## 인공지능/데이터

| No | 분야 | 솔루션명 | 라이선스 | 기술지원 | 홈페이지 | 제품개요 |
|---|---|---|---|---|---|---|
| 1 | Graph Database | FlockDB | Apache-2.0 | community | https://github.com/twitter/flockdb | Twitter에서 사용된 것으로 알려진, 분산 그래프 데이터 보관용 데이터 베이스 |
| 2 | Graph Database | neo4j | GPL-3.0, AGPL-3.0 or commercial | prof/community | https://neo4j.com/ | 네이티브 그래프 저장 및 처리 기능을 갖춘 ACID를 준수하는 트랜잭셔널 데이터베이스 |
| 3 | Graph Database | OrientDB | Apache-2.0 | Prof | http://orientdb.org/ | Java로 작성된 오픈 소스 NoSQL 데이터베이스 관리 시스템 |
| 4 | NoSQL (wide column) | Apache Accumulo | Apache-2.0 | community | https://accumulo.apache.org/ | 클러스터 전체에서 대규모 데이터 세트를 저장하고 관리 |
| 5 | NoSQL (wide column) | Cassandra | Apache-2.0 | community | http://cassandra.apache.org | Ruby, Perl, Python, Scala, Java, PHP, C++, C# 등 다양한 언어를 지원하고 Facebook, Twitter 등에 사용된 분산데이타베이스 시스템 |
| 6 | NoSQL (wide column) | Hbase | Apache-2.0 | community | http://hbase.apache.org | HDFS에 구현된 분산 클럼 기반이며 대규모 데이타셋에 실시간으로 랜덤 액세스가 가능한 분산 데이타베이스 시스템 |
| 7 | NoSQL(Document) | Apache CouchDB | Apache-2.0 | community | http://couchdb.apache.org | 스케일러블 아키텍처를 쉽게 이용하고 보유하는데 초점을 둔 오픈 소스 데이터베이스 소프트웨어 |
| 8 | NoSQL(Document) | Mongodb | AGPL-3.0, Apache- v2.0 | community | https://www.mongodb.com/ | JSON 형태의 문서 콜렉션으로 데이터를 저장하는 오픈 소스 문서 지향 데이타베이스 |
| 9 | NoSQL(Document) | couchbase | Apache-2.0 | prof | https://developer.couchbase.com/open-source-projects | NoSQL DB로 분류되며 DB모델은 key-value이고, Schema-less모델로 문서지향형 |
| 10 | NoSQL(Key-Value) | Redis | BSD | prof/community | https://redis.io/ | 인메모리 기반 NoSQL 데이터베이스 |
| 11 | NoSQL(Key-Value) | Riak | Apache-2.0 | prof/community | https://riak.com/ | 고 가용성, 내결함성, 운영 단순성 및 확장 성을 제공 하는 분산 NoSQL 키-값 데이터 저장소 |
| 12 | NoSQL(Key-Value) | Memcached | BSD-3-Clause | wiki, chat | http://www.memcached.org | 분산 메모리 캐싱 시스템, 데이터베이스 부하를 줄여 웹 애플리케이션 성능향상도움 |
| 13 | 데이터 분석 | ankus | GPL-3.0 | prof/community | http://openankus.org | 데이터 마이닝 / 기계 학습(Machine Learning) 지원 분석 도구 |
| 14 | 데이터 분석 | Apache Spark | Apache-2.0 | community | http://spark.apache.org | 메모리상에서 동작하는 분산 데이터 분석 시스템 대규모 데이터 처리를위한 통합 분석 엔진. |
| 15 | 데이터 분석 | 아파치 제플린 (Zeppelin) | Apache-2.0 | community | https://zeppelin.apache.org | SQL, Scala등을 사용하여 데이터 중심의 대화형 데이터 분석 및 협업 문서를 지원하는 웹기반 노트북 |
| 16 | 데이터 분석 | Pentaho | Apache-2.0 | prof/community | https://www.hitachivantara.com/go/pentaho.html | 다양한 데이터 배포 위한 엔터프라이즈 급 오픈 소스 기반 데이터 통합 및 분석 플랫폼빅데이터 통합, 데이터 웨어하우징, 워크 플로우 통합, 데이터 과학 적용, 데이터 탐색, 데이터 시각화, 보고서 및 대쉬보드 포함한 분석 플랫폼 |
| 17 | 데이터 분석 라이브러리 | pandas | BSD | community | http://pandas.pydata.org | "관계형"또는 "레이블이있는"데이터로 쉽고 유연하며 표현력있는 데이터 구조를 제공하는 Python 패키지 |
