

OASIS-X(D-Stream) v1.0

OASIS-X(D-Stream) v1.0



목차



1. PCN OASIS-X 는?
2. 시스템 구성
3. 서비스 흐름
4. 주요 기능

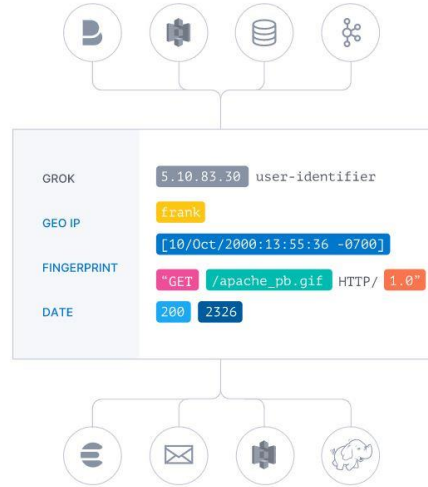


1. OASIS-X는?

개요

무료 오픈 소스 서버의 데이터 처리 파이프라인인 Logstash는 다양한 소스에서 데이터를 수집하여 변환한 후 자주 사용하는 저장소로 전달합니다.
PCN 기업부설연구소는 Logstash 기반의 데이터 수집/저장 프로세스를 쉽게 구성 할 수 있도록 Web 기반의 솔루션을 개발 하였으며 다양한 플러그인을 추가하여 기능을 확장 하였습니다.

구성도



특장점



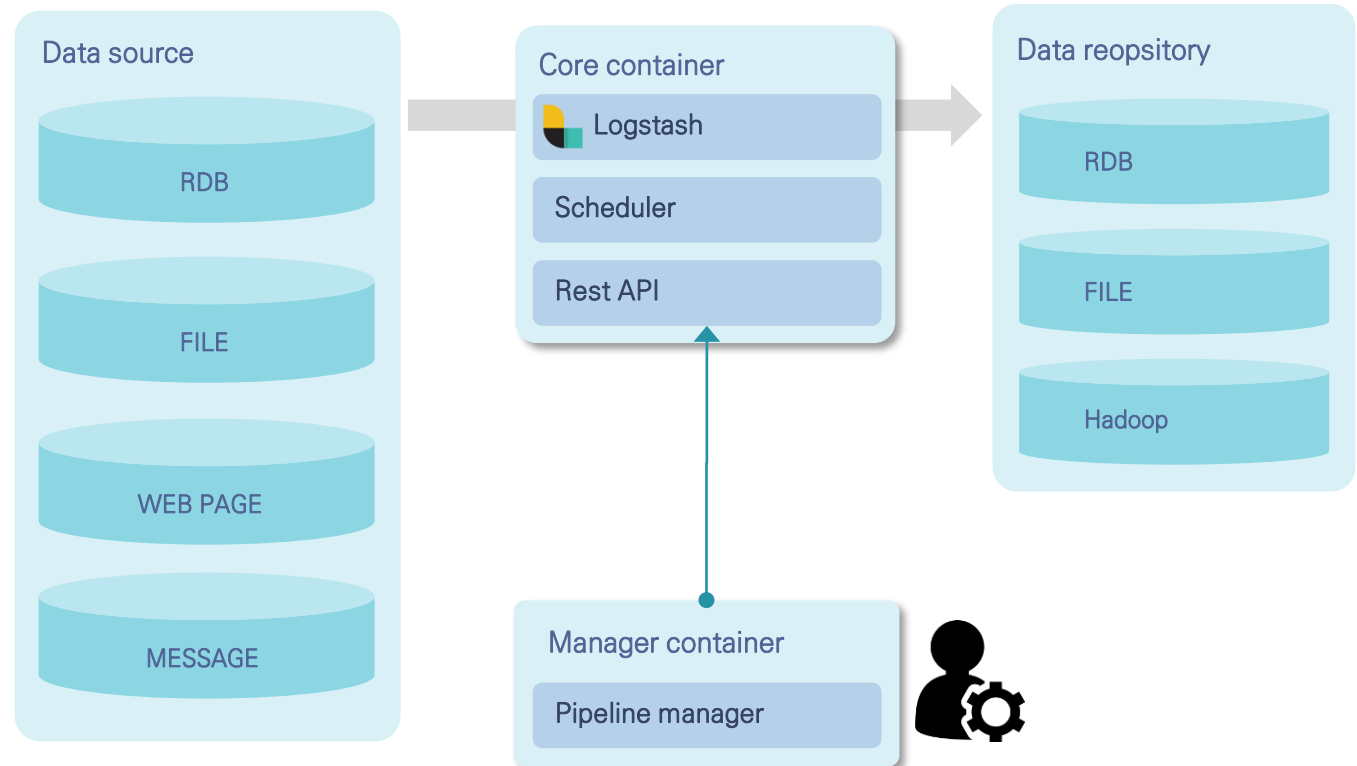
- 다양한 형태의 데이터 소스 지원
- 다양한 형태의 데이터 저장소 지원
- 필터를 제공하여 다양한 형태로의 데이터 가공
- 플러그인 지원으로 확장성 보장
- Web UI 지원으로 손쉬운 구성
- 수집/저장 현황의 모니터링
- 스케줄러 지원으로 배치 작업 수행

