** Proof of Concept**

**on**

**ADVERSE\_FOOD\_EVENTS**

*By – Suhani Bhargava*

*(*[*suhanibhargava009@gmail.com*](mailto:suhanibhargava009@gmail.com)*)*

*Bachelor of Technology in Computer Science and Engineering,*

*Government Women Engineering College Ajmer*

**POC #: Adverse\_Food\_Events**

About 90000 product-related user-reported adverse medical events taken from FOOD AND DRUG ADMINISTRATION(FDA).

Public DATASET available at below website

https://www.kaggle.com/fda/adverse-food-events

**Industry: Medical**

Data: Publicly available dataset with attributes like

*RA\_Report # - Unique report ID.*

*RA\_CAERS Created Date - Date the report was created.*

*AEC\_Event Start Date - Date the event started.*

*PRI\_Product Role - Role the product played in the event.*

*PRI\_Reported Brand/Product Name - Full product name (in all-caps).*

*PRI\_FDA Industry Code - Industry product code.*

*PRI\_FDA Industry Name - Industry name.*

*CI\_Age at Adverse Event - If known, the victim's age at the time of incident.*

*CI\_Age Unit - The unit for the previous value (usually years).\*

*CI\_Gender - Victim's gender.*

*AEC\_One Row Outcomes - Outcomes from the event. Hand-coded.*

*SYM\_One Row Coded Symptoms -Outcomes from the event, system-coded*.

**Problem Statement:**

***Problem Statement is to -***

***1) Analyse the number of males and females affected due to adverse food events and thus compare them.***

***2) Find out the type of industry due to products of which people were affected the most the most and also count the times they were affected by the same.***

