



# Edu **Big** Data

## Education BigData Analysis System

By automating the collection of data, the regularization of various data formats, and database construction, you can understand the collection and storage technologies of big data. You can also enhance the value of your data through visualization of data and high-level statistical analysis

# 목 차

1.

EduData 서비스 개요

2.

EduData 기능 및 시스템 구성

3.

EduData 활용

4.

Company 소개

# EduData 서비스 소개

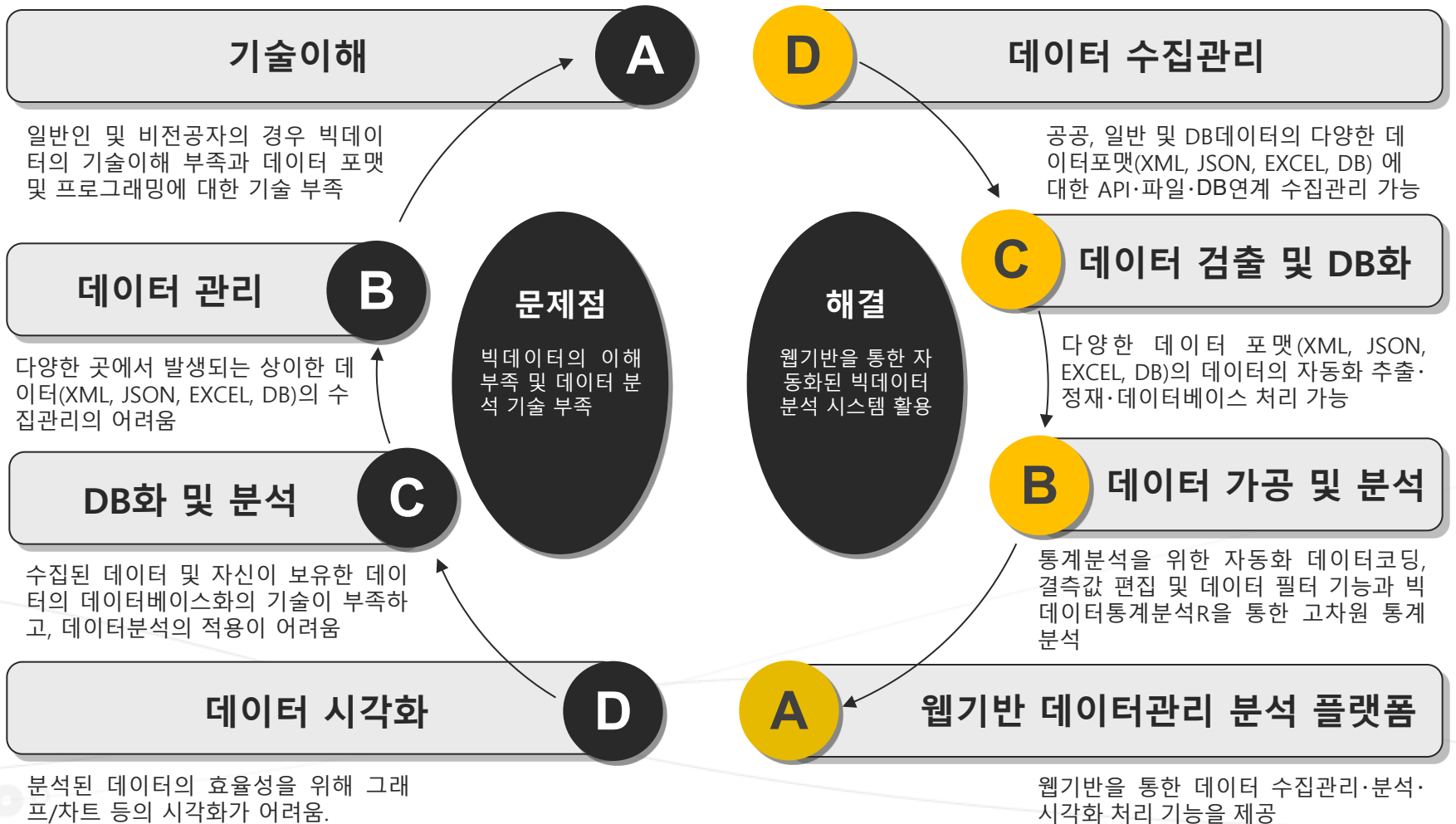
## EduData ?

1. EduData는 "Bigdata for Education"으로 전공자가 아닌, 일반인들과 비전공 학생들에게 빅데이터 분석을 이해할 수 있도록 개발되었습니다.
2. EduData는 데이터의 수집관리, 유형별(XML/JSON/EXCEL/CSV) 데이터를 자동 추출 및 정제가 가능합니다.
3. EduData는 빅데이터 기반의 데이터 저장기술로 자신만의 자동화된 데이터 베이스를 생성을 지원합니다.
4. EduData는 공공데이터 개방 및 활용을 위한 API관리를 지원합니다.
5. EduData는 데이터셋들간의 결합으로 시각화(그래프)생성 기능을 지원하고 있습니다.
6. EduData는 빅데이터 통계R을 접목하여 고차원 통계분석 기능을 지원합니다.
7. SNS 데이터의 수집과 기계학습 및 자연언어처리 기반의 평판분석을 서비스를 활용함으로써 빅데이터분석에 대한 이해도와 활용도를 높일 수 있습니다.

# EduData 서비스 소개

## 데이터 라이프 사이클 지원

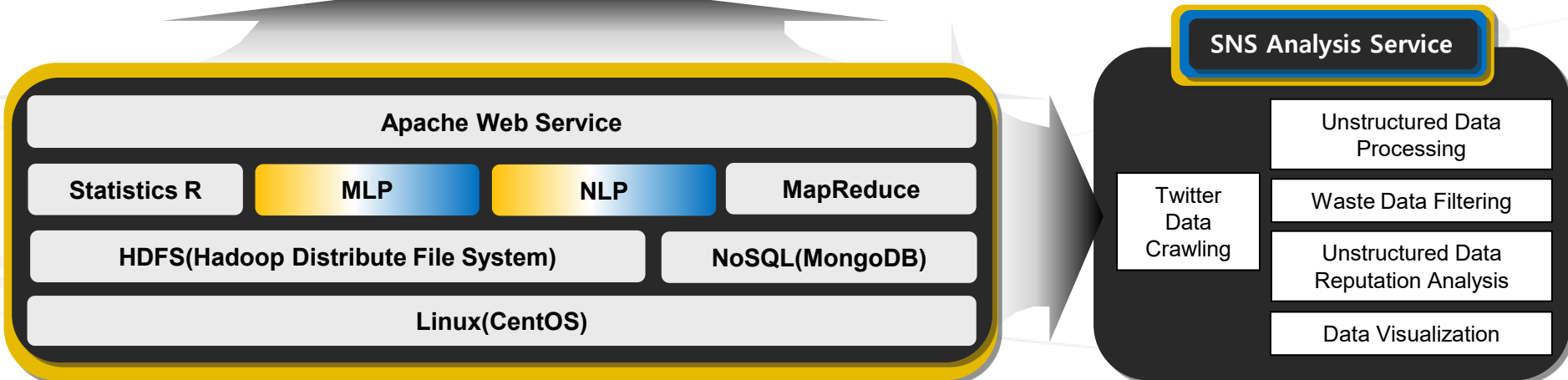
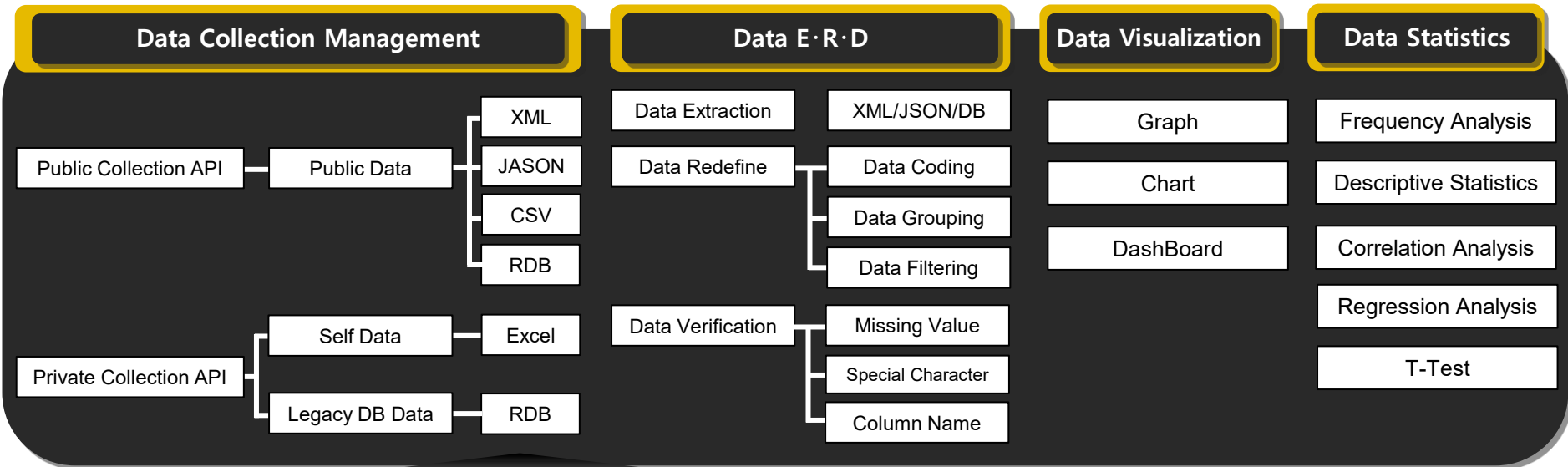
클라우드 웹 기반의 데이터 수집·저장·분석·활용의 전체 데이터 라이프 사이클을 지원하고, 통계·모니터링·시각화 등 데이터 기초분석이 가능하다.



