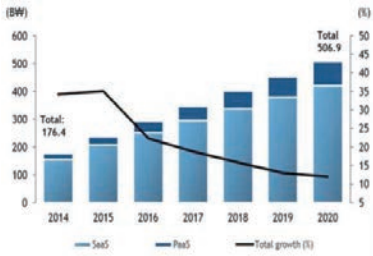


[다차원 빅데이터 분석 기술 (TUPIX) 아키텍처]

기술/시장 동향



[국내 빅데이터 기술 시장 현황]



[클라우드 시장 현황]

활용분야 및 권리현황

| 기술 동향 |

- 기업 전반에서 빅데이터 자체에 대한 관심과 화두가 되던 시기를 지나, 특정 업종 또는 워크로드별로 실질적 인프라를 구현해 나가는 시기로 접어들
- 하둡(Hadoop)과 같은 오픈소스 빅데이터 처리 프레임워크가 비용과 확장성 측면에서 시장의 신뢰를 얻으며, 기존에 보유한 데이터 분석 인프라를 확장하는 방식으로 수요가 구체화 되고 있음
- 전통적 데이터베이스 관리 시스템(DBMS)이나 빅데이터 처리를 위해 개발된 하둡 프레임워크 등은 Array 형태의 다차원 데이터 처리에 비효율적이므로, 수직 모델 및 기계학습 모델 등을 빠르게 분석할 수 있는 기술이 필요

| 시장 동향 |

- 국내 빅데이터 기술 시장 규모 증가
 - 2015년 기준 국내 빅데이터 시장규모는 2014년 대비 30% 성장한 2,632억 원 규모이며, 한국과학기술정보 연구원에 의하면 국내 빅데이터 시장은 2020년까지 8억 9천만 달러(약 1조 원) 규모까지 성장할 것으로 예측
- 정부 투자 증가
 - 이터 관련 정부투자 또한 2013년 230억원에서 2015년 기준 698억원으로 세배 이상 증가
- 클라우드 SW 시장 규모 증가
 - 국내 클라우드 SW 시장 성장은 마이크로소프트(MS), SAP, 세일즈포스닷컴 등 외국계 클라우드 소프트웨어 서비스 회사와 더존비즈온, 한글과컴퓨터 등 국내 기업이 주도하고 있는 것으로 나타났고, 한국 IDC는 향후 5년간 국내 클라우드 SW 시장이 연평균 16.3%의 성장세를 이어가며 2020년 5천억원을 넘어설 것으로 전망

| 기술활용분야 |

기술 수요처	적용처
빅데이터 처리/분석 비즈니스	제조, 물류, 금융, 공공, 지리정보, 영상, 의료, 사회 연결망 등 빅데이터 분석
데이터 기반 과학 응용	위성영상, 기상, 핵융합, 유전체, 화합물, 분자역학, 에너지, 지진 및 해일 등 과학 연구

| 권리현황 |

- 국내 등록특허 3건

발명의 명칭	특허번호	비고
다차원 데이터를 관리하기 위한 장치 및 그 방법	10-1757253	등록
질의를 처리하기 위한 장치 및 방법	10-1757249	등록
집계질의를 처리하기 위한 장치 및 그 방법	10-1757251	등록

추가기술정보

기술분류	빅데이터기술
기술문의	박경석 책임연구원 (슈퍼컴퓨팅클라우드센터) 042-869-1716 gspark@kistire.kr 윤신혜 행정원 (성과확산실 기술이전 담당) 042-869-1832 shyoon@kistire.kr