***3) Find out the product that affected maximum of ladies of age 20-25.***

***4) Name the product which was suspected most of the times to affect people.***

***5) How many events were reported on 27 October 2003?***

***6) How many food events were recorded that affected the children of age less than 5 years?***

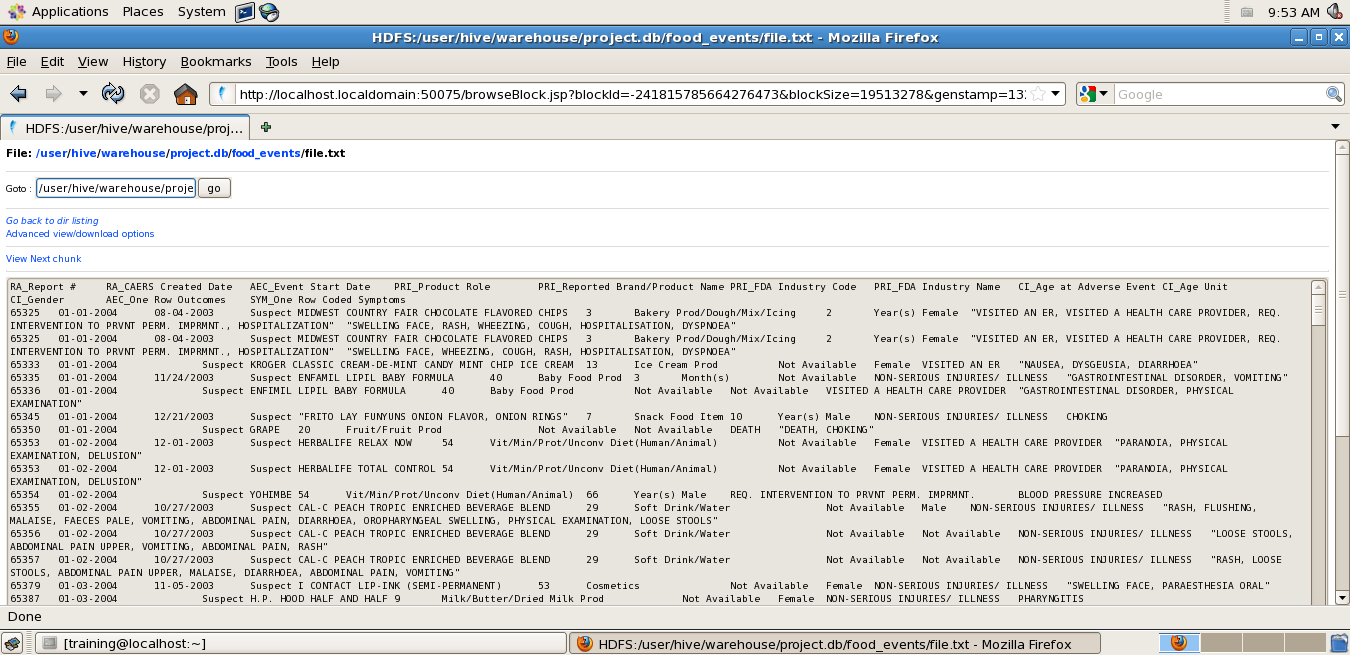
**Shell Script**

**GET THE FILE FROM THE LOCAL FILE SYSTEM TO HDFS**

Hadoop fs -put /home/training/Desktop/file.txt /user/training/project.txt

