# Big Data Hadoop Training

Session 8 Assignment 2 Solution:

Task

**Create a sample dataset and implement the below Pig commands on the same dataset.**

**Sample dataset (used) –**

/My\_Downloads/Assignments\_BDH/BDDS8A2/student\_details.txt

/My\_Downloads/Assignments\_BDH/BDDS8A2/duplicate\_student\_details.txt



**Pig Commands:**

1. Concat

**Pig Latin Script**:

student\_details = LOAD './student\_details.txt' USING PigStorage(',') as (Id:int, firstName:chararray, lastName:chararray, age:int, phone\_no:chararray, gpa:int);

concat\_output = FOREACH student\_details GENERATE CONCAT (firstName,'\*',lastName);

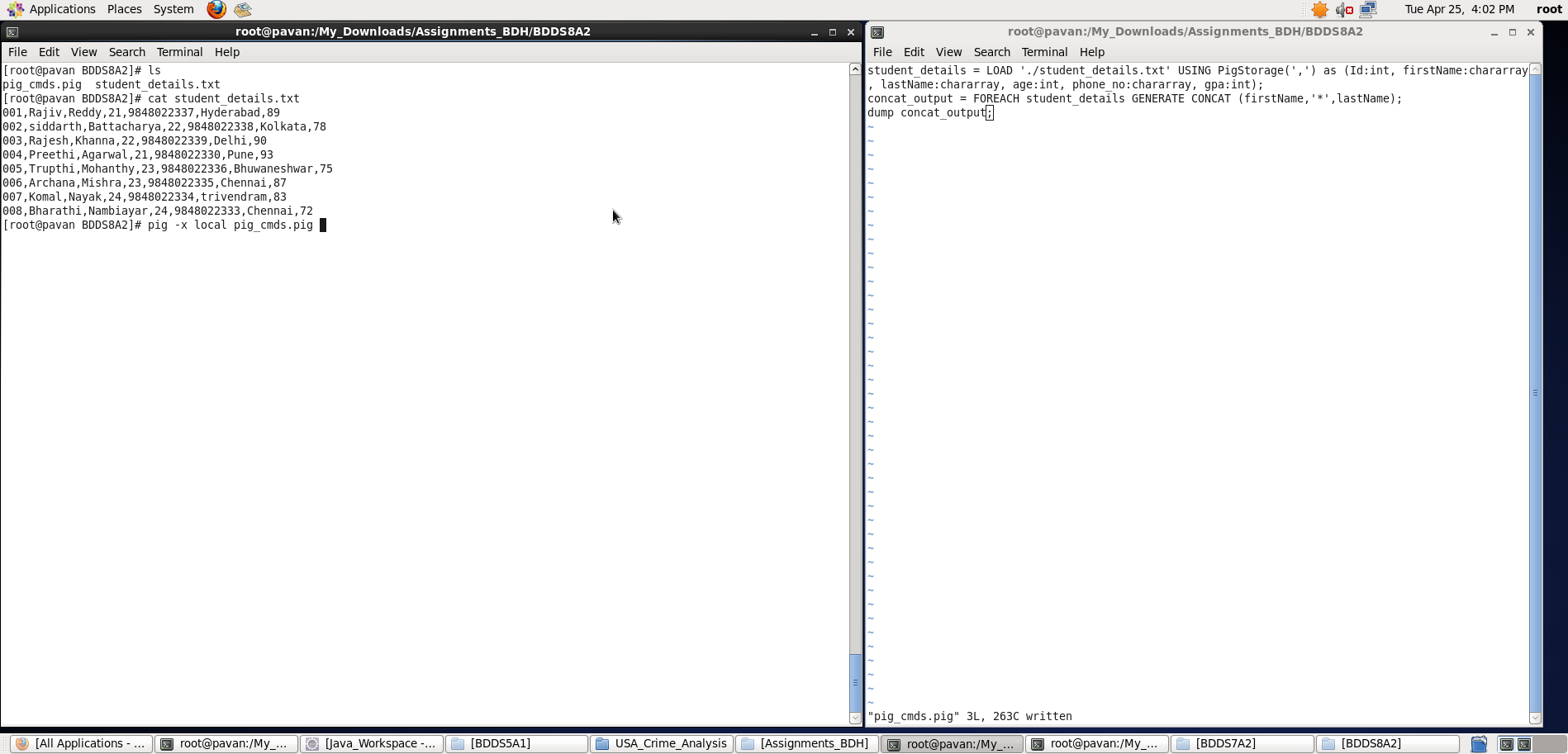
dump concat\_output;

**Explanation:**

* Loaded data from student\_details.txt into “student\_details” relation. Here ‘,’ is the delimiter and schema is provided accordingly
* Then, for all records I have generated a new relation which contains concatenated result of the firstName and lastName using CONCAT function. Here ‘\*’ is kept in between the firstName and lastName and all the three expressions are concatenated.
* Outputting the concat\_output

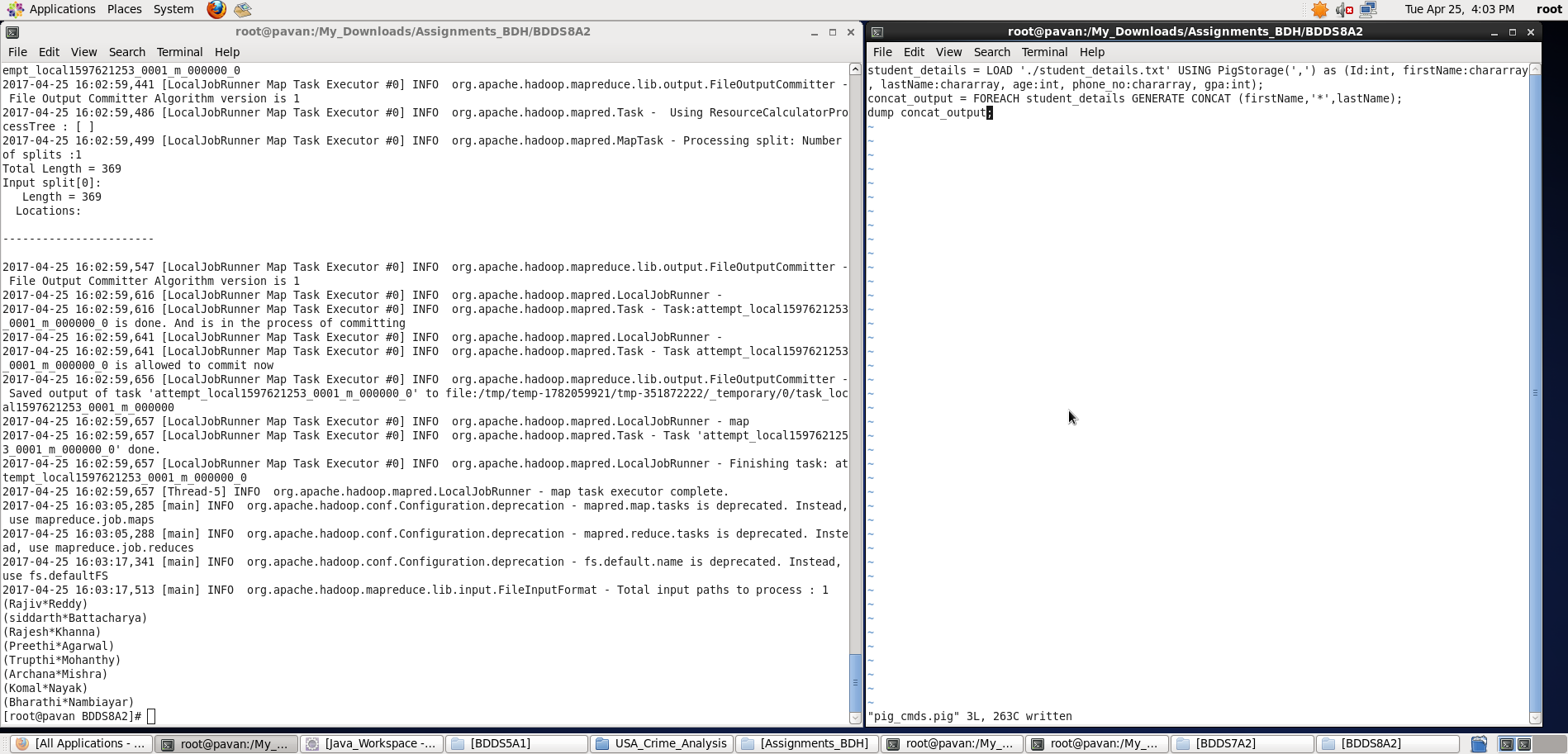
**Screenshots**

Running **Pig Latin Script in Local Mode** inside /My\_Downloads/Assignments\_BDH/BDDS8A2/ directory

****

Output displayed: **<concat\_output>**

**Note : We can store the output if required in a file using “store” command.**

****

1. Tokenize

**Pig Latin Script**:

student\_details = LOAD './student\_details.txt' USING PigStorage(',') as (Id:int, firstName:chararray, lastName:chararray, age:int, phone\_no:chararray, gpa:int);

concat\_output = FOREACH student\_details GENERATE CONCAT (firstName,'\*',lastName);

tokenize\_output = FOREACH concat\_output GENERATE TOKENIZE ($0);

