

자바쌤~!

https://cafe.naver.com/javassem

검색

63기-자바 & 빅데이터 >

[스쿱] 개념



자바쌤 1:1 채팅

2020.06.24. 21:25 조회 16

댓글 0 URL 복사



1. 스쿱의 개요

RDBMS와 HDFS 간의 데이터 전송 지원 도구

외부 시스템의 데이터를 HDFS로 가져오면
Hive 등 에코시스템에서 사용하고
다양한 파일 형태로 저장도 가능하다

JDBC와 호환되는 DB에서 사용 가능

1. 스쿱의 개요

하둡에코시스템에서 플룸과 같이 수집 소프트웨어로 많이 사용된다.

카페정보 나의활동



매니저 자바쌤
since 2013.03.15.
카페소개

새싹2단계

437

초대하기

즐거찾는 멤버

64명

게시판 구독수

4회

우리카페업 수

0회

카페 글쓰기

카페 채팅

즐거찾는 게시판

전체글보기

4,080

KOSMO

63기-자바 & 빅데이터 N

팁

- 심심문제
- 기타등등
- JAVA
- Oracle/Mysql
- JSP
- WebUI
- 파이썬
- Spring
- 리눅스
- 하둡+eco N
- 프로젝트

공유해요

- 알고리즘도전
- 기술면접 준비하기
- 실존 면접
- 기회는 준비된자에게
- 프로젝트 팁

쉬어가기

청년아카데미

전문가과정

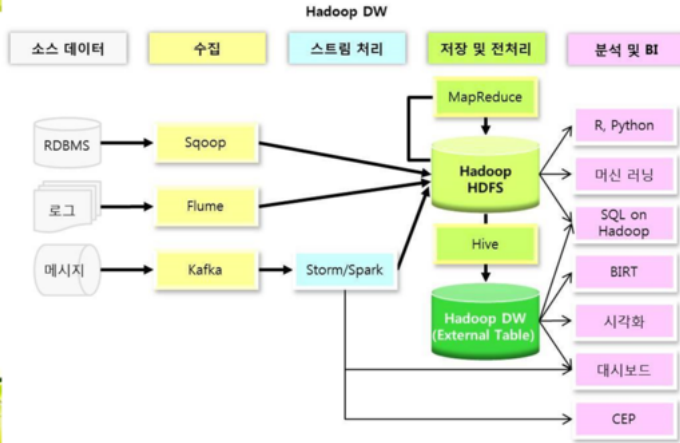
공지사항
카페회칙

카페탈퇴하기

다른 시스템에 접근하는 것으로 접근할 때 가장 중요한 권한 문제를 사전에 준비해야 한다

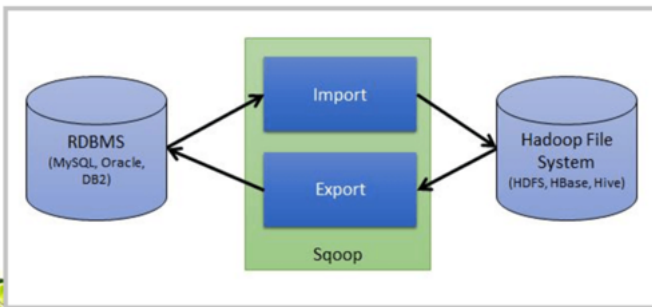


1. 스쿱의 개요



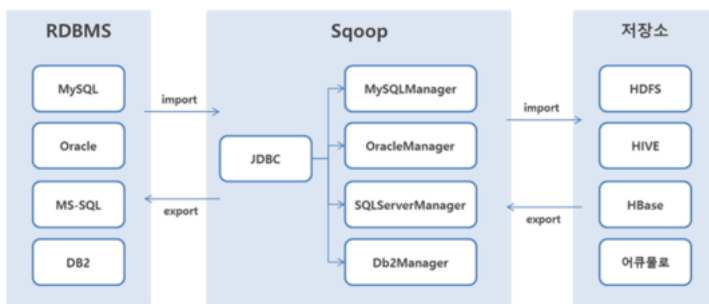
출처: 테라데이터 데이터베이스 중심(이윤성 저)
Source: Teradata Database Centered (Author Yoon-Seong Lee)