Hadoop fs –ls

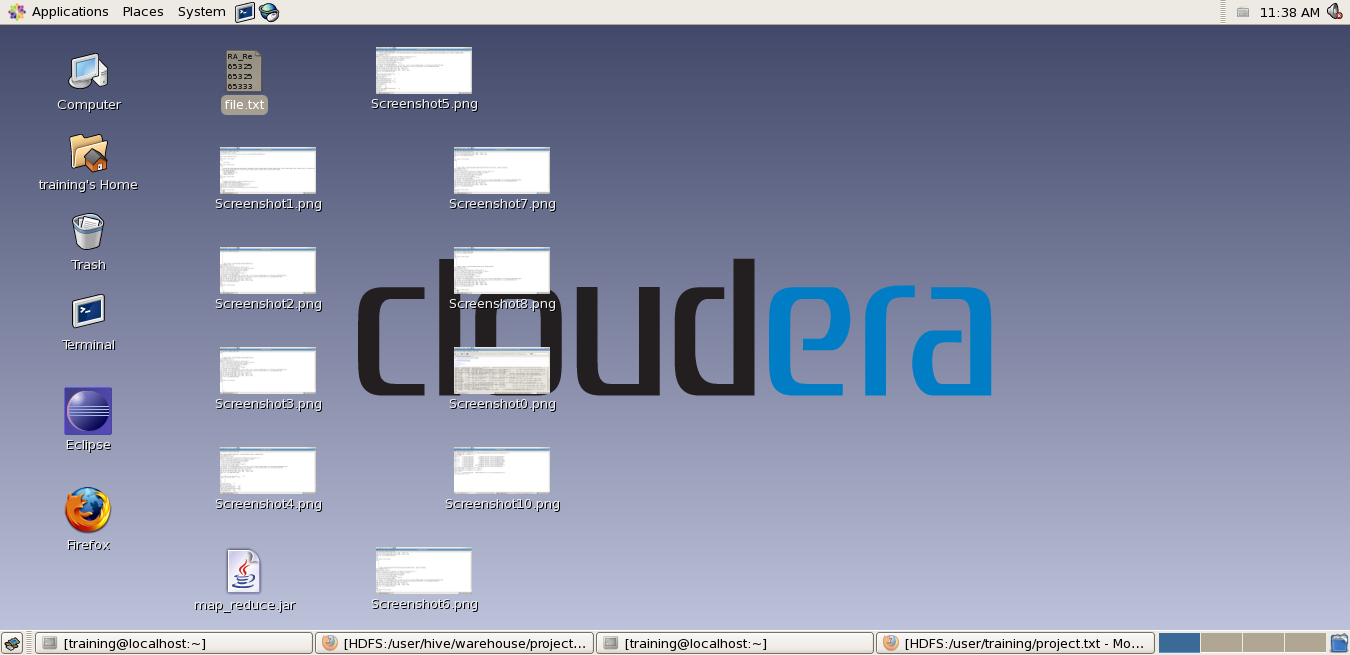
Hadoop fs -ls project.txt



**GET THE FILE FROM THE LOCAL FILE SYSTEM TO HDFS**

hadoop fs -get /user/training/project.txt /home/training/Desktop/file.txt

hadoop fs -ls

****

**Input File Format –** TAB separated values (txt file)

*RA\_Report #*

*Unique report ID.*

*RA\_CAERS Created Date*

*Date the report was created.*

*AEC\_Event Start Date*

*Date the event started.*

*PRI\_Product Role*