2. 시스템 구성

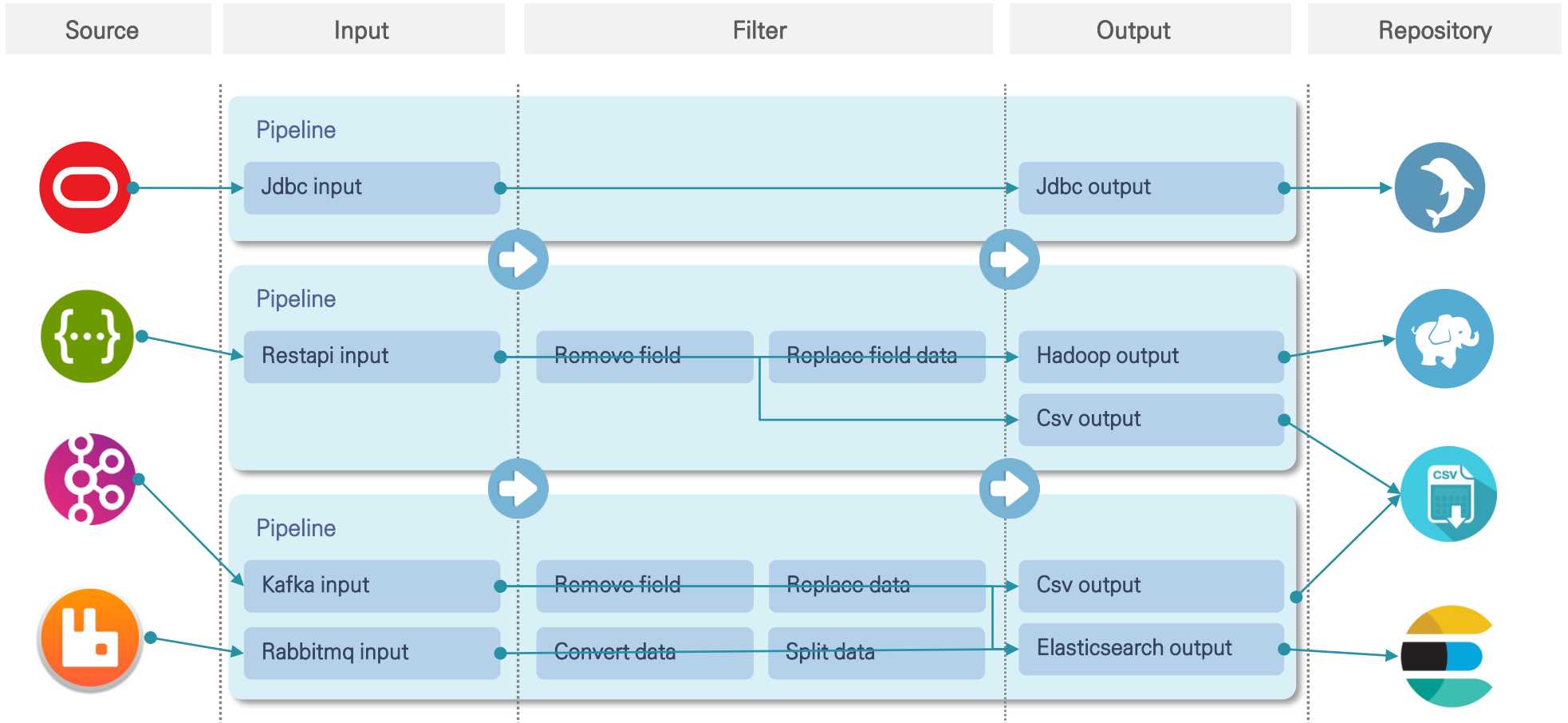
개요

시스템은 데이터수집 실행기 와 Core 어플리케이션으로 구성된 Core 컨테이너와 Web 환경의 Manager 컨테이너로 구성되어 있으며 Docker 컨테이너로 설치 하여 OS를 비롯한 서버 플랫폼의 제한을 받지 않습니다.

구성도



3. 서비스 흐름



4. 주요 기능 > 4.1 데이터 스트림 처리(1/3)

개요

RDB, web page, openapi 등 다양한 형태의 데이터를 수집하여 저장 할 수 있습니다. 또한 서버의 로그 파일 등의 데이터까지도 수집이 가능합니다.

