R·I·T

Binary and CSV (Text) I/O
ISTE-121 – Homework 03

Problem Statement:

An extract of some names and personal information has been made from a baseball database. This assignment is to read the data from a text comma separated value formatted file (CSV), write the data into a binary file, then read it back in writing the data to the screen showing that the data in the binary file is intact and correct.

For any given filename on the command line, the program is to write a binary file with the same filename, but with an extension of ".bin". Example: Given filename BaseballNames2.csv, the output file is named BaseballNames2.bin. Present a user friendly and informative message when no filename was specified, or if the filename does not exist. A file with no extension, is processed the same way

Input Datafile layout:

Two text CSV data files are provided for testing, BaseballNames1.csv and BaseballNames2.csv. Each contains a header line as the first line of the file. A blank line may or may not follow. Your program must account for the header, a blank or non-blank line (record) regardless of what the sample file shows. The program must also account for leading and trailing spaces or tabs in the data, along with other errors in BaseballNames2.csv.

Field name	Data type
Name First	Text
Name Last	Text
Birth Day	Whole number
Birth Month	Whole number
Birth Year	4-digit year
Weight	Whole number
Height	Number with decimal part

BaseballNames1.csv contain no errors, and your program must be able to process this one cleanly before you move to the next data file, BaseballNames2.csv.

BaseballNames2.csv <u>may</u> have a value for each expected field, or values may be missing. Since this was a database extract, some missing (null) fields show up in the data. Missing data has a 'null' between two commas for numeric fields. Do not compare against the string "null" to detect invalid data, try to convert the expected number and handle the exception. Some records may have one or more missing fields.

Write errors to System.err[1] and only complete valid records to BaseballNames2.bin. When an error is encountered in a line read from BaseballNames2.csv, write the line and first error found to the System.err output stream. Then continue to process the

---

[1] Look up System.err

rest of the file. Your program will then read your created *.bin data file and print it to the screen. Reporting multiple errors per record is a bonus, described below.

Correctly writing the above with Javadocs, can gain you full credit for this homework.

Bonus levels:

Bonus level 1 is to reporting <u>all errors</u> in any one record. Level 1 allows printing the original record for each error reported. See sample output for Bonus Level 1.

Bonus level 2 is to report all errors, but only display the original record <u>once</u>. See sample output for Bonus Level 2. Level 2 adds to the points from level 1. You may find bonus level 2 easier than level 1. Only hand in your one best program.

Output datafile layout:

The binary output (.bin) file must have the fields in the same order as the input. No headings are written in the output file, only the data. The appropriate binary I/O data types must be used: text must be written as the appropriate text, numbers must be written as the appropriate numeric data type. You must use the appropriate methods of DataInputStream and DataOutputStream for all I/O to the ".bin" file.

Output report:

At the end of reading the csv file, display how many records were read from what filename, and how many records were written to which filename.
See output samples. Remember to count the heading and blank lines as records read.

Use formatted printing, so the output report looks impressive. Spacing does not have to be exactly as shown, but it should be close.

Hints:

Dates are provided in the western format of "month, day, year".

Use of the "continue;" statement may come in handy.

Suggest getting BaseballNames1.csv to work first. Save a backup copy of the code, then start on BaseballNames2.csv. Make sure you hand in something before the due date, otherwise your bonus points won't make up for the late deductions.

This is NOT a GUI application.

Dropbox:

Send to the dropbox your <u>one best version</u> of the program. Do NOT submit more than one version, as the wrong one may be graded, lowering your grade.
This homework is due into the dropbox by the date/time stated in the content area assignment.

R·I·T

Rochester Institute of Technology
Golisano College of Computing and Information Sciences
Department of Information Sciences & Technology

Sample outputs:

Note: The bolded **"$ java**…" line shows how the program is run on command line.
This is not printed as part of the program execution.

BaseballNames1.csv – no errors. Use this for initial testing only.

This reports that 22 records are read, this includes 2 records for heading and blank line.

```
$ java Baseball BaseballNames1.csv
Read in 22 records from BaseballNames1.csv
Wrote out 20 records to BaseballNames1.bin
```

```
First & Last name      Birthdate   Weight   Height

Starlin Castro         3/24/1990      190     72.8
Madison Bumgarner       8/1/1989      215     76.0
Jason Heyward           8/9/1989      240     77.3
Ruben Tejada          10/27/1989      160     71.2
Jenrry Mejia          10/11/1989      160     72.5
Mike Stanton           11/8/1989      235     77.9
Dayan Viciedo          3/10/1989      240     71.1
Chris Sale             3/30/1989      170     77.2
Freddie Freeman        9/12/1989      225     77.7
Clayton Kershaw        3/19/1988      225     75.4
Travis Snider           2/2/1988      235     72.0
Elvis Andrus           8/26/1988      200     72.0
Trevor Cahill           3/1/1988      220     76.0
Rick Porcello         12/27/1988      200     77.0
Brett Anderson          2/1/1988      235     76.5
Fernando Martinez     10/10/1988      200     73.0
Jhoulys Chacin          1/7/1988      215     75.2
Chris Tillman          4/15/1988      200     77.8
Neftali Feliz           5/2/1988      215     75.5
Craig Kimbrel          5/28/1988      205     71.6
```

R·I·T

**Rochester Institute of Technology
Golisano College of Computing and Information Sciences
Department of Information Sciences & Technology**

