Contents

Flight Data Summary Report - Answers

This report uses:

- Map Reduce Code in Python
- Pig Scripts run in batch
- Hive Scripts run through the command line

Code is contained in the Appendices. Detailed notes and results are contained in the Detail Report and summary responses are contained here in the Summary Report.

Question 1
Average Number of Flights cancelled per year

142,121

Question 2
Month with the least number of flight cancellations on average

November

Question 3
Month where a cancellation due to weather most likely to occur

May

Question 4
Has the proportion of flight cancellations increased since 9/11?

There was a jump up in cancellations after September 2001 as expected after which the flight cancellation percentage fell below the figures recorded prior to the event, perhaps reflecting tighter security procedures. See report for more detailed analysis of cancellations between 2001 and 2002. Therefore, no.

Question 5
Are there any other interesting cancellation features of interest?
If so, provide details or one of these.

There were significantly increased cancellations due to Security (Code 'D') in August 2003 and February 2006. These were related to specific airline terror threats reported internationally and these are discussed in the detailed report.

Question 6
Suppose that we define a flight to have departed late if DepDelay ≥ 10, then which airline had the lowest proportion of flights departing late?

Hawaiian Airlines (HA) at 6.3% of their flights with DepDelay > = 10 minutes.

Question 7
Now suppose we define a flight to arrive late if ArrDelay ≥ 10 then which airline had the highest proportion of late arriving flights?

Express Jet Airlines (EV) with 31.1% of flights with ArrDelay > = 10 minutes.

Question 8
Which day of the week is best to fly in order to avoid delays?

Saturdays

Question 9
Post 2004 which month has the largest total NAS delay time on average?

June

Question 10
Any other things of interest?

Looking at the delays in the other categories, the highest months are as follows:

- Weather delays   July
- Carrier delays   December
- Aircraft delays   July
- Security delays   August

Security delays peaked on 10 August 2006 at 43,906 in line with the data related to Security Cancellations. See detailed report and Appendix 3 and 4 for details.

Question 11
How many flights left in 2007 on each day of the week respectively?

Answer in shaded column

Day	Flights	Didn't Depart	Total
Monday	1,090,187	22,287	1,112,474
Tuesday	1,052,983	25,579	1,078,562
Wednesday	1,062,432	26,426	1,088,858
Thursday	1,072,592	25,146	1,097,738
Friday	1,077,472	24,217	1,101,689
Saturday	917,348	15,990	933,338
Sunday	1,019,453	21,103	1,040,556
Total	7,292,467	160,748	7,453,215
		Check total	7,453,215

Flight Analysis Detailed Report Section 1

Code is shown in Appendix 1

QUESTION 1

Average Number of Flight Cancellations Per Year

Method Note:

Map Reduce Code using Python Streaming to calculate the total cancelled flights for all years 1999 to 2008. Average = total flights divided by 10 years. Cancelled flights is index 21.

Answer

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/cancelled/par*
1421214
142121
```

 Average flights cancelled per year (10 years) = 142,121

QUESTION 2

Which month has the least number of flight cancellations on average?

Method Note:

Map Reduce Code using Python Streaming to calculate the total cancelled flights by month between 1999 and 2008. Cancelled flights is index 21 from the data.

Answer

Month with the least cancellations on average –

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/monthcanc/par*
(172158, 147823, 114537, 82940, 82962, 115618, 114602, 109996, 192180, 75755, 67
340, 145303)
(17215, 14782, 11453, 8294, 8296, 11561, 11460, 10999, 19218, 7575, 6734, 14530)
```

November (67,340 cancellations over the ten years which is equal to 6,734 per year on average)

QUESTION 3

Which month is a cancellation due to weather most likely to occur?

Method Note:

Map Reduce Code using Python Streaming to calculate the total cancelled flights due to weather by month between 1999 and 2008. Cancellation due to weather is index 22. Category B indicates weather is the reason for cancellation.

Answer

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/weathercanc/par*
(42642, 38234, 23179, 9085, 8326, 14478, 15567, 16660, 27524, 11459, 11031, 4886
9)
```

Month with the most cancellations due to weather is May at 8,326

QUESTION 4

Has the proportion of flight cancellations increased since 9/11?

Method Note:

Map Reduce Code using Python Streaming to calculate 1. the total cancelled flights by year and 2. total flights per year to then calculate the proportion. Cancelled flights is index 21.

Answer

Cancelled Flights

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/proportcanc/par*
(154311, 187490, 231198, 65143, 101469, 127757, 133730, 121934, 160748, 137434)
```

Total Flights

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/flightcount/par*
(5527884, 5683047, 5967780, 5271359, 6488540, 7129270, 7140596, 7141922, 7453215
, 7009728)
```

Cancelled as Proportion of Total

Year	Cancelled	Total	%
1999	154,311	5,527,884	2.8%
2000	187,490	5,683,047	3.3%
2001	231,198	5,967,780	3.9%
2002	65,143	5,271,359	1.2%
2003	101,469	6,488,540	1.6%
2004	127,757	7,129,270	1.8%
2005	133,730	7,140,596	1.9%
2006	121,934	7,141,922	1.7%
2007	160,748	7,453,215	2.2%
2008	137,434	7,009,728	2.0%
All years	1,421,214	64,813,341	2.2%

Flight cancellations dropped significantly after 2001. A closer look at 2001 figures using Map Reduce to show cancellations by month and using the same mapper/reducer code using in Q2 for 2001 and 2002, showed the following:

Flight Cancellations Total for 2001 and 2002 By Month

Month	2001	2002
January	19,891	7,301
February	17,448	4,323
March	17,876	6,033
April	11,414	4,513
May	9,452	4,442
June	15,509	7,666
July	11,286	6,260
August	13,318	5,339
September	**99,324**	3,686
October	6,850	4,549
November	4,497	3,675
December	4,333	7,356
Total	231,198	65,143

There is a clear jump in flight cancellations as expected in September 2001 as all flights were grounded following the terrorist attacks. Following this, the cancellations appeared to be a lot lower than they were before the event itself. It is unclear why this is the case, although increased security, tighter procedures, reduced 'carrier' cancellations and a general heightened awareness may have contributed to this. Further investigation would be needed to establish the exact cause.

QUESTION 5

Are there any other interesting cancellation features of interest? If so, provide details or one of these.

I decided to look at the other cancellation reasons by year over the ten years. The list of codes for cancellations are:

Code	A	B	C	D
Reason	Carrier	Weather	NAS	Security

Method Note:

Map Reduce Code using Python Streaming, the following information was gained.

Answer

Carrier Cancellation

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/reason1/par*
(0, 0, 0, 0, 22501, 59357, 59350, 55655, 66779, 54330)
```