수집 필터의 추가 및 커스터마이징으로 지원하지 않는 데이터의 수집이 가능합니다.

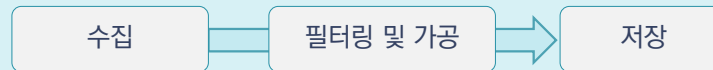
특징 및 장점

• 다양한 데이터 원본으로부터의 수집

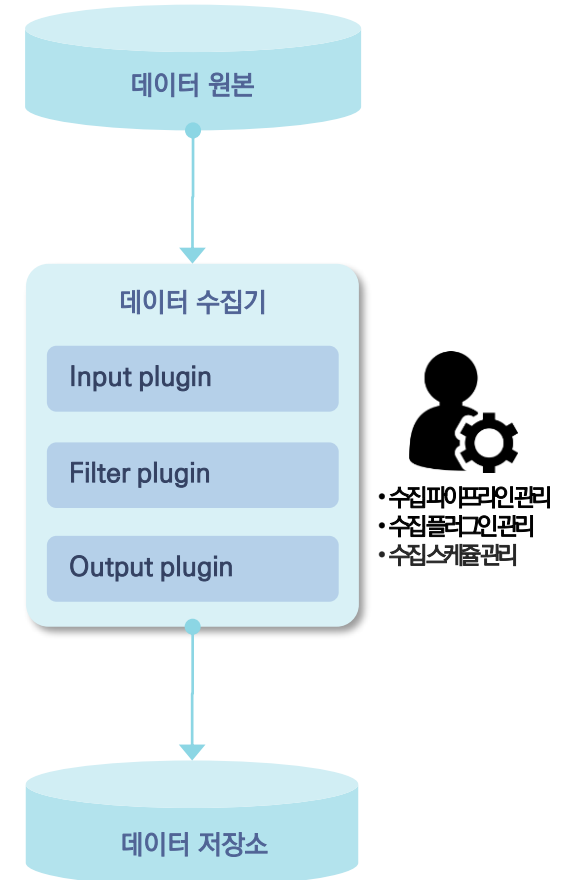
- ❖ oracle, mysql, sql server 등 범용 RDB
- ❖ web page, openapi 등의 http 데이터
- ❖ Server log file
- ❖ Message broker polling

