



# 빅데이터 플랫폼



# Architecture

다양한 소스로부터 빅데이터를 수집하여 통합된 빅데이터 플랫폼 저장소에 저장하여 통합 고급 분석을 가능하도록 하는 빅데이터 플랫폼을 구성



## 제품 Vision

IMBEE는 주요 오픈소스 Hadoop, Flume 등의 엔진을 이용한 엔터프라이즈 레벨의 다양한 기능을 제공합니다.

- 01 빠른 속도의 대용량 로그 수집이 가능하다.
- 02 Hadoop 기반의 빅데이터 플랫폼과 기술적인 호환성을 가진다. (Map Reduce, HDFS 지원)
- 03 이기종 시스템, 장비, 프로토콜을 이용한 데이터 수집 기능을 지원한다.
- 04 실시간 조건으로 데이터 검색이 가능하도록 메타데이터와 인덱스를 관리한다.
- 05 사용자가 직접 쉽게 데이터를 조회할 수 있는 SQL on Hadoop 쿼리 문법을 지원한다.
- 06 저장된 빅데이터를 분석 전용으로 재가공할 수 있는 통계함수 기반의 분석 엔진 기능을 제공한다.
- 07 빅데이터 시각화를 위한 기본 템플릿과 Sample을 제공한다.
- 08 빅데이터 플랫폼 시스템의 상태 관리 모니터링 기능을 제공한다.
- 09 개발자 전용 API 제공으로 커스텀 UI 개발이 가능하도록 한다.
- 10 Linux와 Java 환경에서 손쉽게 설치되고 최적화된 버전 업데이트가 쉽게 가능하도록 한다.