# EduData 서비스 소개

## EduData 시스템 구성

EduData는 클라우드 빅데이터 분석 플랫폼 기반으로 데이터 수집·관리, 데이터 E·R·D, 데이터 검증·필터, 데이터 시각화, 고차원 통계분석 및 SNS 데이터의 수집과 분석의 활용 서비스를 제공합니다.



# EduData 서비스 소개

## 데이터 라이프 사이클 지원

클라우드 웹 기반의 데이터 수집·저장·분석·활용의 전체 데이터 라이프 사이클을 지원하고, 통계·모니터링·시각화 등 데이터 기초분석이 가능하다.

1

### 수집·관리

- XML API 관리
- JSON API 관리
- FILE 관리
- Legacy DB 관리

#### Data 관리

다양한 데이터 수집 사이트로부터의 서로 상이한 API와 파일 및 DB와의 연동 통합관리 기능을 제공합니다.

2

### E·R·D

- Data Extraction
- DB Automation
- Data Redefine

#### Data 가공

유형별 데이터 타입 (XML, JSON)의 자동 파싱과 재가공 및 자동 Database 생성 기능을 제공합니다.

3

### 시각화

- Graph Edit
- Graph Overlay
- DashBoard
- Graph Download

#### Data 시각화

데이터의 그래프생성과 2데이터셋간의 그래프 오버레이 및 대시보드 기능 등을 제공합니다.

4

### 통계분석

- 기술통계분석
- 관계분석
- T검정

#### 고차원 통계분석

통계R을 통한 빅데이터의 고차원 통계분석이 가능합니다.

5

### SNS 활용서비스

- SNS데이터 수집
- SNS 평판분석
- 데이터 시각화

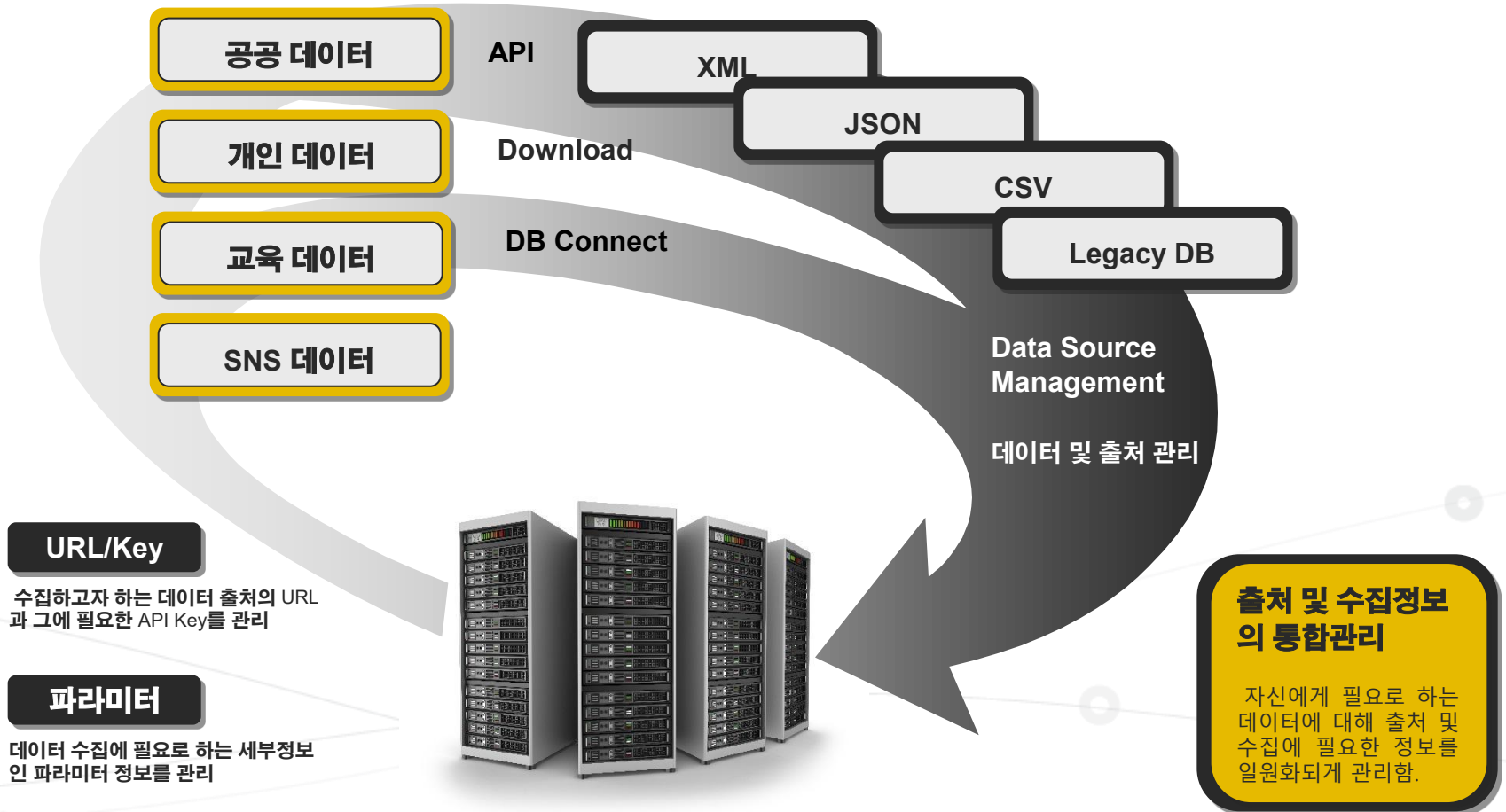
#### 빅데이터 서비스 활용

SNS상의 데이터를 키워드별로 수집, 평판 분석(기계학습, 자연언어처리)하여 빅데이터의 서비스를 활용할 수 있도록 한다.