1. 스쿱의 개요



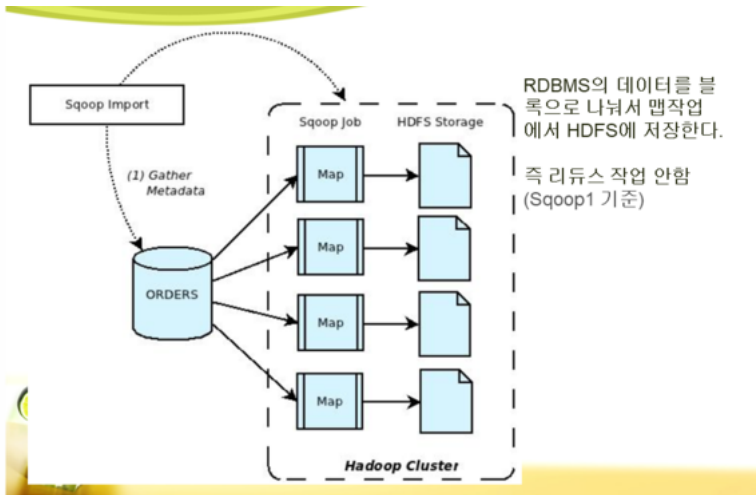
hdfstutorial.com

1. 스쿱의 개요

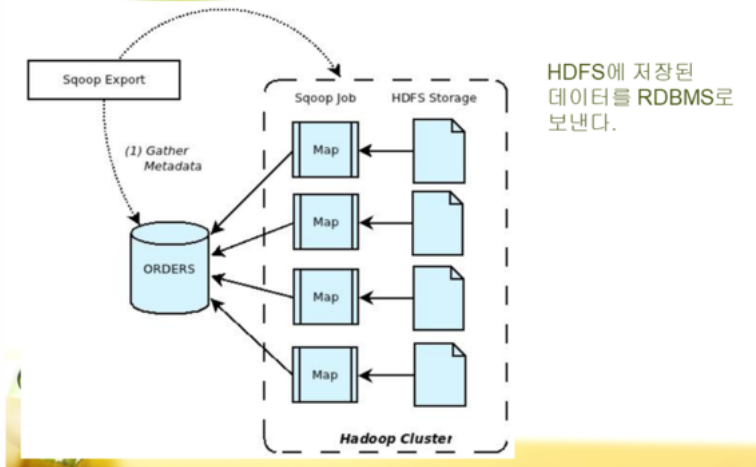


시작하세요 하둡 프로그래밍

1. 스쿱의 개요



1. 스쿱의 개요



2. 스쿱의 버전

버전 2에서는 버전 1의 복잡함을 간소한 것이다.
그래서 버전1과 버전 2가 같이 릴리즈되고 있고
버전1이 아직까지 많이 사용한다.

기능	스쿱 버전 1	스쿱 버전 2
모든 주요 RDBMS를 위한 커넥터	- 지원함	- 지원하지 않음 - 범용 JDBC 커넥터를 사용함
Kerberos 보안 통합	- 지원함	- 지원하지 않음
RDBMS에서 하이브나 HBase로 데이터 전송	- 지원함	- 지원하지 않음 - 우선 RDBMS에서 HDFS로 데이터를 가져온 후, 데이터를 다시 하이브나 HBase에 수동으로 가져와야 함
하이브나 HBase에서 RDBMS로 데이터 전송	- 지원하지 않음 - 우선 하이브나 HBase에서 HDFS로 데이터를 내보낸 다음, 다시 스쿱으로 내보내야 함	

3. sqoop help

\$ sqoop help

\$ sqoop help import

\$ sqoop help export



3. sqoop help - data import

옵션	내용
-username	데이터베이스 접속 계정
-password	데이터베이스 접속 암호
-connect	JDBC 접속 URL
-table	임포트 대상 테이블
-columns	특정 칼럼만 임포트할 경우 칼럼명을 설정. 이 옵션을 설정하지 않을 경우 전체 칼럼을 임포트
-target-dir	임포트 결과를 저장할 HDFS 디렉터리. 별도로 설정하지 않을 경우 <code>*/user/계정명/테이블명</code> 에 데이터가 저장됨
-query	특정 테이블을 설정하지 않고 질의문 실행 결과를 임포트. 단 이 옵션을 사용할 경우 <code>\$CONDITIONS</code> 를 반드시 WHERE절에 사용해야 함. <code>\$CONDITIONS</code> 는 스콧이 자동으로 생성하는 WHERE 조건을 나타내는 키워드임
-m, -num-mappers	실행할 맵 태스크 개수, 데이터베이스에 부하를 줄 수 있으므로 적절한 태스크 개수를 설정해야함.

3. sqoop help - data import

-where	임포트할 때 사용할 WHERE절
-z, -compress	임포트 결과를 압축
-compression-codec	하둡 압축 코덱, 기본값으로 Gzip을 사용함
-as-textfile	임포트 결과를 텍스트 파일로 저장. 스콧은 텍스트 파일을 기본값으로 사용함
-as-sequencefile	임포트 결과를 시퀀스파일로 저장
-null-string<널값문자열>	문자열 칼럼에서 널 값 대신 사용할 문자열
-null-non-string<널값 문자열="">	문자열이 아닌 칼럼에서 널 값 대신 사용할 문자열
-direct	고속 커넥터를 이용할 경우 사용

3. sqoop help - data export

옵션	내용
-export-dir	익스포트 대상 HDFS 디렉터리
-m, -num-mappers	실행할 맵 태스크 개수, 데이터베이스에 부하를 줄 수 있으므로 적절한 태스크 개수를 설정해야함
-table	데이터베이스 최종 대상 테이블
-staging-table	데이터베이스 중간 데이터 저장용 테이블
-clear-staging-table	중간 데이터용 테이블 데이터 삭제 여부
-null-string<널값 문자열="">	문자열 칼럼에서 널 값 대신 사용할 문자열
-null-non-string<널값문자열>	문자열이 아닌 칼럼에서 널 값 대신 사용할 문자열
-direct	고속커넥터를 이용할 경우 사용





댓글

댓글알림

김용선

댓글을 남겨보세요

등록

자바쌤~! <https://cafe.naver.com/javassem>