• 플러그인 추가로 미지원 형식의 원본 데이터 수집

• 수집 파이프 라인의 구성



• 스케줄 수집, 수동 즉시 수집, 실시간 수집

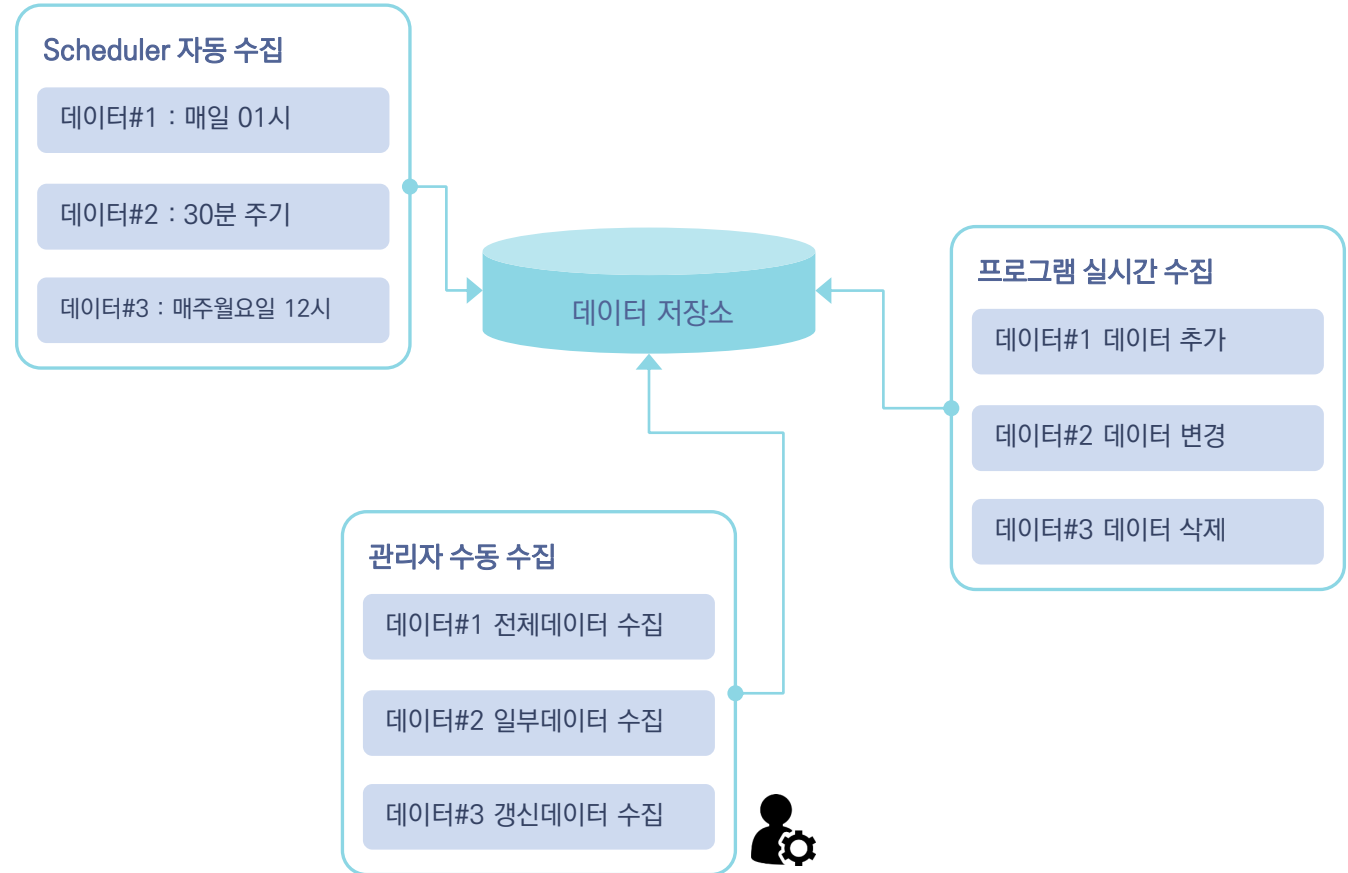


4. 주요 기능 > 4.1 데이터 스트림 처리(2/3)

개요

데이터 수집은 스케줄에 의한 정기적 수집, 관리기능을 통한 수동 수집, 변경된 데이터의 즉시 반영을 위한 실시간 수집이 가능 합니다. 수집 파이프라인의 구성으로 수집할 데이터는 언제든지 변경이 가능 합니다.

