

Name - Harshit Parsai
Roll no. - 06
Section - KM054
Registration no. - 11702448
Subject - INT324

Informatica Data Quality

Academic Task-2 (Report)

Question 6)

Using Employees Dataset Create a mapping that parse the data by splitting the "EmployeesName" into Firstname and Lastname into the target file

Solution)

We are given with the dataset in which Name of the employee is EmployeesName and we have to split into the Firstame and Lastame. For doing this question we need to have Informatica Developer tool installed and Our source and target file I am attaching the snapshot of the steps that have to be followed and following the same step I have done the practical so after the snapshot of the process I am attaching the screenshot of the practical and the result

Solution

To do this question we have to import Source and Target files into our Developer tool.

- To import a file.

- Right click on Physical Data Object → select Flat Files → select the file location → Next → select Delimiter and heading column. → Finish.

- Click on the file go to advance and change the Source File and Target File Directories Location

- Create Mapping.

- Right click on Folder → New → Mapping → Give name → open it by Double clicking

- Drag and drop the source & target file.

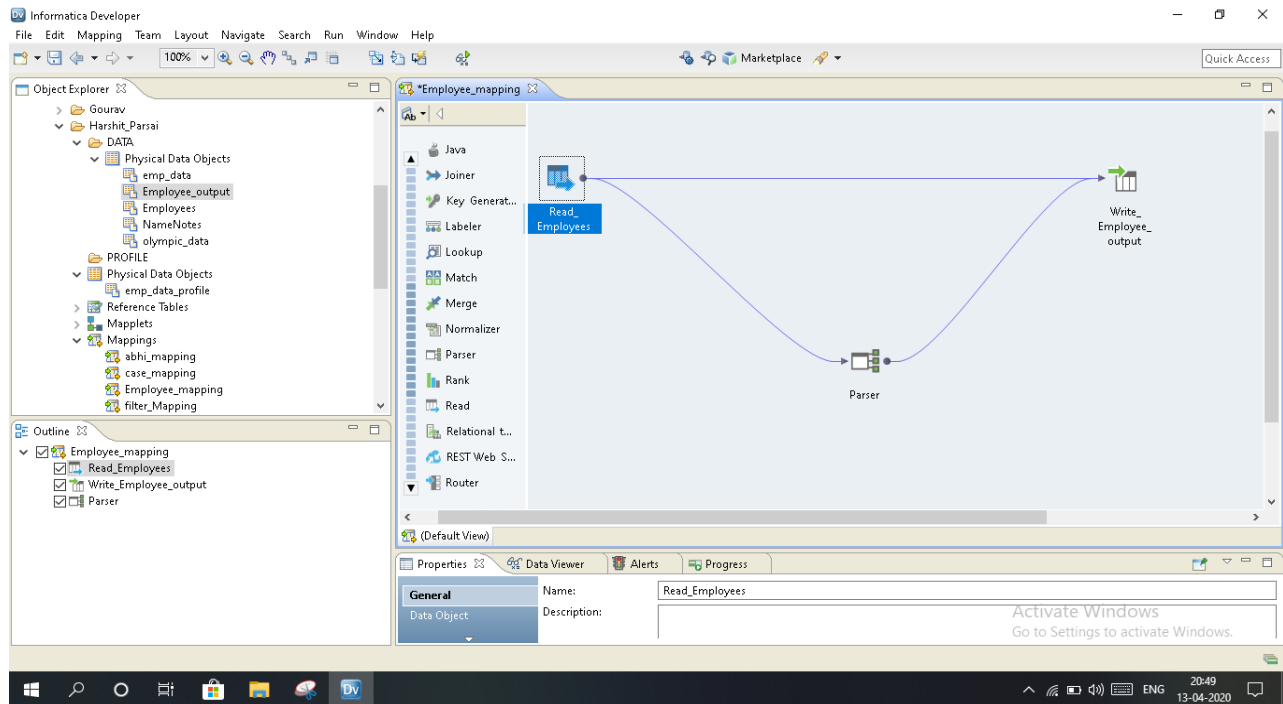
- Add a Parser transformation.

- Source - FullName drag and drop into Input to Parser

- from Properties of Parser → select strategies
→ Create a New Parser → Give a name →
select input FullName → select Delimiters
as space ↘ → click next → select Parse
using Token set → select undefined → select
Output.
- Map these output to target .
- Run the mapping → Right click → Run mapping
- Run Data Viewer on the output file you
clearly see that in Target file FirstName
is perfectly splitted into Firstname and Lastname
I have attached the screenshots of this.

Thankyou.

