



Next Generation Big Data Platform SW

# Contents

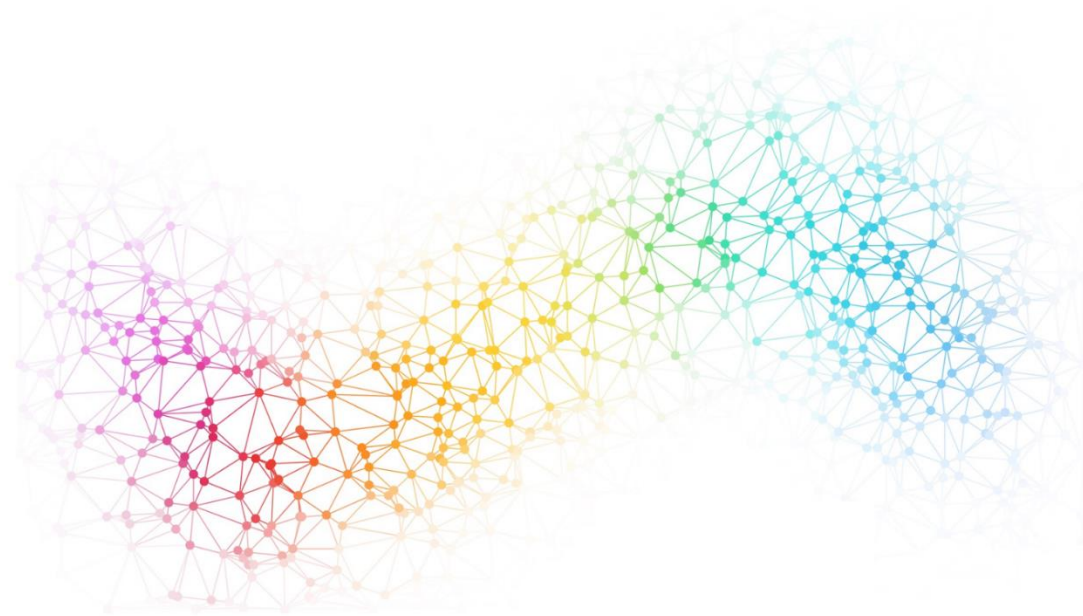
가. NDAP 개요

나. NDAP 소프트웨어 구성

다. NDAP 주요기능

라. NDAP 특.장점

마. 구축사례

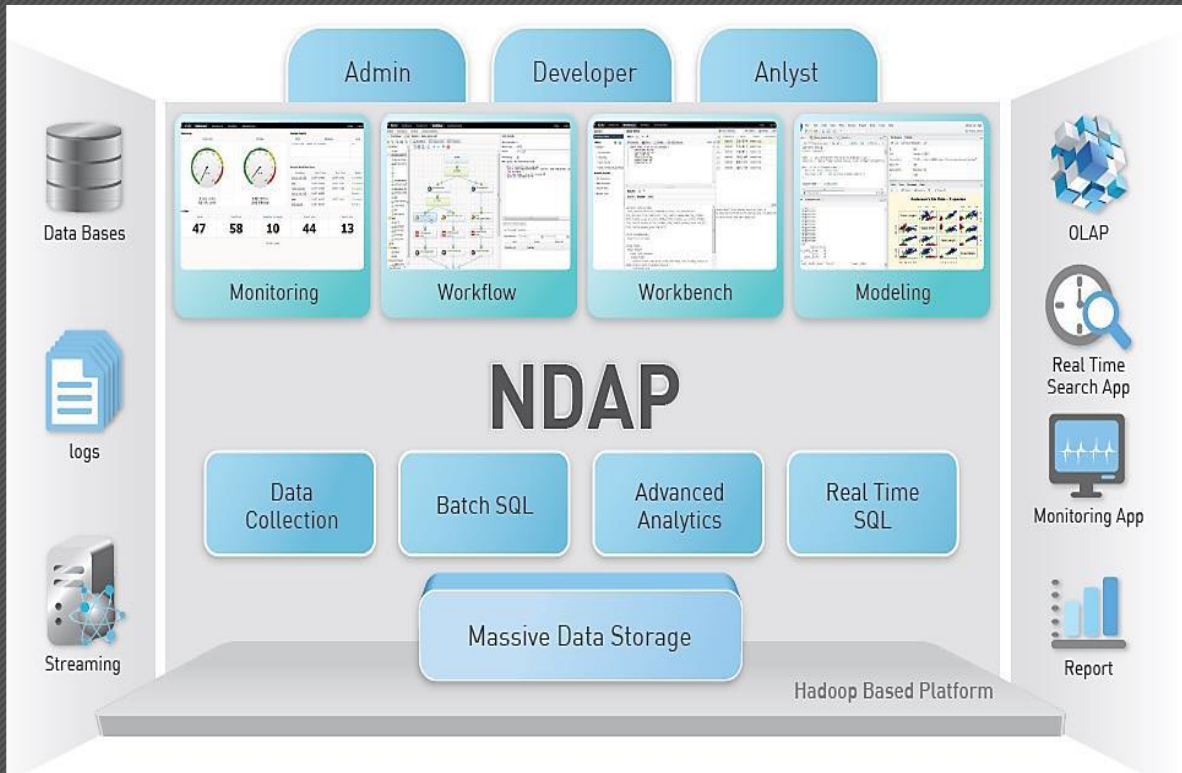


# 가. NDAP 개요

## Data Analytic Platform

Data기반의 비즈니스 의사 결정을 지원하는 빅데이터 솔루션

빅데이터 처리를 위한 **All-In-One 소프트웨어**

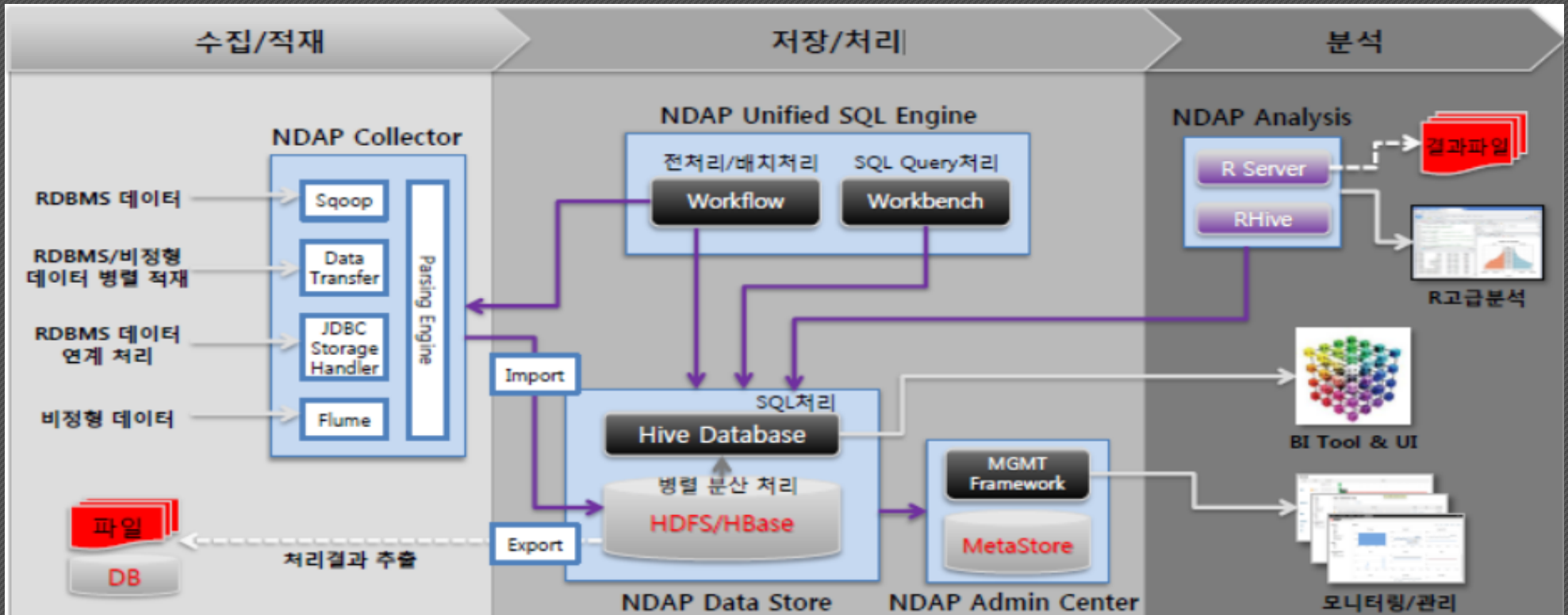


- 1 오픈소스 Hadoop 생태계 기반 **Global Standard**
- 2 x86 기반의 **비용 효율적**인 적스텀
- 3 엔터프라이즈 레벨의 **서비스 가용성**
- 4 용량 및 성능 확장성 보장, 성능 **병목 제거**