특징 및 장점

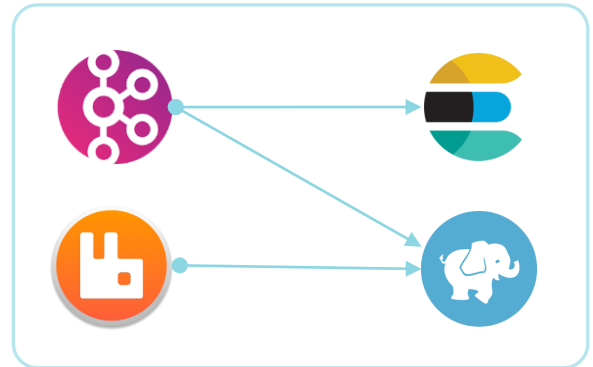
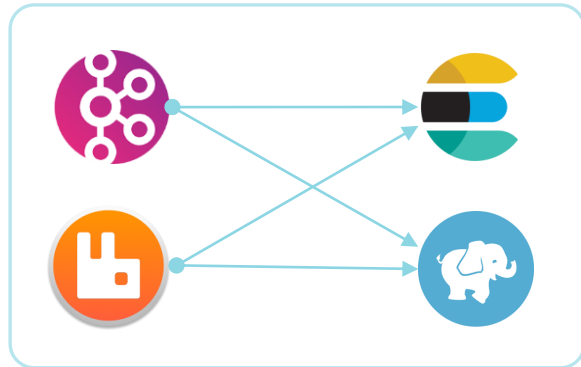
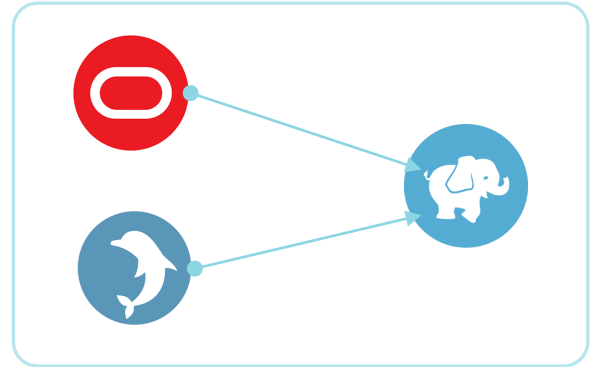
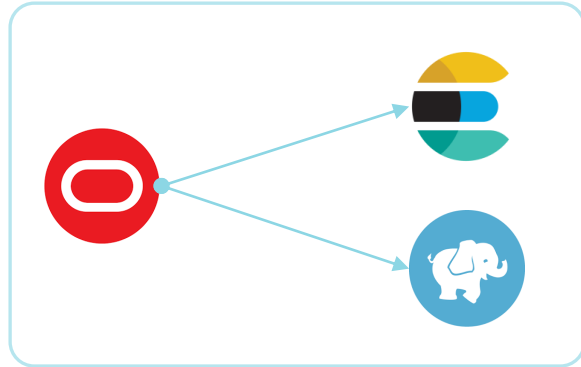


4. 주요 기능 > 4.1 데이터 스트림 처리(3/3)

개요

하나의 소스 데이터를 다수의 저장소에 저장 할 수 있고, 다수의 소스 데이터를 하나의 저장소에 저장 할 수 있습니다. 또한 다수의 소스 데이터를 다수의 저장소에 구분하여 저장 할 수 있습니다.