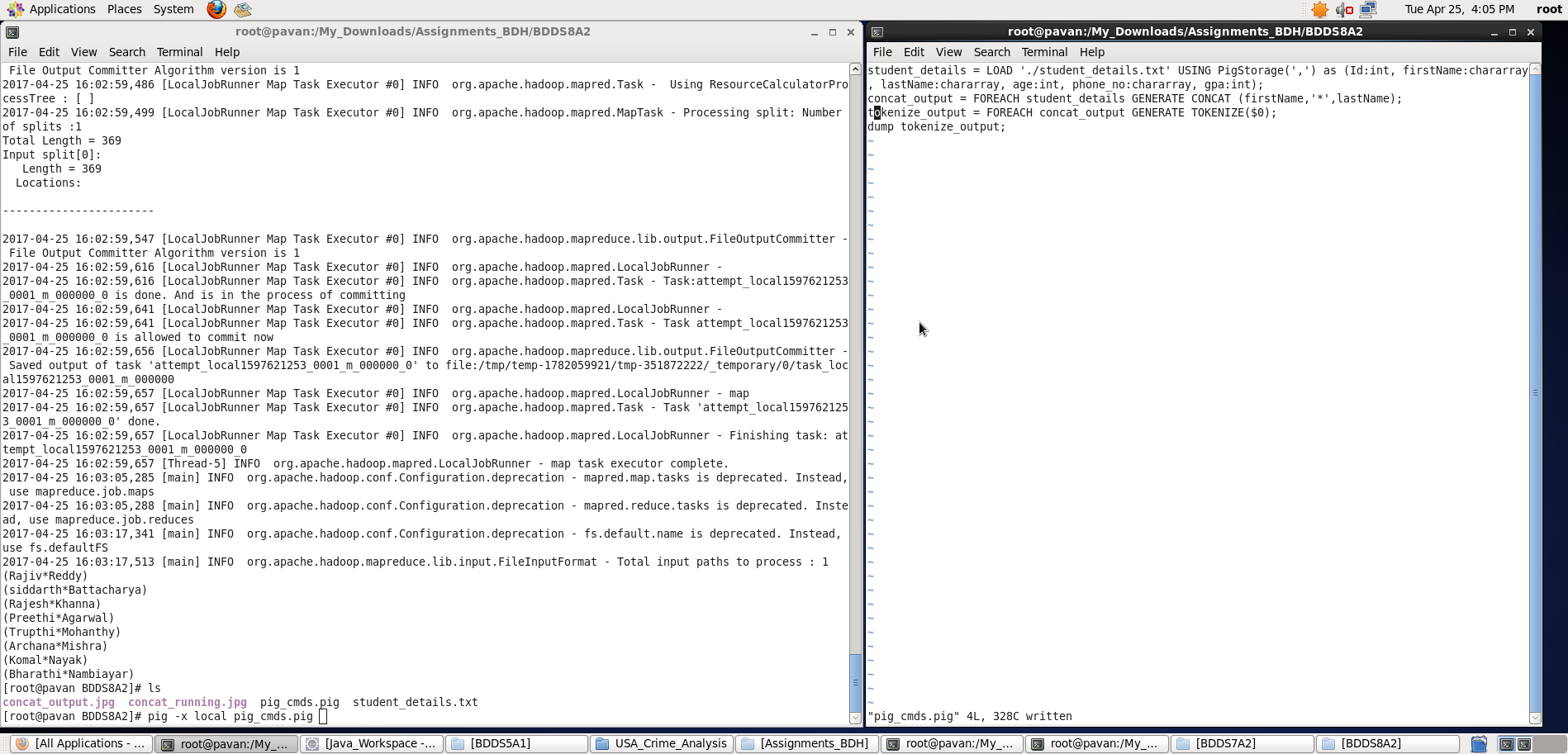
dump tokenize\_output;

**Explanation:**

* Loaded data from student\_details.txt into “student\_details” relation. Here ‘,’ is the delimiter and schema is provided accordingly
* Then, for all records I have generated a new relation which contains concatenated result of the firstName and lastName using CONCAT function. Here ‘\*’ is kept in between the firstName and lastName and all the three expressions are concatenated.
* Now, using TOKENIZE function , split the concat\_output (which contains a group of words) in a single tuple and returns a bag which contains the output of the split operation.
* Outputting the tokenize\_output

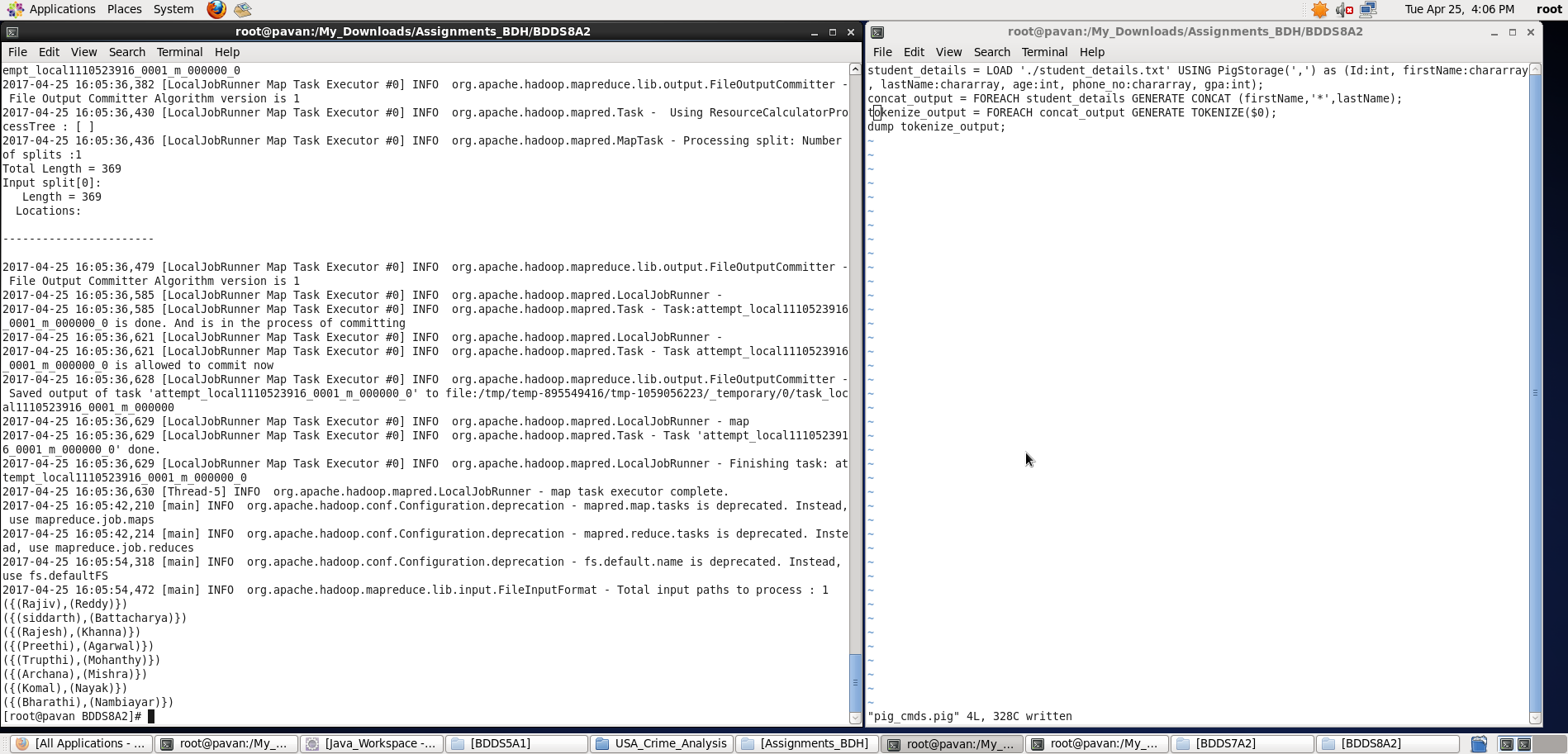
**Screenshots**

Running **Pig Latin Script in Local Mode** inside /My\_Downloads/Assignments\_BDH/BDDS8A2/ directory



Output displayed: **<tokenize\_output>**

**Note : We can store the output if required in a file using “store” command.**



1. Sum

**Pig Latin Script**:

student\_details = LOAD './student\_details.txt' USING PigStorage(',') as (Id:int, firstName:chararray, lastName:chararray, age:int, phone\_no:chararray, gpa:int);

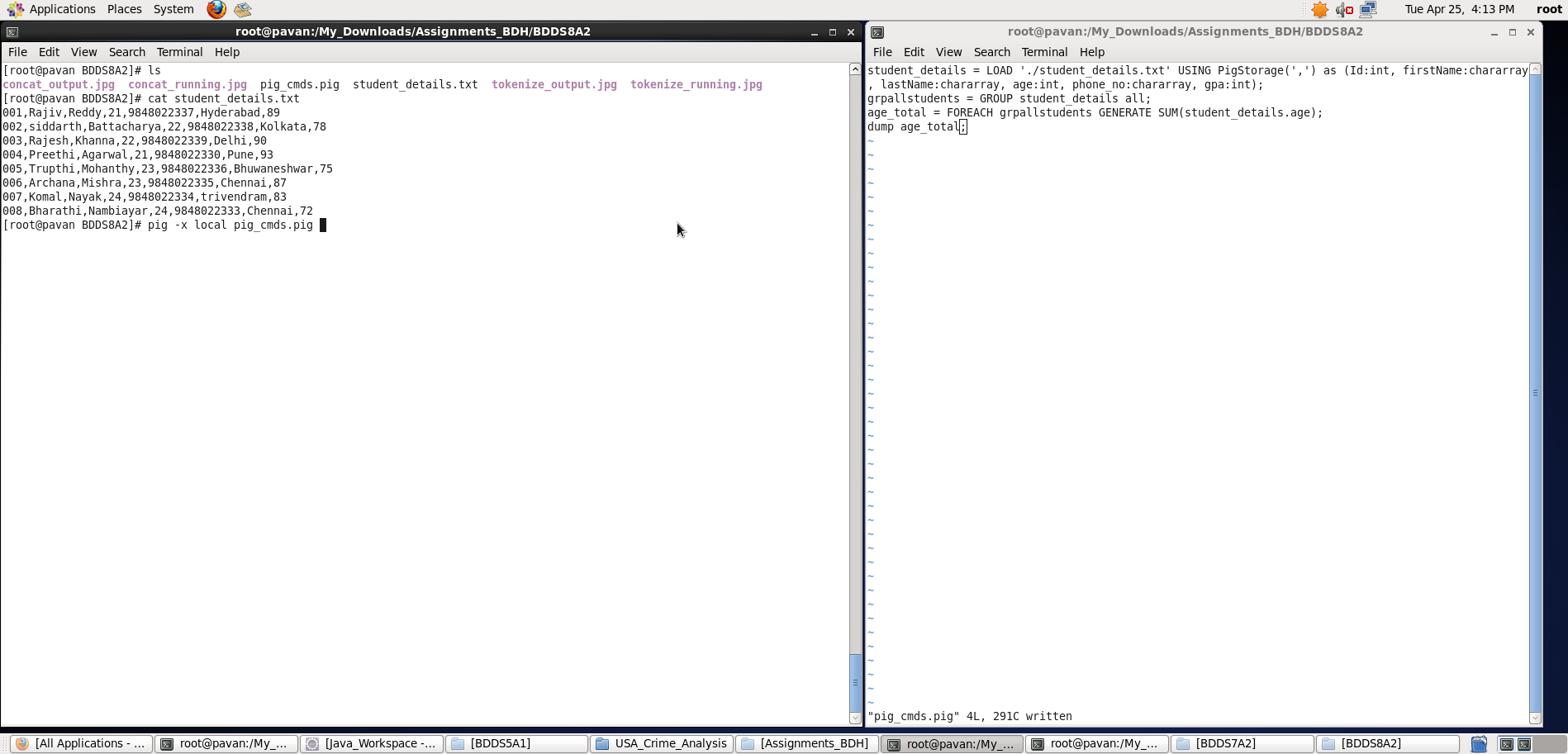
dump concat\_output;