Weather Cancellation

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/reason2/par*
(0, 0, 0, 0, 16537, 44559, 51205, 37913, 61936, 54904)
```

NAS Cancellation

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/reason3/par*
(0, 0, 0, 0, 13870, 23733, 23084, 28209, 31995, 28188)
```

Security Cancellation

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/reason4/par*
(0, 0, 0, 0, 186, 115, 92, 157, 39, 12)
```

The results are contained in the table below. A check total was added to the final column using figures obtained from Question 4. The total codes column does not equal the check total column of all cancelled flights (shown as shaded in the table as it is likely the coding system to split cancellations across the four categories was introduced in 2003. From 2003 the total codes column equals the total cancelled, so the analysis across the codes agrees to the total cancelled flights except for 2005.

Cancellation Reasons by Year

Year	Carrier	Weather	NAS	Security	Total Codes	Total Cancelled
1999	0	0	0	0	0	154,311
200	0	0	0	0	0	187,490
2001	0	0	0	0	0	231,198
2002	0	0	0	0	0	65,143
2003	22,501	16,537	13,870	186	53,095	101,469
2004	59,357	44,559	23,733	115	127,764	127,757
2005	59,350	51,205	23,084	92	87,654	133,730
2006	55,655	37,913	28,209	157	121,934	121,934
2007	66,779	61,936	31,995	39	160,749	160,748
2008	54,330	54,904	28,188	12	137,434	137,434
Total	317,972	267,054	149,079	601	734,707	1,424,214

Should match

A Closer Look at the Security Codes

The Security code D' was of interest because there appear to be two larger figures in 2003 and 2006 of 186 and 157 respectively. This was analysed more closely splitting each year down over months using Mapper and Reducer Codes shown as Mapper and Reducer 2 in the Appendix.

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/cancelbymonth/par*
(0, 0, 0, 0, 0, 3, 0, 170, 2, 4, 3, 4)
```

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/cancelbymonth/par*
(0, 78, 13, 32, 0, 2, 2, 5, 2, 1, 19, 3)
```

Security (Code D) By month for 2003 and 2006

Month	2003	2006
January	0	0
February	0	78
March	0	13
April	0	32
May	0	0
June	3	2
July	0	2
August	170	5
September	2	2
October	4	1
November	3	19
December	4	3

1. August 2003 had a huge jump in Security cancellations.
2. February 2006 also had a high recorded figure for cancellations due to Security.

Terrorist Threats - 2003

A web search for August 2003, identified that there had been significant terror threats related to Al Qaeda at this time. Appendix 2 shows a news article.

'The State Department said the government had received indications of terrorist threats aimed at American and Western interests, including the targeting of transport and civilian aircraft.' (News.bbc.co.uk, 2003).

Terror Bomb Plot - 2006

A review of news articles did not identify any particular issue in February 2006. However, in August 2006, three people were convicted for being part of an Al Qaeda cell planning the Trans-Atlantic Bomb Plot. The plot involved using liquid explosives on 7 passenger flights from the UK to the US and Canada and could have killed up to 10,000 people.

The flight cancellations figures are for domestic flights in the USA, but it is likely that the airlines had received security threat information throughout this year, relating to a possible terror incident and were therefore on high alert in relation to all flights, international or domestic.

Analysis by Day

Mapper Reducers were written to analyse by day over the months in question.

2003

Running the mapper code for 2003 in the shell for August identified that there were three days where all the flight cancellations had been recorded, on 14th, 15th and 16th.

```
[cloudera@quickstart ~]$ cat '2003.csv'|python Question5cont_mapper.py
```

Running a reducer code for those three days showed the pattern of cancellations.

```
[cloudera@quickstart ~]$ cat '2003.csv'|python Question5cont_mapper.py|python Qu
estion5cont_reducer.py
(45, 123, 2)
```

168 of the total cancellations due to security of 186 in 2003 (90%) occurred over two days in August (14th and 15th)

<u>2006</u>

In 2006, the majority of the cancellations were in February and they all occurred on 22nd.

Flight Analysis Detailed Report Section 2

Code is shown in Appendix 3

QUESTION 6

Suppose that we define a flight to have departed late if DepDelay $\geq$ 10, then which airline had the lowest proportion of flights departing late?

Method Note

The analysis was done using Pig Script. Code was tested in the Grunt shell and then run through batch processing. Proportion of flights defined by count of departure delays >= 10 minutes by airline over total flights by airline.

Answer

Count of delay by airline and count of total flights by airline (with header rows excluded)