# EduData 서비스 소개

## 1 데이터 통합 수집·관리

공공 및 SNS 등에서 다양한 형태(XML, JSON, CSV, DB)로 데이터 제공하고 있으며, 이러한 다양한 포맷 데이터의 출처 및 수집에 필요한 정보를 통합관리가 가능하다.



# EduData 서비스 소개

## 2 데이터 E·R·D(Extraction, Redefine, Database)

프로그램 기술적 이해도가 낮은 비전문가들을 위해 수집된 데이터의 추출과 데이터의 재가공 및 자동화된 데이터베이스를 구축하여 시간과 장소에 관계없이 데이터의 정제 및 자산화가 가능하다.

### DATA E·R·D

#### 데이터 추출(Extraction)

##### 데이터 호출(Collection)

기존 OpenAPI, FileData, 레거시DB 데이터를 통해 수집관리된 실데이터를 데이터의 유형별(XML, JSON, File, DB)로 데이터를 호출.

##### 데이터 추출(Extraction)

XML, JSON, CSV 데이터를 웹UI에서 출력하고, 마우스 클릭으로 XML태그와 JSON키값을 선택하여 필요한 데이터만을 추출 가능.

##### 2차원 데이터표현(DataView)

추출된 데이터를 2차원 테이블형태로 변환하고, 추출에서 선택된 태그가 테이블의 컬럼으로 정의.

#### 데이터 재가공(Redefine)

##### 데이터 변환(Conversion)

- 데이터 코딩변환  
통계분석의 전처리작업으로 데이터값을 수치변환(코딩) 가능
- 데이터 그룹핑  
선택항목에 대해 데이터를 그룹핑가능
- 데이터 필터  
선택적 필요데이터와 조건값에 따라 데이터를 필터링 가능

##### 검증(Verification)

- 데이터 검증  
데이터내의 결측값 또는 특수문자로 인한 통계분석 시 오류를 사전 검증
- 컬럼명 검증  
컬럼명내의 특수문자, 공백, 심표, 첫문자숫자 등으로 인한 통계분석 시 오류를 사전에 검증

#### 데이터 베이스(DataBase)

##### 자동 DB생성 (Creation)

추출된 데이터 및 재가공된 데이터를 웹UI에서 DB를 자동 생성.

##### 데이터 자산화(Capitalization)

DB화된 데이터는 시간과 장소에 관계없이 이용 및 활용이 가능.

##### 데이터 다운로드(Data Download)

DB데이터의 CSV파일로 다운로드하여 엑셀 등의 어플리케이션에서 활용 가능.

# EduData 서비스 소개

## 3 데이터 시각화

수집 및 가공된 데이터를 선택하여 사용자에게 맞는 그래프를 생성할 수 있으며, 데이터셋의 그래프를 오버레이한 교차 그래프의 생성과 그래프 및 데이터의 다운로드가 가능하다. 또한 관련 그래프의 대시보드 기능을 제공한다.

### 웹UI

웹UI상에서 데이터View, 그래프 설정View와 그래프 생성View의 레이아웃구성으로 편리한 사용성을 제공

### 데이터 설정

수집된 데이터에 대해 필요한 데이터만을 선택하여 데이터셋을 설정할 수 있음.

### 그래프 설정

2차원 그래프 생성을 위해 x축과 y축에 대한 항목선택, 그래프유형, 라벨 설정 등이 가능.

### 대시보드

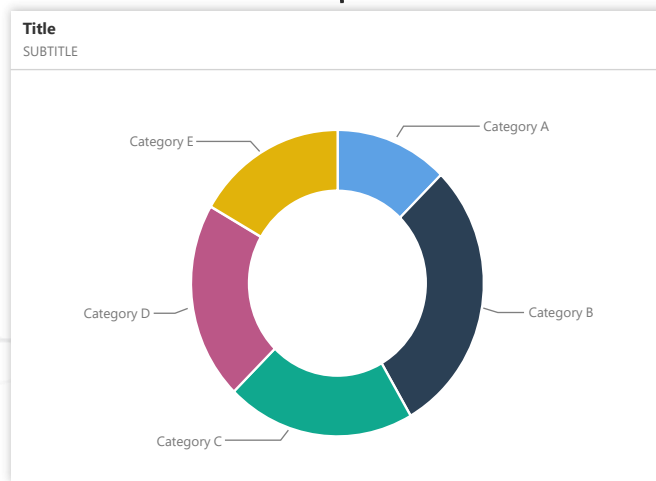
생성된 시각화 그래프에 대해 메인화면상에서의 대시보드 관리가 가능.