Screenshot of the final mapping that we have got -

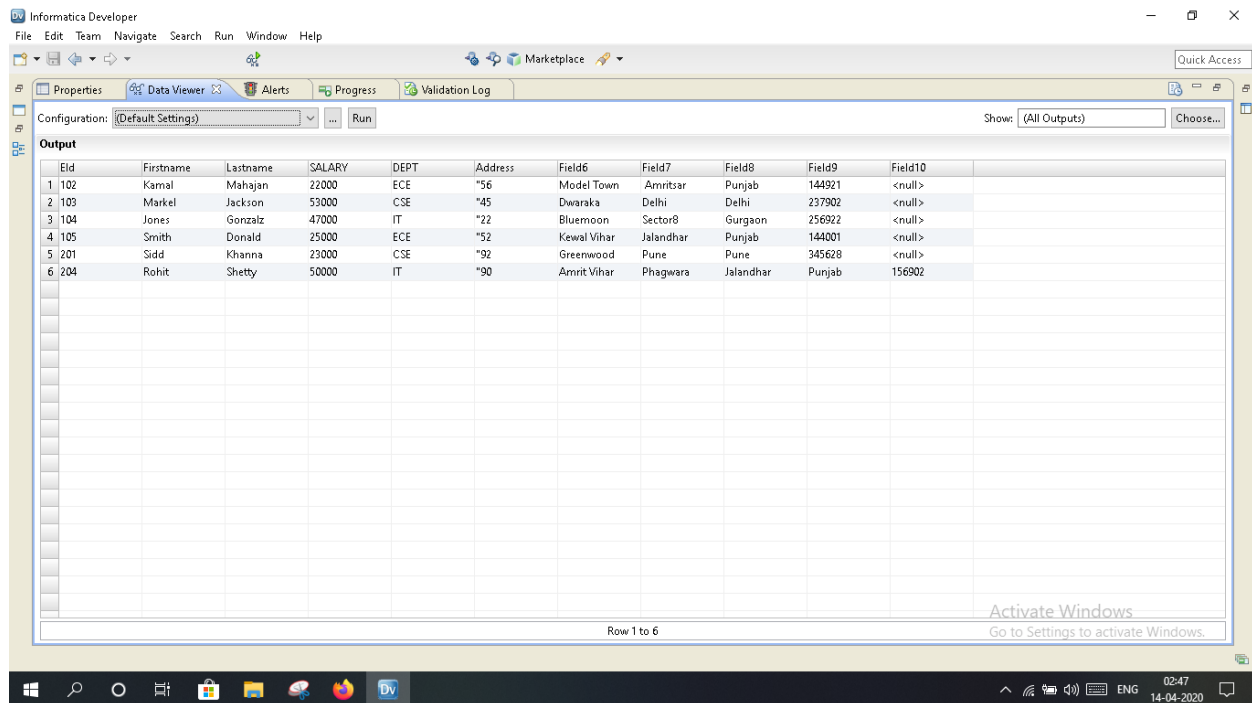


Screenshot of the data in the source file we can see that FULL_NAME

The screenshot shows the 'Data Viewer' tab in Informatica Developer, displaying the data from the 'Employee_mapping' mapping. The table has 10 columns: 'Eld', 'FULL_NAME', 'SALARY', 'DEPT', 'Address', 'Field6', 'Field7', 'Field8', 'Field9', and 'Field10'. The data is as follows:

Eld	FULL_NAME	SALARY	DEPT	Address	Field6	Field7	Field8	Field9	Field10
1 101	Rahul Gupta	10000	CSE	"98	Green Avenue	Ludhiana	Punjab	145678	<null>
2 102	Kamal Mahajan	22000	ECE	"56	Model Town	Amritsar	Punjab	144921	<null>
3 103	Markel Jackson	53000	CSE	"45	Dwaraka	Delhi	Delhi	237902	<null>
4 104	Jones Gonzalz	47000	IT	"22	Blumoon	Sector8	Gurgaon	256922	<null>
5 105	Smith Donald	25000	ECE	"52	Kewal Vihar	Jalandhar	Punjab	144001	<null>
6 201	Sidd Khanna	23000	CSE	"92	Greenwood	Pune	Pune	345628	<null>
7 204	Rohit Shetty	50000	IT	"90	Amrit Vihar	Phagwara	Jalandhar	Punjab	156902

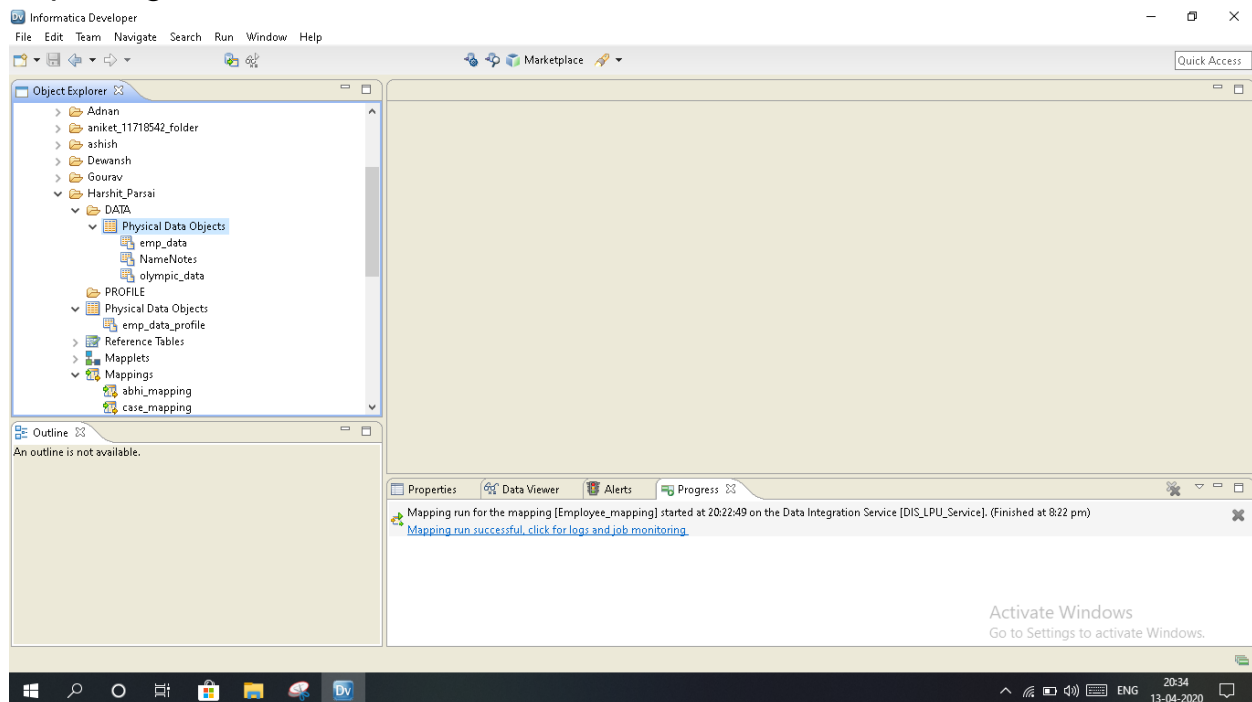
Screenshot of the result we got after running the mapping We can see that the FULL_NAME column is perfectly splitted into the Firstname and Lastname