```
(AA,1468370)  (AA,7011039)
(F9,77111)    (F9,336958)
(HA,17154)    (HA,274265)
(OH,333922)   (OH,1464176)
(OO,590557)   (OO,3090853)
(9E,93020)    (9E,521059)
(HP,314384)   (HP,1418057)
(US,1136002)  (US,5364830)
(B6,206049)   (B6,811341)
(CO,624627)   (CO,3325304)
(DL,1276696)  (DL,6826616)
(EV,495361)   (EV,1697172)
(FL,314105)   (FL,1265138)
(NW,847820)   (NW,4817229)
(TZ,34896)    (TZ,208420)
(AS,416454)   (AS,1583287)
(UA,1301433)  (UA,5869005)
(MQ,913432)   (MQ,3954895)
(XE,467500)   (XE,2350309)
(AQ,19238)    (AQ,154381)
(DH,153363)   (DH,693047)
(TW,149402)   (TW,787682)
(WN,2691678)  (WN,10134222)
(YV,211744)   (YV,854056)
```

Total number of flights with delays >= 10 minutes =   14,154,318

Total number of flights by airline =   64,813,341

Delays Greater than equal to 10 minutes by Airline All Years			
Airline	Delays	Total	%
HA	17,154	274,265	6.3%
AQ	19,238	154,381	12.5%
TZ	34,896	208,420	16.7%
NW	847,820	4,817,229	17.6%
9E	93,020	521,059	17.9%
DL	1,276,696	6,826,616	18.7%
CO	624,627	3,325,304	18.8%
TW	149,402	787,682	19.0%
OO	590,557	3,090,853	19.1%
XE	467,500	2,350,309	19.9%
AA	1,468,370	7,011,039	20.9%
US	1,136,002	5,364,830	21.2%
DH	153,363	693,047	22.1%
UA	1,301,433	5,869,005	22.2%
HP	314,384	1,418,057	22.2%
OH	333,922	1,464,176	22.8%
F9	77,111	336,958	22.9%
MQ	913,432	3,954,895	23.1%
YV	211,744	854,056	24.8%
FL	314,105	1,265,138	24.8%
B6	206,049	811,341	25.4%
AS	416,454	1,583,287	26.3%
WN	2,691,678	10,134,222	26.6%
EV	495,361	1,697,172	29.2%
Total	14,154,318	64,813,341	21.8%

Airline Code HA (Hawaiian Airlines) had the lowest proportion of flight delays greater than 10 minutes. With 17,154 delays out of 274,265 flights over the ten-year period, 6.3% of flights were delayed compared to an average for all airlines of 21%.

QUESTION 7

Now suppose we define a flight to arrive late if ArrDelay $\geq$ 10 then which airline had the highest proportion of late arriving flights?

Method Note

The analysis was done using Pig Script. Code was tested in the Grunt shell and then run through batch processing. Proportion of flights defined by count of arrival delays >= 10 minutes by airline over total flights by airline.

Answer

Flight Delays Count per airline and Count of Total flights per airline calculated in Q6 above (with header rows excluded).

```
(AA,1860667)   (AA,7011039)
(F9,90533)     (F9,336958)
(HA,25651)     (HA,274265)
(OH,395635)    (OH,1464176)
(OO,699033)    (OO,3090853)
(9E,122408)    (9E,521059)
(HP,412323)    (HP,1418057)
(US,1431651)   (US,5364830)
(B6,242480)    (B6,811341)
(CO,881632)    (CO,3325304)
(DL,1794830)   (DL,6826616)
(EV,528598)    (EV,1697172)
(FL,370752)    (FL,1265138)
(NW,1250633)   (NW,4817229)
(TZ,52567)     (TZ,208420)
(AS,476618)    (AS,1583287)
(UA,1635217)   (UA,5869005)
(MQ,1072101)   (MQ,3954895)
(XE,647074)    (XE,2350309)
(AQ,21532)     (AQ,154381)
(DH,176488)    (DH,693047)
(TW,194554)    (TW,787682)
(WN,2458861)   (WN,10134222)
(YV,239935)    (YV,854056)
```

Arrival Delays Greater than equal to 10 minutes by Airline All Years

Airline	Delayed	Total	Percent
EV	528,598	1,697,172	31.1%
AS	476,618	1,583,287	30.1%
B6	242,480	811,341	29.9%
FL	370,752	1,265,138	29.3%
HP	412,323	1,418,057	29.1%
YV	239,935	854,056	28.1%
UA	1,635,217	5,869,005	27.9%
XE	647,074	2,350,309	27.5%
MQ	1,072,101	3,954,895	27.1%
OH	395,635	1,464,176	27.0%
F9	90,533	336,958	26.9%
US	1,431,651	5,364,830	26.7%
AA	1,860,667	7,011,039	26.5%
CO	881,632	3,325,304	26.5%
DL	1,794,830	6,826,616	26.3%
NW	1,250,633	4,817,229	26.0%
DH	176,488	693,047	25.5%
TZ	52,567	208,420	25.2%
TW	194,554	787,682	24.7%
WN	2,458,861	10,134,222	24.3%
9E	122,408	521,059	23.5%
OO	699,033	3,090,853	22.6%
AQ	21,532	154,381	13.9%
HA	25,651	274,265	9.4%
Total	17,081,773	64,813,341	26.4%

Express Jet Airlines (EV) has the highest proportion of their flights with an Arrival Delay > = 10 minutes. With 528,598 flights delayed out of 1,697,172 total flights over ten years, 31.1% of flights were delayed compared to 26.8% average for all airlines.

QUESTION 8

Which day of the week is best to fly in order to avoid delays?

Method Note

The analysis was done using Pig Script. Code was tested in the Grunt shell and then run through batch processing.

Assumption - avoiding delays was taken to be either flights arriving with **no delay or arriving early**, therefore index [14] of zero or a negative number (indicating flight was early). Day codes (1 = Monday, 2= Tuesday etc.). These flights were then expressed over the total number of flights recorded on that day.

Answer

Flights early or zero arrival delay by day and total flights by day