### 그래프 유형

Bar, Line, Area, Donut, Step, Pie

### 그래프 오버레이

데이터셋A와 데이터셋B에서 생성된 각각의 그래프를 오버레이하여 상호관계성을 시각화



### 대시보드

사용자별 생성된 그래프에 대해 메인 화면의 대시보드를 설정

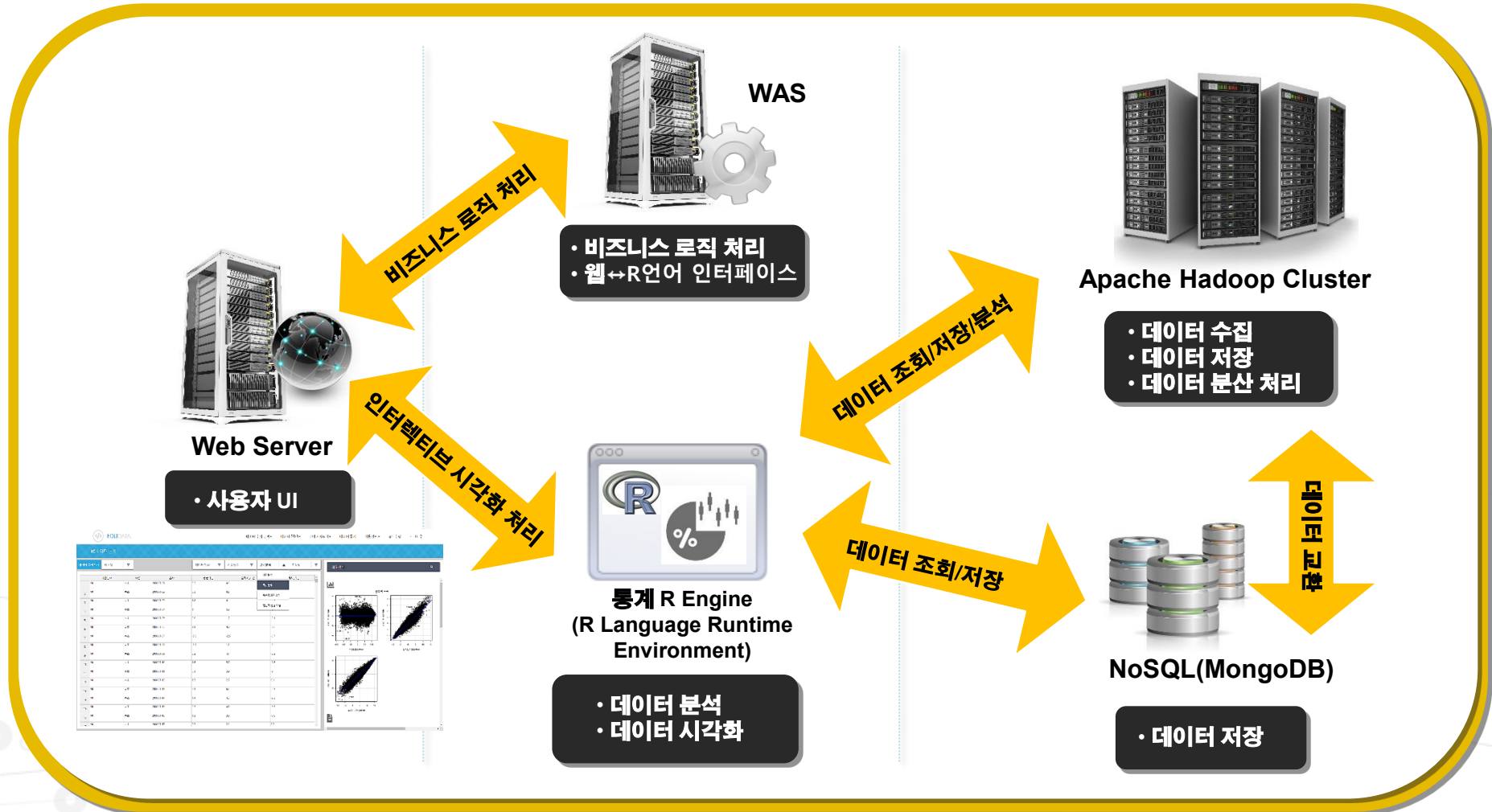
### 기타

컬럼설정, 격자 설정, 라벨설정, 보조그래프, Zoom설정

# EduData 서비스 소개

## 4 고차원 빅데이터 통계분석

웹UI를 통해 수집된 데이터의 고차원통계분석(빈도, 상관, 회귀, 검정 등)이 가능하며 결과값과 그래프를 다운로드 가능하다.



# EduData 서비스 소개

## 5 SNS 데이터 분석 활용

SNS 오피니언 마이닝 기술로, 빅데이터 플랫폼 기반으로 SNS API를 통해 해당 키워드에 대한 비정형 데이터를 자동으로 수집·저장하고, 자연언어처리 및 기계학습기법을 통하여 비정형 데이터의 평판을 분석하도록 합니다.

Machine Learning Analysis



Reporting



Natural Language Analysis



Sensitivity Analysis



MapReduce/Aggregation Search



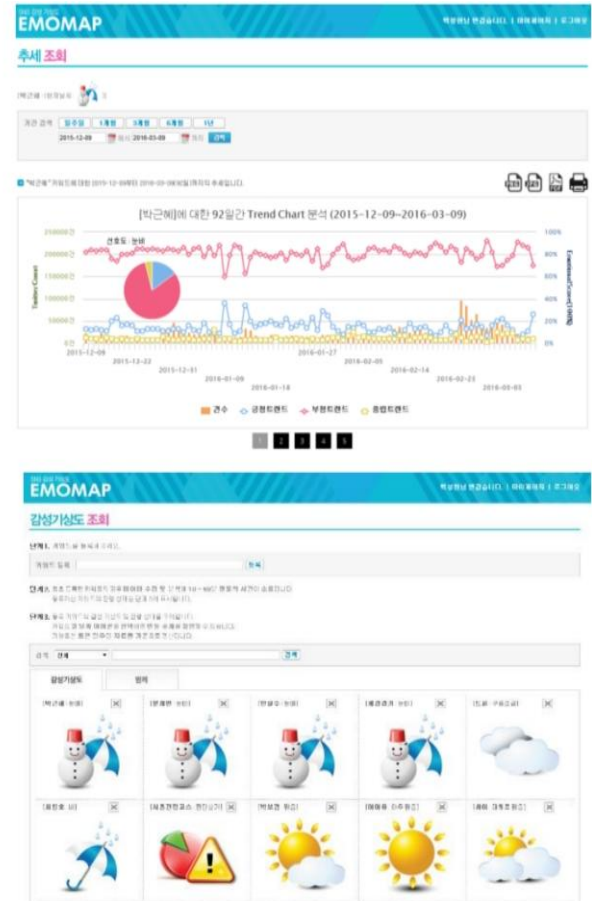
Mongo DB



Crawler API



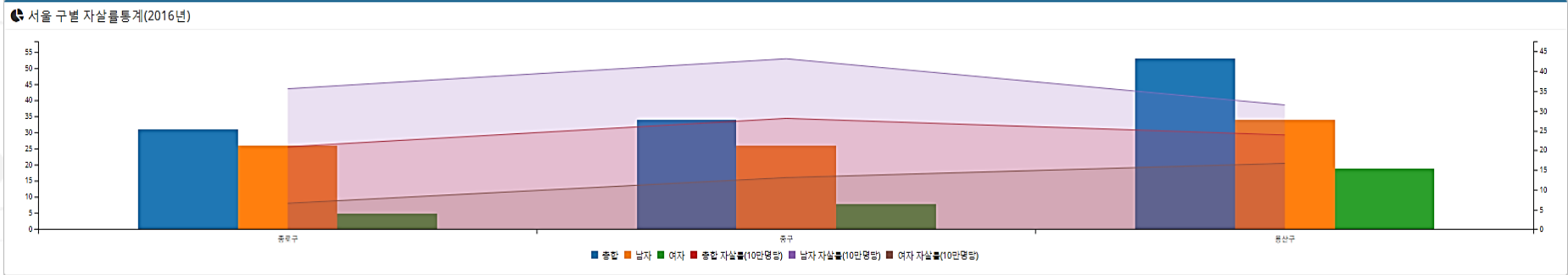
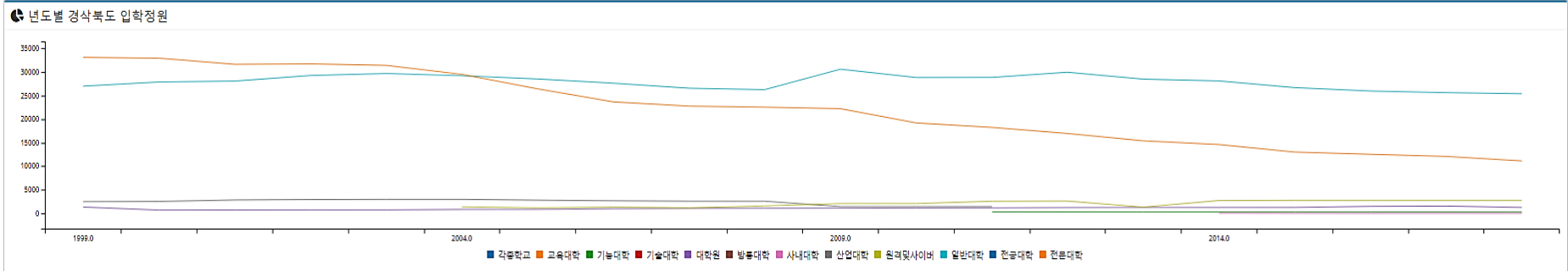
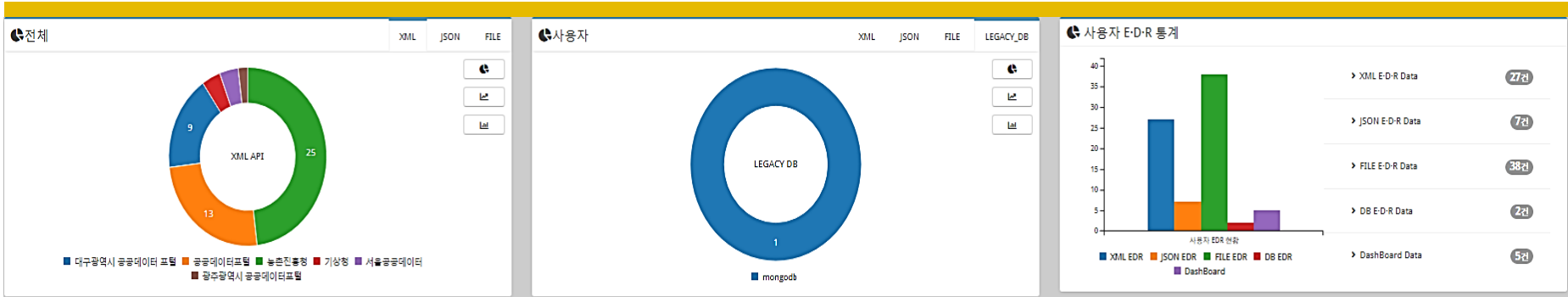
HDFS  
(Hadoop Distributed File System)