- 5 높은 **연동성** 및 **개발 편의성** 제공
- 6 최적화된 **빅데이터 고급 분석** 기능
- 7 다양한 구성 요소, 서비스에 대한 **통합 관리**

# 가. NDAP 개요

## 시스템 기능도

All-in-One 솔루션으로써, 빅데이터 분석 플랫폼을 제공합니다. BDAP의 Collector로 수집, Data Store에서 저장/처리, R + RHive를 통해 분석을 실시하고 고객에게 보다 정확한 분석 자료를 전달 할 수 있습니다.



# 나. NDAP 소프트웨어 구성

## NDAP Software Stack 구성

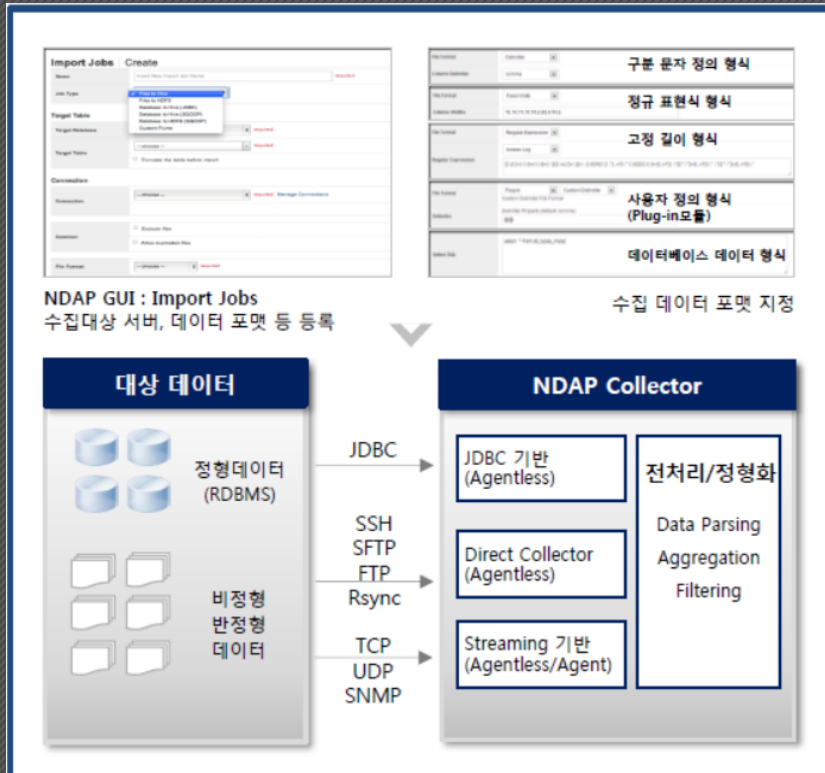
All-in-One 솔루션으로써, 빅데이터 분석 플랫폼을 제공합니다. BDAP의 Collector로 수집, Data Store에서 저장/처리, R + RHive를 통해 분석을 실시하고 고객에게 보다 정확한 분석 자료를 전달 할 수 있습니다.