```
(1,5205660)  (1,9540688)
(2,5366853)  (2,9452110)
(7,4950728)  (7,9035392)
(6,4909107)  (6,8217585)
(5,4741359)  (5,9548997)
(4,4861536)  (4,9520086)
(3,5222475)  (3,9498483)
```

Early flights or those with zero delay time for each day of the week, over the ten-year period are shown in the table below.

Saturday is the best day to fly to avoid delays as 60% of flights arrive early or with zero delay. This was validated by a web search. Fridays were the worst day for delays and Saturdays the best day to travel as supported by this investigation. **(Medina, 2019). See detailed article in** Appendix 3 (Question 8).

Early or Zero Arrival Delay Flights By Day			
Day	Flights	Total	%
Monday	5,205,660	9,540,688	55%
Tuesday	5,366,853	9,452,110	57%
Wednesday	5,222,475	9,498,483	55%
Thursday	4,861,536	9,520,086	51%
Friday	4,741,359	9,548,997	50%
Saturday	4,909,107	8,217,585	60%
Sunday	4,950,728	9,035,392	55%
Total	35,257,718	64,813,341	54%

QUESTION 9

Post 2004 which month has the largest total NAS delay time on average?

Method Note

The analysis was done using HIVE. A table was created using all the columns and the 'all_years.csv' file loaded into it. See code in the Appendix. The total lines were then counted as a check and agreed as 64813351 (including header rows)

```
64813351
Time taken: 144.281 seconds, Fetched: 1 row(s)
hive> ■
```

Running a Select Statement on 3 rows, shows the data and columns

```
1999    1     28     4      2016    1908    2126    2005    US      1244    N
955VJ   70    57     41     81      68      RIC     PHL     198     19      1
0       0     NA     0      NULL    NULL    NULL    NULL    NULL
```

Answer

Total NAS delay for four years: Highest is in June

```
1          11146515
2          10458160
3          10877894
4          8713740
5          10039315
6          15396612
7          15318469
8          13118355
9          8820485
10         10281274
11         11058647
12         15134604
```

NAS Delay By Month Post 2004 (mins)

Month	NAS (4years)	Average/yr
June	15,396,612	3,849,153
July	15,318,469	3,829,617
December	15,134,604	3,783,651
August	13,118,355	3,279,589
January	11,146,515	2,786,629
November	11,058,647	2,764,662
March	10,877,894	2,719,474
February	10,458,160	2,614,540
October	10,281,274	2,570,319
May	10,039,315	2,509,829
September	8,820,485	2,205,121
April	8,713,740	2,178,435

This code includes delays under the control of the National Airspace System (NAS) including:

- Non- extreme weather condition
- Airport operations
- Heavy traffic volume
- Air traffic control   (Aspmhelp.faa.gov, 2019)

The three highest total delays were in June, July and December, all very popular times for air travel (Christmas and school holidays) so higher delays reflects this heavier traffic volume plus impacts on basic airport operations and air traffic control of more flights taking off and landing at this busier time of year.

QUESTION 10

What Else is of Interest?

Appendix 4 shows detailed tables for each delay type including Total Minutes (over 4 years), Total Number of Incidents (over 4 years), Average Minutes per Year, Number of Incidents and the Minutes per Incident all derived from running Hive queries using the script shown in Appendix 3.

The highest percentage of delays in terms of total minutes delay is because of aircraft delays but the highest number of delays is due to NAS checks.
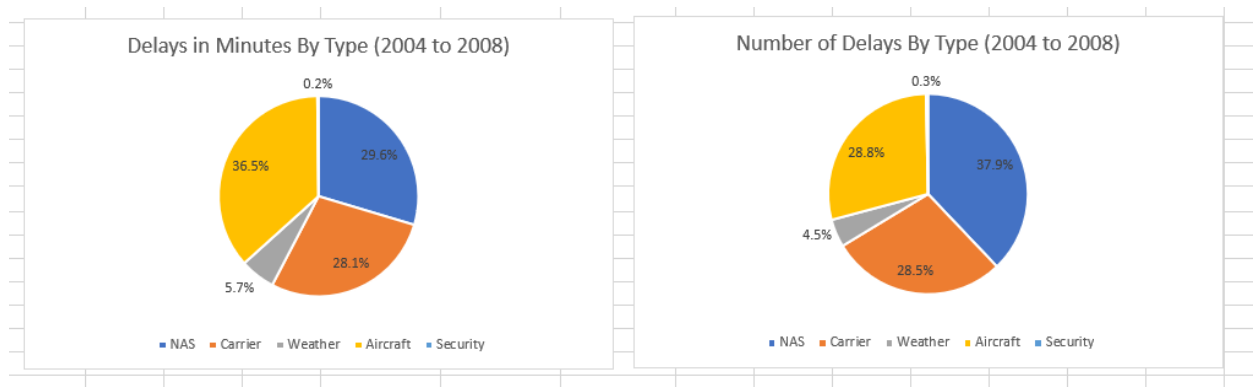

Points of interest from the table:

- Carrier delays are highest in December and July likely due to heavy volumes.
- Weather delays are highest in July (see below).
- Late Aircraft is highest in July again likely due to volumes.
- Security delays were highest in August but the highest minutes per incident were in April, Easter being another key time for travel.

Weather Delays

The higher delays in July for 'Weather' was interesting as it is expected weather delays would be highest in winter. However, research showed delays do indeed peak over the summer months mostly due to thunderstorms and heavy rain.

What type of weather causes the most delay?
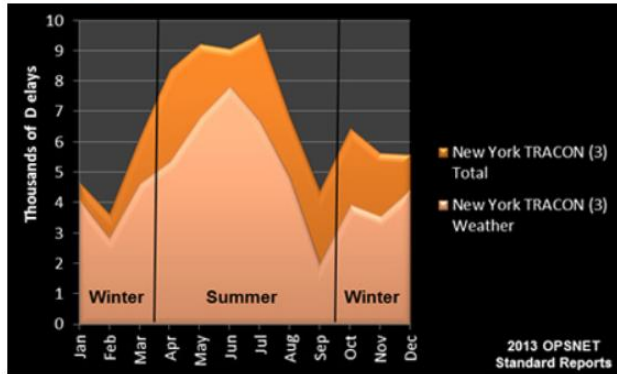
Weather-related delays compared to total delays at Newark, LaGuardia and Kennedy airports by month over the course of a year (2013).

Source: (FAA.Gov, 2013)

Security

There is a huge jump in August, likely mirroring the Security Cancellations peaks in 2006 noted earlier in this report. A series of Hive scripts were run (Appendix 3) which narrowed the results down to confirm this. Looking at the total number of incidents, they peaked in August 2006.
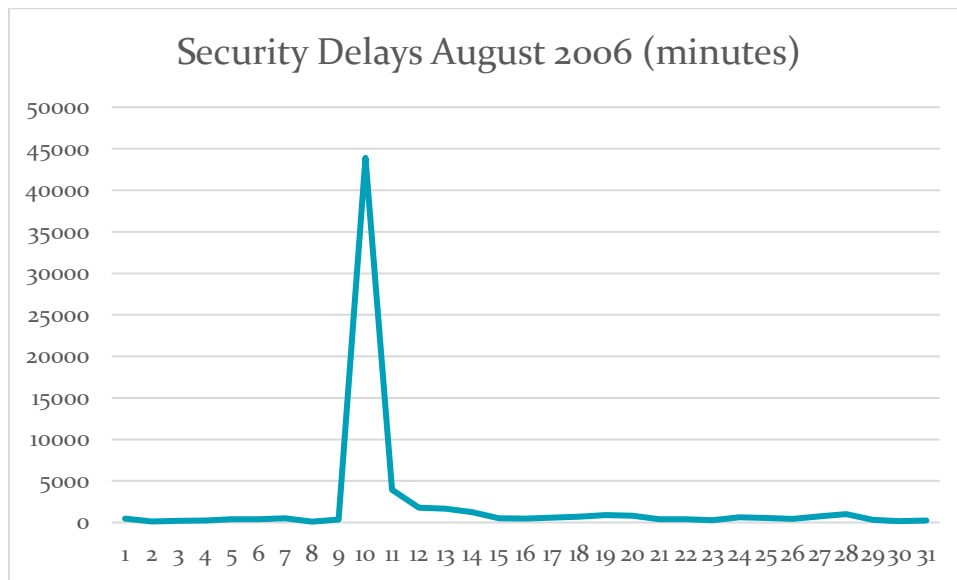
QUESTION 11

How many flights left in 2007 on each day of the week respectively?

Method Note

Map Reduce Code was used for this question and the script is contained in Appendix 3.