*Role the product played in the event. Almost always "suspect".*

*PRI\_Reported Brand/Product Name*

*Full product name (in all-caps).*

*PRI\_FDA Industry Code*

*Industry product code.*

*PRI\_FDA Industry Name*

*Industry name.*

*CI\_Age at Adverse Event*

*If known, the victim's age at the time of incident.*

*CI\_Age Unit*

*The unit for the previous value (usually years).*

*CI\_Gender*

*Victim's gender.*

*AEC\_One Row Outcomes*

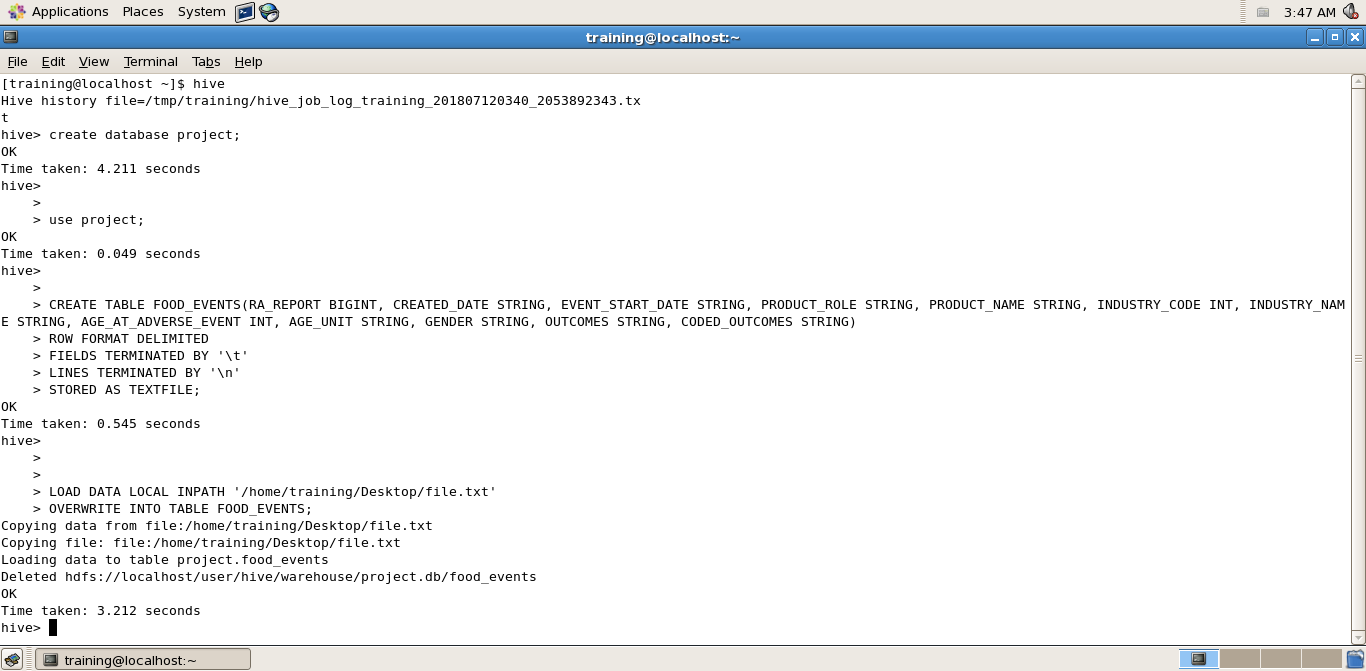
*Outcomes from the event. Hand-coded.*