분석	<b>NDAP Rhive</b> Integration of R and Hive	HDFS에 저장된 전수 데이터를 대상으로 분산 병력처리 방식의 고급 통계 분석 가능
처리	<b>NDAP Unified SQL Engine</b> ANSI SQL Distribute data processing	Ansi-SQL 기반으로 Map Reduce 처리 가능한 분산 처리 엔진
저장	<b>NDAP Data Store</b> HDFS, Map-Reduce, HBase	HDFS 및 Hbase 저장소를 이용한 분산 처리 및 실시간 처리 가능
수집	<b>NDAP Collector / Flume / Sqoop</b> Source data collector Check-pointing for agentless method	다양한 소스의 정형 / 반정형 / 비정형 데이터에 대한 실시간 및 스케줄 방식의 데이터 수집 기능 제공
관리	<b>NDAP Admin Center</b> Distributed coordinator NDAP management, Dashboard	손쉽고 효율적인 빅데이터 인프라 관리 및 분석 / 처리를 위한 직관적인 통합 Dashboard

# 다. NDAP 주요기능

## NDAP Collector - 데이터 수집

데이터 원천으로부터 정형, 반정형, 비정형 등 다양한 형태의 데이터를 수집하며, 처리된 결과 데이터들을 외부로 전달하는 인터페이스 기능을 제공합니다.



## 주요기능

- 다양한 유형의 데이터 수집 (정형, 반정형, 비정형)
- 스트리밍 데이터 수집, Agent/Agentless 선택 구성
- GUI 상에서 수집 작업 등록, 관리, 삭제 수행 지원
- 고성능 분산병렬 처리 및 유연한 Scale-out 구조
- 수집 가용성 환경 제공
- 다양한 수집 프로토콜 지원 (JDBC, SSH, SFTP, FTP, Rsync, TCP, UDP, SNMP)
- SSL 기반의 전송 구간 암호화 지원
- 다양한 포맷 지원을 위한 plug-in 구조의 확장성을 보장하는 전처리 기능
- RDBMS 등 외부 시스템으로의 데이터 전달

# 다. NDAP 주요기능

## NDAP Store - 데이터 저장

대용량 데이터의 저장 및 활용에 적합한 HDFS 및 NoSQL 기반의 통합 분산 저장소 및 저장된 데이터를 편리하게 관리하는 GUI 기반의 통합 관리 기능을 제공합니다.

NDAP GUI : Create New Table  
단일 인터페이스로 테이블 생성 시 목적 별로 Storage Type 선택 : HDFS, HBase 등

**Hive**

- HBase Table
- HDFS Table
- JDBC Table

**HBase**

**MapReduce**

**HDFS**

DataNode

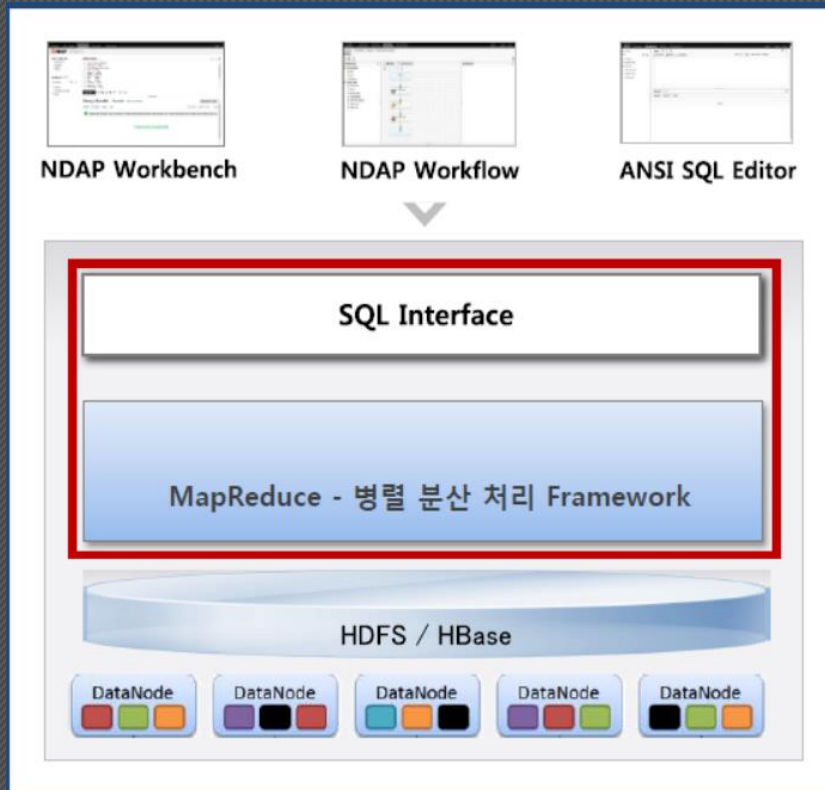
## 주요기능

- Hadoop 2.x(YARN) 기반 데이터 저장 및 처리 최적화
- 비용 효율적인 x86 서버에 블록 단위 분산 저장
- 3-replication 을 통한 데이터 안정성
- 실시간 조회, 처리를 위한 NoSQL 제공
- HDFS, NoSQL 단일 GUI 인터페이스로 제공
- 노드 추가/제거/교체 시 데이터 자동 재분배
- 저장 효율화를 위한 압축 지원
- 표준 보안 알고리즘인 Seed 기반의 데이터 암호화 지원 (128,256 비트 방식)
- 테이블/컬럼에 암호화 설정 지원
- DB 접근 제어를 위해 ACL 기반의 인증 기술 적용

# 다. NDAP 주요기능

## NDAP Unified SQL Engine - 데이터 처리

빅데이터 플랫폼의 집계성 데이터, 예측데이터 등을 다양한 목적에 따라 분석하고 결과를 조회하기 위한 표준 SQL 기반의 사용자 인터페이스를 제공합니다.