**Explanation:**

* Loaded data from student\_details.txt into “student\_details” relation. Here ‘,’ is the delimiter and schema is provided accordingly
* Then, for all records I have generated a new relation which contains concatenated result of the firstName and lastName using CONCAT function. Here ‘\*’ is kept in between the firstName and lastName and all the three expressions are concatenated.
* Outputting the concat\_output

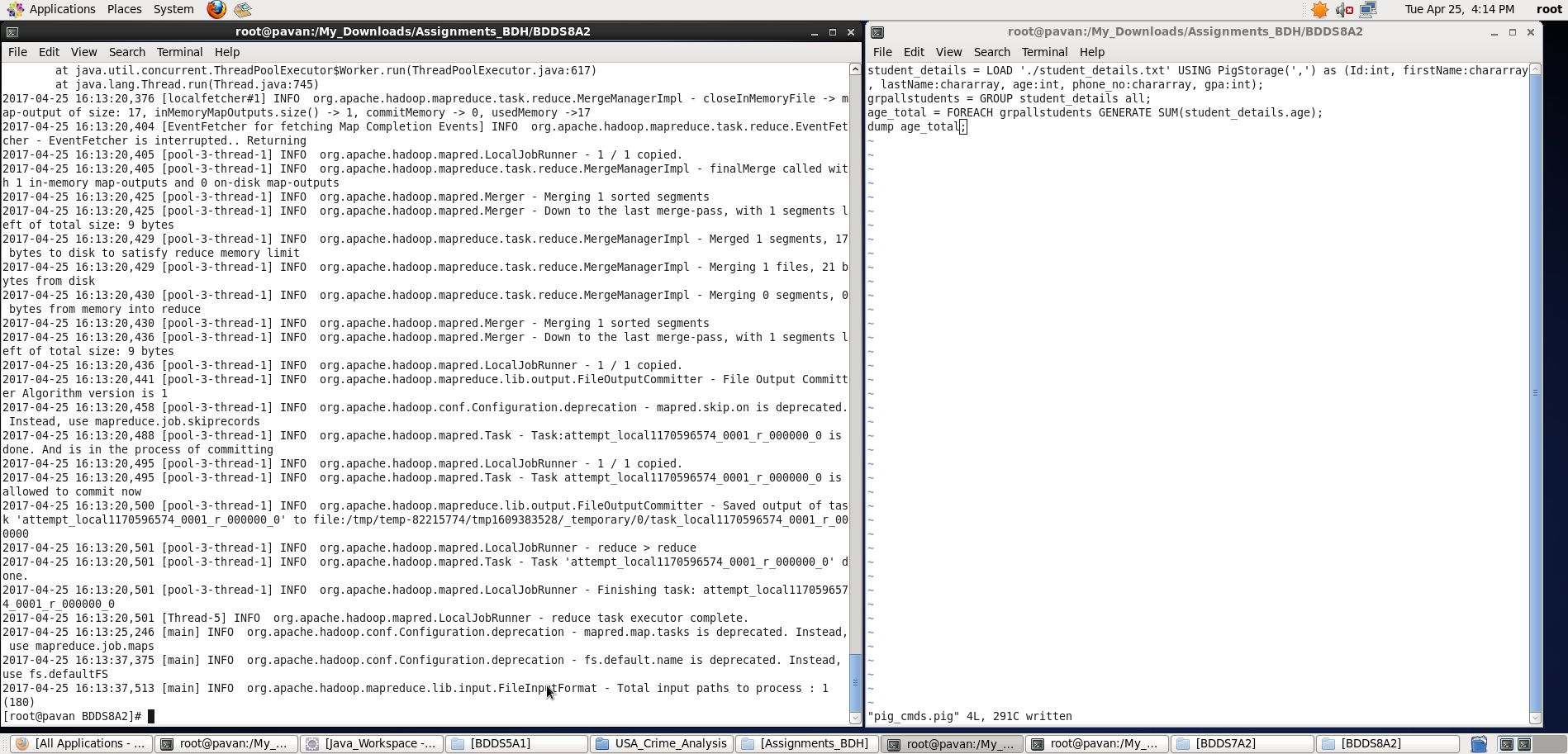
**Screenshots**

Running **Pig Latin Script in Local Mode** inside /My\_Downloads/Assignments\_BDH/BDDS8A2/ directory



Output displayed: **<concat\_output>**

**Note : We can store the output if required in a file using “store” command**



1. Min

**Pig Latin Script**:

student\_details = LOAD './student\_details.txt' USING PigStorage(',') as (Id:int, firstName:chararray, lastName:chararray, age:int, phone\_no:chararray, gpa:int);

concat\_output = FOREACH student\_details GENERATE CONCAT (firstName,'\*',lastName);

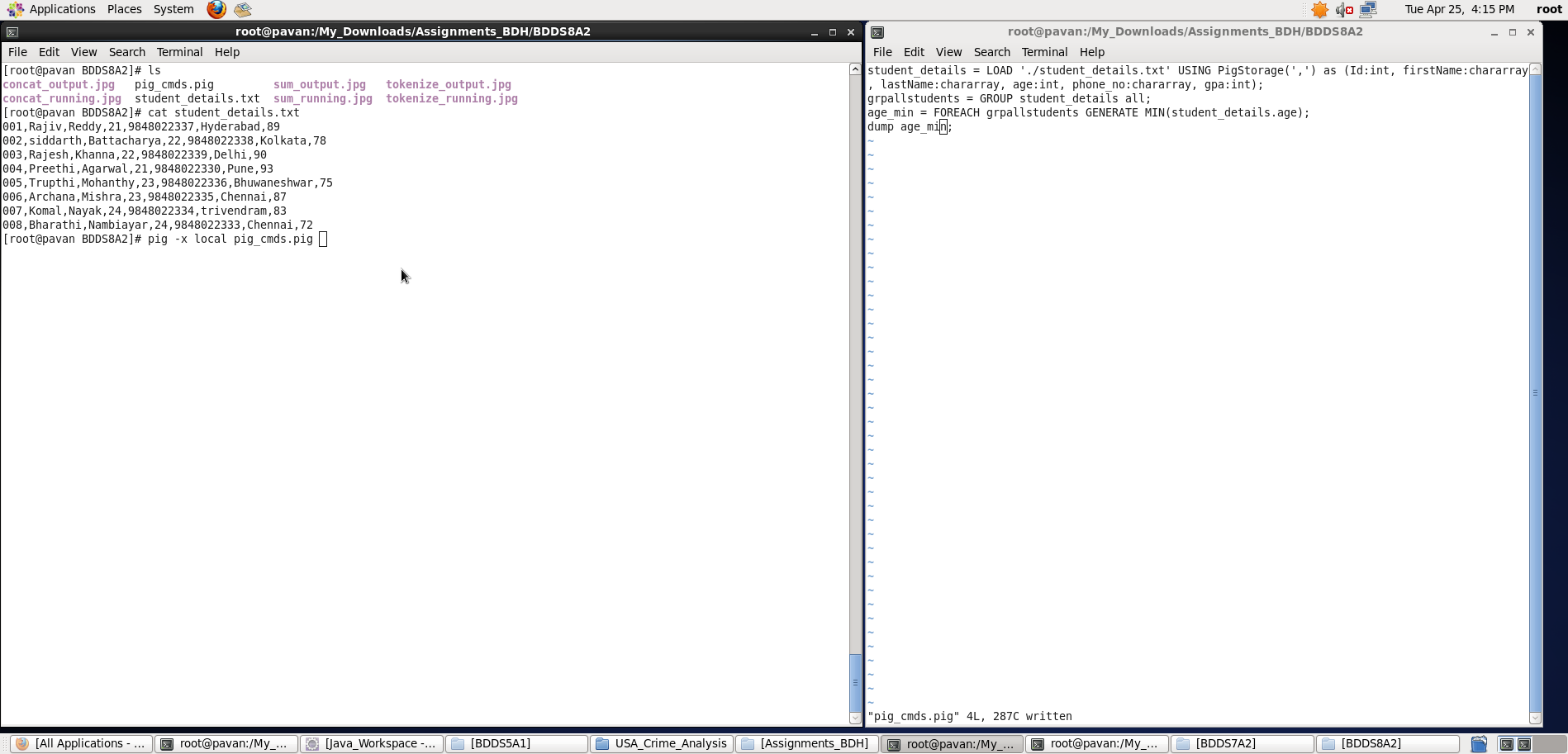
dump concat\_output;

**Explanation:**

* Loaded data from student\_details.txt into “student\_details” relation. Here ‘,’ is the delimiter and schema is provided accordingly
* Then, for all records I have generated a new relation which contains concatenated result of the firstName and lastName using CONCAT function. Here ‘\*’ is kept in between the firstName and lastName and all the three expressions are concatenated.
* Outputting the concat\_output