# EduData 기능 및 시스템 구성(DASH-BOARD)

## DASH-BOARD

- 공공 및 개인데이터 들에 대한 API와 파일 등의 등록 현황 통계
- 사용자에게 의해 분석된 데이터의 E·D·R(Extraction/Database/Redefine) 통계
- 데이터 시각화 그래프 뷰



# EduData 기능 및 시스템 구성(데이터 수집·관리)

## API 관리

- 외부(공공데이터·기관데이터)로부터의 데이터를 제공받을 수 있는 API를 등록관리
- 제공받는 데이터의 출처 및 API의 URL/API-KEY/파라미터 등의 상세접속 정보를 등록관리
- 관련 기관의 API 기술 문서 등록 관리
- 해당 포맷(XML/JSON)

The screenshot displays the 'XML API' management interface. A modal form is open for adding a new API. The form fields are as follows:

- 출처**: 공공데이터포털
- 제목**: 지하철도착정보
- 상세기능**: 지하철역별 시간표 목록조회
- 제공기관**: 국토교통부
- URL**: http://openapi.tago.go.kr/openapi/service/SubwayInfoService/getSubwaySttnA
- API KEY**: aa9WwJ0xThr5hw%2F%2FM3mPolyeQ7cVxVCywwKGF45E4I4U6b1tG2v4FVsHTil
- 파라미터** table:

변수명	변수값	삭제
subwayStationId	DGS323	X
dailyTypeCode	01	X
upDownTypeCode	U	X
- API 가이드 문서**: api 가이드 문서를 업로드해주세요 (업로드 버튼)

At the bottom of the form, there are buttons for '저장' (Save), '삭제' (Delete), and '취소' (Cancel).

The background shows a list of existing APIs with columns for No., 출처, and API details. The total count is 52.

# EduData 기능 및 시스템 구성(데이터 수집·관리)

## FILE 관리

- 외부 (공공데이터·기관데이터)로 부터 다운로드 받은 파일(EXCEL/CSV)문서 및 내부(자체데이터)의 파일(EXCEL/CSV)문서를 등록관리
- 제공받는 데이터의 출처 및 데이터 파일의 시스템 업로드 등록관리

**FILE API**

출처	공공데이터	🔍
제목	교통사고 발생 시간	
상세기능	상세기능	
제공기관	제공기관	
API FILE	2017년_도로형태별_시간대별_교통사고.csv	업로드

저장    삭제    취소

No.	출처	상세기능	제공기관	API FILE	수정	다운로드	
37	공공데이터				수정	다운로드	
36	공공데이터				수정	다운로드	
35	공공데이터				수정	다운로드	
34	교육센터				수정	다운로드	
33	교육센터				수정	다운로드	
32	기상청	2010년~2016년 전국 월별 통계	admin		VIEW	수정	다운로드
31	기상청	2015년 전국 월별 기온	admin		VIEW	수정	다운로드
30	기상청	2016년 전국 월별 기온	admin		VIEW	수정	다운로드
29	기상청	이항로지스틱 테스트데이터	admin		VIEW	수정	다운로드
28	기상청	2016년 월 평균 날씨	admin		VIEW	수정	다운로드

Total Count 37

« 1 2 3 4 »

# EduData 기능 및 시스템 구성(데이터 수집·관리)

## Legacy DB 관리

1

- 내외부서버(DB서버)의 접속정보를 등록관리
- RDB(MYSQL·ORACLE) 및 NOSQL(MongoDB) 등의 DBMS 연동 관리



데이터 수집·관리 ▾ 데이터 E·D·R ▾ 데이터 시각화 ▾ 데이터 통계 회원관리 ▾ 정보수정 로그아웃


데이터 수집·관리 >> 레거시 DB 관리


레거시 DB 등록


DBMS 전체 DB 명

  
**레거시 데이터베이스 샘플**  
DBMS mongodb  
DB IP localhost  
등록자 admin

**세션 추가**

  
MY SQL

  
Oracle

  
Mongo DB

취소

# EduData 기능 및 시스템 구성 (데이터 수집·관리)

## Legacy DB 관리

2

- 데이터베이스 접속에 필요한 접속 정보를 등록관리
- 데이터베이스 접속확인을 위해 접속 세션 테스트 기능



데이터 수집·관리 ▾ 데이터 E-D-R ▾ 데이터 시각화 ▾ 데이터 통계 회원관리 ▾ 정보수정 로그아웃

데이터 수집·관리 >> 레거시 DB 관리

레거시 DB 등록

### 세션 정보

DBMS	mongodb
DB 설명	레거시 데이터베이스 샘플
DB IP	localhost
DB PORT	37017
DB ID	db_id
DB PW	db_pw
DB NAME	test

🟢 세션연결에 성공했습니다.

세션 테스트

취소

# EduData 기능 및 시스템 구성(데이터 E·R·D)

## API ERD 관리

1

- 등록된 외부(공공데이터·기관데이터)로 부터의 API호출
- 호출된 데이터를 웹페이지에 노출
- 해당 포맷(XML/JSON)



데이터 수집·관리 ▾ 데이터 E·D·R ▾ 데이터 시각화 ▾ 데이터 통계 회원관리 ▾ 정보수정 로그아웃

데이터 E·D·R >> XML E·D·R

XML 불러오기 XML Extraction 데이터 다운로드 XML VIEW 설정

데이터 변환 ▼ 저장

### XML API 호출

서비스 명 (신)등내예보정보조회서비스 - 초단기예보조회

XML API 리스트

파라미터

변수 명	변수 값	삭제
base_date	20180111	X
base_time	0200	X
	88	X

파라미터 추가

**A** 등록된 API 호출

xml 호출

취소

최저임금.시간, 0.9862228550485758

2014 2016

최저임금.시간,

0.9862228550485758

1

# EduData 기능 및 시스템 구성(데이터 E·R·D)

## API EDR 관리

2

- 호출된 데이터의 추출(Extraction)
- 반정형 데이터의 태그 또는 KEY 항목으로 데이터 추출
- 태그 또는 KEY로 추출된 데이터의 데이터베이스 자동화
- 해당 포맷(XML/JSON)



데이터 수집·관리 ▶ 데이터 E·D·R ▶ 데이터 시각화 ▶ 데이터 통계 회원관리 ▶ 정보수정 로그아웃

데이터 E·D·R >> XML E·D·R

XML 불러오기 XML Extraction 데이터 다운로드 XML VIEW 설정

데이터 변환 ▼ 저장

```
<?xml version="1.0" encoding="utf-8" ?>
<rss version="2.0">
  <channel>
    <title>기상청 육상 중기예보</title>
    <description>
      기상청 날씨 웹서비스
    </description>
    <language>
      ko
    </language>
    <generator>
      기상청
    </generator>
    <pubdate>
      2018년 10월 28일 (일)요일 18:00
    </pubdate>
    <item>
      <author>
        기상청
      </author>
      <category>
        육상중기예보
      </category>
      <title>전국 육상 중기예보 - 2018년 10월 28일 (일)요일 18:00 발표</title>
      <guid>
        http://www.kma.go.kr/weather/forecast/mid-term_01.jsp
      </guid>
      <description>
        <header>
          <title>전국 육상중기예보</title>
          <tm>
            201810281800
          </tm>
          <wf>
            이번 예보기간에는 고기압의 영향으로 대체로 맑겠습니다. <br /> </>기온은 평년(최저기온: 0~11℃, 최고기온: 13~20℃)보다 낮겠으나, 후반에는 비슷하겠습니다. <
          </wf>
        </header>
      </description>
    </item>
  </channel>
</rss>
```