## 주요기능

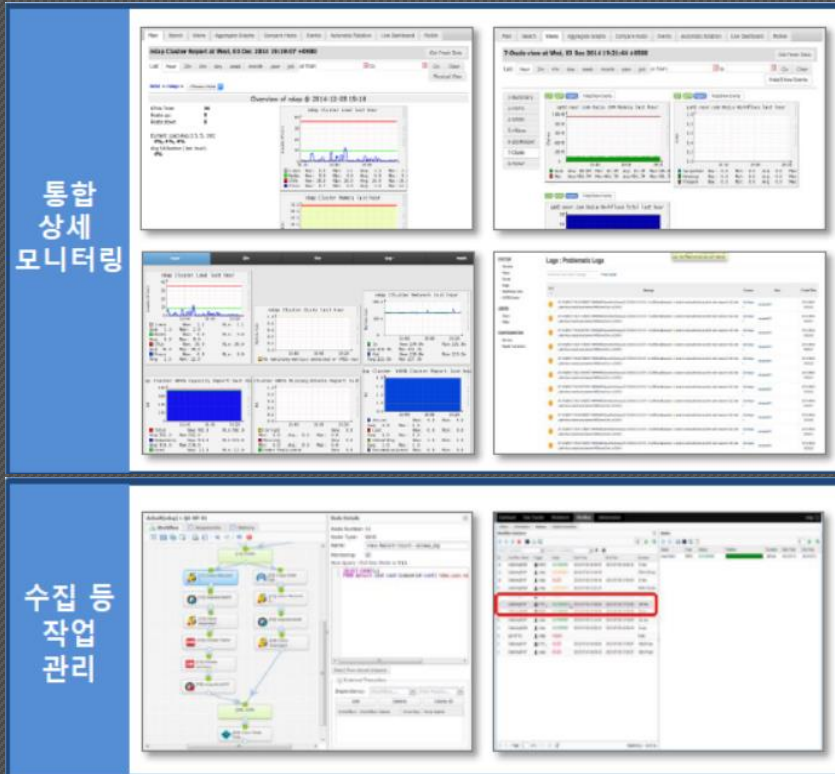
- MapReduce 기반의 병렬 분산 처리
- Hive 기반 표준 SQL 처리 제공으로 개발 편의성 향상
- HDFS / NoSQL 통합 SQL 인터페이스 제공
- 다양한 사용자 정의 함수(UDF) 제공
- GUI 기반 SQL 개발 도구 제공
- GUI 기반 워크플로우 연계 처리 및 작업 스케줄링 도구 제공
- 저장소가 HDFS, Hbase에 상관없이, SQL Interface로 데이터를 처리할 수 있는 Unified SQL Engine 제공 : Hive Storage Handler (Hbase를 Hive 기반 쿼리로 쉽게 저장, 처리할 수 있는 API) 제공



# 다. NDAP 주요기능

## NDAP Admin Center - 모니터링/관리

서버 상태를 포함한 빅데이터 플랫폼의 구성요소와 데이터 수집 및 처리상태들을 모니터링 및 관리하는 기능을 제공합니다.



## 주요기능

### [상태 모니터링 & 관리]

- Ganglia 기반의 Hadoop eco System 모니터링
- 각 시스템에서 제공하는 Metric 정보 제공
- NDAP Service Metric 정보 제공
- CPU/Memory/Disk/Network 모니터링
- 에러/이벤트 등에 대한 통합 필터링 검색

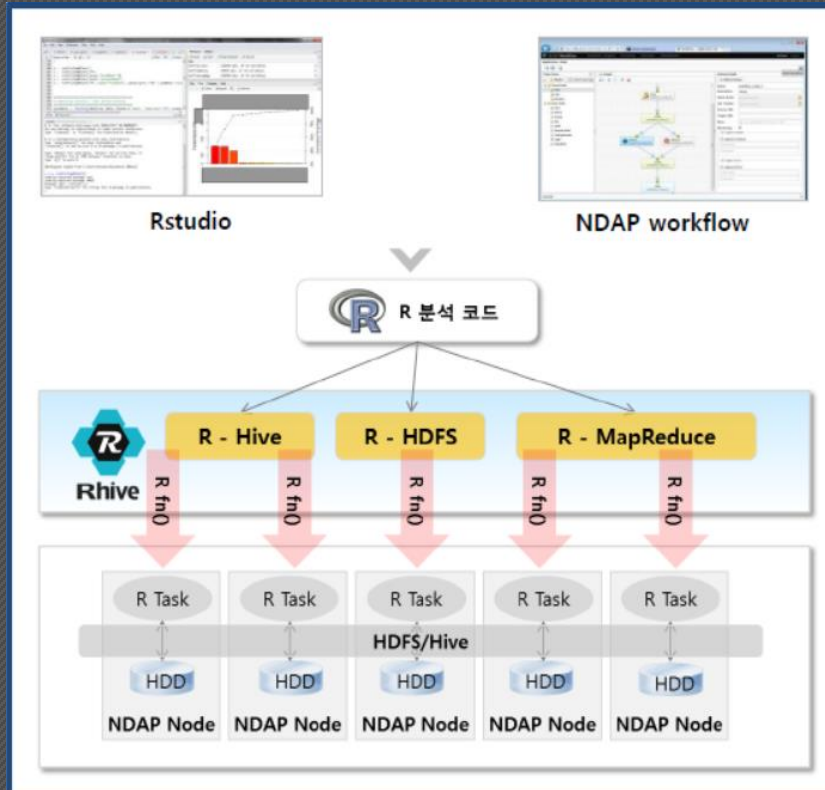
### [수집 등 작업 모니터링 & 관리]

- 단위 작업 간 관계 설정, 순차 실행 관리(Workflow)
- 작업 스케줄링, 소요시간, 에러 상태 관리/모니터링

# 다. NDAP 주요기능

## NDAP R Hive - 고급 통계 분석 (1/3)

다양하고 강력한 통계 분석 및 데이터 마이닝 알고리즘을 제공하는 R과 Hive 기반의 빅데이터 병렬분산처리 기능을 제공하는 고급통계분석 솔루션을 제공합니다.