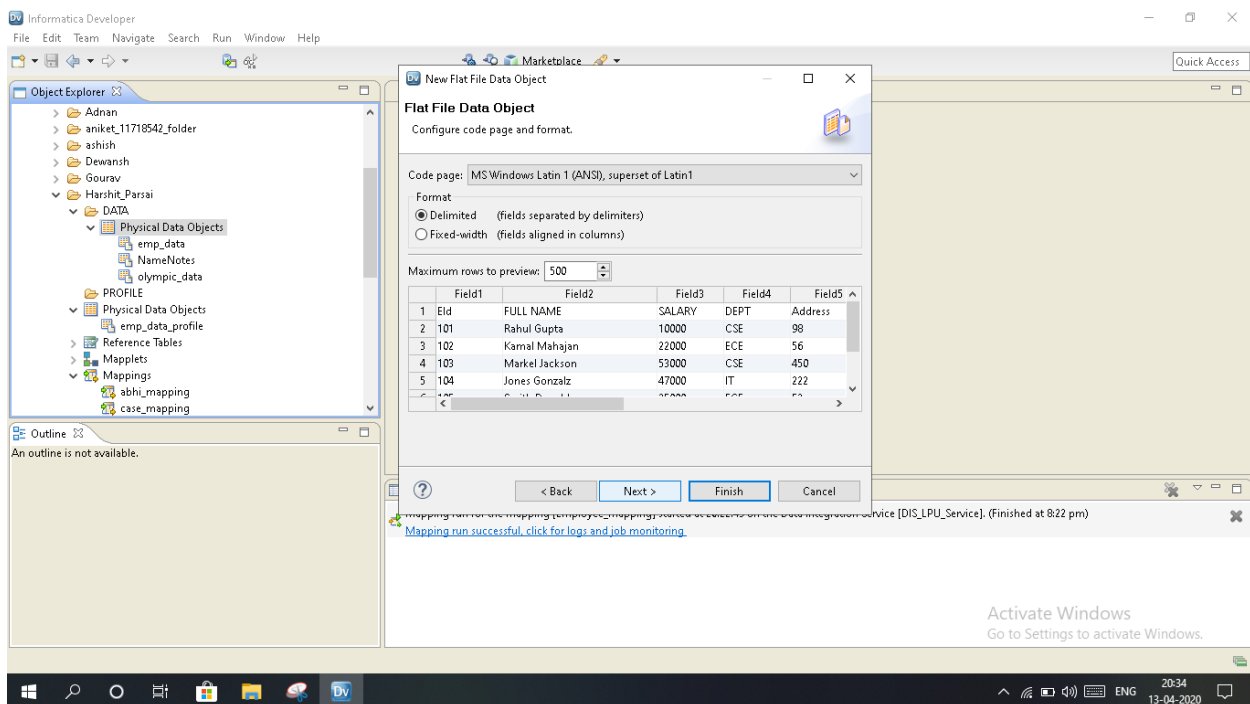
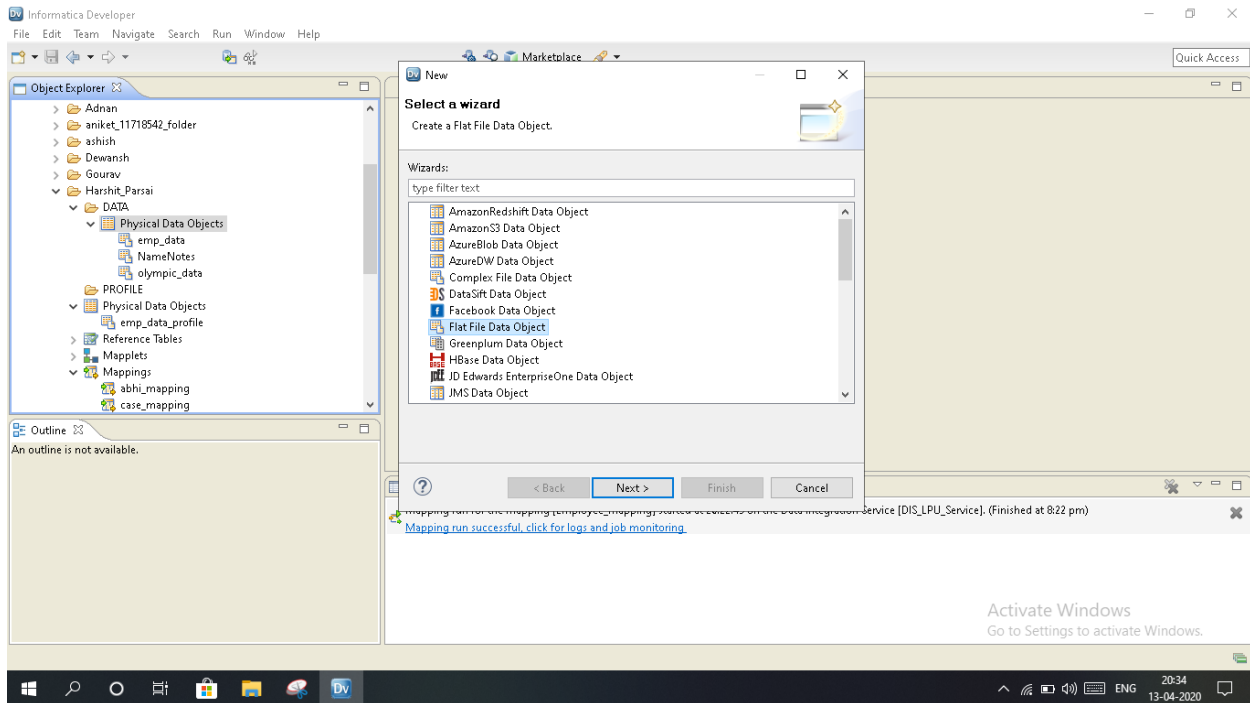
The screenshot shows the Informatica Developer Data Viewer window. The 'Output' tab is selected, displaying a table with 11 columns: Eid, Firstname, Lastname, SALARY, DEPT, Address, Field6, Field7, Field8, Field9, and Field10. The table contains 6 rows of data. The status bar at the bottom indicates 'Row 1 to 6'.

