

빅데이터 활용을 위한 통합 프레임워크 개발

NIPA 공개SW 개발지원사업 kick-off 워크샵



2013.06.05

과제 책임자: 이성준

어니컴(주) / 전략사업팀 / 부장

ONYCOM

목차

- 과제 개요
- 개발 목표
- 세부 개발 내용
- 추진 전략
- 커뮤니티 운영/활동
- 사업화 방안

ONYCOM

공개 S/W 기반
빅데이터 시스템
운영 및 지능 분석

운영
/분석

Big
Data

개발
/구축

Cloudine

Apache Hadoop 기반
빅데이터 시스템
개발 및 최적화

전문
컨설팅

현업의 빅데이터
시스템 도입을 위한
이론/실무 중심 컨설팅

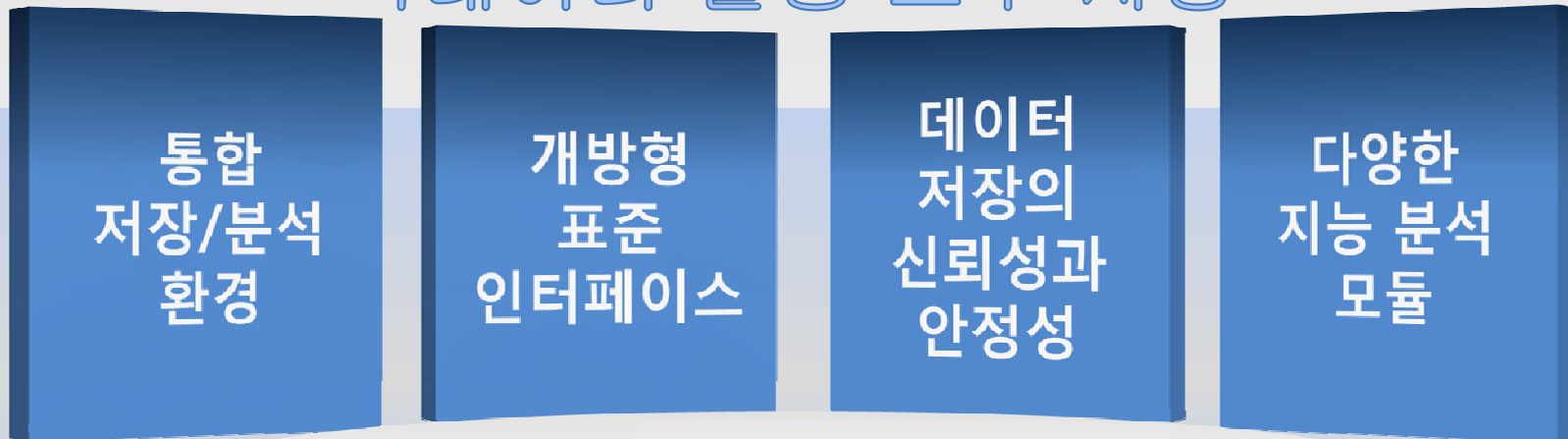
빅데이터 전문가
컨소시엄



과제 개요

- 기업의 빅데이터 기술 도입 비용 감소 및 활용 능력을 증대시킬 수 있는 **빅데이터 고급 분석 기술 확보 및 프레임워크 개발/제공**

빅데이터 활용 요구 사항

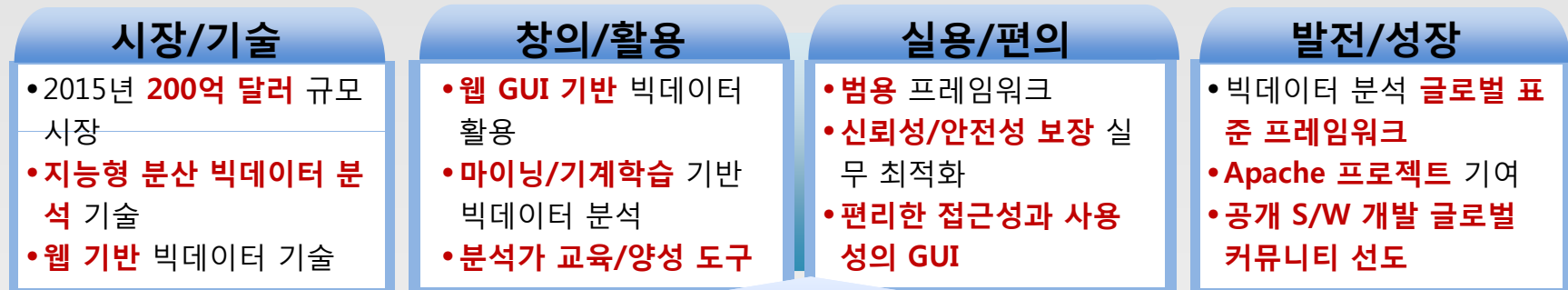


공개 SW 기반 빅데이터 활용 통합 프레임워크

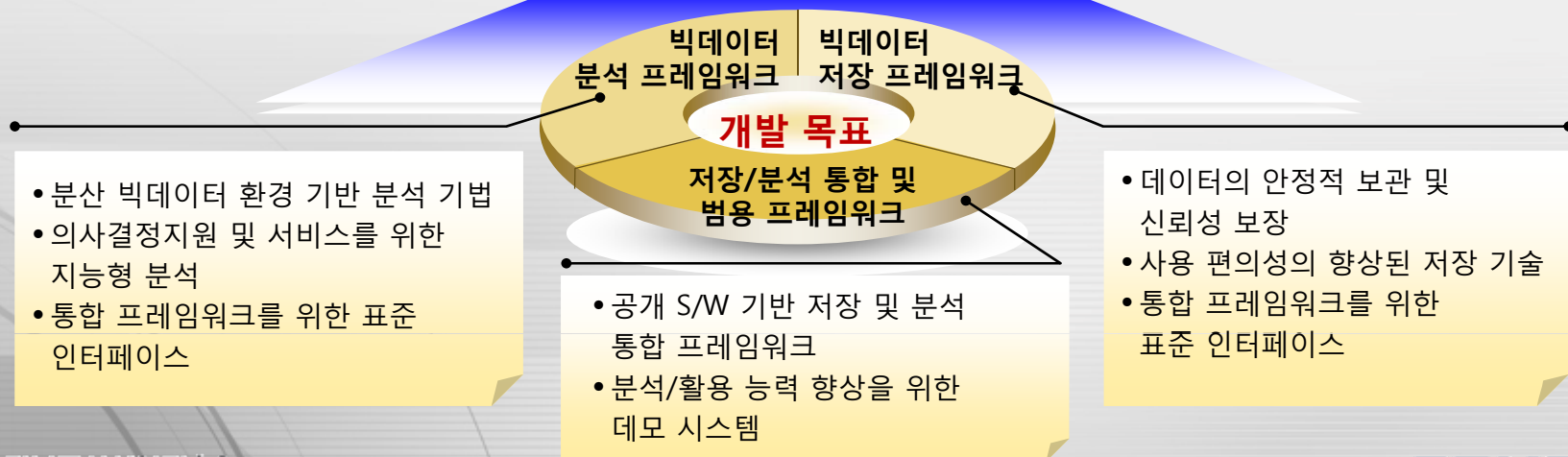
기업의 빅데이터 활용 능력 증대, 글로벌 시장에서의 경쟁력 확보

개발 목표 (1/2)