```
[cloudera@quickstart ~]$ cat '2007.csv'|python quest10_mapper.py|python quest10_reducer.py
(1090187, 1052983, 1062432, 1072592, 1077472, 917348, 1019453)
```

Answer

Day	Flights
Monday	1,090,187
Tuesday	1,052,983
Wednesday	1,062,432
Thursday	1,072,592
Friday	1,077,472
Saturday	917,348
Sunday	1,019,453
Total	7,292.467

Bibliography

Aspmhelp.faa.gov. (2019). *Types of Delay - ASPMHelp*. [online] Available at: https://aspmhelp.faa.gov/index.php/Types_of_Delay [Accessed 11 Jan. 2019].

Aspmhelp.faa.gov. (2019). *Types of Delay - ASPMHelp*. [online] Available at: https://aspmhelp.faa.gov/index.php/Types_of_Delay [Accessed 11 Jan. 2019].

Duncan Gardham, S. (2019). *Airline terror trial: The bomb plot to kill 10,000 people*. [online] Telegraph.co.uk. Available at: https://www.telegraph.co.uk/news/uknews/terrorism-in-the-uk/6153243/Airline-terror-trial-The-bomb-plot-to-kill-10000-people.html [Accessed 9 Jan. 2019].

Medina, J. (2019). *Best Days, Times to Fly to Avoid Flight Delays and Cancellations*. [online] Flightbucks.com. Available at: http://www.flightbucks.com/blog/best-days-times-to-fly-to-avoid-flight-delays-cancellations [Accessed 11 Jan. 2019].

News.bbc.co.uk. (2019). *BBC NEWS | UK | BA suspends flights to Saudi Arabia*. [online] Available at: http://news.bbc.co.uk/1/hi/uk/3148067.stm [Accessed 9 Jan. 2019].

Appendix 1

Streaming Code

Python Streaming Code Used for Running Map Reduce Jobs on Q 1-5

```
`
```

```
hadoop \
jar /usr/lib/hadoop-0.20-mapreduce/contrib/streaming/hadoop-streaming-2.6.0-mr1-cdh5.13.0.jar \
-file /home/cloudera/canc_mapper.py /home/cloudera/canc_reducer.py \
-mapper "python canc_mapper.py" \
-reducer "python canc_reducer.py" \
-input /user/cloudera/all_years.csv \
-output /user/cloudera/cancelled
```

| Data source file | Output file | Mapper and Reducer Codes |

Code to view output:

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/cancelled/par*
```

Question 1

Mapper and Reducer Code

```python
#!/usr/bin/env python

import sys

flight_count = 0
canc_count = 0

for line in sys.stdin:
    line1 = line.replace('"','')
    flight = line1.split(",")
    canc_count = canc_count + 1


    if flight[21] == '1':
        cancelled = flight[21]
        if cancelled.isdigit:
            year = flight[0]
            print(year,cancelled)
```

```python
#!/usr/bin/env python
import sys


canc_flight = 0

for line in sys.stdin:
    line3 = line.replace("'","")
    line2 = line3.replace("(","")
    line1 = line2.replace(")","")

    flights = line1.split(',')
    cancelled = int(flights[1])
    year = flights[0]

    canc_flight = canc_flight + cancelled

print(canc_flight)
print(canc_flight/10)
```