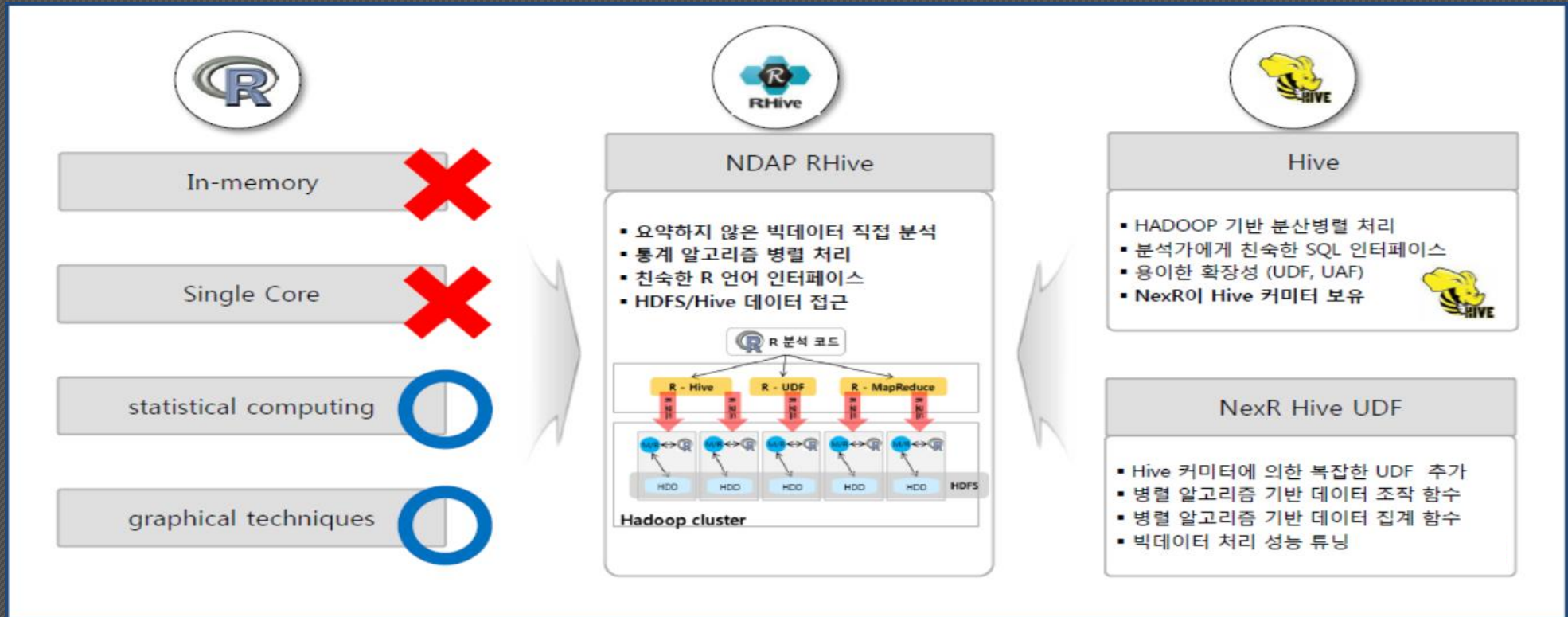
## 주요기능

- 통계분석도구 R 및 ANSI SQL 기반
- 빅데이터 대상 통계 분석, 마이닝 지원
- 통계 알고리즘의 분산 병렬 처리 지원 : 용량 및 처리 시간 한계 극복
- HDFS 및 Hive Query 인터페이스 제공
- R 상에서 MapReduce 프로그래밍 지원
- SSO 연동 지원하는 Rstudio 제공
- NDAP 상의 workflow와의 연동 지원

# 다. NDAP 주요기능

## NDAP R Hive - 고급 통계 분석 (2/3)

NDAP R Hive는 오픈소스 분석 소프트웨어인 R의 약점을 보완하여, 멀티노드 형태로 대용량의 빅데이터 분석이 가능하도록 만든 빅데이터 분석 솔루션입니다.



# 다. NDAP 주요기능

## NDAP R Hive - 고급 통계 분석 (3/3)

빅데이터 플랫폼의 고급통계분석 솔루션인 NDAP R Hive를 통해 대용량 데이터에 대한 패턴 발굴, 예측 등의 고급 분석 환경을 제공합니다.

구분	항목	특징	장점
분석 처리 성능	고속 대용량 연산	분석 알고리즘이 Map/Reduce 기반 분산 환경에서 수행	<ul style="list-style-type: none"> <li>• 복잡한 연산 알고리즘이 병렬 처리</li> <li>• 대량 데이터의 연산의 속도가 빠름</li> </ul>
	빅데이터 처리 능력	빅데이터 용량의 데이터 처리가 가능함	<ul style="list-style-type: none"> <li>• 빅데이터 모델 분석 가능</li> </ul>
모델의 성능	분산 병렬 모델	Map/Reduce 기반의 다양한 분석 알고리즘 지원	<ul style="list-style-type: none"> <li>• R에서 불가능한 빅데이터 데이터 분석이 가능</li> </ul>
	R 호환성	R에서 사용할 수 있는 모든 알고리즘 100% 수용	<ul style="list-style-type: none"> <li>• R의 분석 모델 사용 가능</li> </ul>
	Mahout 통합	분산 병렬 알고리즘 라이브러리인 Mahout 인터페이스	<ul style="list-style-type: none"> <li>• Mahout의 분석 모델 사용 가능</li> </ul>
확장성	분석 모델 이관	PMML을 지원하는 분석 도구 (SAS, SPSS 등), BI 도구 (MicroStrategy 등)의 마이닝 모델의 이관	<ul style="list-style-type: none"> <li>• 기존 분석 도구에서의 모델이 분산 병렬 알고리즘으로 전환 가능함</li> </ul>
모델의 자산화	모델의 자산화	Reproducible Research, Workflow, Model Template Framework 등의 지원	<ul style="list-style-type: none"> <li>• 분석 프로세스의 표준화, 자산화, 공유화 용이</li> </ul>

# 라. NDAP 특.장점 : Domain 적응성

## 사용자 친화적인 기능

NDAP의 다양한 사용자 친화적 기능들은 새로운 Domain에 플랫폼을 적용하고 비즈니스 로직을 구현할 때, Raw Open source 대비 2배 이상의 생산성 향상을 기대할 수 있습니다.