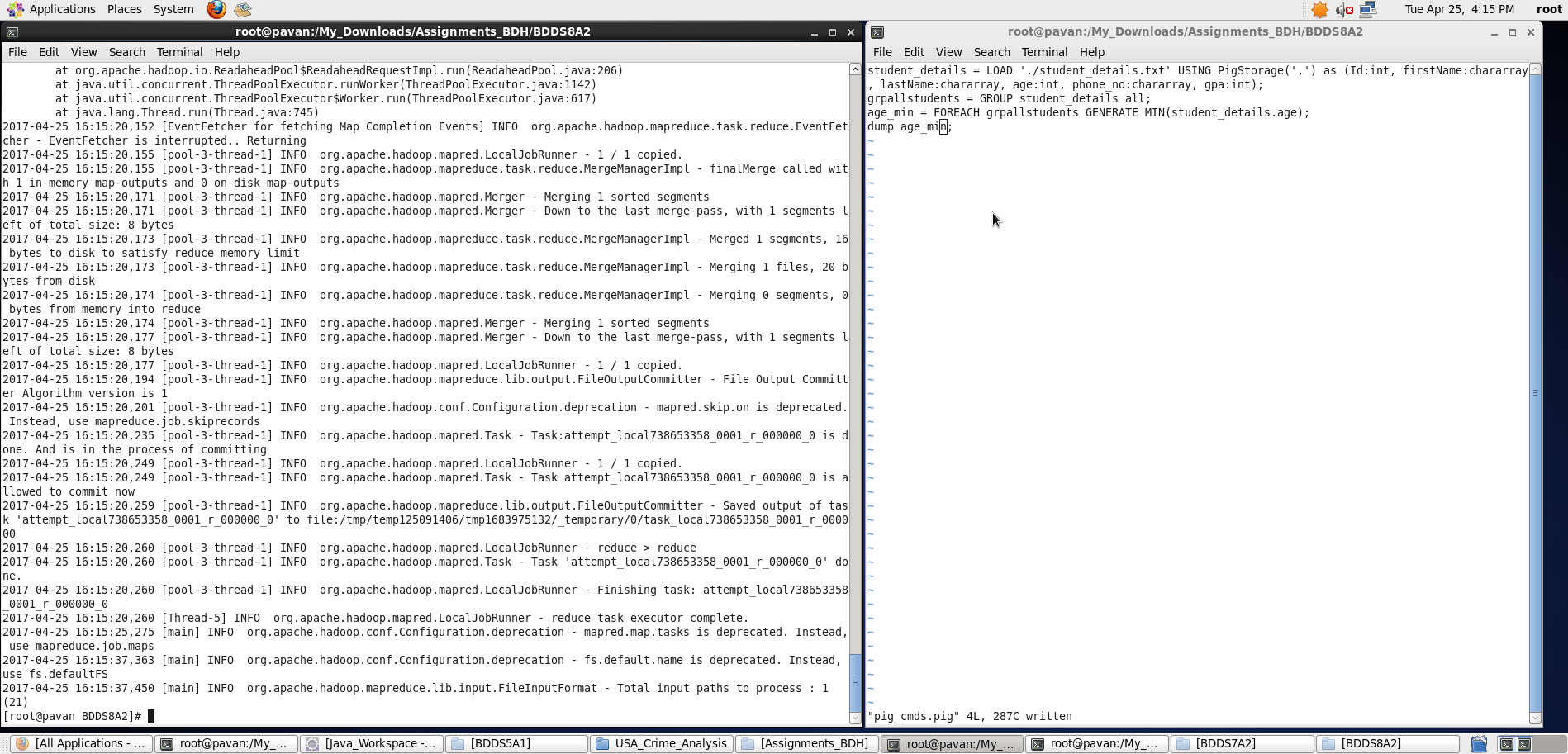
**Screenshots**

Running **Pig Latin Script in Local Mode** inside /My\_Downloads/Assignments\_BDH/BDDS8A2/ directory



Output displayed: **<concat\_output>**

**Note : We can store the output if required in a file using “store” command**



1. Max

**Pig Latin Script**:

student\_details = LOAD './student\_details.txt' USING PigStorage(',') as (Id:int, firstName:chararray, lastName:chararray, age:int, phone\_no:chararray, gpa:int);

concat\_output = FOREACH student\_details GENERATE CONCAT (firstName,'\*',lastName);

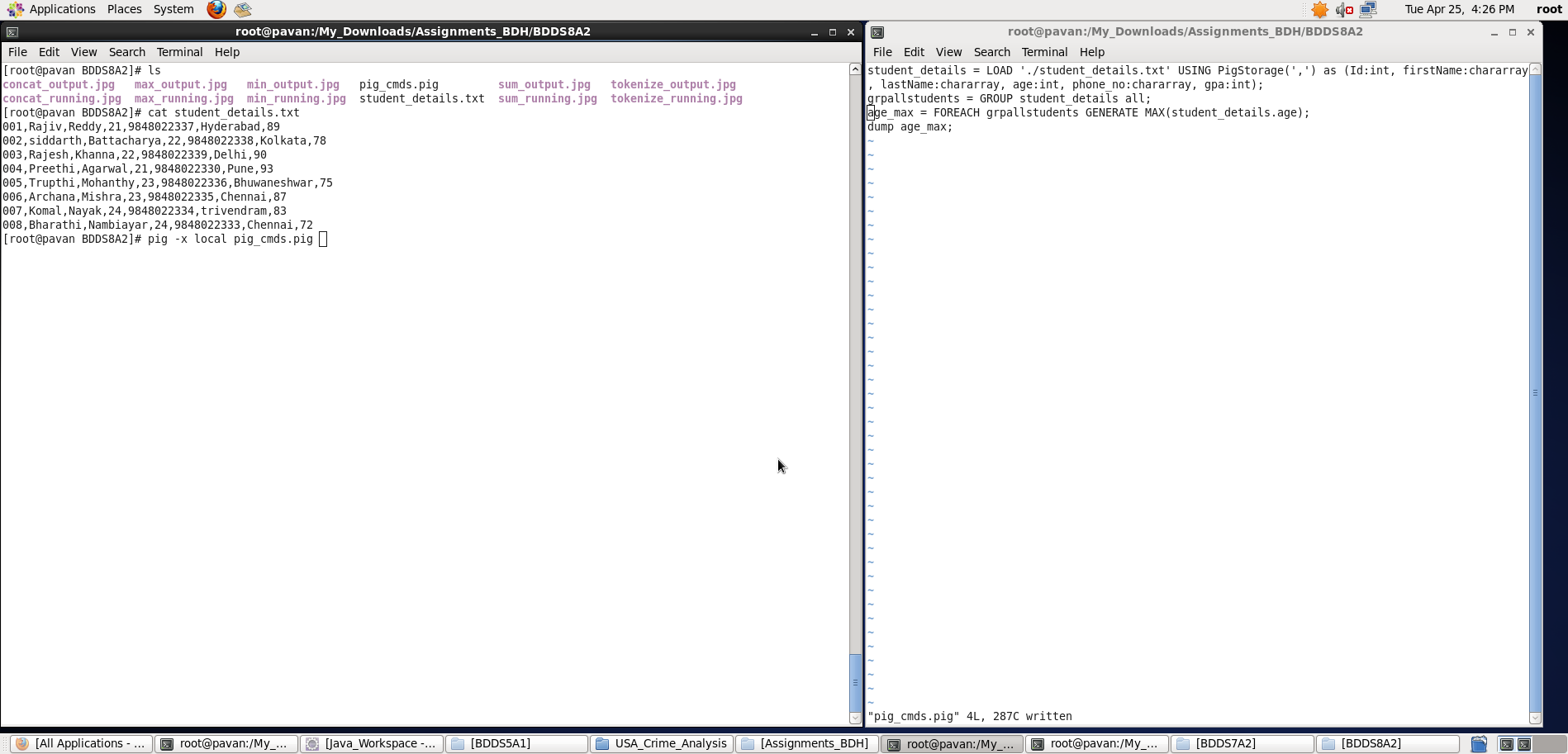
dump concat\_output;

**Explanation:**

* Loaded data from student\_details.txt into “student\_details” relation. Here ‘,’ is the delimiter and schema is provided accordingly
* Then, for all records I have generated a new relation which contains concatenated result of the firstName and lastName using CONCAT function. Here ‘\*’ is kept in between the firstName and lastName and all the three expressions are concatenated.
* Outputting the concat\_output

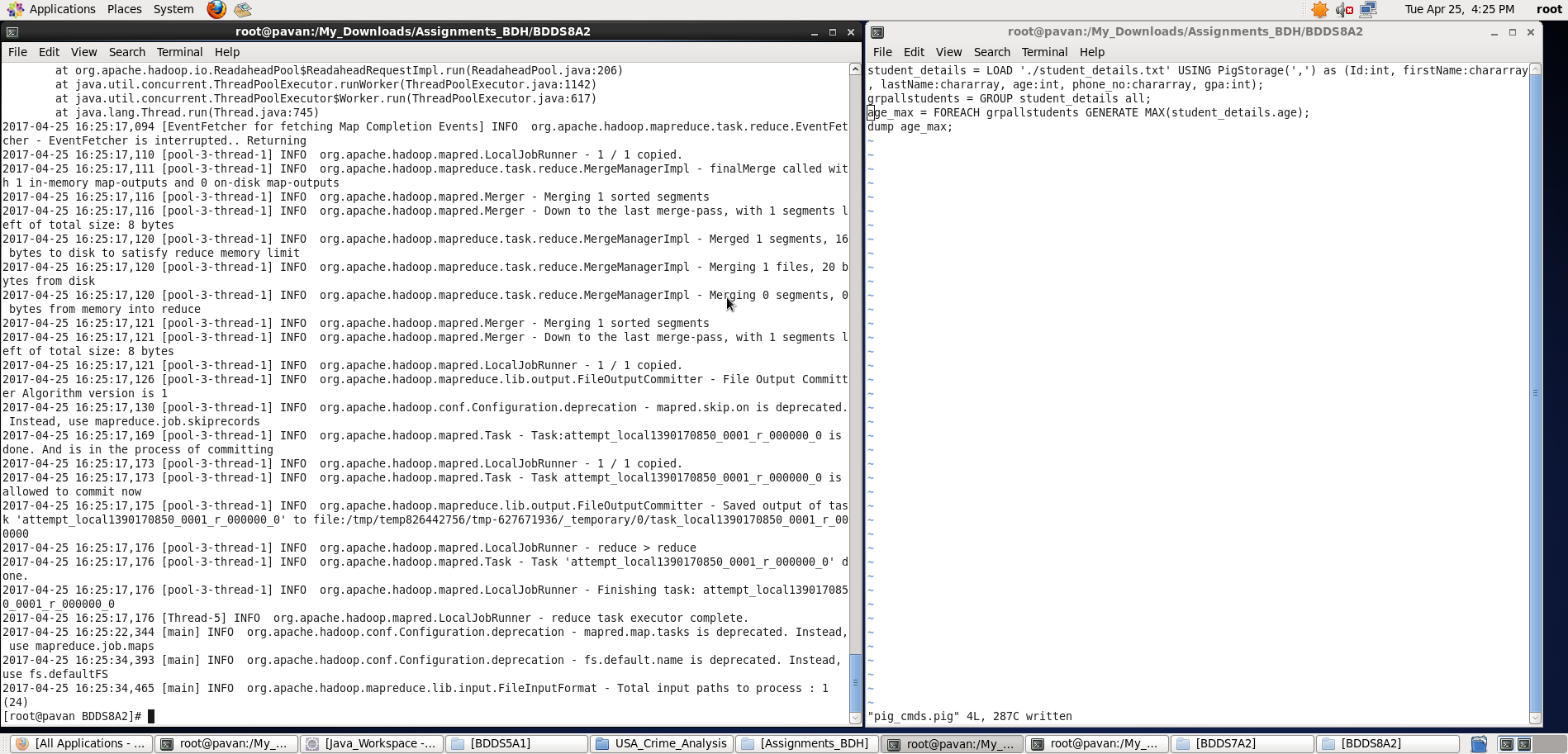
**Screenshots**

Running **Pig Latin Script in Local Mode** inside /My\_Downloads/Assignments\_BDH/BDDS8A2/ directory