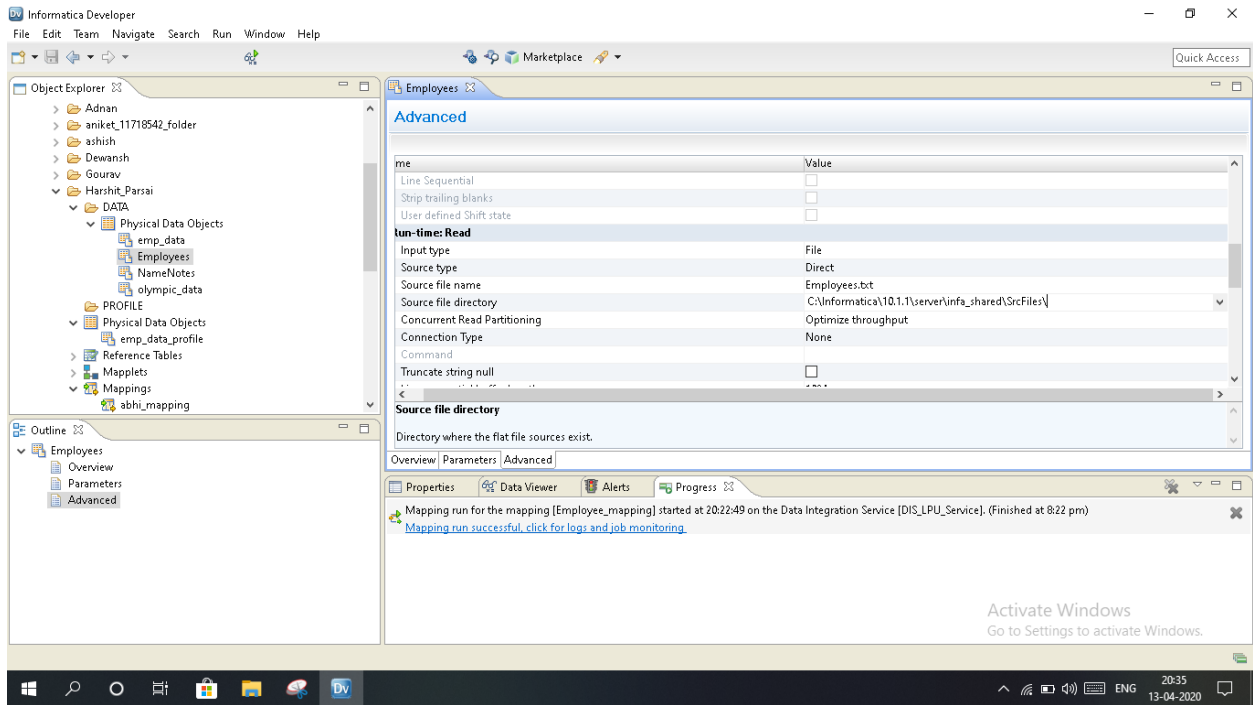
Eid	Firstname	Lastname	SALARY	DEPT	Address	Field6	Field7	Field8	Field9	Field10
1 102	Kamal	Mahajan	22000	ECE	*56	Model Town	Amritsar	Punjab	144921	<null>
2 103	Markel	Jackson	53000	CSE	*45	Dwaraka	Delhi	Delhi	237902	<null>
3 104	Jones	Gonzalz	47000	IT	*22	Bluemoon	Sector8	Gurgaon	256922	<null>
4 105	Smith	Donald	25000	ECE	*52	Kewal Vihar	Jalandhar	Punjab	144001	<null>
5 201	Sidd	Khanna	23000	CSE	*92	Greenwood	Pune	Pune	345628	<null>
6 204	Rohit	Shetty	50000	IT	*90	Amrit Vihar	Phagwara	Jalandhar	Punjab	156902

Importing a File - >

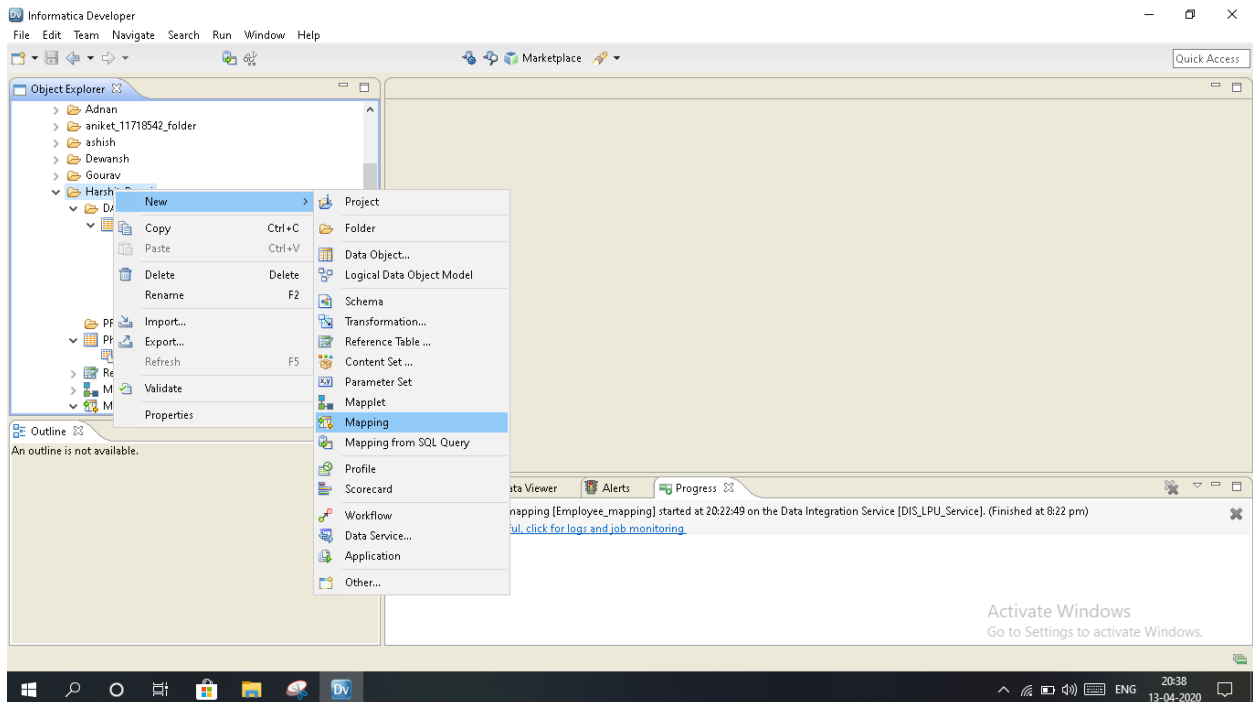


The screenshot shows the Informatica Developer interface. The 'Object Explorer' on the left displays a tree structure of objects, including 'Physical Data Objects', 'PROFILE', 'Reference Tables', 'Mapplets', and 'Mappings'. The 'Mappings' folder is expanded, showing 'abhi_mapping' and 'case_mapping'. The 'Outline' window at the bottom left shows the message 'An outline is not available.' The 'Data Viewer' window at the bottom right displays a message: 'Mapping run for the mapping [Employee_mapping] started at 20:22:49 on the Data Integration Service [DIS_LPU_Service]. (Finished at 8:22 pm) Mapping run successful, click for logs and job monitoring.'