**Job 스케줄 관리**  
스케줄링 Job들에 대한 모니터링 및 상세 상태 확인

**Job 모니터링**  
스케줄링 Job 모니터링 처리 이력 확인

**System Audit**  
기존 Query 수정

**Import Jobs**  
수집대상 서버, 데이터 포맷 등 등록

**수집 등록**  
수집 노드 등록 및 상태 확인

**전체 Job처리 이력관리**  
전체 스케줄링 Job에 대한 이력 관리 및 로그 확인

**수집 데이터 포맷 지정**  
구분 문자 정의 형식  
평균 표현식 형식  
고정 길이 형식  
사용자 정의 형식 (Tag-in/아웃)  
데이터베이스 데이터 형식

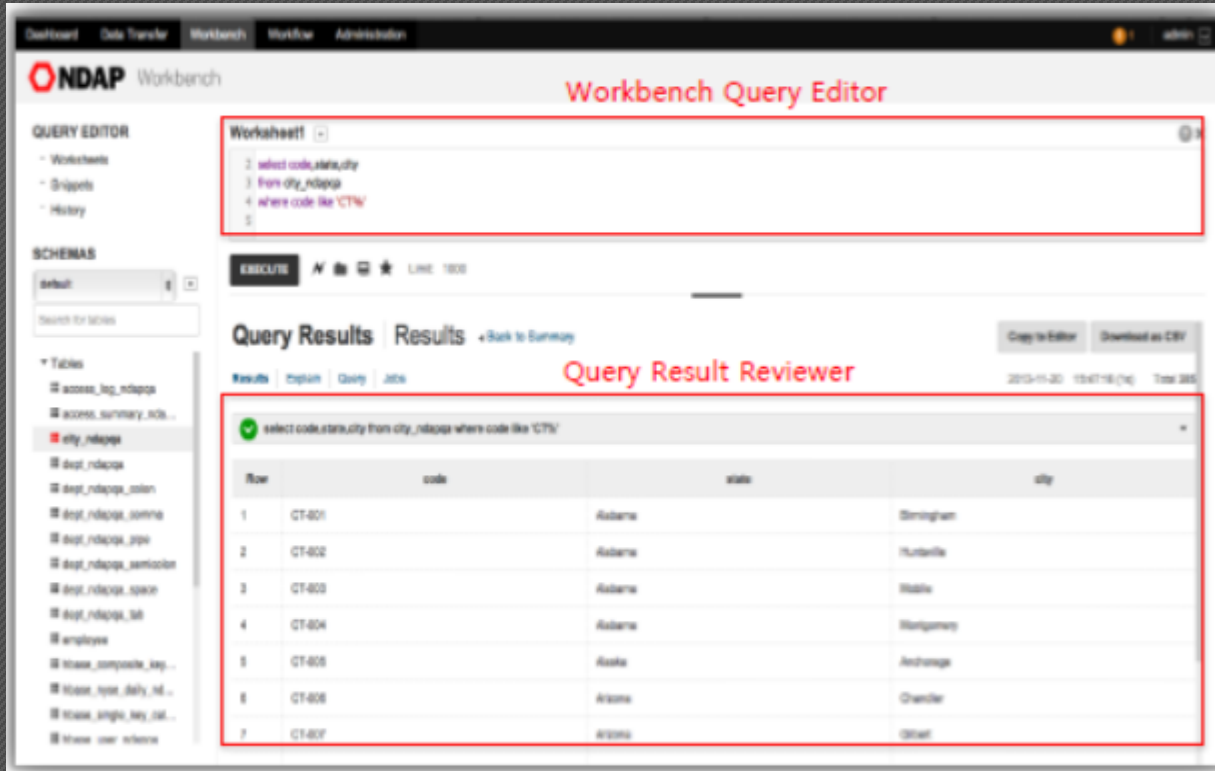
**집속 정의**  
정형/비정형 집속 생성

**데이터 변환 정의**  
수집시 정형/비정형 데이터 매핑 설정

# 라. NDAP 특.장점 : 운영 편의성

## ANSI SQL 처리를 위한 Workbench

표준 SQL 기반의 데이터 처리를 위한 GUI 환경을 제공하며, 저장소의 종류에 상관없이 Hive를 사용할 수 있는 기능을 제공합니다.



- 개발 편의성을 고려해 표준 SQL 기반의 **Hive**를 데이터 처리를 위한 기본 인터페이스로 사용
- Query Editor에서 **Syntax check** 기능 및 **Result Review** 기능 제공
- 데이터를 SQL 기반으로 처리할 수 있는 **Unified SQL Engine : Hive Storage Handler** (HDFS, Hbase, JDBC Connect를 Hive 기반으로 쉽게 처리할 수 있는 API) 제공
- 활용 레벨에서의 유연성을 위해 수십 여 개의 사용자 정의 함수(UDF)를 **API** 형식으로 제공

# 라. NDAP 특.장점 : 운영 편의성

## Workflow 디자이너

데이터 변환 처리를 위해 NDAP Workflow 디자이너에서 분리/병합/필터링등의 다양한 변환을 제공하며, 세부 노드별로 오퍼레이션을 사용자 정의할 수 있습니다.