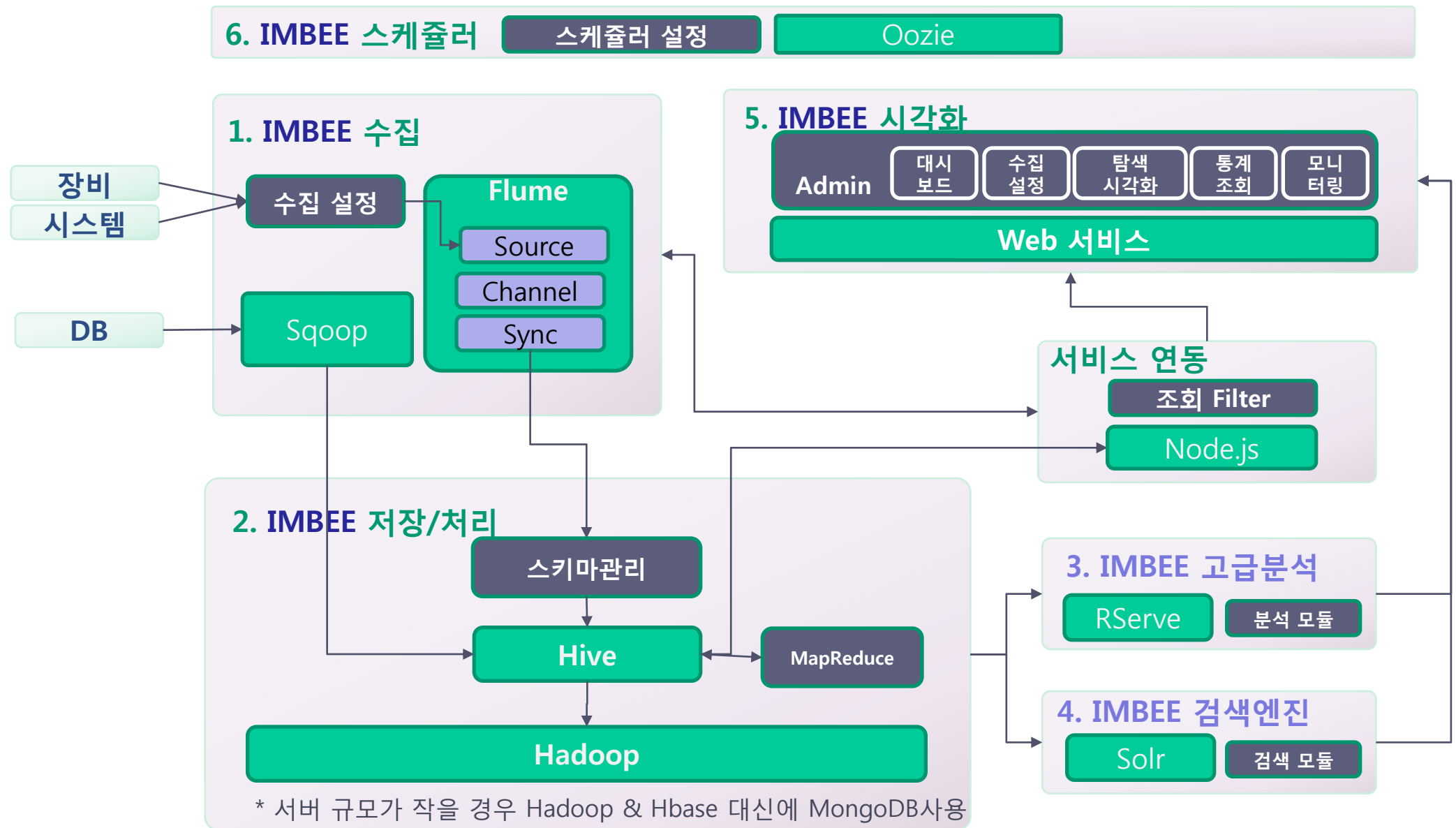
## Ecosystem 구성요소

빅데이터에 대한 수집/저장/처리/분석/관리를 위한 빅데이터 플랫폼의 모듈이 오픈소스 기반의 IMBEE S/W 엔진이 구성됩니다.

구분	Open Source	S/W 모듈	주요 기능
데이터 수집	<ul style="list-style-type: none"> <li>Flume</li> <li>Sqoop</li> </ul>	<ul style="list-style-type: none"> <li>IMBEE Collector</li> </ul>	<ul style="list-style-type: none"> <li>다양한 종류의 로그에 대한 수집 및 수집 설정 가능</li> <li>RDB로부터 데이터 수집 처리 기능 지원</li> </ul>
데이터 저장	<ul style="list-style-type: none"> <li>Hadoop</li> <li>MongoDB</li> </ul>	<ul style="list-style-type: none"> <li>IMBEE Storage</li> </ul>	<ul style="list-style-type: none"> <li>원본 데이터 및 정형화된 데이터 저장</li> <li>HDFS/NoSQL 저장소에 대한 일원화된 인터페이스 제공</li> <li>빅데이터에 대한 대용량 병렬 처리 (Scale out)</li> </ul>
데이터 처리	<ul style="list-style-type: none"> <li>Hive</li> </ul>	<ul style="list-style-type: none"> <li>IMBEE Query</li> </ul>	<ul style="list-style-type: none"> <li>Hive를 통해서 ANSI SQL 쿼리 문법 지원</li> <li>MapReduce 분산 병렬 데이터 처리 지원</li> </ul>
데이터 검색	<ul style="list-style-type: none"> <li>Solr</li> </ul>	<ul style="list-style-type: none"> <li>IMBEE Search</li> </ul>	<ul style="list-style-type: none"> <li>Solr 를 이용한 데이터 크롤링 및 검색 엔진 구축 가능</li> </ul>
데이터 분석	<ul style="list-style-type: none"> <li>R</li> </ul>	<ul style="list-style-type: none"> <li>IMBEE Analysis</li> </ul>	<ul style="list-style-type: none"> <li>통합 보안 데이터에 대한 R 기반 분산 병렬 처리</li> <li>다양한 종류의 로그에 대한 통합 분석</li> <li>로그들에 대한 연관분석 가능</li> </ul>
관리	<ul style="list-style-type: none"> <li>ZooKeeper</li> </ul>	<ul style="list-style-type: none"> <li>IMBEE Monitoring</li> </ul>	<ul style="list-style-type: none"> <li>GUI 기반 통합 관리/모니터링 기능 제공</li> <li>빅데이터 플랫폼 시스템 모니터링</li> </ul>
스케줄러	<ul style="list-style-type: none"> <li>Oozie</li> </ul>	<ul style="list-style-type: none"> <li>IMBEE Scheduler</li> </ul>	<ul style="list-style-type: none"> <li>빅데이터 플랫폼 처리 각 단계별 스케줄러 설정</li> <li>데이터 처리에 대한 Workflow 기능 지원</li> </ul>
시각화	<ul style="list-style-type: none"> <li>Node.js</li> </ul>	<ul style="list-style-type: none"> <li>IMBEE Visualization</li> </ul>	<ul style="list-style-type: none"> <li>시각화 서비스 인터페이스 모듈 제공</li> </ul>