특징 및 장점



4. 주요 기능 > 4.2 Input 플러그인

Input 플러그인



플러그인	개요
Jdbc	Oracle, mysql, mssql을 포함한 대부분의 RDB 데이터를 수집
Open api	Rest 방식의 api 를 호출하여 xml, json 형태의 데이터를 수집
File	특정 경로의 파일을 지속적으로 polling하여 파일 내용을 수집
Message broker	Kafka topic, rabbitMq exchange의 이벤트 수집
Elasticsearch	Elasticsearch 인덱스 데이터 수집
Http poller	특정 url 을 http 요청하여 수신된 응답 데이터 수집
Http	Http 로 수신 되는 json, plain 형태의 데이터 수집
NoSQL	MongiDB, Redis 등 NoSQL 데이터 수집
Cloud storage/event	AWS S3, Google cloud storage 등 클라우드 스토리지 파일이나 Google pub/sub, Azure event hubs 등 이벤트 스트리밍 데이터 수집
Hadoop	HDFS에 있는 대용량 text형식 데이터를 수집
플러그인은 지속적으로 업데이트 예정	

4. 주요 기능 > 4.3 Output 플러그인(1/2)

Output 플러그인



플러그인	개요
File	Csv, txt 형태의 파일로 저장
Elasticsearch	Elasticsearch index 로 저장
Jdbc	Oracle, mysql, mssql을 포함한 대부분의 RDB로 저장
Message broker	Kafka topic, rabbitMq exchange로 메시지 전송
Elasticsearch	Elasticsearch 인덱스로 저장
Socket	TCP/UDP 네트워크 데이터 송신
Http	Http/Https Endpoint 로 데이터 전송
Hadoop	WEBHDFS를 이용한 Hadoop 로 저장
Cloud storage	AWS S3, Google cloud storage 등 클라우드 스토리지로 저장
Cloud event	Google pub/sub, Azure event hubs 등 클라우드 이벤트 스트리밍 데이터 전송
NoSql	MongoDB 로 저장

4. 주요 기능 > 4.3 Output 플러그인(2/2)

Output 플러그인



플러그인	개요
Time Series Database	OpenTSDB, InfluxDB, Graphite 등의 시계열 DB로 저장
Email	수집한 데이터를 E-mail 로 전송
플러그인은 지속적으로 업데이트 예정	

4. 주요 기능 > 4.4 Binary File 전송 처리

Output 플러그인



플러그인	개요
FTP File	FTP를 이용하여 HWP, DOC, ZIP, Image 등의 이진 파일을 읽어 Target(File)으로 전달
Local File	Local File System에 있는 HWP, DOC, ZIP, Image 등의 이진 파일을 읽어 Target(File)으로 전달
플러그인은 지속적으로 업데이트 예정	

4. 주요 기능 > 4.4 파이프라인 구성 설정

개요

Web 환경에서 쉽게 파이프라인 구성을 설정하고 실행 및 중지 할 수 있으며 현재 실행 상태를 모니터링 할수 있습니다.

주요 화면 (실행 및 모니터링)

상태	생성	변경	최근 실행	관리	실행
STANDBY COMPLETED	2020-06-10 10:56:20	2020-06-18 10:38:50	2020-06-18 10:39:27	변경 삭제 복사	실행
STANDBY COMPLETED	2020-06-03 17:39:53	2020-06-18 10:47:59	2020-06-29 11:10:41	변경 삭제 복사	실행
PROCEEDING	2020-06-24 10:14:36	2020-06-29 11:10:29	2020-07-03 09:14:08	변경 복사	중지
STANDBY COMPLETED	2020-05-25 11:31:14	2020-06-18 12:53:56	2020-06-30 18:10:10	변경 삭제 복사	실행
STANDBY COMPLETED	2020-05-27 09:15:14	2020-06-18 12:54:06 Not applied	2020-06-03 16:58:55	변경 삭제 복사	실행
STANDBY COMPLETED	2020-06-05 10:02:45	2020-06-18 12:54:10	2020-06-30 18:10:10	변경 삭제 복사	실행
STANDBY FAILED	2020-06-08 15:21:56	2020-06-18 14:41:29	2020-06-30 18:10:10	변경 삭제 복사	실행