Single error reported per record (BaseballNames2.csv)

Expected to receive full credit

```
$ java Baseball BaseballNames2.csv
Error found in:  Aroldis   , Chapman    ,null,2,1988,185,76
Offending item is: Birth day

Error found in:  Ben       , Revere     ,3,null,1988,175,69
Offending item is: Birth month

Error found in:  Hank      , Conger     ,29,1,null,220,73
Offending item is: Birth year

Error found in:  Osvaldo   , Martinez   ,7,5,1988,null,70
Offending item is: Weight

Error found in:  Justin    , Upton      ,25,8,1987,205,null
Offending item is: Height

Error found in:  Jay       , Bruce      ,null,null,1987,225,75
Offending item is: Birth day

Error found in:  Shairon   , Martis     ,null,null,null,225,73
Offending item is: Birth day

Error found in:  Conor
Not enough fields in the record.

Error found in:  Ryan      , Perry      ,13,2,1987,200,null
Offending item is: Height

Error found in:  Gerardo   , Parra      ,6,5,1987,null,null
Offending item is: Weight

Error found in:  Mat       , Latos      ,null,null,null,null,null
Offending item is: Birth day

Read in 16 records from BaseballNames2.csv
Wrote out 4 records to BaseballNames2.bin


First & Last name      Birthdate  Weight  Height

Ryan Kalish            3/28/1988    205     73.0
Cameron Maybin         4/4/1987     210     75.0
Sean O'Sullivan        9/1/1987     230     74.0
Josh Reddick           2/19/1987    180     74.0
```

Multiple errors per record (Bonus level 1)
Partial sample, notice the original record prints for each error per person.
For space considerations, not all error records are shown in this sample.

```
$ java Baseball BaseballNames2.csv
Error found in:  Jay        , Bruce      ,null,null,1987,225,75
Offending data: Birth day

Error found in:  Jay        , Bruce      ,null,null,1987,225,75
Offending data: Birth month




Error found in:  Gerardo    , Parra      ,6,5,1987,null,null
Offending data: Weight

Error found in:  Gerardo    , Parra      ,6,5,1987,null,null
Offending data: Height


Read in 16 records from BaseballNames2.csv
Wrote out 4 records to BaseballNames2.bin


First & Last name      Birthdate  Weight  Height

Ryan Kalish            3/28/1988    205    73.0
Cameron Maybin         4/4/1987     210    75.0
Sean O'Sullivan        9/1/1987     230    74.0
Josh Reddick           2/19/1987    180    74.0
```

Multiple errors per record, original record shown once
All output shown in the sample (Bonus level 2)

```
$ java Baseball BaseballNames2.csv
Error found in:  Aroldis    , Chapman    ,null,2,1988,185,76
Offending data: Birth day

Error found in:  Ben        , Revere     ,3,null,1988,175,69
Offending data: Birth month

Error found in:  Hank       , Conger     ,29,1,null,220,73
Offending data: Birth year

Error found in:  Osvaldo    , Martinez   ,7,5,1988,null,70
Offending data: Weight

Error found in:  Justin     , Upton      ,25,8,1987,205,null
Offending data: Height

Error found in:  Jay        , Bruce      ,null,null,1987,225,75
Offending data: Birth day, Birth month

Error found in:  Shairon    , Martis     ,null,null,null,225,73
Offending data: Birth day, Birth month, Birth year

Error found in:  Conor
Not enough fields in the record.

Error found in:  Ryan       , Perry      ,13,2,1987,200,null
Offending data: Height

Error found in:  Gerardo    , Parra      ,6,5,1987,null,null
Offending data: Weight, Height

Error found in:  Mat        , Latos      ,null,null,null,null,null
Offending data: Birth day, Birth month, Birth year, Weight, Height

Read in 16 records from BaseballNames2.csv
Wrote out 4 records to BaseballNames2.bin


First & Last name       Birthdate  Weight  Height

Ryan Kalish             3/28/1988    205    73.0
Cameron Maybin           4/4/1987    210    75.0
Sean O'Sullivan          9/1/1987    230    74.0
Josh Reddick            2/19/1987    180    74.0
```

R·I·T

Rochester Institute of Technology
Golisano College of Computing and Information Sciences
Department of Information Sciences & Technology

ISTE-121 – Homework 3 – Binary I/O   Name: _____

Item	Possible points	Earned points
Accept filename from command line. User informative error message and exit if omitted	5	
Parses input file data properly	10	
Handles heading and blank line as expected	5	
Correctly writes binary file (and with proper name)	10	
End of writing to binary file, reports input counts & filename, output counts & filename	5	
Reads binary file using appropriate DataInputStream methods	10	
Writes formatted report to screen, including dates formatted as shown Report is clean and properly aligned	15	
Complete Javadoc and comments in code	10	
All file I/O operations are properly included in the code	10	
All errors and exceptions are handled with messages to System.err (BaseballNames2.csv for these points)	10	
Code works as expected	**10**	
Bonus points levels: Bonus 1 – Report all errors that occur on a record Bonus 2 – Report the original record once for all reported errors found in that record (includes points from bonus 1)	+5 +5	
Deductions: • Late • Coding standard violations, etc. • Submitted something other than a zip file (-5) • Duplication of code that has the same functionality		
Total:	100	

Comments: