1. Prepare pentaho transformation that will read data from excel file and dump into tables employee, project and balance with following transformations performed.

Note:

1. Extra spaces in name should be trimmed.
2. Capitalize name.
3. Gender should be dumped as M and F.
4. Add new field named is\_active with default value as ‘Y’ for all employees.
5. Calculate age based on date\_of\_birth and years\_of\_joining based on joining\_date
6. Extract first\_name and last\_name.
7. Calculate level as
   1. If years\_of\_joining<3, level=Beginner
   2. If years\_of\_joining>=3 and <=5, level=Intermediate
   3. If years\_of\_joining>5, level=Advanced
8. All date fields should have format yyyy-mm-dd
9. Only consider employees whose project\_code is known

Final table will contain following fields

|  |
| --- |
| Table: employee |
| Full\_name |
| First\_name |
| Last\_name |
| Gender |
| Date\_of\_birth |
| Age |
| Joining\_date |
| Level |
| Project\_code |
| Salary |
| Is\_active |

|  |
| --- |
| Table: project |
| Project\_code |
| Project\_name |

|  |
| --- |
| Table: balance |
| account\_number |
| tran\_date |
| balance |

1. Prepare pentaho transformation to find following:
   1. Total number of male and female in company
   2. Total number of male and female in each project of company

Output of 2.a. should look like

|  |  |
| --- | --- |
| Gender | Count |
| Male |  |
| Female |  |

Output of 2.b. should be like

|  |  |  |
| --- | --- | --- |
| Project | Gender | count |
| A+ Property | Male |  |
| A+ Property | Female |  |
| A+ Auto | Male |  |
| A+ Auto | Female |  |
| ICDPP | Male |  |
| ICDPP | Female |  |
| Datahub | Male |  |
| Datahub | Female |  |
|  |  |  |
|  |  |  |

Note:

If count is null in some, replace it with 0.

1. Prepare transformation to get following output from above question

Note: Make use of row normalizer and row denormalizer

|  |  |  |
| --- | --- | --- |
| Project | gender | Count |
| A+ Property | Male |  |
| A+ Property | Female |  |
| A+ Auto | Male |  |
| A+ Auto | Female |  |
| ICDPP | Male |  |
| ICDPP | Female |  |
| Datahub | Male |  |
| Datahub | Female |  |

|  |  |  |
| --- | --- | --- |
| Project | Male | Female |
| A+ Property |  |  |
| A+ Auto |  |  |
| ICDPP |  |  |
| Datahub |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gender | A+ Property | A+ Auto | ICDPP | Datahub |
| Male |  |  |  |  |
| Female |  |  |  |  |

1. Prepare transformation to calculate following
   1. Sum, min, max, average salary earned by employees in company
   2. Sum, min, max, average salary earned by employees of each gender in company
   3. Sum, min, max, average salary earned by employees in each project in company
   4. Sum, min, max, average salary earned by employees of each gender in each project in company

Output of 4.a. should be like:

|  |  |  |  |
| --- | --- | --- | --- |
| sum\_salary | min\_salary | max\_salary | average\_salary |

Output of 4.b. should be like

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gender | sum\_salary | min\_salary | max\_salary | average\_salary |
| Male |  |  |  |  |
| Female |  |  |  |  |

Output of 4.c. should be like:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| project | sum\_salary | min\_salary | max\_salary | average\_salary |
| A+ Property |  |  |  |  |
| A+ Auto |  |  |  |  |
| ICDPP |  |  |  |  |
| Datahub |  |  |  |  |
| Content Management |  |  |  |  |

Output of 4.d. should be like:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| project | gender | sum\_salary | min\_salary | max\_salary | average\_salary |
| A+ Property | Male |  |  |  |  |
| A+ Property | Female |  |  |  |  |
| A+ Auto | Male |  |  |  |  |
| A+ Auto | Female |  |  |  |  |
| ICDPP |  |  |  |  |  |
| ICDPP |  |  |  |  |  |
| Datahub |  |  |  |  |  |
| Datahub |  |  |  |  |  |
| Content Management |  |  |  |  |  |
| Content Management |  |  |  |  |  |

1. Prepare transformation to find balance of current day and previous 2 days

Output should be:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| account\_number | tran\_date | balance | balance\_before\_1\_day | balance\_before\_2\_days |

1. Prepare a pentaho script that will iterate through each project id and dump employees working on that project to respective text file delimited by |#^#|.

Result of all employees working for first project, should be dumped into first\_project\_name.csv.

Result of all employees working for second project, should be dumped into second\_project\_name.csv.

1. Prepare pentaho job that will execute above transformations and jobs one after another. If process fails at any point, send failure mail to certain email address.