주요 화면 (파이프라인 설정)

● 파이프라인 정보

파이프라인 명: LHTML 뉴스 수집(Once) 설명: LHTML 뉴스 수집(Once)

↳ 입력

httpOnce

+ 입력 추가

⇄ 필터 및 가공

htmlParse
html 에서 뉴스 목록 추출하여 news 필드 생성 및 원본 html 삭제

split
news 필드를 개별 이벤트로 변환

http
추출한 href 링크를 호출하여 html 수집

↳ 출력

stdout

file

+ 출력 추가

4. 주요 기능 > 4.5 스케줄링

개요

스케줄링을 통해 구성된 파이프라인을 원하는 시간 혹은 일정 간격으로 실행 할 수 있습니다.

주요 화면

스케줄 목록

파이프라인	타입	시간 값	관리
법률 용어	CRON	0 10 * * * ?	<input checked="" type="checkbox"/> 변경 <input type="checkbox"/> 삭제
법률 용어	CRON	0 0 0 * * ?	<input checked="" type="checkbox"/> 변경 <input type="checkbox"/> 삭제
음식점	CRON	0 10 * * * ?	<input checked="" type="checkbox"/> 변경 <input type="checkbox"/> 삭제
REST API 호출 파이프라인	CRON	0 10 * * * ?	<input checked="" type="checkbox"/> 변경 <input type="checkbox"/> 삭제

스케줄 설정

Modify Schedule

Name	Desc	Pipeline	Type	Value
법률용어 주기실행2	법률용어 색인 실행_매일	법률 용어	CRON	0 0 0 * * ?

Cancel OK

4. 주요 기능 > 4.6 모니터링

개요

수집되고 있는 데이터 현황을 모니터링 할 수 있습니다.

실행 현황 및 리소스 모니터링

파이프라인



0

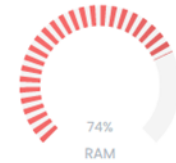
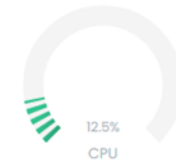


