

07. 사용자 정의 옵션

2020년 7월 18일 토요일 오후 9:00

1. 사용자 정의 옵션

1) 헬퍼 클래스

- a. org.apache.hadoop.util 패키지에 다양한 헬퍼 클래스를 제공
- b. 맵리듀스 잡을 실행하면 잡 객체는 사용자가 설정한 Configuration 객체를 이용해 org.apache.hadoop.mapred.JobConf 객체를 생성
- c. JobConf 객체는 하둡의 환경설정 파일과 하둡 명령어에서 입력한 파라미터를 참조해 모든 태스크에 이 정보를 제공

2) GenericOptionsParser

- a. 하둡 콘솔 명령어에서 입력한 옵션을 분석
- b. 사용자가 하둡 콘솔 명령어에서 네임노드, 잡트래커 추가 구성 자원 등을 설정할 수 있는 각종 옵션을 제공

옵션	기능
-conf<파일명>	명시한 파일을 환경설정에 있는 리소스 정보에 추가
-D<옵션=값>	하둡 환경설정 파일에 있는 옵션에 새로운 값을 설정
-fs<네임노드 호스트: 네임노드 포트>	네임노드를 새롭게 설정
-jt<잡트래커 호스트: 잡트래커 포트>	잡트래커를 새롭게 설정
-files<파일1, 파일2, ..., 파일n>	로컬에 있는 파일을 HDFS에서 사용하는 공유 파일 시스템으로 복사
-libjars <JAR파일1, JAR파일2,...,JAR파일n>	로컬에 있는 JAR파일을 HDFS에서 사용하는 공유 파일 시스템으로 복사하고, 맵리듀스의 태스크 클래스패스에 추가
-archives <아카이브파일1, 아카이브파일2,...,아카이브파일n>	로컬에 있는 아카이브 파일을 HDFS에서 사용하는 공유 파일 시스템으로 복사한 후 압축을 풉니다.

- c. GenericOptionsParser 클래스는 내부적으로 Configuration 객체를 만들어 생성자에서 전달받은 환경설정 정보를 설정함
- d. 맵리듀스 프로그램을 개발할 때는 GenericOptionsParser 만 단독으로 사용하기 보다 GenericOptionsParser 가 사용하는 Configuration 객체를 상속받는 Tool 인터페이스와 GenericOptionsParser 를 내부적으로 선언한 ToolRunner 클래스를 이용함

3) Tool

- a. Tool 인터페이스는 GenericOptionsParser의 콘솔 설정 옵션을 지원하기 위한 인터페이스
- b. Tool 인터페이스는 Configuration 클래스를 상속받으며, 내부적으로 run 메서드가 정의되어 있음

4) ToolRunner

- a. ToolRunner 는 Tool 인터페이스를 구현한 클래스의 실행을 도와주는 헬퍼 클래스
- b. ToolRunner 는 GenericOptionsParser 를 사용해 사용자가 콘솔 명령어에서 설정한 옵

션을 분석하고, Configuration 객체에 설정함

- c. Configuration 객체를 Tool 인터페이스의 구현 클래스에 전달한 후, run 메서드를 실행