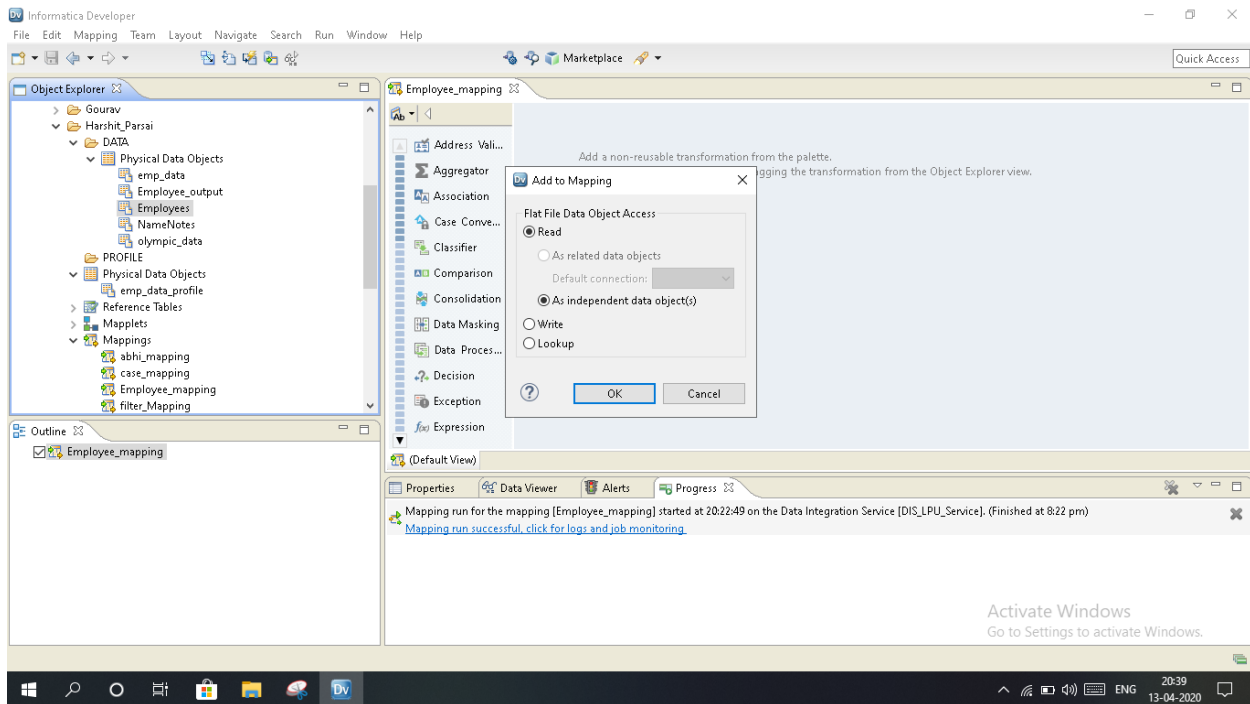




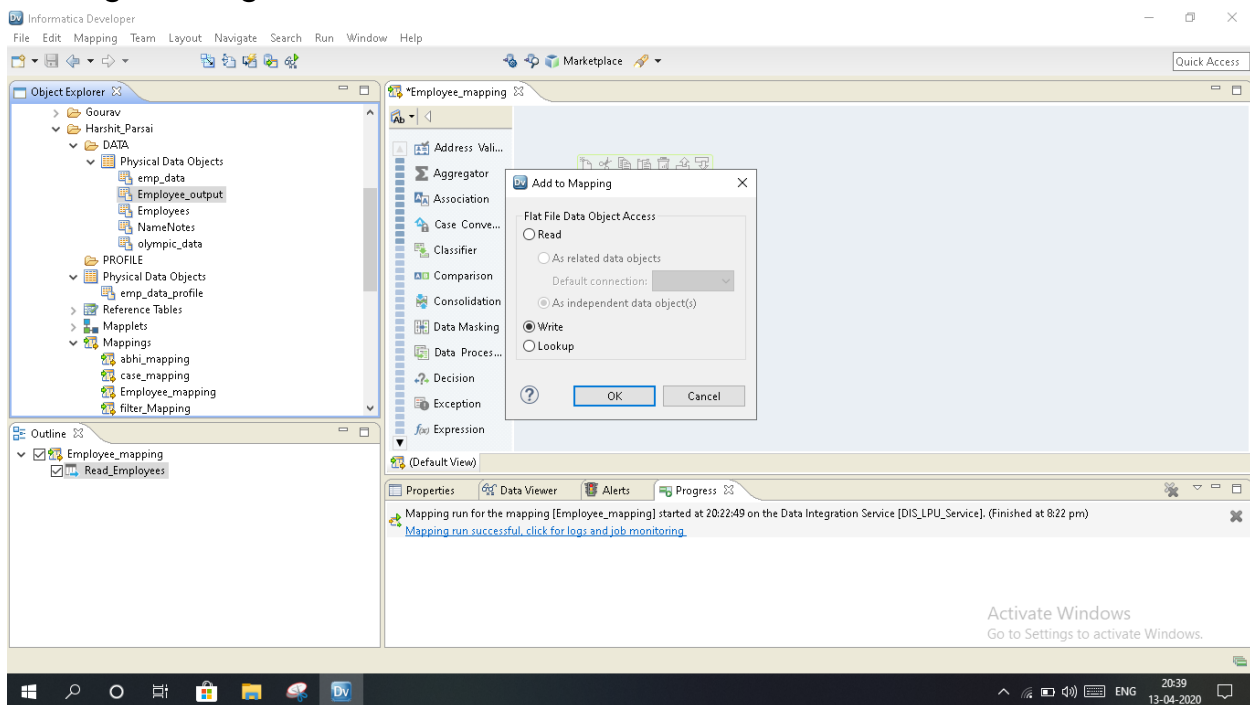
Adding The Mapping



Adding the source file in the read mode



Adding the target file in write mode



Adding Parser

Informatica Developer

File Edit Mapping Team Layout Navigate Search Run Window Help

100% Marketplace Quick Access

Object Explorer

- Gourav
- Harshit_Parsai
- DATA
- Physical Data Objects
 - emp_data
 - Employee_output
 - Employees
 - NameNotes
 - olympic_data
- PROFILE
- Physical Data Objects
 - emp_data_profile
- Reference Tables
- Mapplets
- Mappings
 - abhi_mapping
 - case_mapping
 - Employee_mapping
 - filter_Mapping

Outline

- Employee_mapping
 - Read_Employees
 - Write_Employee_output
 - Parser

Properties

General

Name: Parser

Description:

Activate Windows
Go to Settings to activate Windows.

