

Data processing Challenge

Problem Definition

1. Project People Information

In this challenge you will have to receive the people assigned to a given project (given as project code), from a CSV file, and export it (the report) into a JSON File.

The [CSV](#) file is a report from an Oracle PBCS server that is being run periodically. For practical effects, you can use just the CSV file attached to this exercise to retrieve the data.

The file that should be generated is as follows:

```
{
  "Staffs": [
    {
      "startDate": "2015-01-01",
      "endDate": "2020-12-31",
      "projectCode": "INPIS001"
      "referenceNumber": 5628
    }
    ...
  ]
}
```

Namely:

Fields:

Name	Type	Notes
startDate	String	Start date of the assignation of the person to this project (if is not possible to get the exact day, the day should be set as first day of the month)
endDate	String	End date of the assignation of the person to this project (if is not possible to get the exact day, the day should be set as the end day of the month)
projectCode	String	Project code given as parameter (command line), which identifies the project in the report, as well.
referenceNumber	Number	HR Identifier of the team member.

The dates can be inferred from the report ,as

- **Start date:** by the earliest date the team member has been assigned (reported time against the project) .
- **End date:** by the latest forecast date for the team member to be assigned (forecast time against the project).

On the other hand the report contains the HR ID, but not the first name, last name and user name. Therefore, the HR ID on this report should be crossed with HCM system to obtain this data.

2. Source Report

The source report contains the hours either actual and forecast that a team member has in their reports, therefore it allows to infer the start

dates and end dates of an employee in a project. You will have to find a way to group the team members of a given project in order to expose them in the services, since the project code is the parameter for invoking the service.

The file **ampas_ExportedData_Staff.1.1.csv** has the following structure:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Staff	Home Base	Committee	Relocated	Committee	STR Days	Committee	Point-of-V	Data Load	Cube Name				
2	HR7607	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	CLD	No Locati	No Role	ISDIT999
3	HR7607	1	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	CLD	No Locati	No Role	ISDIT999
4	HR7607	8	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	CLD	No Locati	No Role	ISDIT999
5	HR7607	4	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	CLD	No Locati	No Role	ISDIT999
6	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
7	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
8	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
9	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
10	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
11	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
12	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
13	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
14	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
15	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
16	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
17	HR7661	10	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	MDD	No Locati	No Role	M&AIMP
18	HR67206	4	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	IND - Disc	No Locati	No Role	M&AIMP
19	HR67206	5	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	IND - Disc	No Locati	No Role	M&AIMP
20	HR67206	6.125	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	IND - Disc	No Locati	No Role	PSPAY999
21	HR67206	3.125	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	IND - Disc	No Locati	No Role	PSRCP999
22	HR67206	0.5	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	IND - Disc	No Locati	No Role	AJGDV018
23	HR67206	0.375	#missing	#missing	#missing	#missing	#missing	"No Grad	No Entity	Default C	IND - Disc	No Locati	No Role	AJGDV018

- The first color (in blue gray) is the HR ID of the **team member**.
- The column in green is the **project code**.
- Column in light blue is the **fiscal year**. This column, combined with the Period (in orange) and the month (in yellow) will allow to obtain both start date and end date of the team member on the project.
- Remaining columns are not relevant to the information we want to retrieve to project Staff.

3. Retrieval of Project People Data

To retrieve the set of team members for a given project ID, the report should follow the next rules:

1. The latest period registered will be used for the dates calculation.
2. For each group of employee + project + financial year there could be different start/end dates.
3. The start date will be calculated based on the latest period registered by retrieving the lowest month for that period.
4. The end date will be calculated based on the latest period registered by retrieving the highest month for that period.
5. For months between Jul-Dec, 1 year is subtracted from the financial year to get the date.
6. For months between Jan-Jun, the year for the date is the same as the financial year.
7. The DAY_OF_MONTH for start/end dates is always 1.
8. If there's a gap between months (January is registered as the first month for the period, February is not, March is registered) there would be a new start/end date.

Note: Fiscal years begin in July and end in Jun for the next year. As an example the Fiscal year 2017 starts in July 2016 and ends in June 2017.

Project Setup

1. Configure your environment setup according to [Java Environment Setup](#)
2. Clone the repository:

```
git clone https://[user]@bitbucket.endava.com/scm/bd/java-rampup.git
```

3. Create a new branch for your challenge:

```
git checkout -b [endavauser]-dataprocessingchallenge
```

4. Import the **dataprocessing-challenge** gradle project as is described in [IntelliJ Import Project](#).
5. Push your changes when you finish:

```
git push origin -b [endavauser]-dataprocessingchallenge
```