A XML 데이터 뷰

XML Extraction

location	2 - Depth	X
province	3 - Depth	X
city	3 - Depth	X
data	3 - Depth	X
mode	4 - Depth	X
tmef	4 - Depth	X
wf	4 - Depth	X
tmn	4 - Depth	X
tmx	4 - Depth	X
reliability	4 - Depth	X

B 데이터 추출

확인

# EduData 기능 및 시스템 구성(데이터 E·R·D)

## API EDR 관리

3

- 추출/DB화된 데이터 뷰
- 코딩·결측값·필터·추출등을 통한 데이터 정재(Redefine)
- 정재(Redefine)된 데이터 다운로드
- 해당 포맷(XML/JSON)



데이터 수집·관리 ▾ 데이터 E·D·R ▾ 데이터 시각화 ▾ 데이터 통계 회원관리 ▾ 정보수정 로그아웃

데이터 E·D·R >> XML E·D·R

### C 데이터 다운로드

XML 불러오기 XML Extraction 데이터 다운로드 XML VIEW 설정

	gigan	jachigu	balsaeng_1	geomgeo_1	balsaeng_2	geomgeo_2	balsaeng_3	geomgeo_3	balsaeng_4	geomgeo_4
1	2016	합계	117839	84421	148	147	262	257		
2	2016	종로구	4459	4298	6	5	12	14		
3	2016	중구	4584	3068	3	3	8	7		
4	2016	용산구	4137	2835	4	4	9	10		
5	2016	성동구	3026	2392	4	5	3	5	104	

데이터 변환 ▲ 저장

- 코딩변경
- 결측값 편집
- 데이터 필터
- 데이터 추출

### B 데이터 정재

코딩 변경

gigan  
jachigu  
balsaeng\_1

geomgeo\_3  
balsaeng\_4

변환 유형

기존 변수로 코딩  새로운 변수로 코딩

새로운 변수명

설명

데이터 타입

코딩 변수명

변수 타입

이전 다음

### A 데이터 뷰

# EduData 기능 및 시스템 구성(데이터 E·R·D)

## FILE EDR 관리

1

- 파일 수집관리에서 등록된 FILE 데이터 호출
- 해당 포맷(EXCEL/CSV)

The screenshot displays the 'FILE API LIST' interface within the EduData system. At the top, there are search filters for '등록자' (Registered User), '출처' (Source), '제공기관' (Providing Agency), and '제목' (Title). Below the filters is a table with the following columns: 'No.', '출처', '제공기관', and '제목'. The table lists 10 entries, with the first entry (No. 37) highlighted. A callout box labeled 'A' points to the '등록된 파일 호출' (Registered File Retrieval) button, which is located in the middle of the table. At the bottom of the table, there is a 'Total Count' of 37 and a pagination control showing page 1 of 4. A '취소' (Cancel) button is located at the bottom center of the interface.

No.	출처	제공기관	제목
37	공공데이터		교통사고 발생 시간
36	공공데이터	경찰청	지역별범죄건수(2016년)
35	공공데이터	국토교통부	산업단지현황
34	교육센터		년도별인구수
33	교육센터	교육센터	년도별 입학정원
32	기상청		2010년~2016년 전국 월별 통계
31	기상청		2015년 전국 월별 기온
30	기상청		2016년 전국 월별 기온
29	기상청		이항로지스틱 테스트데이터
28	기상청		2016년 월 평균 날씨

Total Count 37

« 1 2 3 4 »

취소

# EduData 기능 및 시스템 구성 (데이터 E·R·D)

## FILE EDR 관리

2

- 호출된 데이터 뷰
- 코딩·결측값·필터·추출등을 통한 데이터 정재(Redefine)
- 정재(Redefine)된 데이터 다운로드
- 데이터 자동 DB화
- 해당 포맷(EXCEL/CSV)



데이터 수집·관리 > 데이터 E·D·R > 데이터 시각화 > 데이터 통계 > 회원관리 > 정보수정 > 로그아웃

데이터 E·D·R >> FILE E·D·R

파일 불러오기

데이터 다운로드

**C** 데이터 다운로드

	도로형태	시간대	발생건수	사망자수	부상자수	중상	경상	소상
1	터널안	00시-02시	23	0	46	13	33	
2	터널안	02시-04시	57	2	121	31	78	
3	터널안	04시-06시	47	3	96	25	61	
4	터널안	06시-08시	32	2	69	20	44	
5	터널안	08시-10시	49	3	94	21	71	2
6	터널안	10시-12시	53	3	114	27	78	9
7	터널안	12시-14시	42	1	87	18	63	6
8	터널안	14시-16시	41	2	78	7	63	8
9	터널안	16시-18시	40	1	78	28	79	1
10	터널안	18시-20시	41	1	81	29	97	5
11	터널안	20시-22시	49	1	95	28	59	8
12	터널안	22시-24시	60	1	173	47	109	17
13	교량위	00시-02시	37	4	89	26	61	2
14	교량위	02시-04시	79	5	162	32	123	7
15	교량위	04시-06시	88	2	167	30	123	14
16	교량위	06시-08시	91	3	172	42	118	12

**A** 데이터 뷰

데이터 변환 ▲ 저장

코딩변경

결측값 편집

데이터 필터

데이터 추출

**D** 데이터 DB화

코딩 변경

도로형태  
시간대  
발생건수

**B** 데이터 정재

변환 유형

기존 변수로 코딩

새로운 변수로 코딩

새로운 변수명

설명

데이터 타입

코딩 변수명

변수 타입

이전

다음

# EduData 기능 및 시스템 구성 (데이터 E·R·D)

## DB EDR 관리

1

- 등록된 Legacy DB 호출
- 해당 DBMS 유형(RDB 및 NoSQL)

EDUDATA.

데이터 수집·관리 ▾ 데이터 E·D·R ▾ 데이터 시각화 ▾ 데이터 통계 회원관리 ▾ 정보수정 로그아웃

데이터 E·D·R >> 레거시

레거시 DB ED

No.	DB
2	
1	

Total Count 2

데이터베이스

DBMS 전체 DB 명

레거시 데이터베이스 샘플

DBMS mongodb  
DB IP localhost  
등록자 admin

A 레거시 DB호출

Total Count 1

Close

# EduData 기능 및 시스템 구성 (데이터 E·R·D)

## DB EDR 관리

2

- 선택된 레거시 데이터베이스 및 테이블 호출
- 추출할 컬럼항목 선택 및 뷰
- 코딩·결측값·필터·추출등을 통한 데이터 정재(Redefine)
- 정재된 데이터 DB화

EDUDATA.