Output displayed: **<concat\_output>**

**Note : We can store the output if required in a file using “store” command**



1. Limit

**Pig Latin Script**:

student\_details = LOAD './student\_details.txt' USING PigStorage(',') as (Id:int, firstName:chararray, lastName:chararray, age:int, phone\_no:chararray, gpa:int);

concat\_output = FOREACH student\_details GENERATE CONCAT (firstName,'\*',lastName);

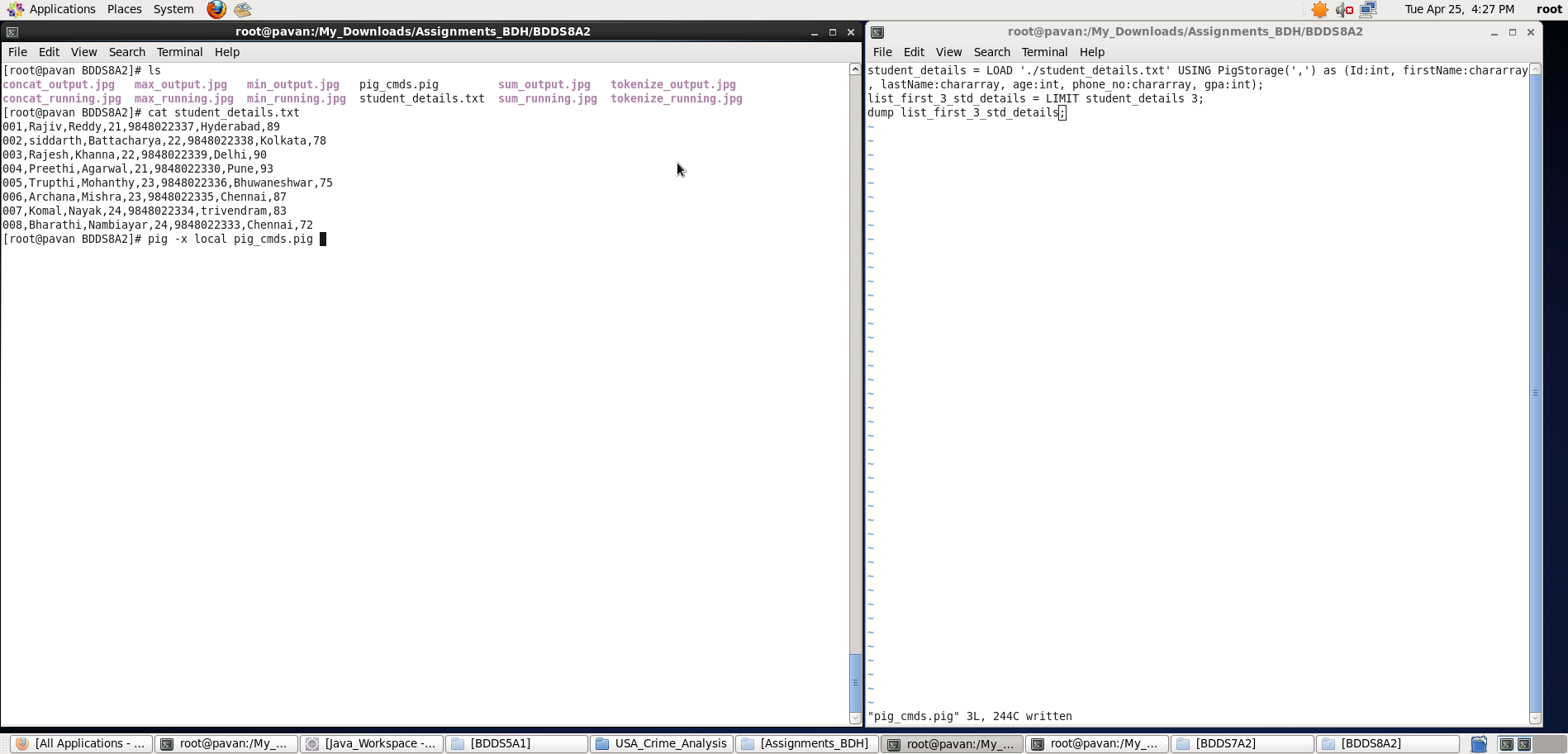
dump concat\_output;

**Explanation:**

* Loaded data from student\_details.txt into “student\_details” relation. Here ‘,’ is the delimiter and schema is provided accordingly
* Then, for all records I have generated a new relation which contains concatenated result of the firstName and lastName using CONCAT function. Here ‘\*’ is kept in between the firstName and lastName and all the three expressions are concatenated.
* Outputting the concat\_output

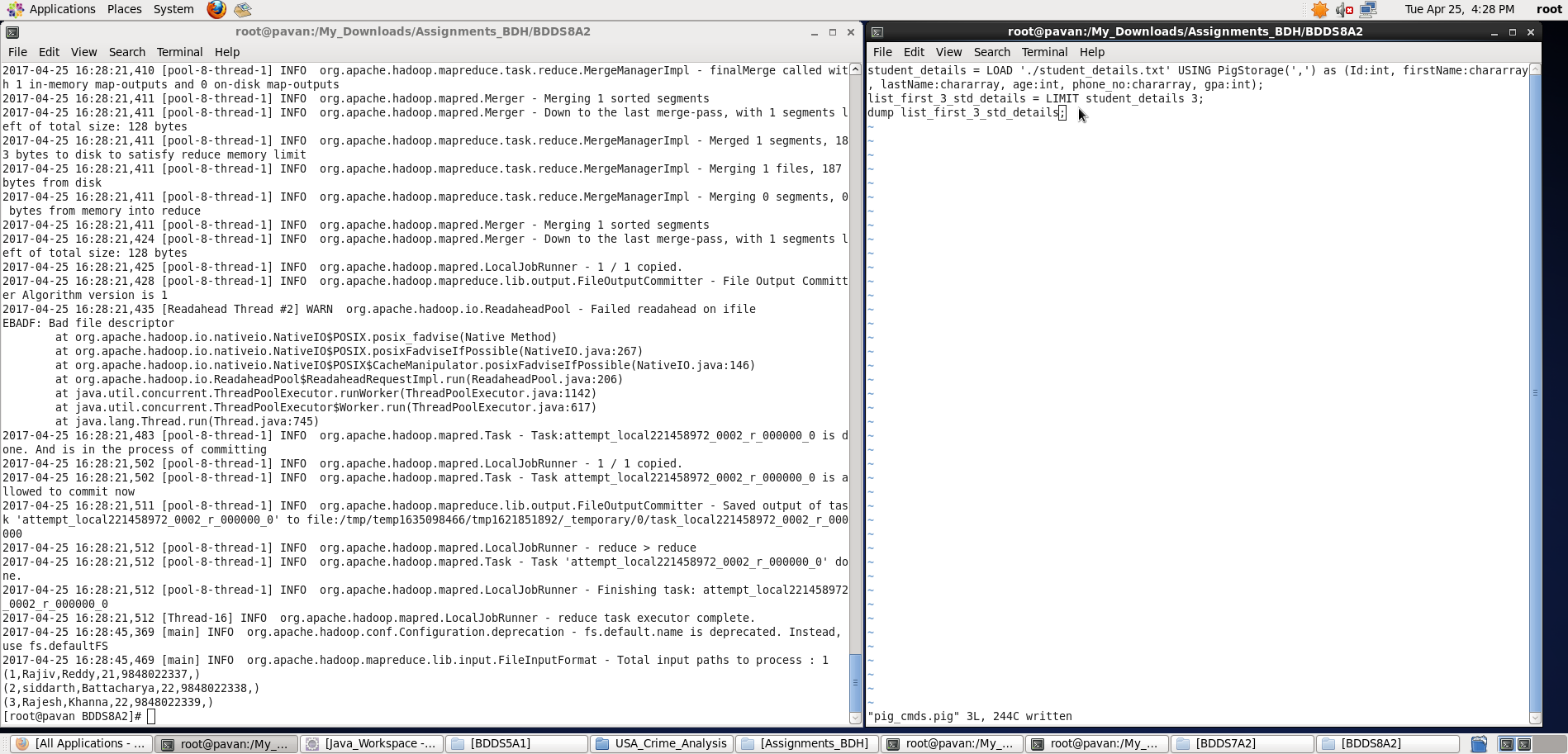
**Screenshots**

Running **Pig Latin Script in Local Mode** inside /My\_Downloads/Assignments\_BDH/BDDS8A2/ directory



Output displayed: **<concat\_output>**

**Note : We can store the output if required in a file using “store” command**



1. Store

**Pig Latin Script**:

student\_details = LOAD './student\_details.txt' USING PigStorage(',') as (Id:int, firstName:chararray, lastName:chararray, age:int, phone\_no:chararray, gpa:int);

concat\_output = FOREACH student\_details GENERATE CONCAT (firstName,'\*',lastName);