Question 2

Mapper and Reducer Code

```python
#!/usr/bin/env python

import sys
for line in sys.stdin:

    line1 = line.replace('"','')
    flight = line1.split(",")
    if flight[21] == '1':
        canc = flight[21]
        month = flight[1]
        print(month,canc)
```

```python
#!/usr/bin/env python
import sys

jan = 0
feb = 0
mar = 0
apr = 0
may = 0
jun = 0
jul = 0
aug = 0
sep = 0
oct = 0
nov = 0
dec = 0

for line in sys.stdin:
    line3 = line.replace("'","")
    line2 = line3.replace("(","")
    line1 = line2.replace(")","")
    flights = line1.split(',')

    if flights[0] == '1':
        jan = jan + 1
    elif flights[0] == '2':
        feb = feb + 1
    elif flights[0] == '3':
        mar = mar + 1
    elif flights[0] == '4':
        apr = apr + 1
    elif flights[0] == '5':
        may = may + 1
    elif flights[0] == '6':
        jun = jun + 1
    elif flights[0] == '7':
        jul = jul + 1

    elif flights[0] == '8':
        aug = aug + 1
    elif flights[0] == '9':
        sep = sep + 1
    elif flights[0] == '10':
        oct = oct + 1
    elif flights[0] == '11':
        nov = nov + 1
    elif flights[0] == '12':
        dec = dec + 1

print(jan,feb,mar,apr,may,jun,jul,aug,sep,oct,nov,dec)
print(jan/10,feb/10,mar/10,apr/10,may/10,jun/10,jul/10,aug/10,sep/10,oct/10,nov/10,dec/10)
```

Question 3

Mapper and Reducer Code

```python
#!/usr/bin/env python

import sys
for line in sys.stdin:

    line1 = line.replace('"','')
    flight = line1.split(",")

    if flight[22] == "B":
        cancelled = flight[21]
        if cancelled.isdigit:
            month = flight[1]
            print(month,cancelled)
```

```python
#!/usr/bin/env python
import sys

jan = 0
feb = 0
mar = 0
apr = 0
may = 0
jun = 0
jul = 0
aug = 0
sep = 0
oct = 0
nov = 0
dec = 0


for line in sys.stdin:
    line3 = line.replace("'","")
    line2 = line3.replace("(","")
    line1 = line2.replace(")","")

    flights = line1.split(',')

    if flights[0] == '1':
        jan = jan + 1
    elif flights[0] == '2':
        feb = feb + 1
```

```
    elif flights[0] == '3':
        mar = mar + 1
    elif flights[0] == '4':
        apr = apr + 1
    elif flights[0] == '5':
        may = may + 1
    elif flights[0] == '6':
        jun = jun + 1
    elif flights[0] == '7':
        jul = jul + 1
    elif flights[0] == '8':
        aug = aug + 1
    elif flights[0] == '9':
        sep = sep + 1
    elif flights[0] == '10':
        oct = oct + 1
    elif flights[0] == '11':
        nov = nov + 1
    elif flights[0] == '12':
        dec = dec + 1


print(jan,feb,mar,apr,may,jun,jul,aug,sep,oct,nov,dec)
```

Question 4 (a)(i)

Mapper (Cancelled Flights by Year)

```
#!/usr/bin/env python

import sys

canc_count = 0

for line in sys.stdin:
    line1 = line.replace('"','')
    flight = line1.split(",")
    canc_count = canc_count + 1


    if flight[21] == '1':
        cancelled = flight[21]
        if cancelled.isdigit:
            year = flight[0]
            print(year,cancelled)
```

Question 4 (a)(ii)

Reducer (Cancelled Flights by Year)

```python
#!/usr/bin/env python
import sys

y1 = 0
y2 = 0
y3 = 0
y4 = 0
y5 = 0
y6 = 0
y7 = 0
y8 = 0
y9 = 0
y10 = 0


for line in sys.stdin:
    line3 = line.replace("'","")
    line2 = line3.replace("(","")
    line1 = line2.replace(")","")

    flights = line1.split(',')

    if flights[0] == '1999':
        y1 = y1 + 1
    elif flights[0] == '2000':
        y2 = y2 + 1
    elif flights[0] == '2001':
        y3 = y3 + 1

    elif flights[0] == '2002':
        y4 = y4 + 1
    elif flights[0] == '2003':
        y5 = y5 + 1
    elif flights[0] == '2004':
        y6 = y6 + 1
    elif flights[0] == '2005':
        y7 = y7 + 1
    elif flights[0] == '2006':
        y8 = y8 + 1
    elif flights[0] == '2007':
        y9 = y9 + 1
    elif flights[0] == '2008':
        y10 = y10 + 1

print(y1,y2,y3,y4,y5,y6,y7,y8,y9,y10)
```

Question 4 (b)(i)

Mapper (Total Flights by Year)

```python
#!/usr/bin/env python

import sys

flight_count = 0

for line in sys.stdin:
    line1 = line.replace('"','')
    flight = line1.split(",")
    flight_count = flight_count + 1
    year = flight[0]

    print(year,flight_count)
```

Question 4 (b)(ii)

Reducer (Total Flights by Year)

```python
#!/usr/bin/env python
import sys

y1 = 0
y2 = 0
y3 = 0
y4 = 0
y5 = 0
y6 = 0
y7 = 0
y8 = 0
y9 = 0
y10 = 0


for line in sys.stdin:
    line3 = line.replace("'","")
    line2 = line3.replace("(","")
    line1 = line2.replace(")","")
```

```
    flights = line1.split(',')

    if flights[0] == '1999':
        y1 = y1 + 1
    elif flights[0] == '2000':
        y2 = y2 + 1
    elif flights[0] == '2001':
        y3 = y3 + 1
    elif flights[0] == '2002':
        y4 = y4 + 1
    elif flights[0] == '2003':
        y5 = y5 + 1
    elif flights[0] == '2004':
        y6 = y6 + 1
    elif flights[0] == '2005':
        y7 = y7 + 1
    elif flights[0] == '2006':
        y8 = y8 + 1
    elif flights[0] == '2007':
        y9 = y9 + 1
    elif flights[0] == '2008':
        y10 = y10 + 1

print(y1,y2,y3,y4,y5,y6,y7,y8,y9,y10)
```