| 18 | 데이터 시각화 | Apache Superset | Apache-2.0 | community | https://superset.apache.org/ | 데이터 탐색 및 시각화 플랫폼 |

| No | 분야 | 솔루션명 | 라이선스 | 기술지원 | 홈페이지 | 제품개요 |
|---|---|---|---|---|---|---|
| 19 | 데이터 시각화 | D3.js | BSD | community | https://d3js.org/ | 웹브라우저 상에서 동적이고 인터렉티브한 정보시각화를 구현하기 위한 자바스크립트 라이브러리 |
| 20 | 데이터 웨어하우스 | Hive | Apache-2.0 | community, wiki | http://hive.apache.org | 대용량 데이터 분석에 적합한 하둡의 상위에 위치한 SQL 기반의 Data Warehouse 구성요소 |
| 21 | 데이터 통합 | Talend Open Studio | Apache-2.0 | community | https://www.talend.com/ | 개발, 테스트, 배포, 데이터관리 및 어플리케이션 통합프로젝트를 관리하기 위한 다기능 공개SW 세트 |
| 22 | Integration Framework | Apache Camel | Apache-2.0 | community | http://camel.apache.org | EIP(Enterprise Integration Patterns) 기반의 오픈소스 통합 연계 프레임워크 |
| 23 | 데이터 테이블 및 스토리지 관리 | Hcatalog | Apache-2.0 | community | https://cwiki.apache.org/confluence/display/Hive/Hcatalog | Pig, Spark SQL 및/또는 사용자 지정 MapReduce 애플리케이션 내의 Hive 메타스토어 테이블에 액세스할 수 있는 도구 |
| 24 | 빅데이터 프레임워크 | Apache Avro | Apache-2.0 | community | http://avro.apache.org | 오픈소스 데이터 직렬화 프레임워크, 이기종간 데이터 타입을 교환할 수 있는 체계를 제공. 다양한 데이터 구조 지원 |
| 25 | 빅데이터 프레임워크 | Apahce Giraph | Apache-2.0 | community | http://giraph.apache.org | 높은 확장 성을 위해 구축 된 반복적 인 그래프 처리 시스템 |
| 26 | 빅데이터 플랫폼 | Apache Impala | Apache-2.0 | prof/community | http://impala.io | Apache Hadoop 용 오픈 소스 네이티브 분석 데이터베이스 |
| 27 | 빅데이터 플랫폼 | Apache Sqoop | Apache-2.0 | community | https://sqoop.apache.org | RDBMS와 아파치 하둡간의 대용량 데이터들을 변환하여 주는 Command-Line Interface 애플리케이션. 데이터 커넥터와 일치하지 않는 데이터 포맷을 사용할 수 있도록 지원 |
| 28 | 빅데이터 플랫폼 | Apache Tajo | Apache-2.0 | community | http://tajo.apache.org | Apache Hadoop을위한 강력한 빅 데이터 관계형 및 분산 데이터웨어 하우스 시스템 |
| 29 | 빅데이터 플랫폼 | Cascading | Apache-2.0 | prof/community | http://www.cascading.org | 다양한 클러스터 컴퓨팅 플랫폼에서 복잡하고 내결함성이있는 데이터 처리 워크 플로를 정의하고 실행하기위한 풍부한 기능의 API |
| 30 | 빅데이터 플랫폼 | Gephi | CDDL, GPL v3 | community | https://gephi.org/ | 1백만 노드 규모의 네트워크와 다양한 크래프 타입을 지원하는 시각적 분석 도구 |
| 31 | 빅데이터 플랫폼 | HPCC (High-Performance Computing Cluster) | Apache-2.1 | prof/community | https://hpccsystems.com | 빅 데이터 문제를 해결하기 위해 엔터프라이즈 용 오픈 소스 슈퍼 컴퓨팅 플랫폼 |
| 32 | 빅데이터 플랫폼 | Hiho | Apache-2.0 | community | https://github.com/sonalgoyal/hiho | 정형 데이터 수집 |
| 33 | 빅데이터 관리 | Apache Ambari | Apache-2.0 | community | http://ambari.apache.org | Apache Hadoop 클러스터 프로비저닝, 관리 및 모니터링을위한 소프트웨어를 개발하여 Hadoop 관리를 단순화하는 것을 목표 |
| 34 | 빅데이터 관리 | Hue | Apache-2.0 | prof/community | http://gethue.com | 하둡을 위한 Web 인터페이스. 웹브라우저를 통한 시스템 접근 및 파일 시스템 열람, 사용자 계정 생성 및 관리, 클러스터 모니터링 등 다양한 기능 제공 |