*SYM\_One Row Coded SymptomsOutcomes from the event, system-coded*

***Processing Logic – Data Analysis in Hadoop Eco-System.***

*.*

**CREATION OF TABLE AND LOADING OF DATA IN IT**

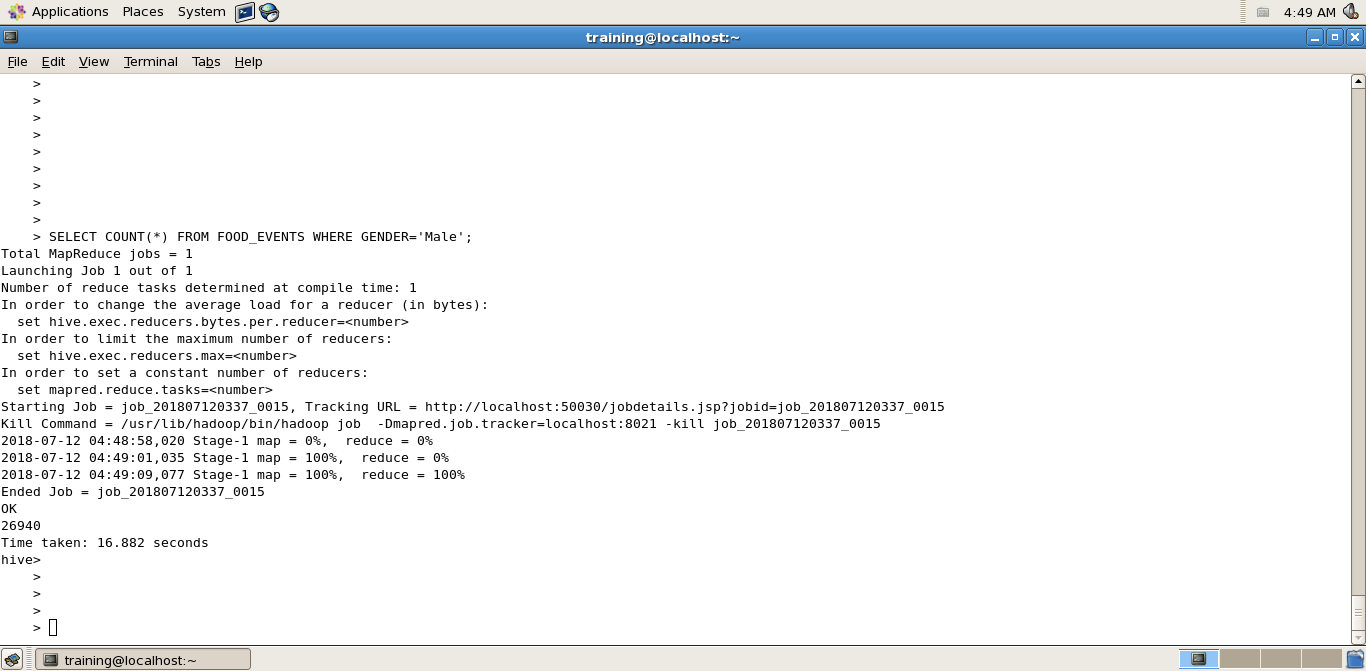
****

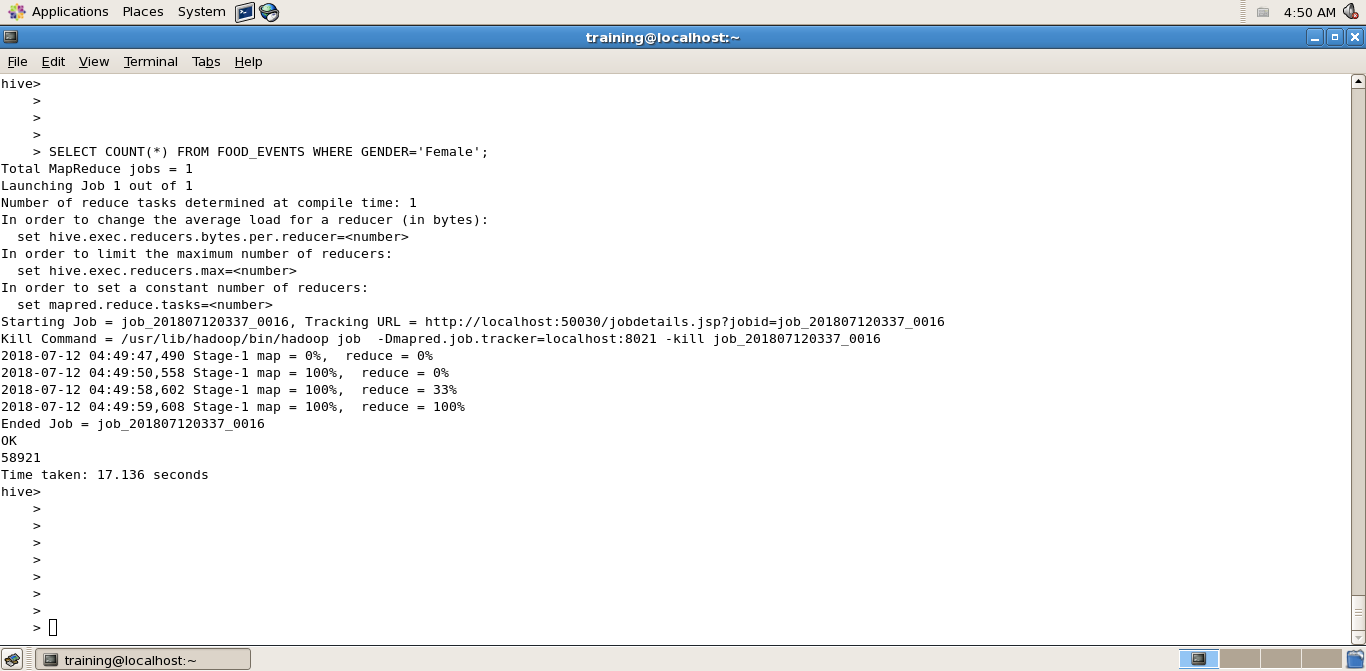
***1) Analyse the number of males and females affected due to adverse food events and thus compare them.***

SELECT COUNT(\*) FROM FOOD\_EVENTS WHERE GENDER=’Female’;

SELECT COUNT(\*) FROM FOOD\_EVENTS WHERE GENDER=’Male’;