Question 5 (a)(i)

Mapper 1 (Analyse Security Code Cancellations by Year)

```
#!/usr/bin/env python

import sys
for line in sys.stdin:

    line1 = line.replace('"','')
    flight = line1.split(",")

    if flight[22] == 'A':
        canc = flight[21]
        if canc.isdigit:
            year = flight[0]
            print(year,canc)
```

Each Security Code, A, B, C, D

was used in turn

Question 5 (a)(i)

Reducer (Analyse Security Code Cancellations by Year)

```python
#!/usr/bin/env python
import sys

y1 = 0
y2 = 0
y3 = 0
y4 = 0
y5 = 0
y6 = 0
y7 = 0
y8 = 0
y9 = 0
y10 = 0


for line in sys.stdin:
    line3 = line.replace("'","")
    line2 = line3.replace("(","")
    line1 = line2.replace(")","")

    flights = line1.split(',')


    if flights[0] == '1999':
        y1 = y1 + 1
    elif flights[0] == '2000':
        y2 = y2 + 1
    elif flights[0] == '2001':
        y3 = y3 + 1
    elif flights[0] == '2002':
        y4 = y4 + 1
    elif flights[0] == '2003':
        y5 = y5 + 1
    elif flights[0] == '2004':
        y6 = y6 + 1
    elif flights[0] == '2005':
        y7 = y7 + 1
    elif flights[0] == '2006':
        y8 = y8 + 1
    elif flights[0] == '2007':
        y9 = y9 + 1
    elif flights[0] == '2008':
        y10 = y10 + 1

print(y1,y2,y3,y4,y5,y6,y7,y8,y9,y10)
```

Question 5 (b)(i)

Mapper 2 (Analyse Security Code Cancellations for 2003 & 2006 by Month)

```python
#!/usr/bin/env python

import sys
for line in sys.stdin:

    line1 = line.replace('"','')
    flight = line1.split(",")

    if flight[22] == 'D':
        security = flight[21]
        if security.isdigit:
            month = flight[1]
            print(month,security)
```

Question 5 (b)(ii)

Reducer 2 (Analyse Security Code Cancellations for 2003 & 2006 by Month)

```python
#!/usr/bin/env python
import sys

jan = 0
feb = 0
mar = 0
apr = 0
may = 0
jun = 0
jul = 0
aug = 0
sep = 0
oct = 0
nov = 0
dec = 0


for line in sys.stdin:
    line3 = line.replace("'","")
    line2 = line3.replace("(","")
    line1 = line2.replace(")","")

    flights = line1.split(',')


    if flights[0] == '1':
        jan = jan + 1
    elif flights[0] == '2':
        feb = feb + 1
    elif flights[0] == '3':
        mar = mar + 1
    elif flights[0] == '4':
        apr = apr + 1
    elif flights[0] == '5':
        may = may + 1
    elif flights[0] == '6':
        jun = jun + 1
    elif flights[0] == '7':
        jul = jul + 1
    elif flights[0] == '8':
        aug = aug + 1
    elif flights[0] == '9':
        sep = sep + 1
    elif flights[0] == '10':
        oct = oct + 1
    elif flights[0] == '11':
        nov = nov + 1
    elif flights[0] == '12':
        dec = dec + 1

print(jan,feb,mar,apr,may,jun,jul,aug,sep,oct,nov,dec)
```

Question 5 (c) (i)

Mapper 3 (Analyse Cancellations for February 2006 and August 2003 by Day)

```python
#!/usr/bin/env python

import sys
for line in sys.stdin:

    line1 = line.replace('"','')
    flight = line1.split(",")

    if flight[0] == '2003'and flight[1] == '3'and flight[22] == 'D':
        security = flight[21]
        if security.isdigit:
            day = flight[2]
            print(day,security)
```

Code used for both 2003 and 2006 changing year and month accordingly

Question 5 (c) (ii)

Reducer 3 (Analyse Cancellations for February 2006 and August 2006 by Day)

```python
#!/usr/bin/env python
import sys


d14 = 0
d15 = 0
d16 = 0

for line in sys.stdin:
    line3 = line.replace("'","")
    line2 = line3.replace("(","")
    line1 = line2.replace(")","")

    flights = line1.split(',')

    if flights[0] == '14':
        d14 = d14 + 1
    elif flights[0] == '15':
        d15 = d15 + 1
    elif flights[0] == '16':
        d16 = d16 + 1


print(d14,d15,d16)
```

Code used for both 2003 and 2006 changing days accordingly

Appendix 2

TERROR THREATS

BBC NEWS

▶ LIVE BBC NEWS CHANNEL

Last Updated: Wednesday, 13 August, 2003, 22:18 GMT 23:18 UK

✉ E-mail this to a friend   🖨 Printable version

BA suspends flights to Saudi Arabia

British Airways has suspended all flights to Saudi Arabia until further notice, in response to a specific terrorist threat to its planes in the country.

The threat, announced on Wednesday, is not thought to be related to the missile arrests in the United States.

Several passengers were stranded at Heathrow by the news

Meanwhile, the US renewed its warning for its citizens to avoid non-essential travel to Saudi Arabia.

The State Department said the government had received indications of terrorist threats aimed at American and Western interests, including the targeting of transport and civilian aircraft.

BA's decision follows advice from the Department for Transport.

