|  |
| --- |
| FICS |
| Hadoop |
| MapReduce |

|  |
| --- |
| SeokRae Kim  2017-7-31 |

내용

[I. MapReduce Application 개발 1](#_Toc489298247)

[1. 항공데이터 분석 1](#_Toc489298248)

1. MapReduce Application 개발
   1. 항공데이터 분석

|  |
| --- |
| **#데이터 내려 받기** |
| **http://stat-computing.org/dataexpo/2009/**    #1987년 ~ 2008년까지 총 12GB의 데이터로 하둡을 테스트하기로 한다. |

|  |
| --- |
| **#다운로드 받기 위한 셸 스크립트 작성** |
| **[root@wikibooks01 dataexpo]# vi downloads.sh**  **#!/bin/sh**  **for((i = 1987; i <= 2008; i++)) ; do**  **wget http://stat-computing.org/dataexpo/2009/$i.csv.bz2**  **bzip2 -d $i.csv.bz2**  **sed -e '1d' $i.csv > $i\_temp.csv**  **mv $i\_temp.csv $i.csv**  **done** |

|  |
| --- |
| **#셸스크립트 실행** |
| **[root@wikibooks01 dataexpo]# chmod 755 downloads.sh**  **[root@wikibooks01 dataexpo]# ./downloads.sh** |

|  |
| --- |
| **#다운로드 확인 및 용량 확인** |
| **[root@wikibooks01 dataexpo]# ls –al**  **[root@wikibooks01 dataexpo]# du -h** |

|  |
| --- |
| **#CSV 파일을 HDFS에 업로드 및 확인 – 오래 걸림** |
| **[root@wikibooks01 dataexpo]# hadoop dfs –mkdir airdata**  **[root@wikibooks01 dataexpo]# hadoop dfs –put /home/hadoop/dataexpo/\*.csv airdata**  **[root@wikibooks01 dataexpo]# hadoop dfs –ls** |