**SCREENSHOT**

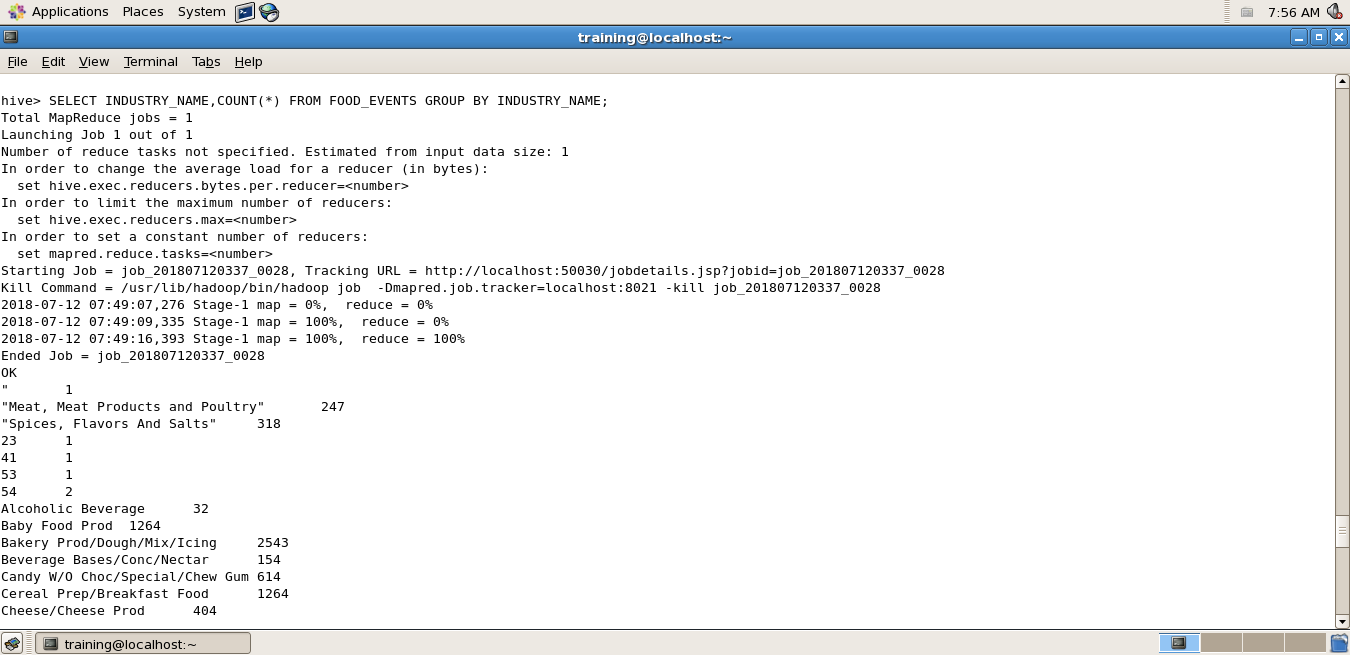
**

**

***2) Find out the type of industry due to products of which people were affected the most the most and also count the times they were affected by the same.***

SELECT INDUSTRY\_NAME,COUNT(\*) FROM FOOD\_EVENTS GROUP BY INDUSTRY\_NAME;