13



1

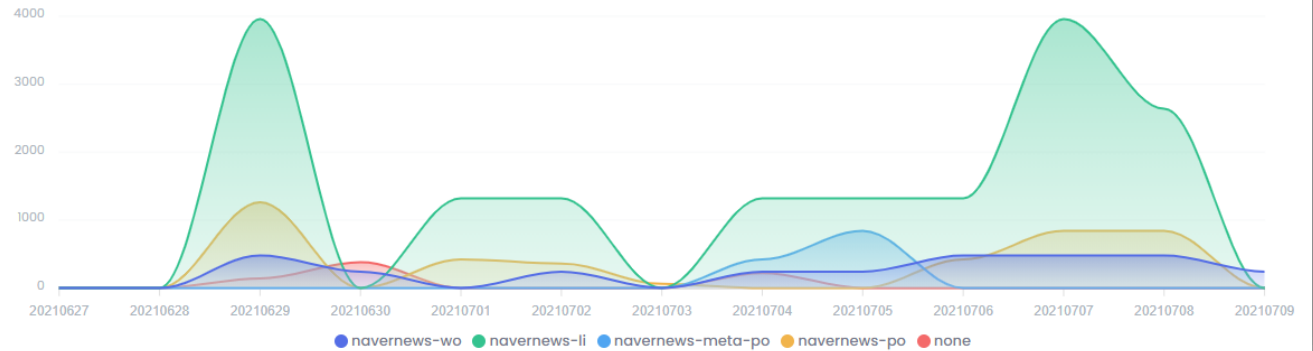
자원 현황



주요 화면

수집 현황 모니터링

수집 현황

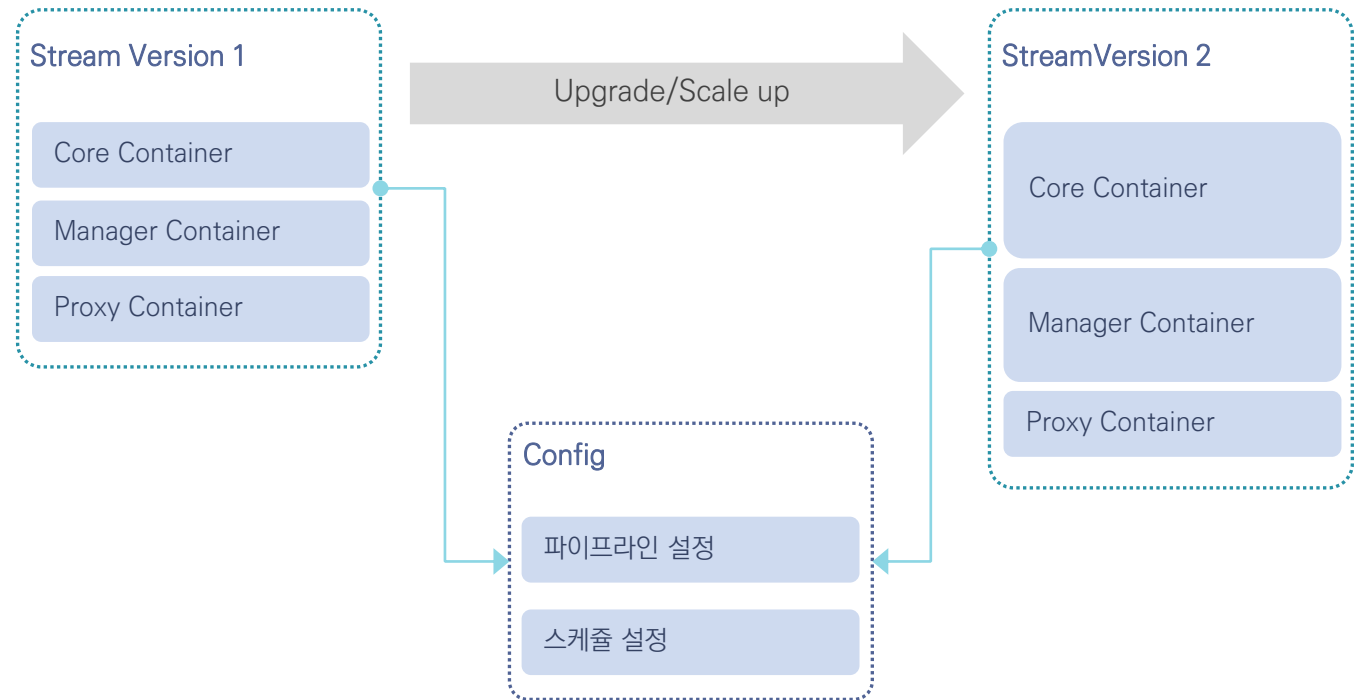


4. 주요 기능 > 4.6 설치 및 유지보수

개요

프로그램의 설치는 Docker 컨테이너를 활용하여 진행 되어 설치 과정이 단순합니다. 또한 별도의 데이터 백업이나 복원 없이 컨테이너의 교체만으로 업그레이드가 가능하여 유지보수 부담이 없고 컨테이너별 자원 할당이 가능하여 자원 상황에 따른 Scale In/Out 이 가능 합니다. 또한 업그레이드가 진행 되는 동안에도 이전컨테이너에서 계속 서비스가 가능하여 무중단 업그레이드가 가능 합니다.

특징 및 장점





No.1 Digital Value Service Company



제조사 : (주)피씨엔

서울특별시 강남구 언주로 135길 33 PCN빌딩

TEL : 02-565-7740, FAX : 02-565-7790

기술문의 : tech@pcninc.co.kr

홈페이지 : solution.pcninc.co.kr / www.pcninc.co.kr