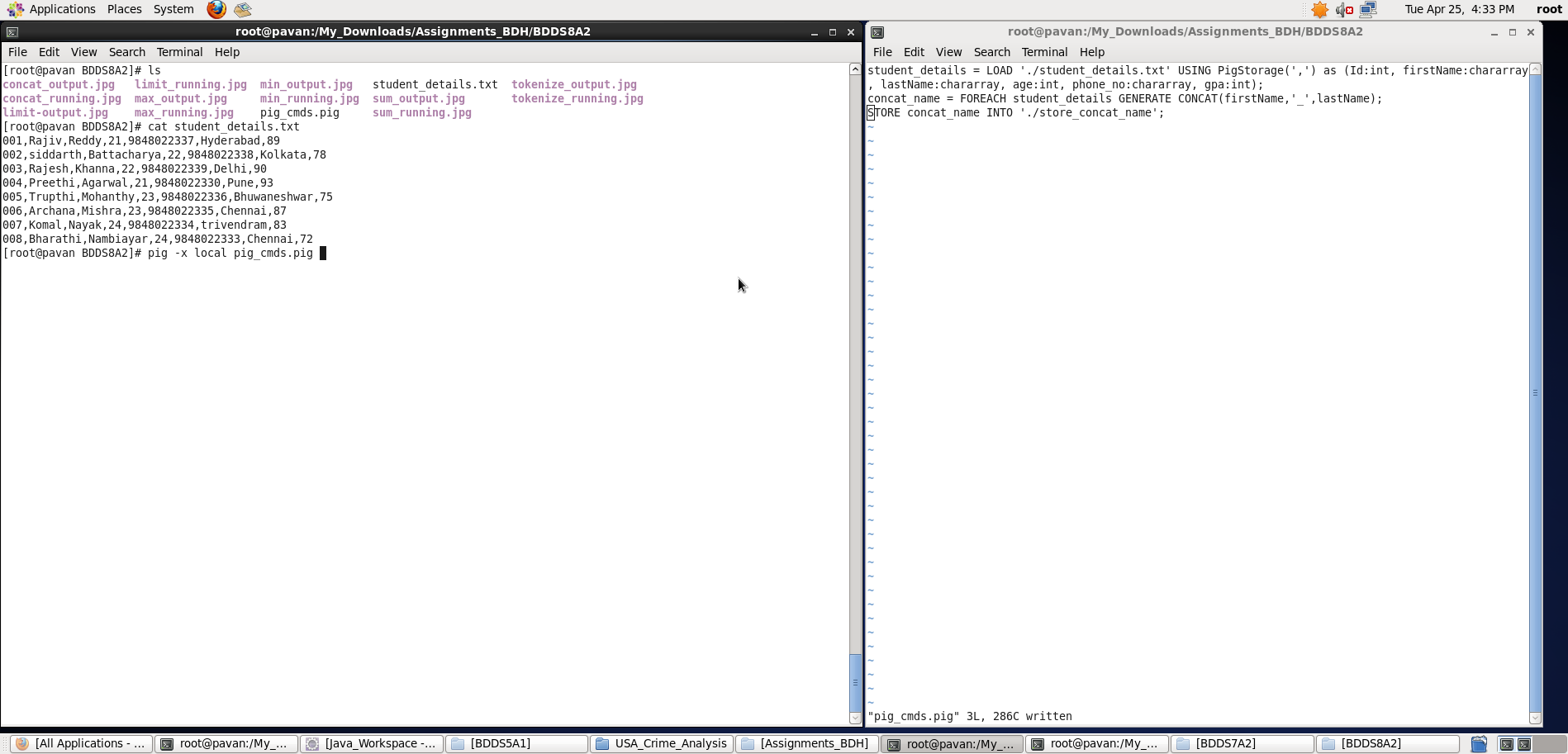
dump concat\_output;

**Explanation:**

* Loaded data from student\_details.txt into “student\_details” relation. Here ‘,’ is the delimiter and schema is provided accordingly
* Then, for all records I have generated a new relation which contains concatenated result of the firstName and lastName using CONCAT function. Here ‘\*’ is kept in between the firstName and lastName and all the three expressions are concatenated.
* Outputting the concat\_output

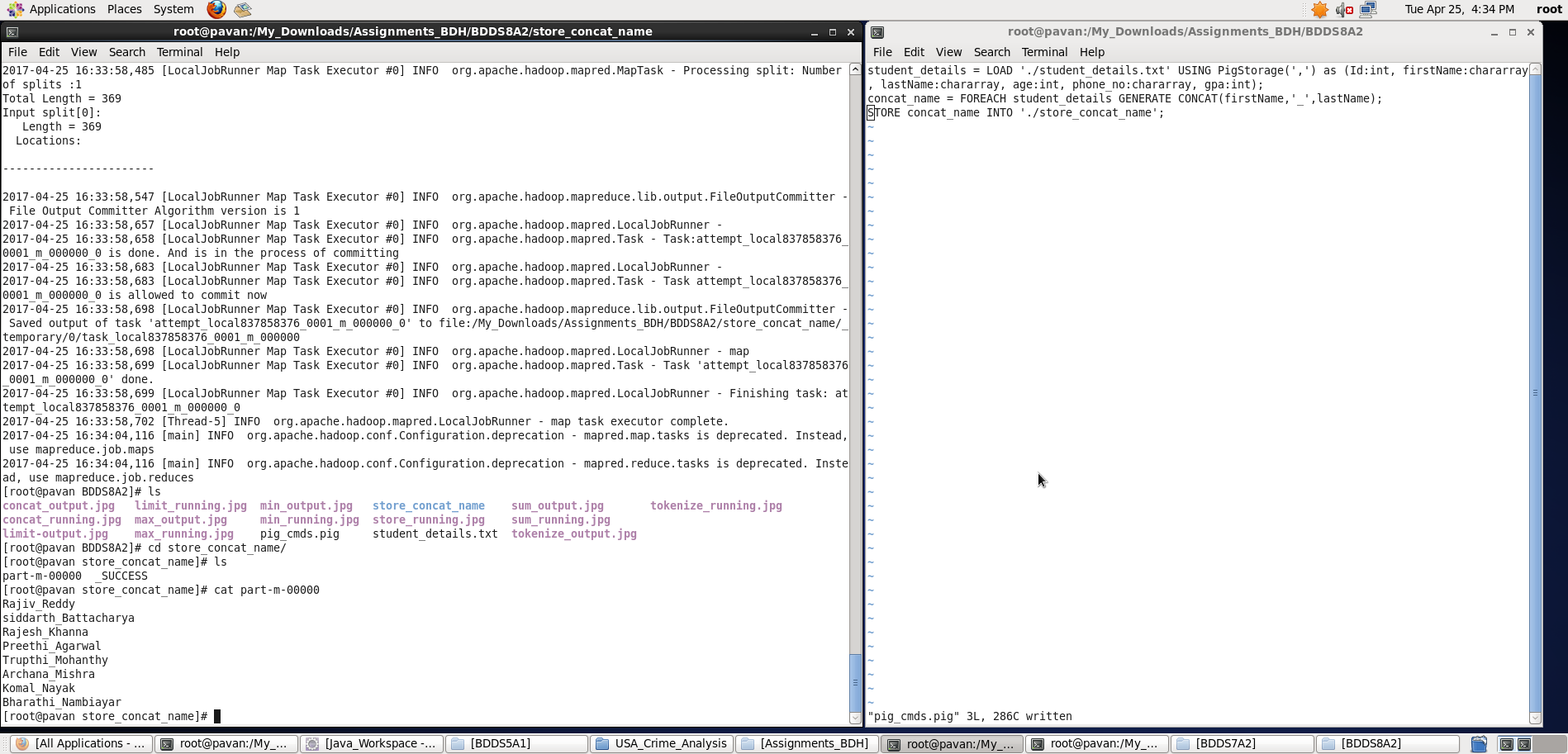
**Screenshots**

Running **Pig Latin Script in Local Mode** inside /My\_Downloads/Assignments\_BDH/BDDS8A2/ directory



Output displayed: **<concat\_output>**

**Note : We can store the output if required in a file using “store” command**

****

1. Distinct

**Pig Latin Script**:

student\_details = LOAD './student\_details.txt' USING PigStorage(',') as (Id:int, firstName:chararray, lastName:chararray, age:int, phone\_no:chararray, gpa:int);

concat\_output = FOREACH student\_details GENERATE CONCAT (firstName,'\*',lastName);

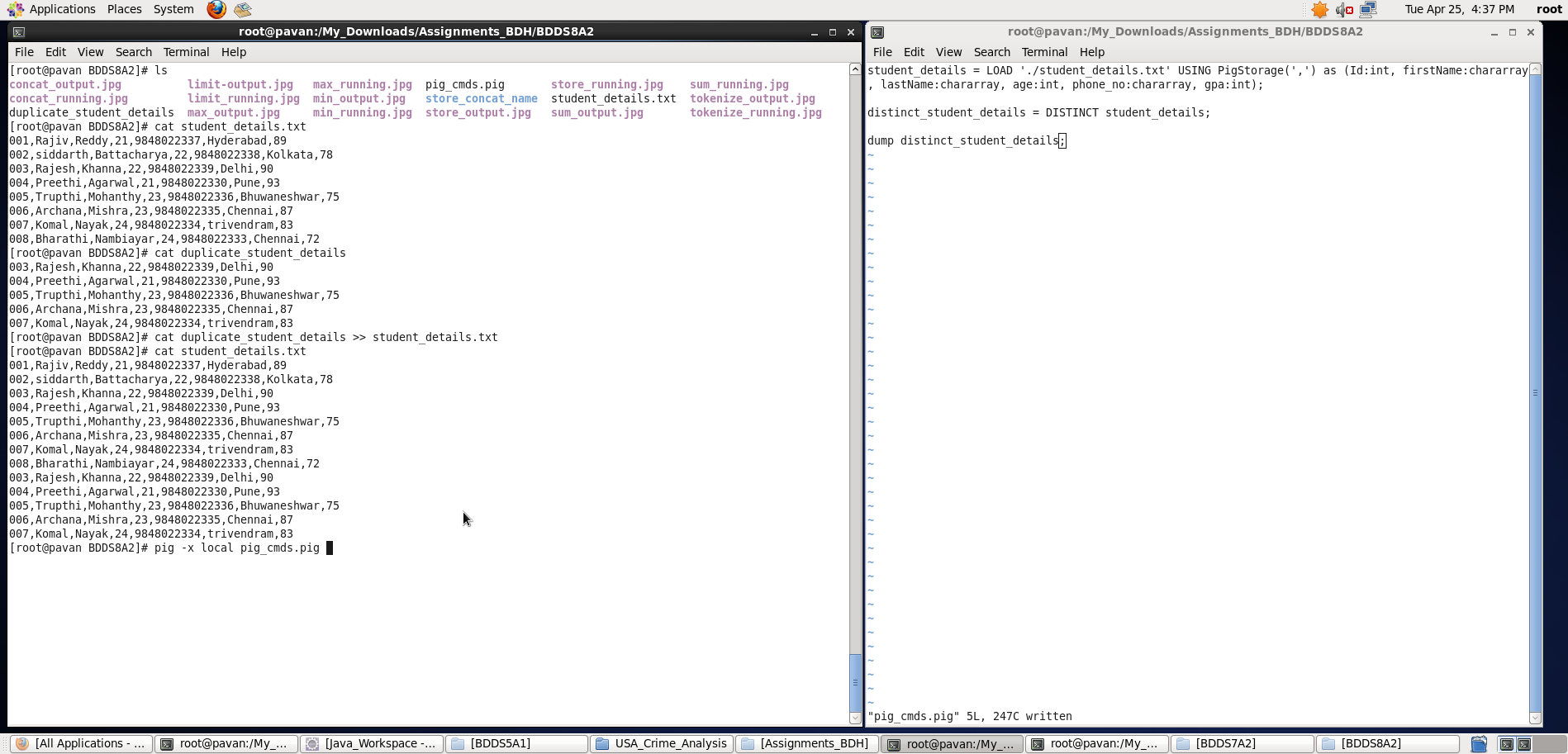
dump concat\_output;