- 국산 공개 S/W 기반의 빅데이터 저장/분석 등 활용을 위한 통합 프레임워크 개발

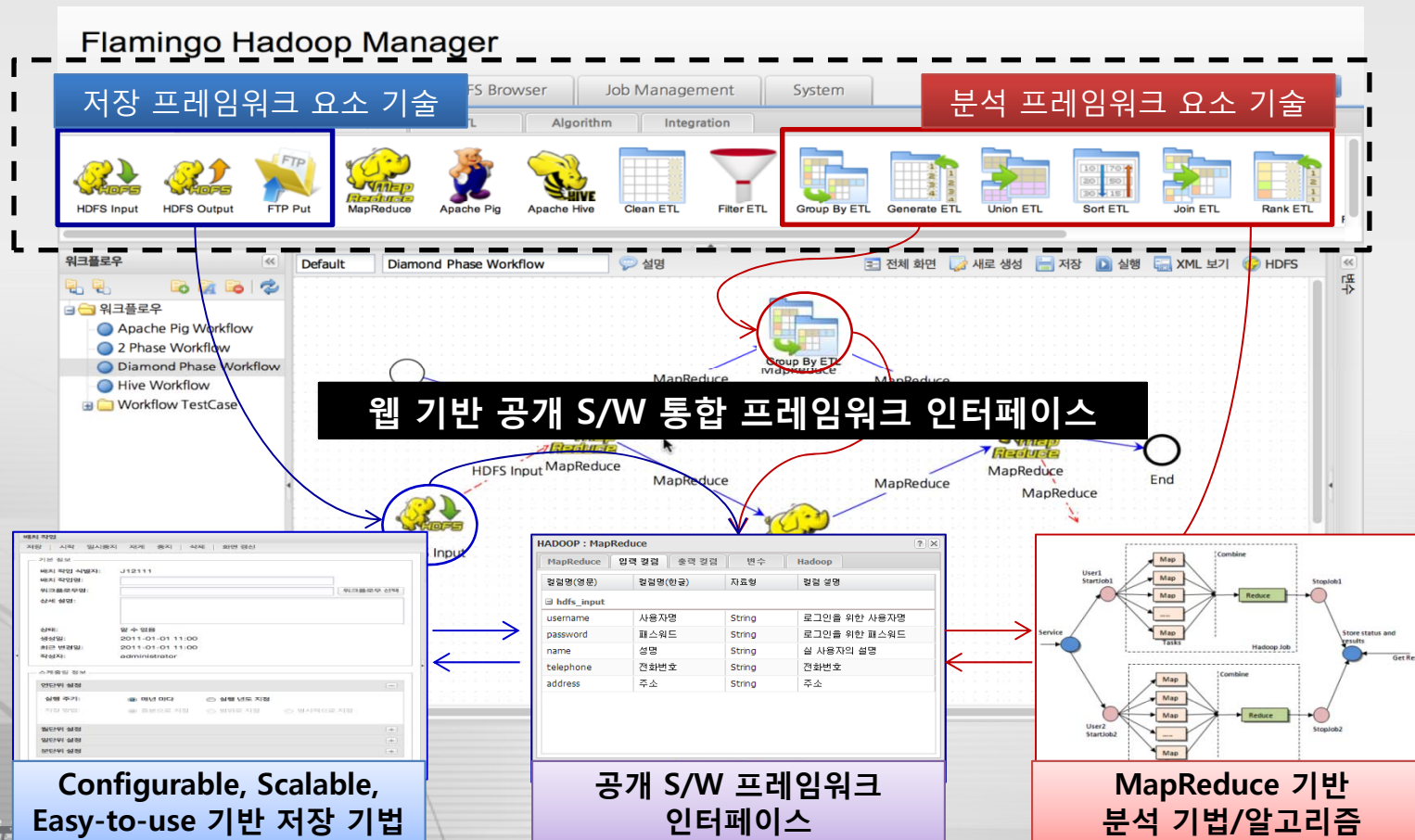


공개 S/W 기반 빅데이터 활용 통합 프레임워크



개발 목표 (2/2)

- 마이닝/기계학습 기반의 **지능형 빅데이터 분석**
- **실무 운용 및 안정성**을 고려한 **빅데이터 저장**
- **웹 기반 GUI**를 활용한 **통합 프레임워크**



세부 개발 내용

• 개발 목표별 세부 개발 내용 및 범위

개발 목표	개발 내용 및 범위
빅데이터 분석 프레임워크 개발	<ul style="list-style-type: none"> ● 통계 기반 데이터 패턴 분석 개발 - 기초 통계 분석 (수치/범주형 데이터 분포 및 통계 분석) ● 특징 벡터 기반 데이터 유사성/거리 분석 개발 - 해밍/자카드/다이스/코사인/피어슨 알고리즘 등 ● 마이닝/기계학습 기반 분석 개발 - 분류/군집 분석 (DT, K-Means, EM 알고리즘 등) ● 실무 서비스 기반 분석 컴포넌트 개발 - 추천 분석 (사용자 기반 협력적 추천, 콘텐츠 내용 기반 추천) ● 분석 결과 가시화 공개 S/W 컴포넌트 연동
빅데이터 저장 프레임워크 개발	<ul style="list-style-type: none"> ● Cron Expression 기반 로그 수집 및 저장 스케줄러 개발 - Scale-out이 가능한 멀티 노드 클러스터링 기반 이중화 수행 및 부하 분산 ● 로그 선별 수집을 위한 패턴 매칭 개발 - : 파일명의 prefix, postfix, 날짜 및 Apache Ant 기반 패턴 매칭 ● 로그 수집/저장 모듈 개발 - 선언적 기법을 활용한 Source(수집) 및 Target(저장) 기반 모듈
저장/분석 통합 및 범용 프레임워크 개발	<ul style="list-style-type: none"> ● 공개 S/W 기반 저장 및 분석 프레임워크 인터페이스 통합 - 기 개발/보유 공개 S/W 기술을 이용한 저장/분석 인터페이스 통합 ● 파일럿 시스템 개발 및 테스트 - Hadoop 클러스터를 이용한 데모 시스템 개발 및 테스트

추진 전략

- 빅데이터 구축/운영/분석 전문성을 보유한 컨소시엄과 외부 그룹 협업

ONYCOM

어니컴(주) - 주관기관

- 빅데이터 분석 프레임워크 개발
- 저장/분석 통합 및 범용 프레임워크 개발
- 마이닝/기계학습 기반 데이터 분석 분야 박사급 전문가 보유
- 웹 기반 빅데이터 관리/운영 S/W 기 개발/보유