데이터 수집·관리 ▶ 데이터 E·D·R ▶ 데이터 시각화 ▶ 데이터 통계 회원관리 ▶ 정보수정 로그아웃

데이터 E·D·R >> 레거시 DB E·D·R

DataBase View

레거시 데이터베이스 샘플

_id	db_conn_name	user_id	db_conn_code
5b6ab20958de6ade0986f607	test_db	admin	3caca51a-ac53-463b-939c-a16ad
5b6ba2d341cdaf10087aeadd	test	admin	088cdfa5-35f5-476a-9c02-ec5803
5b6babe641cdaf10087aeae2	test	admin	67fae2f9-adeef-46c5-a262-0b8a88
5b6bac4441cdaf10087aeae5	test	admin	e01a4927-5099-4707-9ae3-3c3ca
5b6bce0141cdaf10087aeb12	test	admin	4655c634-4ed8-4403-a194-4742a

데이터 뷰

데이터베이스

\_id  
db\_conn\_name  
db\_conn\_code  
db\_conn\_type\_input  
db\_conn\_ip  
db\_conn\_pw  
reg\_dt  
upd\_dt

user\_id  
db\_conn\_id  
db\_conn\_info  
db\_conn\_port

추출할 컬럼 선택

데이터 불러오기 Close

Rows 1-5 of 5

« 1 2 3 4 5 » Rows Columns 다운로드 DB 저장

데이터 DB저장

# EduData 기능 및 시스템 구성(데이터 시각화)

## 데이터 시각화

1

- 대상 데이터셋(XML/JSON/FILE/DB)에서 E·R을 통해 정제된 DB 데이터셋 호출
- 화면구분(왼쪽 : 데이터 시각화 VIEW / 오른쪽 : 데이터셋 VIEW)



데이터 수집·관리 ▾ 데이터 E·R ▾ 데이터 시각화 ▾ 데이터 통계 회원관리 ▾ 정보수정 로그아웃

시각화 관리 >> 데이터 시각화

### A 데이터셋 호출

데이터 불러오기

데이터셋 (A) ▾

다운로드 ▾

저장

데이터 설정

그래프 설정

데이터셋 (A) - 년도별 실업률



### C 데이터 시각화 뷰

### B 데이터 셋 뷰

시점	계	15 - 29세	30 - 59세	60세이상
2000	4.4	8.1	3.5	1.5
2001	4.0	7.9	3.0	1.2
2002	3.3	7.0	2.4	1.1
2003	3.6	8.0	2.5	1.0
2004	3.7			1.2
2005	3.7			1.3
2006	3.5	7.9	2.5	1.4
2007	3.2	7.2	2.5	1.4
2008	3.2	7.2	2.4	1.2
2009	3.6	8.1	2.9	1.6
2010	3.7	8.0	2.8	2.8
2011	3.4	7.6	2.5	2.6
2012	3.2	7.5	2.4	2.4
2013	3.1	8.0	2.3	1.8
2014	3.5	9.0	2.5	2.3
2015	3.6	9.2	2.6	2.5
2016	3.7	9.8	2.5	2.6

# EduData 기능 및 시스템 구성(데이터 시각화)

## 데이터 시각화

2

- 융합 시각화를 위한 데이터셋 선택
- 대상 데이터의 그래프 설정
- 시각화(그래프) 뷰
- 생성된 그래프 및 대상 데이터셋의 다운로드 및 대시보드 저장



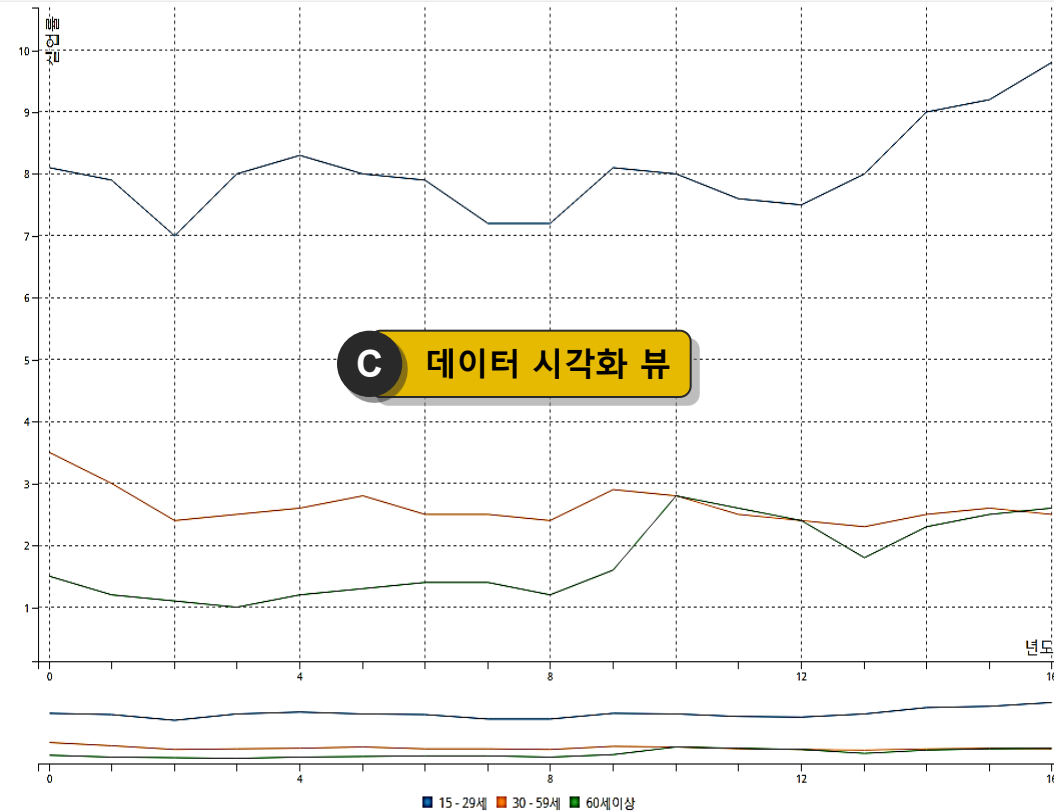
데이터 수집·관리 > 데이터 E·D·R > 데이터 시각화 > 데이터 통계 회원관리 > 정보수정 로그아웃

시각화 관리 >> 데이터 시각화

다운로드

D

데이터셋 (A) - 년도 별 실업률



C 데이터 시각화 뷰

데이터 불러오기

데이터셋 선택

데이터셋 (A)	▲	다운로드	▲	저장
데이터셋 (A) ✓		그래프 다운로드		
데이터셋 (B)		데이터 다운로드		
데이터셋 (C)				

X & Y 컬럼 설정: 설정된 그래프로 표현할 X, Y 축을 설정합니다.

x축 타입: indexed

X축: 시점

Y축: 계

15-29세, 30-59세, 60세이상

↑ ↓

B 그래프 설정

컬럼 설정: 컬럼의 변경과 표현 방식을 설정합니다.

그래프 유형: line

1. 컬럼명: 15-29세

# EduData 기능 및 시스템 구성(데이터 통계)

## 데이터 통계

1

- 빅데이터 통계분석을 위한 분석대상 데이터 호출
- 호출된 데이터뷰와 분석옵션뷰
- 데이터분석 메뉴(데이터변환,기술통계,관계분석,T검정)
- 분석결과 레포팅

EDUDATA. 데이터 시각화 데이터 통계 회원관리 정보수정 로그아웃

데이터분석 메뉴

데이터 불러오기 레포팅

기술통계 관계분석 T검정

**A** 대상 데이터 호출

		15 - 29세	30 - 59세	60세이상
1		8.1	3.5	1.5
2	2001	4.0	7.9	3.0
3	2002	3.3	7.0	2.4
4	2003	3.6	8.0	2.5
5	2004	3.7	8.3	2.6
6	2005	3.7	8.0	2.8
7	2006	3.5	7.9	2.5
8	2007	3.2	7.2	2.5
9	2008	3.2	7.2	2.4
10	2009	3.6	8.1	2.9
11	2010	3.7	8.0	2.8
12	2011	3.4	7.6	2.5
13	2012	3.2	7.5	2.4
14	2013	3.1	8.0	2.3
15	2014	3.5	9.0	2.5
16	2015	3.6	9.2	2.6
17	2016	3.7	9.8	2.6