The screenshot displays the NDAP Workflow Designer interface. On the left, there is a sidebar with a tree view of workflows and nodes. The main workspace shows a workflow diagram with nodes: START, [H] FROM, [H] HIVE, [H] HIVE, [H] HIVE, [H] HIVE, [H] HIVE, and END. A red box highlights the central part of the workflow diagram. On the right, the 'Node Details' panel is open, showing configuration for a node. The 'Web Query' section contains ANSI SQL code for data selection and aggregation. Below the code, there are options for 'ANSI SQL 등록' (ANSI SQL Registration) and 'Interval' (60 minutes).

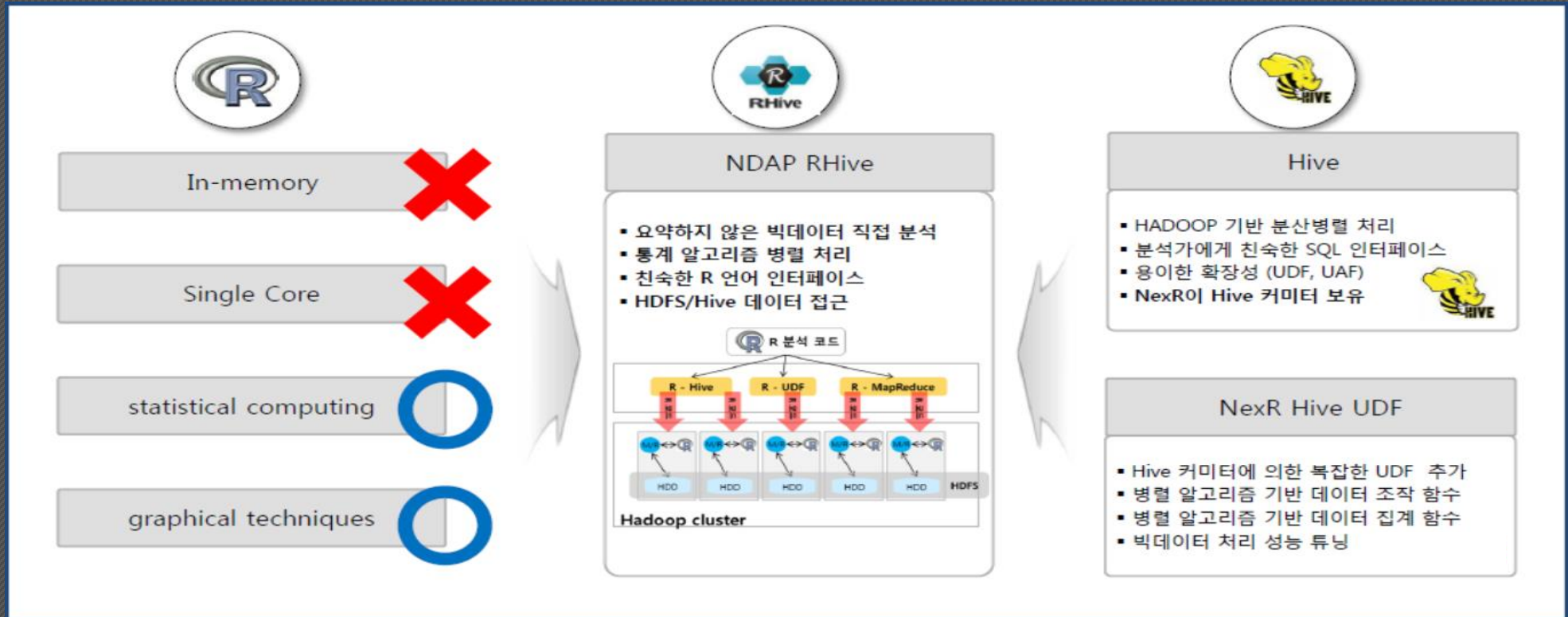
디자인된 Workflow

- Web UI 기반 개발/실행/관리 환경 제공 (Drag&Drop 방식)
- 다양하고 편리한 실행 컴포넌트 제공(Hive, R, R Hive, Shell, External DB)
- Expression Language 지원 자동화 기능 제공
- 프로젝트 단위 Workflow 관리 기능 제공
- 정기 스케줄링, 모니터링, 작업관리 기능 제공
- Workflow 단위 작업으로 ANSI SQL 및 R/RHive 스크립트 등록 가능
- 간편한 데이터 소스 설정(RDBMS, SSH)
- Data Import/Export 기능 제공

# 라. NDAP 특.장점 : 분석 기술

## 대용량 분석 SW NDAP R Hive

NDAP R Hive는 오픈소스 분석 소프트웨어인 R의 약점을 보완하여, 멀티노드 형태로 대용량의 빅데이터 분석이 가능하도록 만든 빅데이터 분석 솔루션입니다.



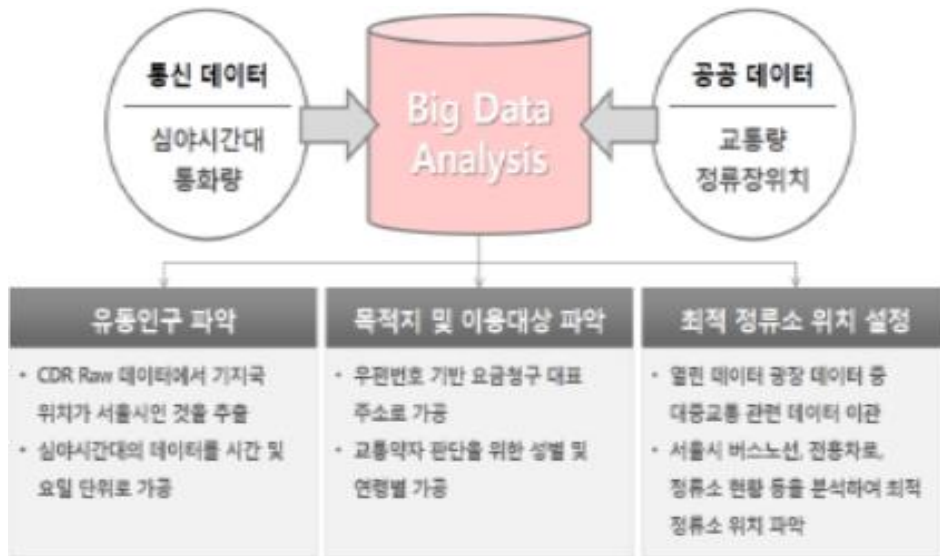


# 마. 구축사례

## 서울시 심야버스 노선수립 지원시스템(1/4)

서울특별시시는 공공 및 민간 데이터를 연계하여 노선을 최적화하여 심야버스 이용 시민의 편의성을 도모하였습니다.