- 데이터 분석 및 시스템 운영 학계/실무 전문가 그룹 및 협의회
- 개발 기술에 대한 이론적/기술적 타당성 검증

빅데이터 전문가 그룹

Clouidine

(주)클라우드인 - 참여기관

- 빅데이터 저장 프레임워크 개발
- 저장/분석 통합 및 범용 프레임워크 개발
- 빅데이터 시스템 운영/개발/관리 분야 국내 최고 전문가 보유
- 웹 기반 빅데이터 관리/운영 S/W 기 개발/보유

- JCO, eGov, Jboss User Group, Apache Group 등
- 국내외 커뮤니티 참여를 통한 기술 확산

공개 S/W 커뮤니티 그룹

사업 추진 체계

역할 및 전문성

커뮤니티 운영/활동

- 체계적 사용자 관리를 통한 운영과 온/오프라인 모임을 통한 활동

커뮤니티 운영 방안

- 소수의 커미터 그룹을 통한 전체 코드를 개발/관리
국내 커뮤니티, 기업, 개인, 학계 등이 참여하여 제품에 대한 평가, 기능 추가, 개발 지원
- 커뮤니티 구성원은 '프로젝트 관리자', '커미터', '릴리즈 관리자', '문서 관리자', '커뮤니티 관리자'로 나누어 **구성원별 역할을 정의함으로써 커뮤니티의 체계적 관리** 도모
- 개발활동은 기본적으로 온라인에서 이루어지며, **1. 이슈 매니지먼트 툴을 통한 요구사항 수집, 기능 개선 및 아이디어 공유, 2. 원격지원 도구를 통한 상호지원 및 의사전달, 3. 페이스북 및 메일링 리스트를 이용한 상호 의견 개진**을 이룸

커뮤니티 활동 계획

- 프로젝트에 관심 있는 다양한 참여자들이 손쉽게 참여할 수 있도록 **오프라인 사용자 세미나를 격월 1회 개최**하여 관련 정보 및 진행 상황 그리고 사용법을 공유
- **1개월 2회 이상의 오프라인 운영 미팅**을 통하여 상호 개발 및 발전방향에 대한 의견 공유
- 웹 2.0 기반의 참여를 확대하기 위한 커뮤니티 웹사이트 기능으로 위키, 블로그, 메일링 리스트, 페이스북을 적극 활용하며 **커미터 주도의 Q/A 및 의견 교환**을 추진

사업화 방안 (1/3)

- '빅데이터 시스템 구축'과 '빅데이터 분석'을 중심으로 하는 사업 모델
 - 시스템 구축 - 시스템 구축, 컨설팅, 교육, 기술지원, 응용 개발
 - 분석 - 분석 개발, 컨설팅, 교육



사업화 방안 (2/3)

- 사업화 추진 전략 및 계획



사업화 방안 (3/3)

- 빅데이터 어플라이언스

©Scanrail/Fotolia

‘어니컴-클라우드인’

맞춤형 빅데이터 어플라이언스 출시!!

‘어니컴-클라우드인’ 맞춤형 빅데이터 어플라이언스는 다릅니다!!

- 합리적인 가격 - 시스템과 운용 지원을 하나로!
Global Vendor 공급 어플라이언스 성능 + 전문가의 풍부한 기술 노하우 지원
- 성능 튜닝 서비스 - 업무/프로젝트 성과를 최대로!
실제 업무/프로젝트에 맞는 맞춤형 구성과 빅데이터 분석 성능 최적화 지원
- 개발/운영인력 실무교육 - 고객사 인력을 실전 전문가로!
전문교육센터에서 제공하는 고급 실무 교육을 현장 인력 대상 무상 지원
- 체계적인 방법론 - 업무/프로젝트 진행 시 제대로!
업무/프로젝트 도입 단계에서 개발 및 운영 방법론 제시

*** Big Data All-in-One Service**
맞춤형 빅데이터 어플라이언스를 도입한 고객사 대상으로 빅데이터 분석 교육-기술지원-유지보수 등의 전문서비스를 지원하는 어니컴-클라우드인만의 차별화된 서비스입니다.

맞춤형 빅데이터 어플라이언스를 도입하시면 'Big Data All-in-One Service'를 제공합니다.

onycom Contact us : bigdata@ony.com

전자신문(2013.04.17) 1면

	마스터 노드(2대)		소프트 웨어	버전	
	프로세서	2 CPU, 1 CPU당 4 Core	Apache Hadoop	1.0.3	
	메모리	64G Memory	Apache Sqoop	1.4.3	
	디스크	300G * 2	Apache Hive	0.9.0	
		12TB(2TB * 6)	Apache Pig	0.10.0	
			Ganglia	3.5.0	
			Flamingo Hadoop Manager (필요시)	0.2	
			OS	Redhat Enterprise Linux	64bit 안정화 버전 제공
				CentOS	
				Asianux	
		* 상세 사양은 별도 제공			

감사합니다.

질의/응답



ONYCOM