**Explanation:**

* Loaded data from student\_details.txt into “student\_details” relation. Here ‘,’ is the delimiter and schema is provided accordingly
* Then, for all records I have generated a new relation which contains concatenated result of the firstName and lastName using CONCAT function. Here ‘\*’ is kept in between the firstName and lastName and all the three expressions are concatenated.
* Outputting the concat\_output

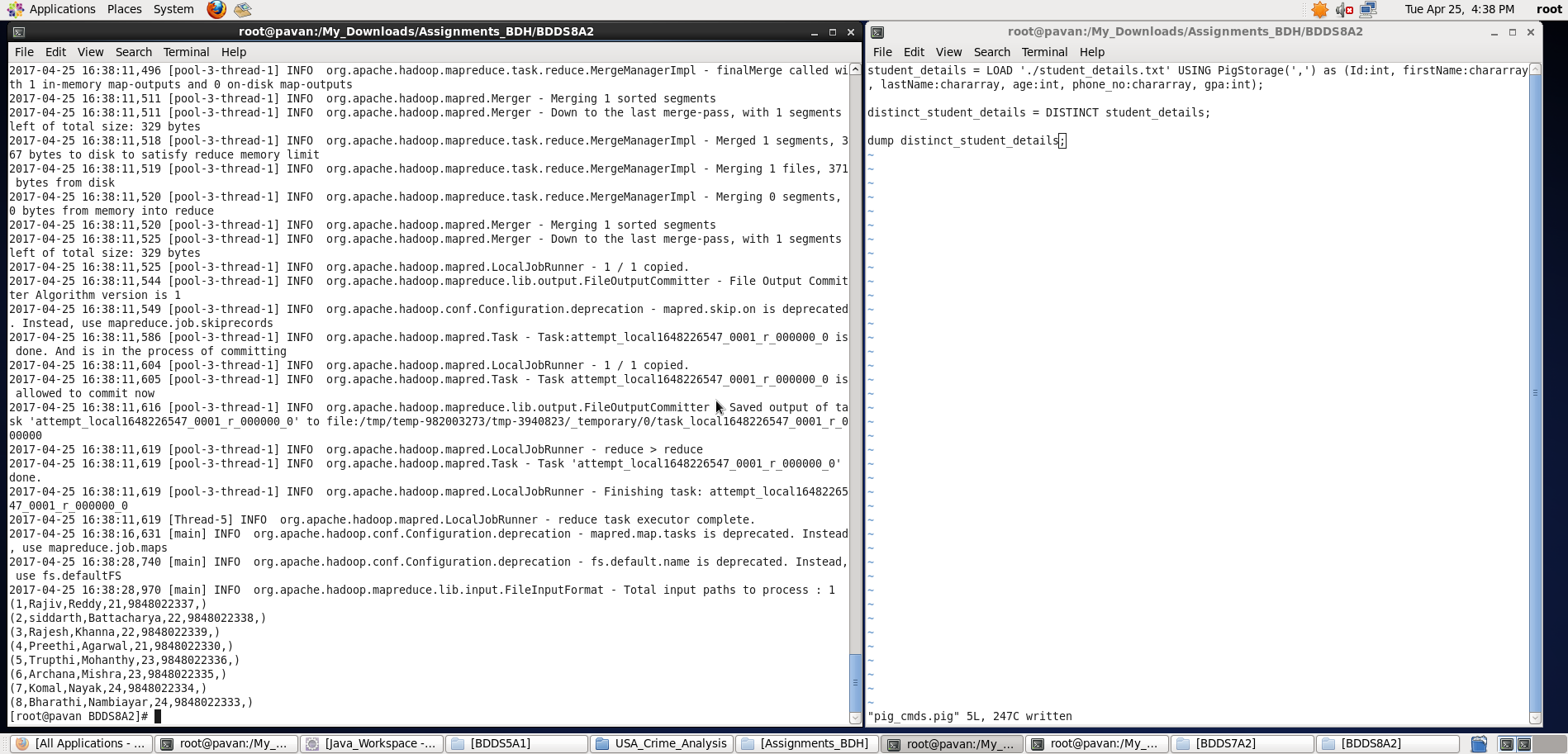
**Screenshots**

Running **Pig Latin Script in Local Mode** inside /My\_Downloads/Assignments\_BDH/BDDS8A2/ directory



Output displayed: **<concat\_output>**

**Note : We can store the output if required in a file using “store” command**



1. Flatten

**Pig Latin Script**:

student\_details = LOAD './student\_details.txt' USING PigStorage(',') as (Id:int, firstName:chararray, lastName:chararray, age:int, phone\_no:chararray, gpa:int);

concat\_output = FOREACH student\_details GENERATE CONCAT (firstName,'\*',lastName);

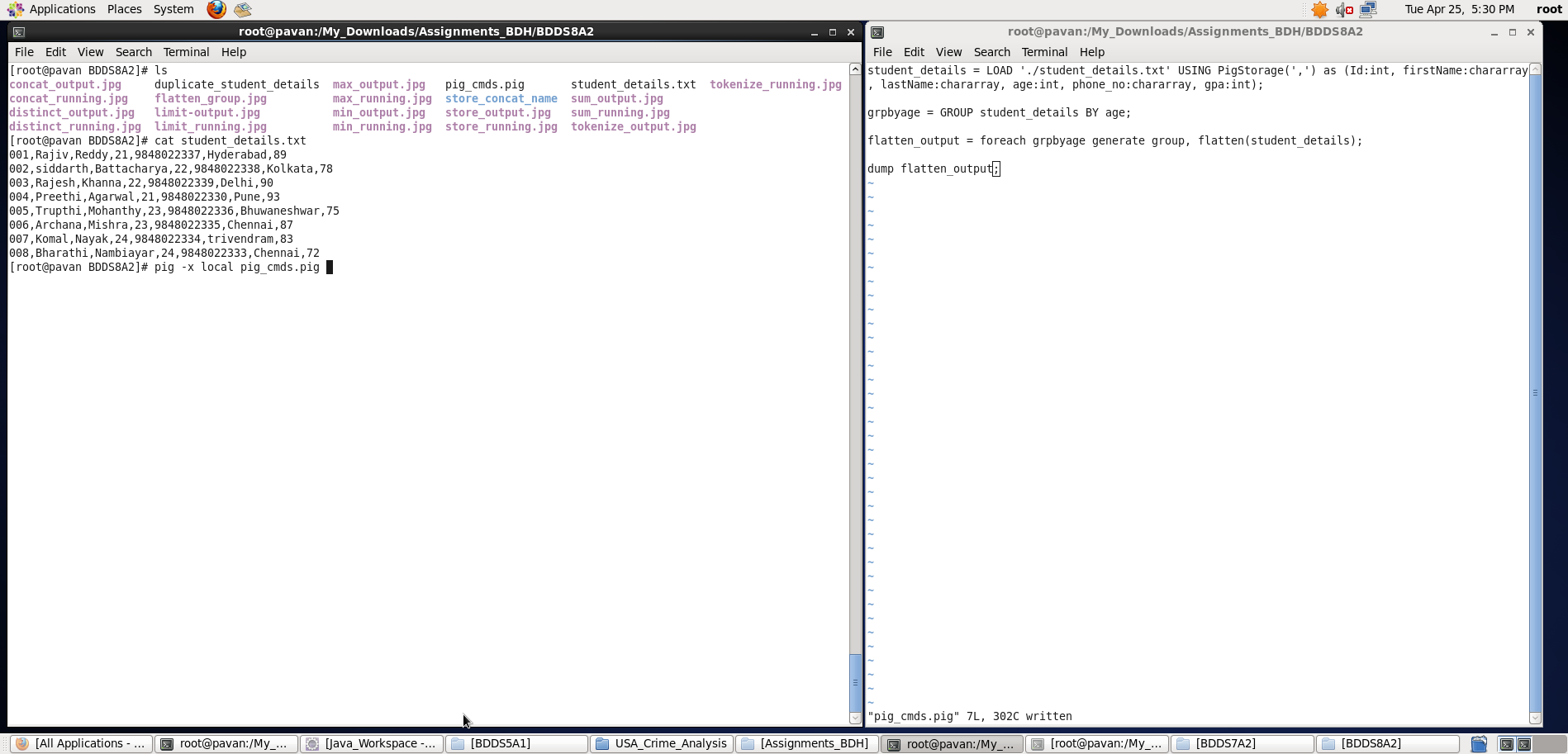
dump concat\_output;