Informatica Developer

File Edit Mapping Team Layout Navigate Search Run Window Help

100% Marketplace Quick Access

Object Explorer

- Gourav
- Harshit_Parsai
- DATA
- Physical Data Objects
 - emp_data
 - Employee_output
 - Employees
 - NameNotes
 - olympic_data
- PROFILE
- Physical Data Objects
 - emp_data_profile
- Reference Tables
- Mapplets
- Mappings
 - abhi_mapping
 - case_mapping
 - Employee_mapping
 - filter_Mapping

Outline

- Employee_mapping
 - Read_Employees
 - Write_Employee_output
 - Parser

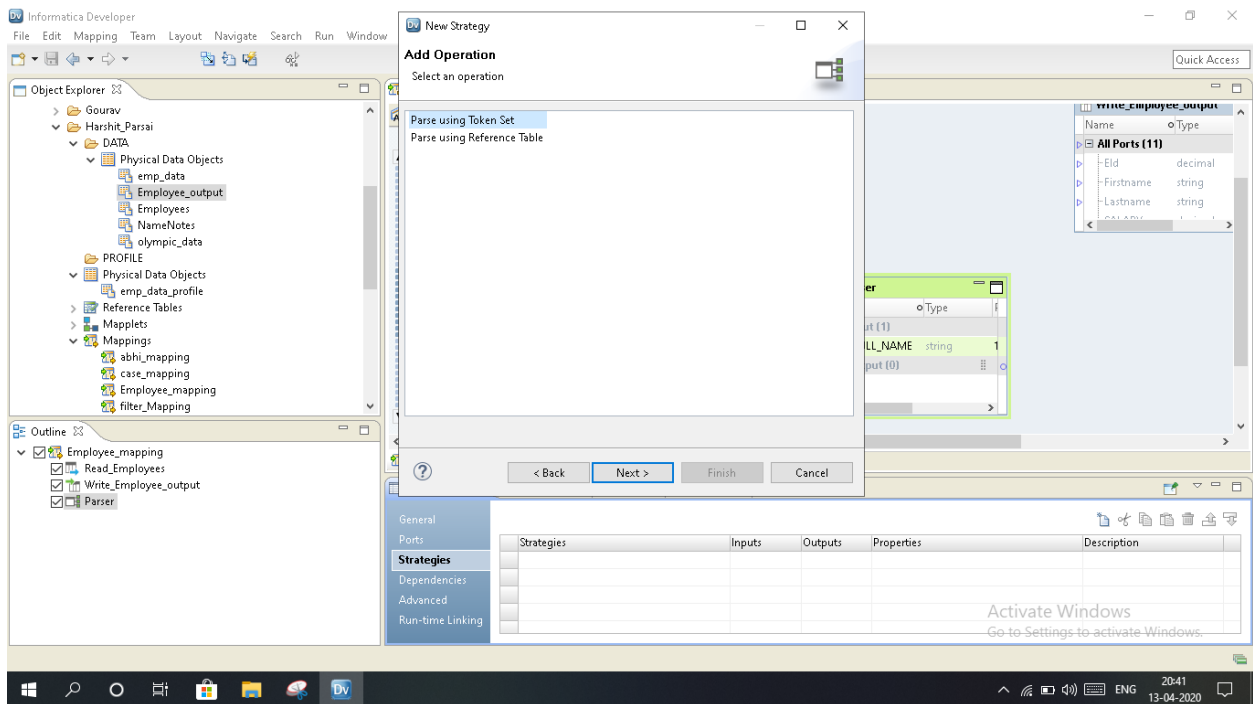
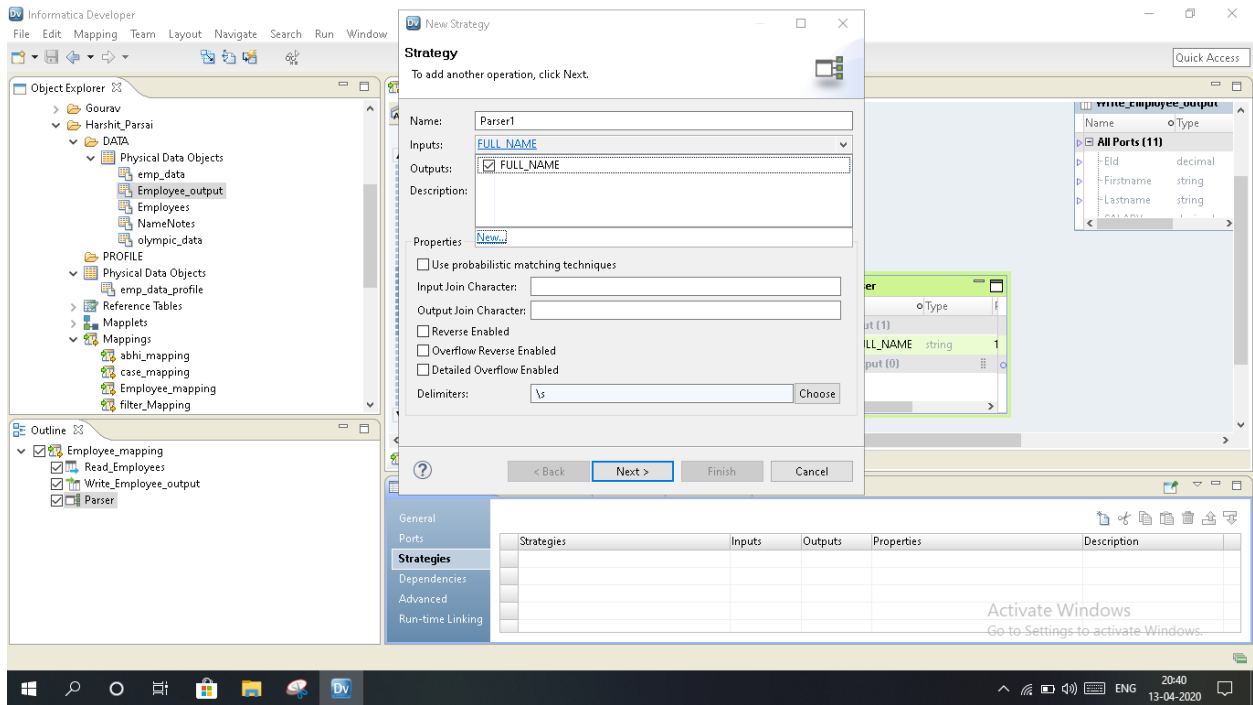
Properties

