a8-shreyasa-review

balaraj

November 17, 2017

- Commit info
- Code review
- Report review
- Execution

Commit info

The code is committed on time. Commi-id : a8875d9953a91807d6fe446c4097fe961b55bf7a is used for review

Code review

- The code is very well written and easy to understand. Both Java and rmd files are concise and to the point.
- The entire process is written in 1 map-reduce call which is ideal.
- [Suggestion] Consecutive conditions can be clubbed using boolean operators.

```
Lines 74 to 77 of Flightrecords.java
if (arrDelay >= 15) {
    if (!arrDel15)
        return;
}

can be written as,
if (arrDelay >= 15 && !arrDel15) return;
```

- Unnecessary multiple casting.

```
arrDel15 = parseBoolean("" + (int) Float.parseFloat(r.get(44)));

private boolean parseBoolean(String boolS) throws NumberFormatException {
        int boolI = Integer.parseInt(boolS);
        if (boolI == 0) return false;
        else if (boolI == 1) return true;
        else throw new NumberFormatException();
}

arrDel15 is converted from String to Float to int to String to Integer
and else is not necessary at the end
The above code can be written as,

arrDel15 = parseBoolean(Integer.parseInt(r.get(44)));

private boolean parseBoolean(Integer boolI) throws NumberFormatException {
  if (boolI == 0) return false;
  else if (boolI == 1) return true;
  throw new NumberFormatException();
}
```

Report review

- Significant improvement in data representation between previous and current version of A4.
- It would have been interesting to see, how much time report generation took for the entire dataset. As major amount of computation is done in reports as well.
- The following analysis should be part of the final conclusion of the report.

```
For the period 1987 to 1996 Hartsfield-Jackson Atlanta International Air- port was the o
ne with most delays however after 1996 Chicago O'Hare International Airport has been the
most prominent in terms of delays out of the top 5 busiest airports.

Denver International Airport was the one with least delays among these till 2010 however
post that Phoenix Sky Harbor International Airport has the minimum delays out of the se
t.
```

- While it is interesting the author has included data for each year, this kind of representation takes a lot of space in the restricted four page report. It would have been better to see different kind of graph representations.

Execution

Code runs as expected on a small dataset, however report can not be generated. As report is hardbinded to full dataset.