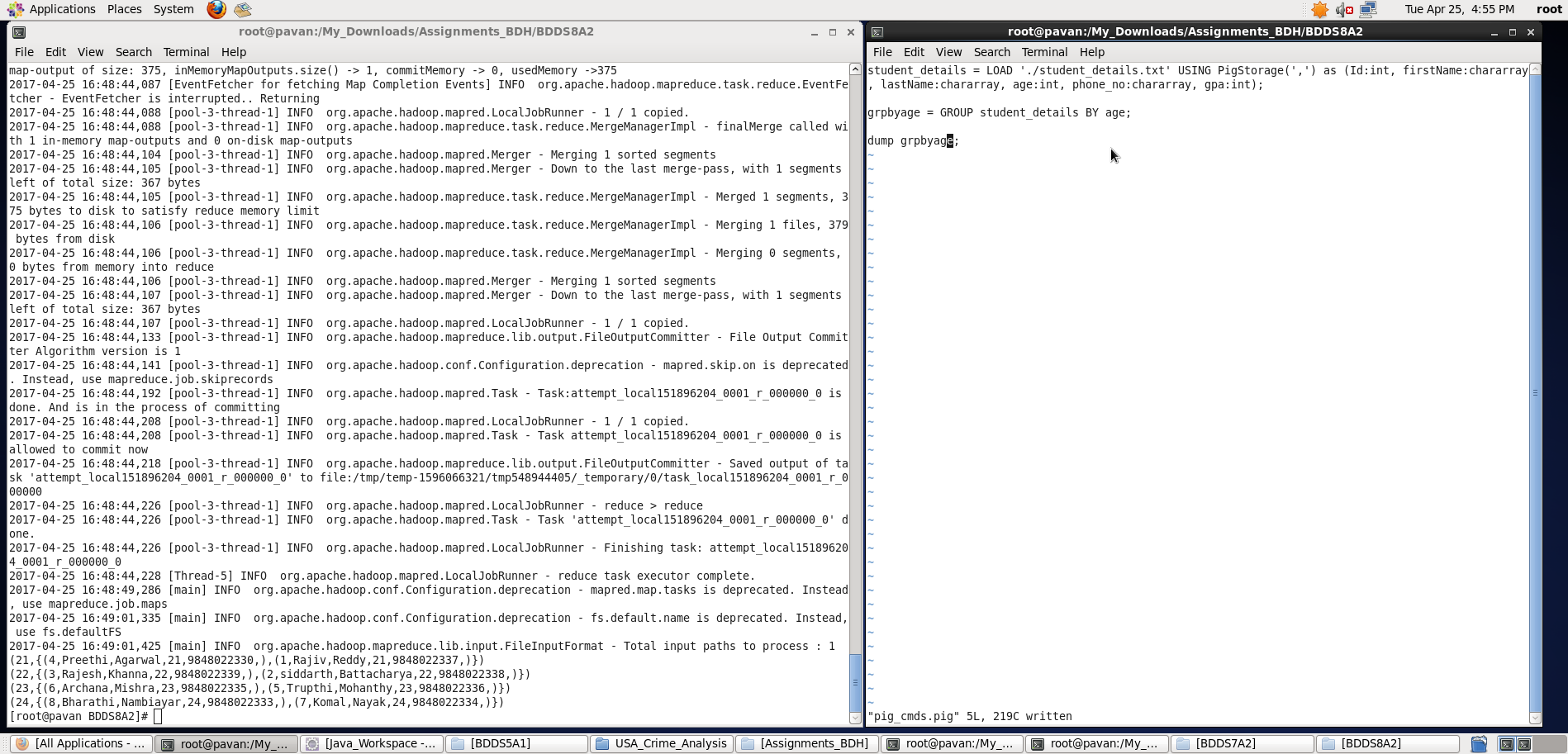
**Explanation:**

* Loaded data from student\_details.txt into “student\_details” relation. Here ‘,’ is the delimiter and schema is provided accordingly
* Then, for all records I have generated a new relation which contains concatenated result of the firstName and lastName using CONCAT function. Here ‘\*’ is kept in between the firstName and lastName and all the three expressions are concatenated.
* Outputting the concat\_output

**Screenshots**

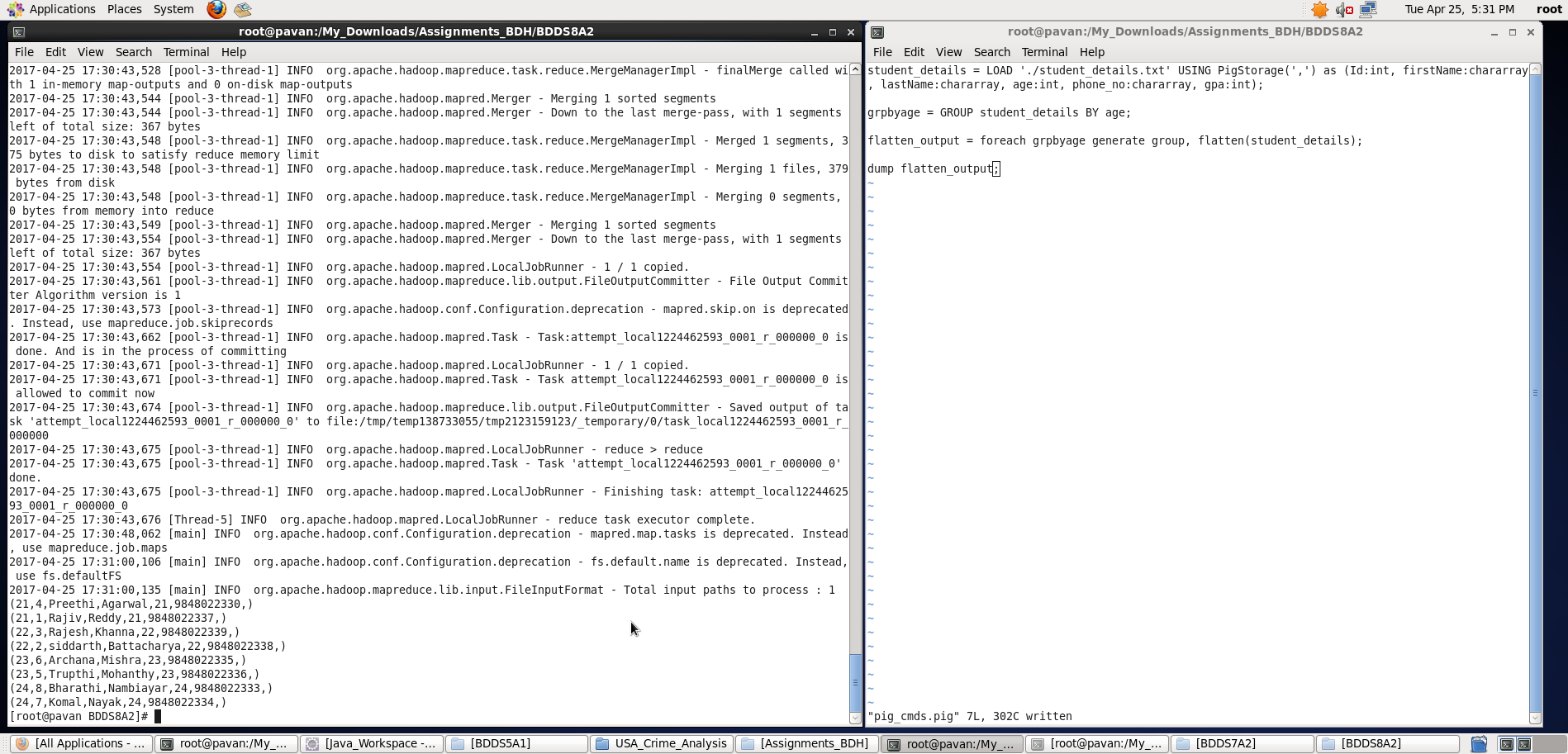
Running **Pig Latin Script in Local Mode** inside /My\_Downloads/Assignments\_BDH/BDDS8A2/ directory





Output displayed: **<concat\_output>**

**Note : We can store the output if required in a file using “store” command**

****

1. IsEmpty

**Pig Latin Script**:

student\_details = LOAD './student\_details.txt' USING PigStorage(',') as (Id:int, firstName:chararray, lastName:chararray, age:int, phone\_no:chararray, gpa:int);

concat\_output = FOREACH student\_details GENERATE CONCAT (firstName,'\*',lastName);

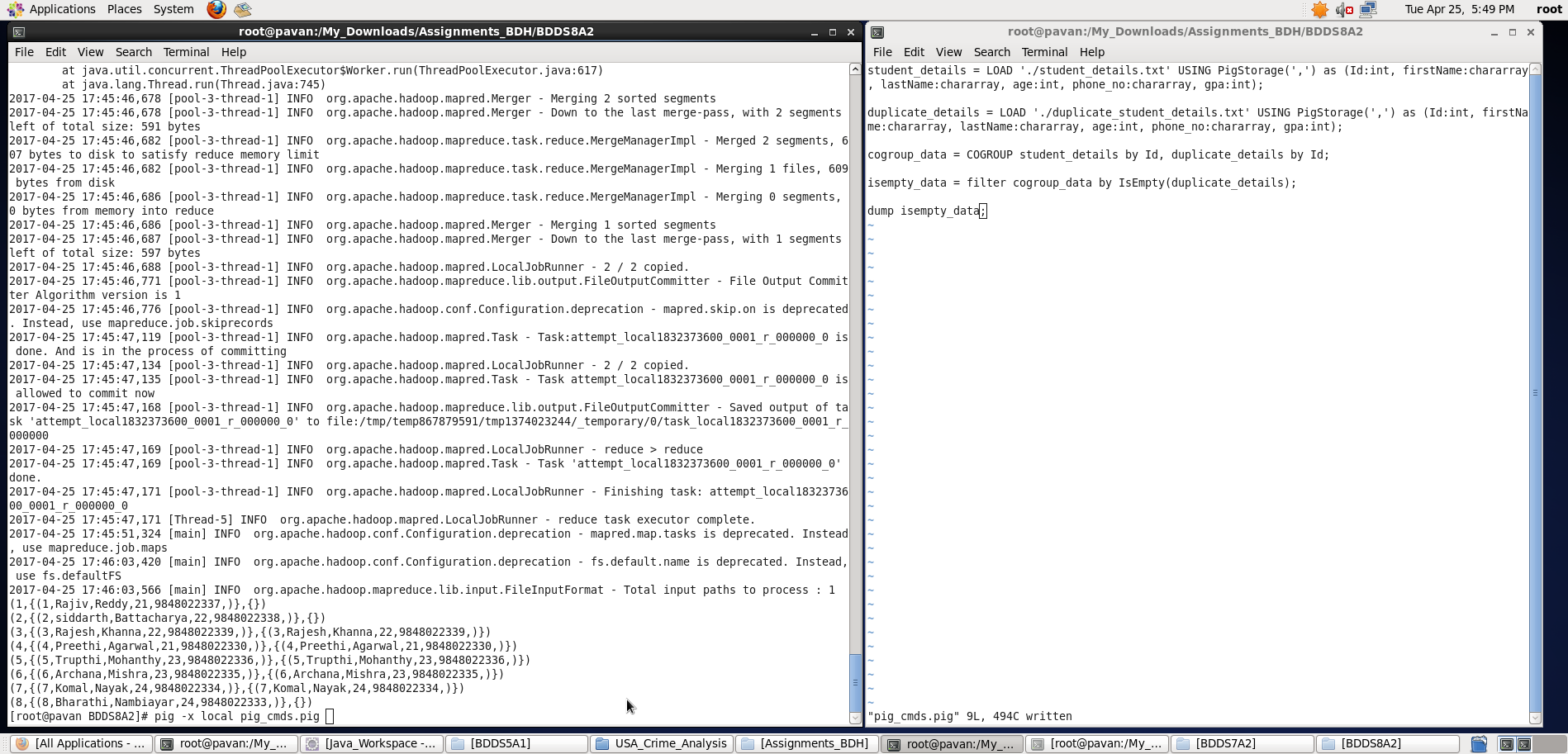
dump concat\_output;

**Explanation:**

* Loaded data from student\_details.txt into “student\_details” relation. Here ‘,’ is the delimiter and schema is provided accordingly
* Then, for all records I have generated a new relation which contains concatenated result of the firstName and lastName using CONCAT function. Here ‘\*’ is kept in between the firstName and lastName and all the three expressions are concatenated.
* Outputting the concat\_output

**Screenshots**

Running **Pig Latin Script in Local Mode** inside /My\_Downloads/Assignments\_BDH/BDDS8A2/ directory



Output displayed: **<concat\_output>**

**Note : We can store the output if required in a file using “store” command**