**B** 데이터뷰·분석결과뷰

년도별	최저임금시간
1	0.9862228550485758
최저임금시간	0.9862228550485758
	1

**C** 데이터분석 메뉴

**D** 분석결과 레포팅

# EduData 기능 및 시스템 구성 (데이터 통계)

## 데이터 통계

2

- 빅데이터 통계분석을 위한 분석대상 데이터 호출
- 호출된 데이터뷰와 분석옵션뷰
- 데이터분석 메뉴(데이터변환, 기술통계, 관계분석, T검정)
- 분석결과 레포팅

데이터 불러오기   레포팅   ▼   기술통계   ▼   관계분석   ▼   T검정   ▼

시점	계	15 - 29세	30 - 59세	60세이상	
1	2000	4.4	8.0	3.5	1.5
2	2001	4.0	7.9	3.0	1.2
3	2002		7.0		
4	2003		8.0		
5	2004		8.3		
6	2005		8.0		
7	2006		7.9		
8	2007	3.2	7.2		
9	2008	3.2	7.2		
10	2009	3.6	8.1		
11	2010	3.7	8.0		1.6
12	2011	3.4	7.6	2.5	2.6
13	2012	3.2	7.5	2.4	2.4
14	2013	3.1	8.0	2.3	1.8
15	2014	3.5	9.0	2.5	2.3
16	2015	3.6	9.2	2.6	2.5
17	2016	3.7	9.8	2.5	2.6

기술통계

- 빈도분석
- 기술통계

관계분석

- 상관분석
- 회귀분석
- 이분형 로지스틱
- 일반화 선형 모형

T 검정

- 일표본 T검정
- 독립표본 T검정
- 대응표본 T검정

# EduData 기능 및 시스템 구성 (데이터 통계)

## 데이터 통계

3

- 빅데이터 통계분석을 위한 분석대상 데이터 호출
- 호출된 데이터뷰와 분석옵션뷰
- 데이터분석 메뉴(데이터변환, 기술통계, 관계분석, T검정)
- 분석결과 레포팅



데이터 수집·관리 ▾ 데이터 E·D·R ▾ 데이터 시각화 ▾ 데이터 통계 회원관리 ▾ 정보수정 로그인

데이터 통계

통계 메뉴 선택 **A**

기술통계 ▲ 관계분석 ▾ T검정 ▾

빈도분석

기술통계

기술통계

변수

시점

계

15 - 29세  
30 - 59세  
60세 이상

옵션

평균

합계

산포도

표준편

최소값

분산

최대값

범위

평균의 오차

분포

정규

왜도

통계분석옵션 **B**

확인

취소

기술통계



	mean	sum	sd	min	se	
15 - 29세	8.047058823529412	136.8	0.7272530913124601	7	0.17638478306107286	0.0280
30 - 59세	2.6294117647058823	44.7	0.2953064211170173	2.3	0.07162232742285836	1.8121
60세 이상	1.7588235294117647	29.9	0.6255585739255475	1	0.15172023972386978	-1.606

통계분석결과VIEW **C**

# EduData 기능 및 시스템 구성 (데이터 통계)

## 데이터 통계

3

- 빅데이터 통계분석을 위한 분석대상 데이터 호출
- 호출된 데이터뷰와 분석옵션뷰
- 데이터분석 메뉴(데이터변환, 기술통계, 관계분석, T검정)
- 분석결과 레포팅



데이터 수집·관리 > 데이터 E-D-R > 데이터 시각화 > 데이터 통계 > 회원관리 > 정보수정 > 로그아웃

데이터 통계

데이터 불러오기 레포팅

기술통계 > 관계분석 > T검정

	지점번호	지점	일자	회계기	회계기
1	90	속초	2000-01-01		회계기
2	90	속초	2000-01-02	5.3	
3	90	속초	2000-01-03	1.8	6
4	90	속초	2000-01-04	1	5.5

통계 메뉴 선택 A

- 상관분석
- 회귀분석
- 이분형 로지스틱
- 일반화 선형 모형

회귀분석

C 통계분석결과VIEW

회귀분석

변수

종속변수: 평균기

독립변수: 지점번호, 일최고\_기온, 일최저\_기온

회귀계수

- 추정값
- 공분산 행렬
- 신뢰구간
- 수준(%) 95

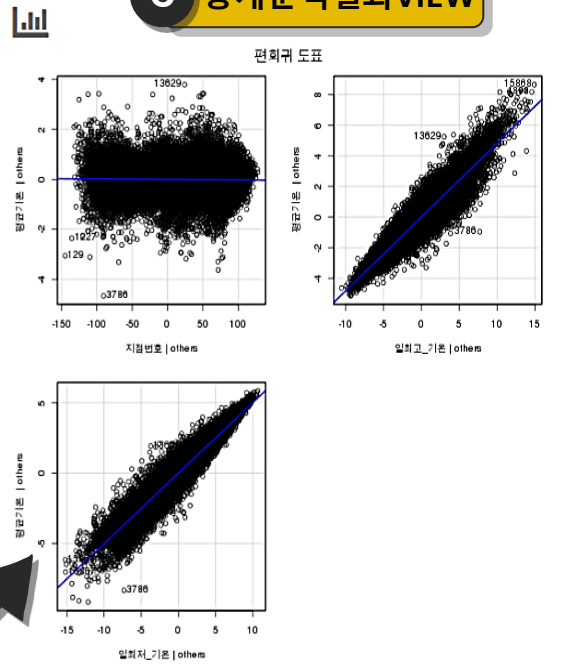
잔차

- durbin watson
- 케이스별잔단

결과 옵션

- 분포적합
- R 제곱량 변화
- 기술통계
- 공선성 진단
- 히스토그램
- 편회귀 도표

통계분석옵션 B



확인 취소

# EduData 기능 및 시스템 구성 (데이터 통계)

## 데이터 통계

3

- 빅데이터 통계분석을 위한 분석대상 데이터 호출
- 호출된 데이터뷰와 분석옵션뷰
- 데이터분석 메뉴(데이터변환, 기술통계, 관계분석, T검정)
- 분석결과 레포팅



데이터 수집·관리 > 데이터 E-D-R > 데이터 시각화 > 데이터 통계 > 회원관리 > 정보수정 > 로그아웃

데이터 통계

데이터 불러오기 레포팅

통계 메뉴 선택 A

T검정

일표본 t검정

독립표본 t검정

대응표본 t검정

일표본 t검정

일표본 통계량

	최소값	최대값	평균값	표준오차
지점번호	90	295	197.6271186440678	62.8323817039107
일최고_기온	-10.8	37.3	18.13363434287302	9.646969418359962
평균기온	-14.5	31.8	12.571399462813744	9.66904767057937
일최저_기온	-20.2	29.1	7.78815411688432	10.374072871423548

통계분석결과VIEW C

일표본 t검정

	t	자유도	유의확률(양측)	평균차이	신뢰구간(하한)	신뢰구간(상한)
지점번호	462.2	21593	0	197.627	196.789	198.465
일최고_기온	276.223	21593	0	18.134	18.005	18.262
평균기온	191.059	21593	0	12.571	12.442	12.7
일최저_기온	110.319	21593	0	7.788	7.65	7.927

일표본 T검정

변수

지점번호  
평균기온  
일최고\_기온  
일최저\_기온

신뢰구간 & 검정값 설정

신뢰구간 수준(%)

95

검정값

0

결측값 처리

결측값 포함시 NA

결측값 모두 제거

통계분석옵션 B

확인

취소

	지점번호	지점	일자	평균기온	일최고_기온	일최저_기온
1	90	속초	2000-01-01	3.2	4.1	2.4
2	90	속초	2000-01-02	5.3	8.8	2.4
3	90	속초	2000-01-03	1.8	6	-2.7
4	90					
5	90					
6	90					
7	90					
8	90					
9	90					
10	90					
11	90					
12	90					
13	90					
14	90					
15	90					
16	90					
17	90	속초	2000-01-17	0.9	2.1	0.2



# THANK YOU