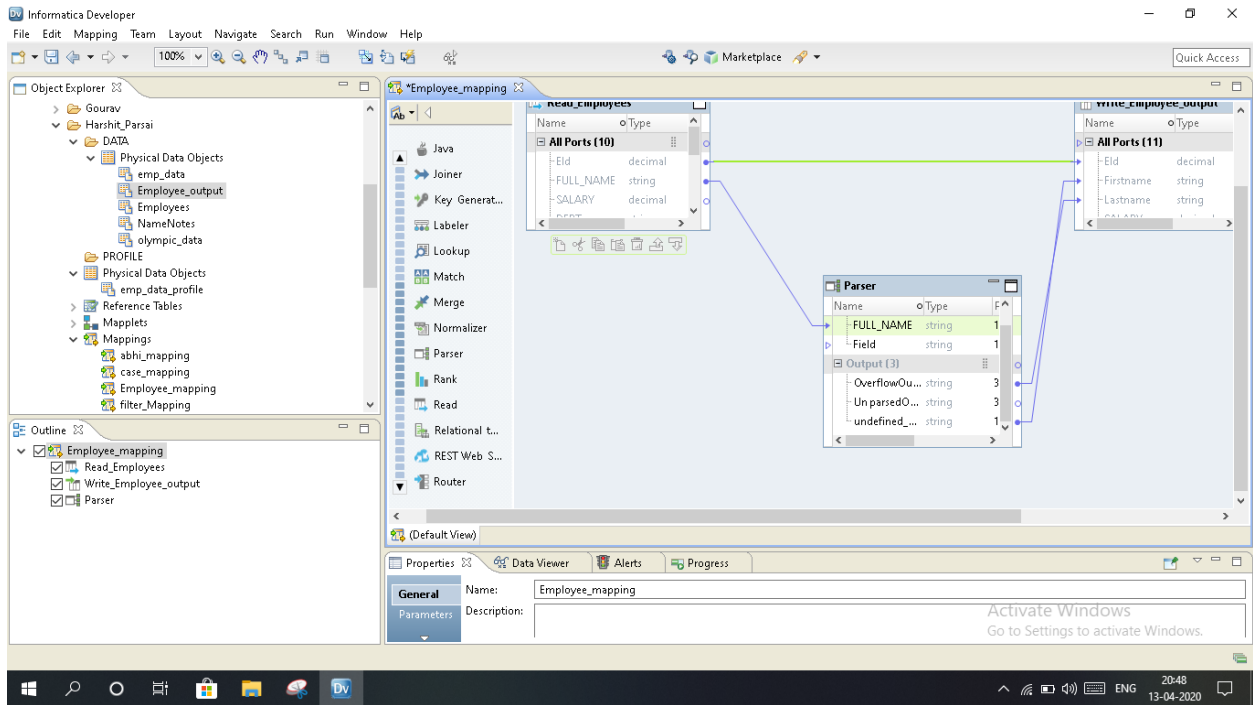
General

Name: Parser

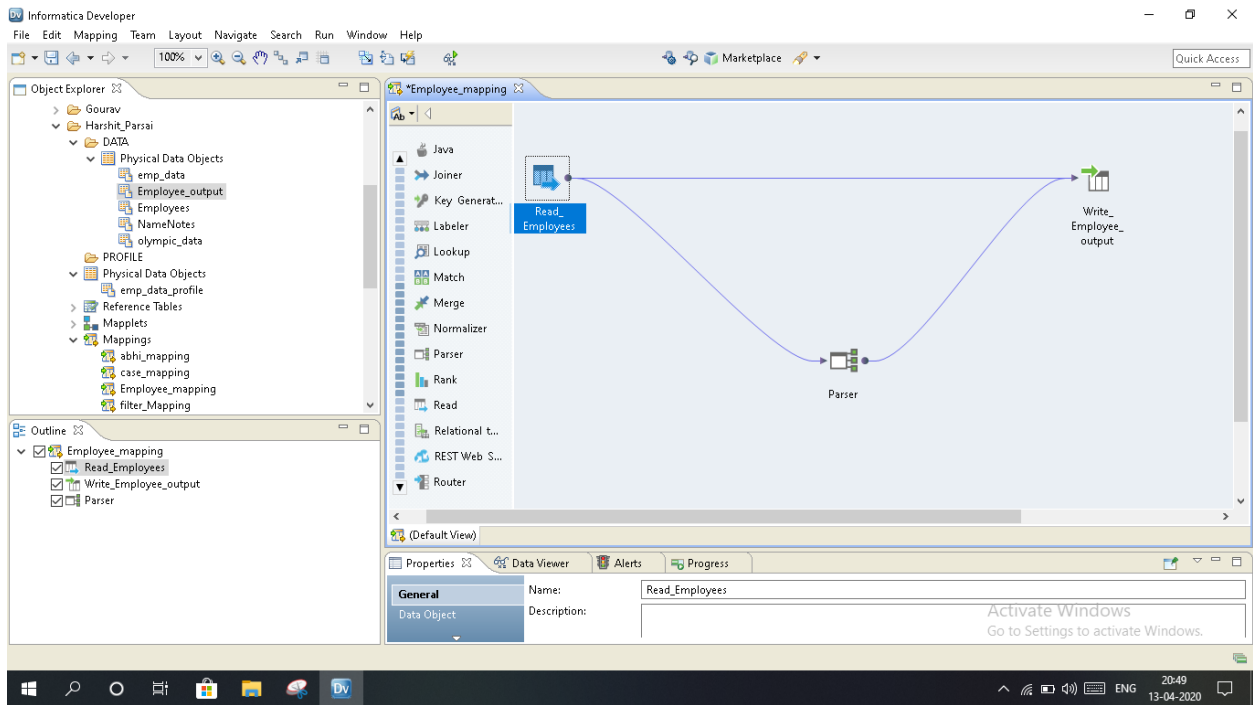
Description:

Activate Windows
Go to Settings to activate Windows.