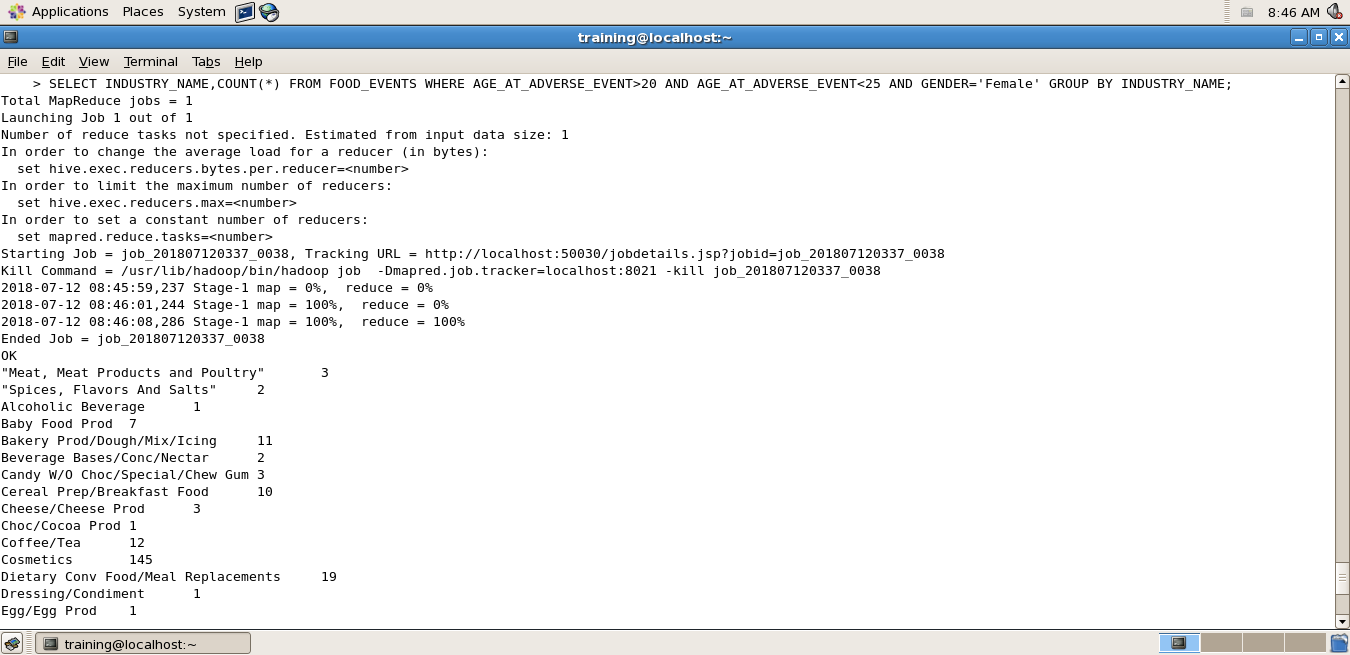
**SCREENSHOT**



***3) Find out the product that affected maximum of ladies of age 20-25.***

SELECT INDUSTRY\_NAME,COUNT(\*) FROM FOOD\_EVENTS WHERE AGE\_AT\_ADVERSE\_EVENT>20 AND AGE\_AT\_ADVERSE\_EVENT<25 AND GENDER=’Female’ GROUP BY INDUSTRY\_NAME;

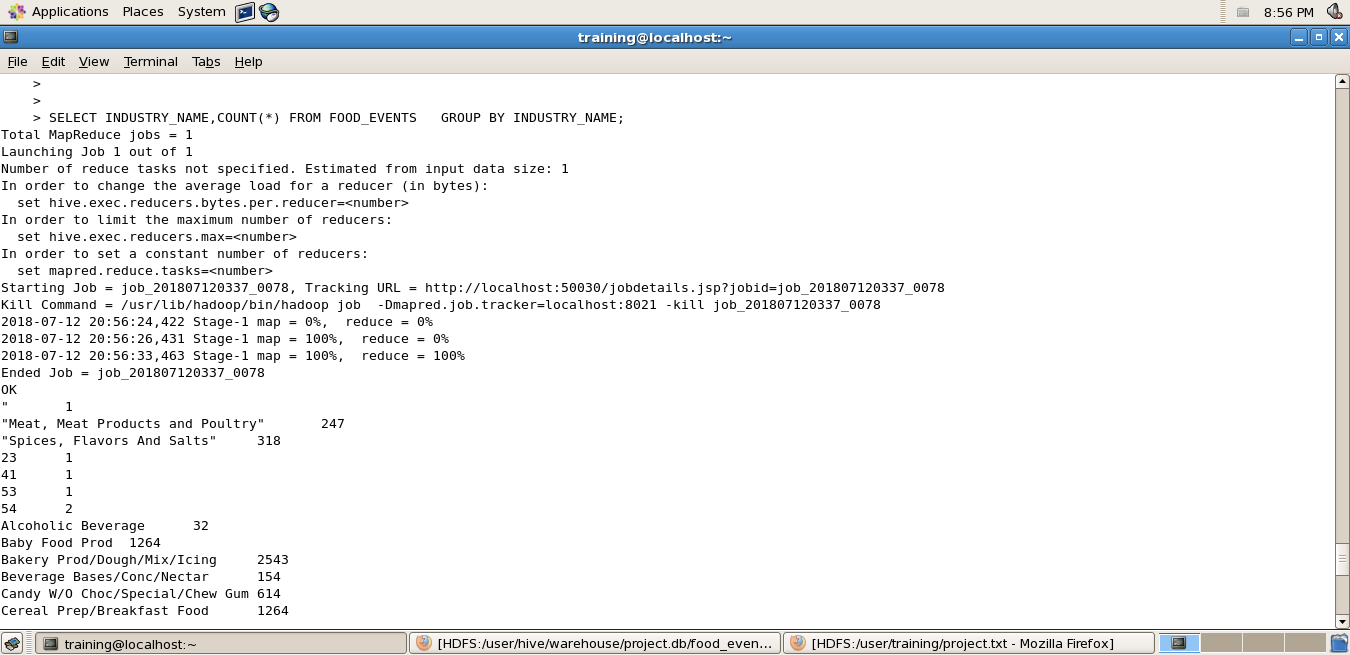
**SCREENSHOT**

****

***4) Name the product which was suspected most of the times to affect people.***

SELECT INDUSTRY\_NAME,COUNT(\*) FROM FOOD\_EVENTS GROUP BY INDUSTRY\_NAME;