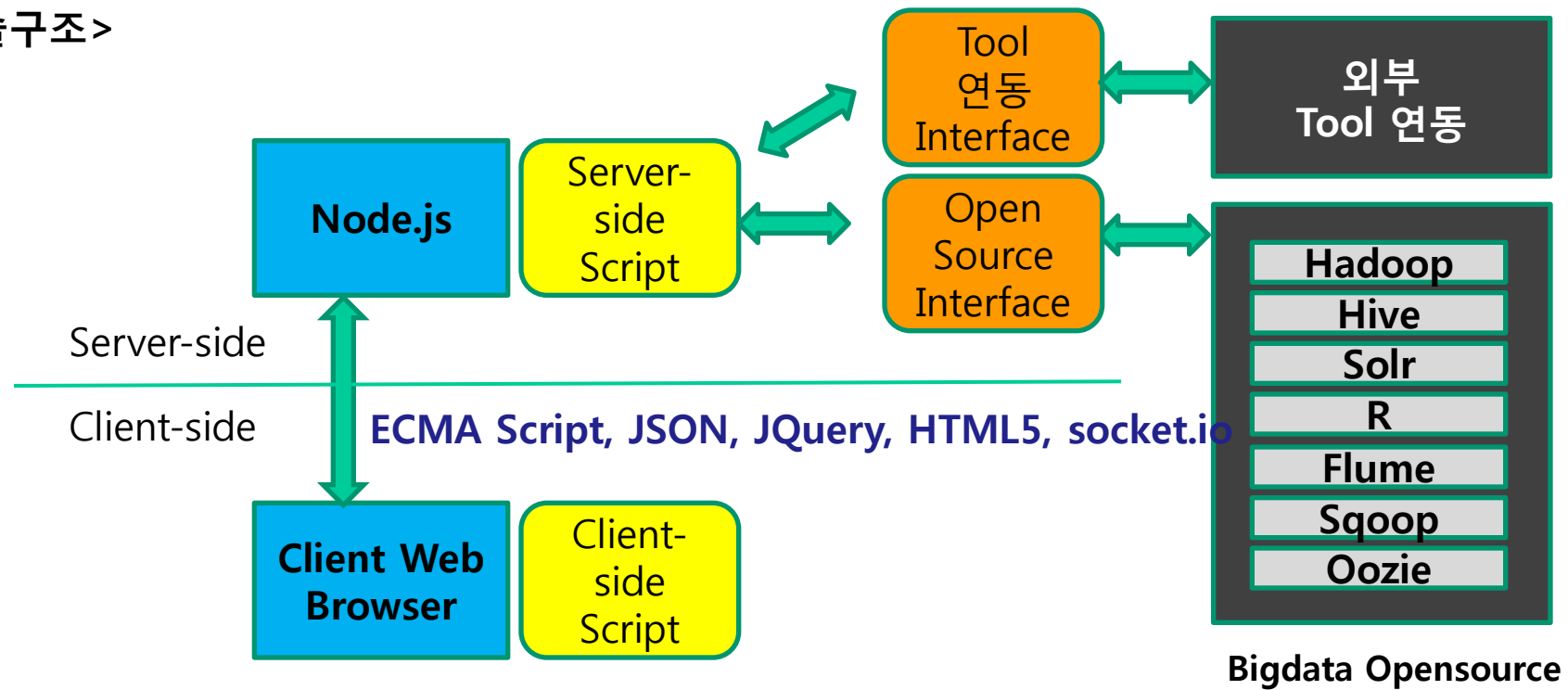
# 플랫폼 동작 프로세스

IMBEE는 오픈소스 기반의 수집,저장,처리 구조를 가지고 시각화된 관리자 화면과 개발자 전용 API를 지원합니다.

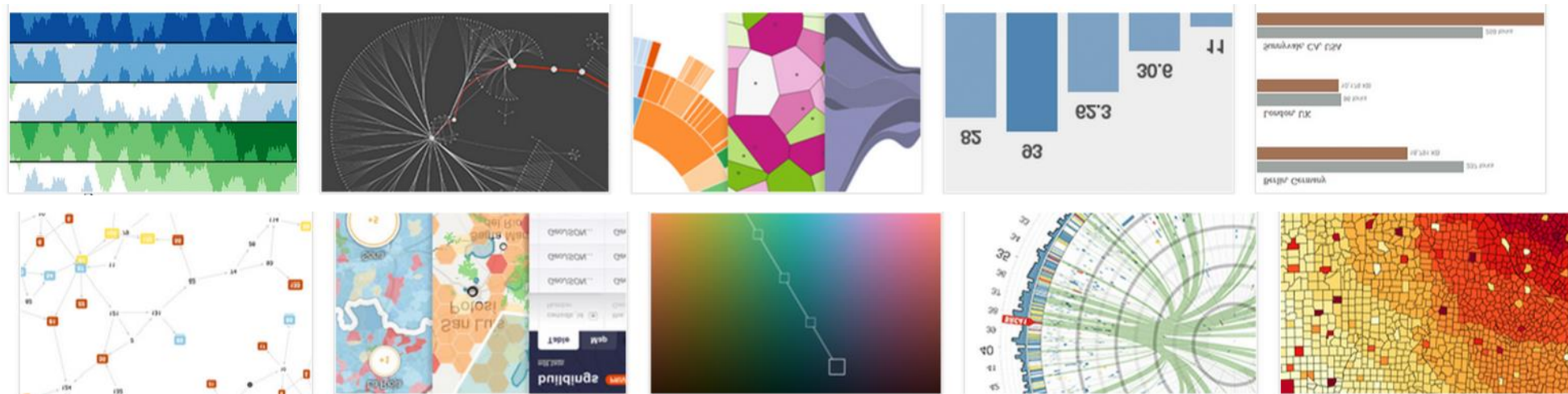


# Node.js 스크립트 개발 환경 (Web 최신 기술 적용)

<기술구조>



<다양한 분석 시각화 표현 방법>



# 빅데이터 플랫폼 관리 서비스

관리자 서비스 제공 화면 - 수집, 저장, 처리 구성 요소에 대한 통합 및 상세 모니터링 기능