### 현황및이슈



### 주요효과

- 객관적 근거로 최적노선임을 입증
- 심야버스 이용 고객 호응 극대화

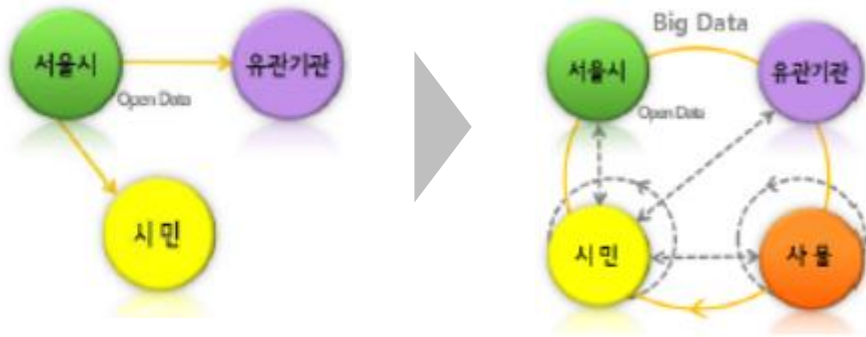


# 마. 구축사례

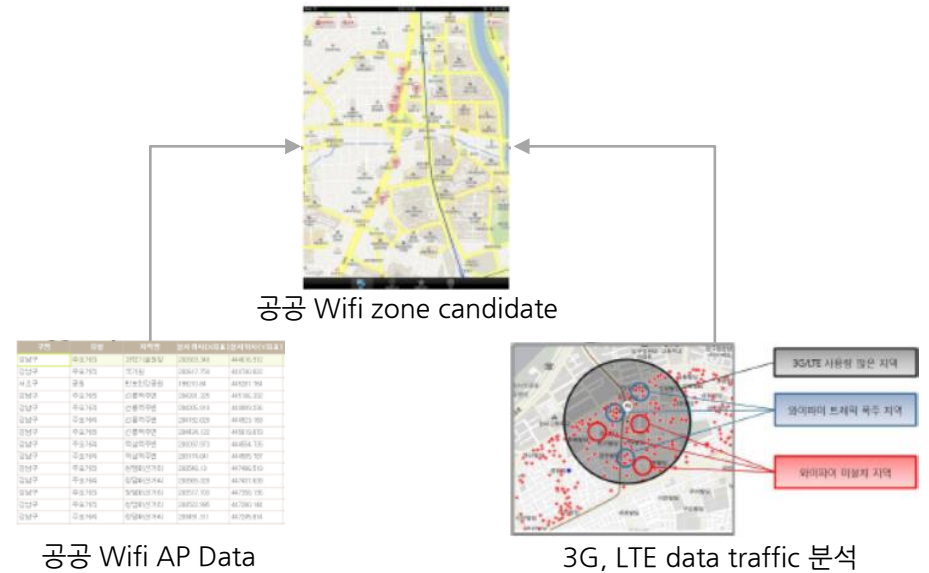
## 서울시 심야버스 노선수립 지원시스템(2/4)

서울특별시시는 시간대별 Wifi, 3G, LTE의 통화량을 분석하여 심야버스 노선을 증가 혹은 변경하여 시민의 안심 귀가 서비스를 제공하고 있습니다.

### 목표



### 방향



# 마. 구축사례

## 서울시 심야버스 노선수립 지원시스템(3/4)

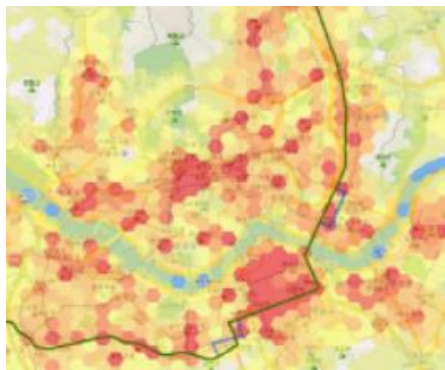
심야시간대의 통화량을 분석하여 유동인구 밀집도를 분석하여 유동인구 기반 노선 최적화를 통한 심야버스의 차량 배차 간격을 조정하였습니다.

### 유동인구 밀집도 분석



- 심야 시간(0시~5시) 통화량 분석
- 구역별 유동인구 밀집도 분석
- 인구 밀집도 색상 시각화

### 유동인구 기반 노선 최적화



- 노선, 시간, 요일별 패턴 분석
- 노선 부근 유동인구 가중치 계산
- 유동인구 기반 노선 최적화

### 유동인구 기반 배차 간격 조정



- 정류장 단위로 통행량 추정
- 통행량을 선의 굵기로 시각화
- 요일별 배차 간격 조정

# 마. 구축사례

## 서울시 심야버스 노선수립 지원시스템(4/4)

심야시간대의 통화량을 분석하여 유동인구 밀집도를 분석하여 유동인구 기반 노선 최적화를 통한 심야버스의 차량 배차 간격을 조정하였습니다.





# 감사합니다



서울시 구로구 디지털로 30길 28 마리오타워7층 701호  
TEL: 02-558-9130 / FAX: 02-558-7868 / [www.in-con.biz](http://www.in-con.biz)