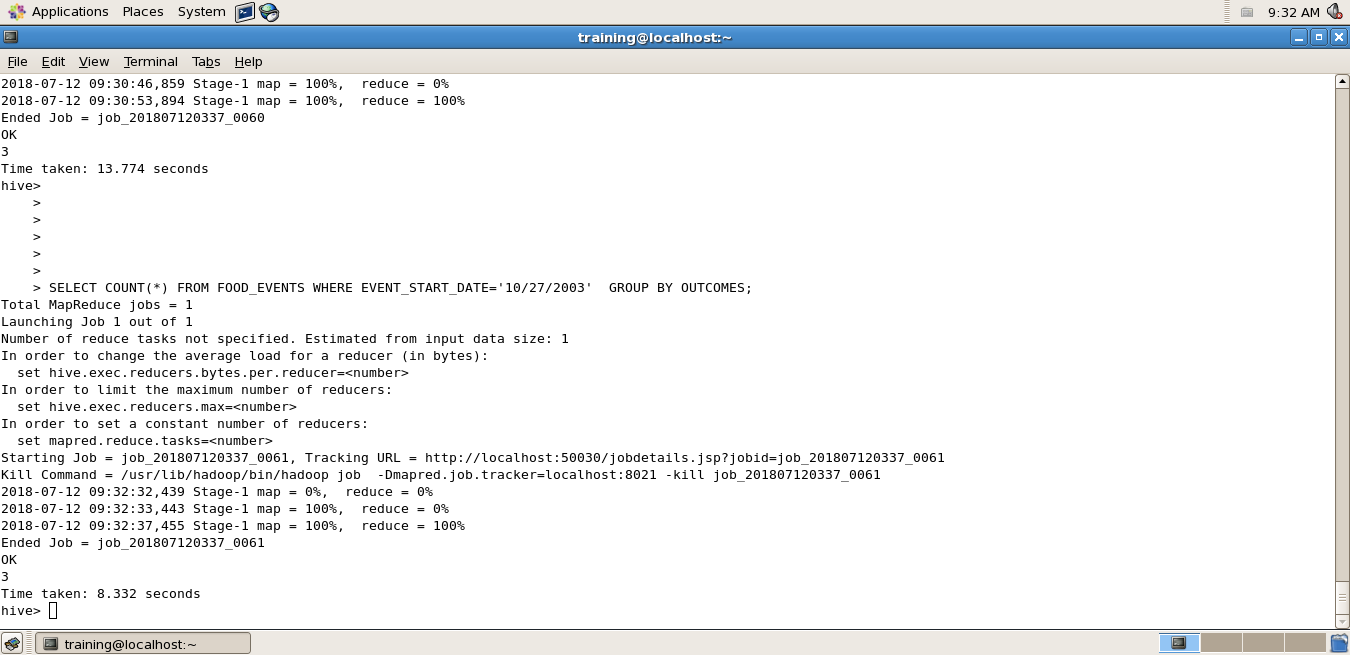
**SCREENSHOT**

****

***5) Show the events were reported on 27 October 2003 with their respective outcomes.***

SELECT COUNT(\*) FROM FOOD\_EVENTS WHERE EVENT\_START\_DATE=’10/27/2003’ GROUP BY OUTCOMES;

**SCREENSHOT**



***6) How many food events were recorded that affected the children of age less than 5 years?***

Select count(\*) from food\_events where age\_At\_Adverse\_event<5;

**SCREENSHOT**