This is what we got after this



We can profile now in parser to see the expected output and the the parser output

Informatica Developer

File Edit Team Navigate Search Run Window Help

Object Explorer

- Gourav
- Harshit_Parsai
- DATA
- Physical Data Objects
- emp_data
- Employee_output
- Employees
- NameNotes
- olympic_data
- PROFILE
- Physical Data Objects
- emp_data_profile
- Reference Tables
- Maplets
- Mappings
- abhi_mapping
- case_mapping
- Employee_mapping
- filter_Mapping

Outline

- Overview
- Definition
- Results

Results

Column Profiling

Output

All 7 rows. Last run on:13 Apr, 2020 8:51:29 PM IST

Column	Distinct Values	% Distinct	Nulls	%
Output				
OverflowOutput	7	100.00	-	
UnparsedOutput	1	14.28	-	
undefined_output	7	100.00	-	

Details

Show: Values

Value	Frequency	Percent	Chi
Smith	1	14.29%	
Sidd	1	14.29%	
Rohit	1	14.29%	
Rahul	1	14.29%	
Markel	1	14.29%	
Karnal	1	14.29%	
Jones	1	14.29%	

Overview Definition Results

Properties Data Viewer Alerts Progress Validation Log

Show: (All Outputs) Choose... Run

Output

Activate Windows
Go to Settings to activate Windows.

Informatica Developer

File Edit Team Navigate Search Run Window Help

Object Explorer

- Gourav
- Harshit_Parsai
- DATA
- Physical Data Objects
- emp_data
- Employee_output
- Employees
- NameNotes
- olympic_data
- PROFILE
- Physical Data Objects
- emp_data_profile
- Reference Tables
- Maplets
- Mappings
- abhi_mapping
- case_mapping
- Employee_mapping
- filter_Mapping

Outline

- Overview
- Definition
- Results

Results

Column Profiling

Output

All 7 rows. Last run on:13 Apr, 2020 8:51:29 PM IST

Column	Distinct Values	% Distinct	Nulls	%
Output				
OverflowOutput	7	100.00	-	
UnparsedOutput	1	14.28	-	
undefined_output	7	100.00	-	

Details

Show: Values

Value	Frequency	Percent	Chi
Shetty	1	14.29%	
Mahajan	1	14.29%	
Khanna	1	14.29%	
Jackson	1	14.29%	
Gupta	1	14.29%	
Gonzalz	1	14.29%	
Donald	1	14.29%	

Overview Definition Results

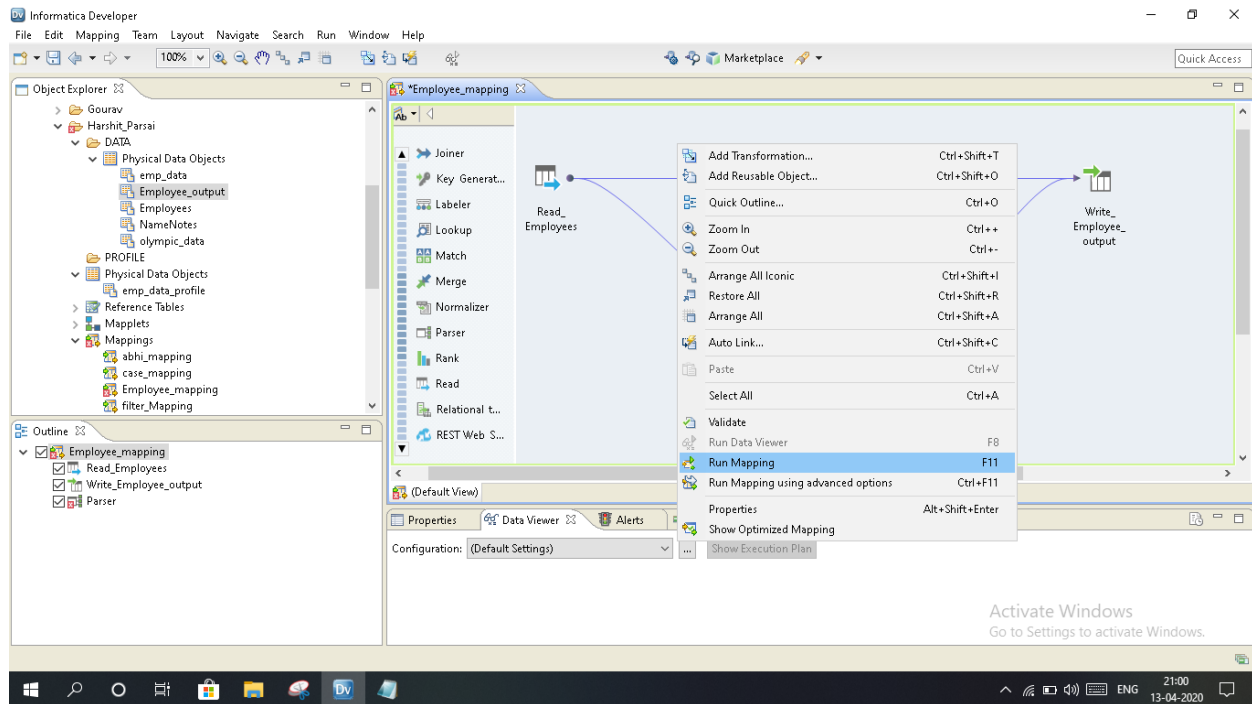
Properties Data Viewer Alerts Progress Validation Log

Show: (All Outputs) Choose... Run

Output

Activate Windows
Go to Settings to activate Windows.

Running Mapping



Final Result in the output file

Informatica Developer

File Edit Team Navigate Search Run Window Help

Properties Data Viewer Alerts Progress Validation Log

Configuration: (Default Settings) Run Show: (All Outputs) Choose...

Output

	Eid	Firstname	Lastname	SALARY	DEPT	Address	Field6	Field7	Field8	Field9	Field10
1	102	Kamal	Mahajan	22000	ECE	"56	Model Town	Amritsar	Punjab	144921	<null>
2	103	Markel	Jackson	53000	CSE	"45	Dwaraka	Delhi	Delhi	237902	<null>
3	104	Jones	Gonzalz	47000	IT	"22	Bluemoon	Sector8	Gurgaon	256922	<null>
4	105	Smith	Donald	25000	ECE	"52	Kewal Vihar	Jalandhar	Punjab	144001	<null>
5	201	Sidd	Khanna	23000	CSE	"92	Greenwood	Pune	Pune	345628	<null>
6	204	Rohit	Shetty	50000	IT	"90	Amrit Vihar	Phagwara	Jalandhar	Punjab	156902

Row 1 to 6

Activate Windows
Go to Settings to activate Windows.

Windows taskbar: 02:47 14-04-2020