| 35 | 프로그래밍 언어 | R | GPL | prof/community | https://www.r-project.org/ | 통계계산 및 시각화를 위한 언어 및 개발환경을 제공 |
| 36 | 프로그래밍 언어 | 피그(Pig) | Apache-2.0 | community, wiki | http://pig.apache.org | 데이터-흐름 기반의 스크립트 프로그래밍 언어 |

| No | 분야 | 솔루션명 | 라이선스 | 기술지원 | 홈페이지 | 제품개요 |
|---|---|---|---|---|---|---|
| 37 | 프로그래밍 언어 | python | Python Software Foundation License | community | https://www.python.org/ | 1991년 프로그래머인 귀도 반 로섬이 발표한 고급 프로그래밍 언어로, 플랫폼 독립적이며 인터프리터식, 객체지향적, 동적타이핑 대화형 언어 |
| 38 | 프로그래밍 언어 | 아나콘다(Anaconda) | New BSD | prof/community | https://www.anaconda.com/ | Python/R 데이터 과학 및 기계 학습을 수행하기위한 개인 실무자를 위해 개발되어 수천 개의 오픈 소스 패키지 및 라이브러리를 사용 |
| 39 | machine learning | Mahout | Apache-2.0 | community | http://mahout.apache.org | 분산 선형 대수 프레임 워크 및 수학적으로 표현 가능한 Scala DSL |
| 40 | machine learning | Tensorflow | Apache-2.0 | community | https://www.tensorflow.org | 분산 데이터를 사용해 머신러닝 및 기타 계산을 수행할 수 있는 오픈소스 프레임워크입니다. |
| 41 | machine learning | PyTorch | BSD | community | https://pytorch.org/ | Python을 위한 오픈소스 머신 러닝 라이브러리, Torch를 기반으로 하며, 자연어 처리와 같은 애플리케이션을 위해 사용 |
| 42 | deep-learning | MxNet | Apache-2.0 | community | https://mxnet.incubator.apache.org | 오픈소스기반의 분산 딥러닝 소프트웨어 프레임워크 (distributed deep-learning framework) |
| 43 | Search | Apache Lucene | Apache-2.0 | community | http://lucene.apache.org | 자바로 개발된 오픈소스 정보검색 라이브러리 |
| 44 | 분산검색엔진 | ElasticSearch | Apache-2.0 | prof/community | https://www.elastic.co/kr/ | 시간이 갈수록 증가하는 문제를 처리하는 분산형 RESTful 검색 및 분석 엔진 |
| 45 | 분산검색엔진 | 아파치 솔라(Apache Solr) | Apache-2.0 | community | https://solr.apache.org/ | 분산 인덱싱, 복제 및 로드 균형 조정 쿼리, 자동 장애 조치 및 복구, 중앙 집중식 구성 등을 제공하는 검색 플랫폼 |
| 46 | 분산파일시스템 | Hadoop | Apache-2.0 | community, wiki | http://hadoop.apache.org | 컴퓨터 클러스터를 이용하여 대용량 데이터 셋의 분산 처리를 지원하는 분산 파일 시스템 |
| 47 | 분산소프트웨어프레임워크 | MapReduce | Apache-2.0 | community | http://hadoop.apache.org | 분산 컴퓨팅을 지원하기 위한 소프트웨어 프레임워크 |
| 48 | 워크플로우 스케쥴러 | Apache Oozie | Apache-2.0 | community | http://oozie.apache.org | Apache Hadoop 작업을 관리하기위한 워크 플로우 스케줄러 시스템 |
| 49 | 분산실시간 연산시스템 | Apache Storm | EPL | community, blog | https://storm.apache.org/ | 클로저 프로그래밍 언어로 작성된 분산형 스트림 프로세싱 연산 프레임워크 |
| 50 | 분산 로그관리 시스템 | Flume | Apache-2.0 | community | https://flume.apache.org/ | 대용량 로그데이터를 중앙집중화된 데이터 저장소로 효율적으로 수집, 통합,저장하기 위한 시스템 |
| 51 | 분산 관리 시스템 | 주키퍼(Zookeeper) | Apache-2.0 | community, wiki | http://zookeeper.apache.org | 분산 환경에서 노드들 간의 분산 동기화, 그룹,서비스, 공유, 락 등을 관리하는 시스템 |
| 52 | 분산 메시지 브로커 | 아파치 카프카(Apache Kafka) | Apache-2.0 | community | https://kafka.apache.org/ | 실시간으로 기록 스트림을 게시, 구독, 저장 및 처리할 수 있는 분산 데이터 스트리밍 플랫폼 |