A spokesman for the department said it had warned airlines there was "credible intelligence of a serious threat to UK aviation interests in area".

The source of the scare lies in the Saudi capital Riyadh, where the authorities have been trying to break up Islamist terror cells, said BBC security correspondent Frank Gardner.

> **By watching Riyadh airport they've learned this week that al-Qaeda sympathisers also had the airport under surveillance**
> The BBC's Frank Gardner

News Front Page
World
UK
England
Northern Ireland
Scotland
Wales
Business
Politics
Health
Education
Science & Environment
Technology
Entertainment
Also in the news
- - - - - - - - - -
Video and Audio
- - - - - - - - - -
Have Your Say
Magazine
In Pictures
Country Profiles
Special Reports

RELATED BBC SITES
SPORT
WEATHER
CBBC NEWSROUND
ON THIS DAY
EDITORS' BLOG

LANGUAGES
NEWYDDION

August 2003

Terror Threats

To Western Airlines in US and UK

Airline terror trial: The bomb plot to kill 10,000 people

Three British Muslims have been convicted of planning a series of co-ordinated suicide bomb attacks on transatlantic airliners, which could have killed up to 10,000 people.

This video isn't encoded for your device

Link to this video
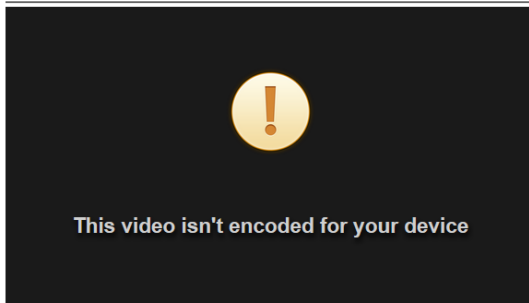
By Duncan Gardham, Security Correspondent
11:00PM BST 07 Sep 2009

The al-Qaeda cell plotted to cause mass murder by detonating home-made liquid explosives on board at least seven passenger flights bound for the US and Canada. The plot had the potential to be three times as deadly as the 9/11 attacks of 2001.

The convictions followed Britain's largest counter-terrorism operation and two criminal trials which, in total, cost an estimated £60million.

All three men convicted on Monday had been found guilty at an earlier trial last year of conspiracy to murder, but prosecutors said it was vital to secure a conviction on another charge of conspiring to blow up the aircraft in order to prove that the threat to air traffic was genuine.

Terrorism in the UK
News » Major News »
UK News »
Telegraph TV »
Duncan Gardham »

In Terrorism In The UK

Law & Order

Defence

Operation Strong Tower

Bomb Plot. Terrorists convicted in August 2006.

High cancellations in February 2006 in US domestic flights could have been due to security alerts relating to activity being planned

Appendix 3

Questions 6 to 10

Question 6

Using Pig Script

Load data and examine first few rows of data set.

```
flights = LOAD 'all_years.csv' USING PigStorage(',');
limited = LIMIT flights 2;
dump limited;
```

Iterate line by line and pull out Airline Code and Delay in minutes, filter to remove header, filter to get flights where the departure delay is greater than equal to 10 minutes and then group by airline. Count the number of flights within each airline code group and dump result to screen. Run in pig map reduce using the Grunt Shell pig -x

```
flight = LOAD 'all_years.csv' USING PigStorage(',');
flight_limit = FOREACH flight GENERATE $8 as airline, $15 as delay,
filtered = FILTER flight_limit BY delay>=10;
grouped = GROUP filtered BY airline;
count = FOREACH grouped GENERATE group,COUNT(filtered);
ordered = ORDER count BY DESC;
dump ordered;
```

To find total flights to calculate proportion:

Filtered to remove 10 header rows from results using '! = UniqueCarrier' in the code.

```
flights = LOAD 'all_years.csv' USING PigStorage(',');
flight_limit = FOREACH flights GENERATE $8 AS airline, $0 AS year;
filtered = FILTER flight_limit By airline != 'UniqueCarrier';
total_grouped = GROUP filtered BY airline;
total_flights = FOREACH total_grouped GENERATE group,COUNT(filtered);
dump total_flights;
```

Question 7

Using Pig Script

Iterate line by line and pull out Airline Code and Delay in minutes, filter to get flights where the arrival delay is greater than equal to 10 minutes and then group by airline. Count the number of flights within each airline code group and dump result to screen. Run in pig map reduce using the Grunt Shell pig -x

```
flight = LOAD 'all_years.csv' USING PigStorage(',');
flight_limit = FOREACH flight GENERATE $8 AS airline, $14 AS arrivaldelay;
filtered = FILTER flight_limit BY arrivaldelay >= 10;
grouped = GROUP filtered BY airline;
count = FOREACH grouped GENERATE group,COUNT(filtered);
dump count;
```

Total flights to calculate proportion. Filtered to remove 10 header rows from results using ! = UniqueCarrier in the code.

```
flights = LOAD 'all_years.csv' USING PigStorage(',');
flight_limit = FOREACH flights GENERATE $8 AS airline, $0 AS year;
filtered = FILTER flight_limit By airline != 'UniqueCarrier';
total_grouped = GROUP filtered BY airline;
total_flights = FOREACH total_grouped GENERATE group,COUNT(filtered);
dump total_flights;
```

Question 8

Using Pig Script

Iterate line by line and pull out Day, Arrival Delay and Departure Delay, filter to get flights where the arrival delay is less than or equal to zero and then group by airline. Count the number of flights for each day and dump result to screen. Run in pig map reduce using the Grunt Shell pig -x

```
flight = LOAD 'all_years.csv' USING PigStorage(',');
flight_limit = FOREACH flight GENERATE $3 AS day, $14 AS arrivaldelay, $15 AS departdelay;
filtered = FILTER flight_limit BY arrivaldelay <= 0;
grouped = GROUP filtered BY day;
count = FOREACH grouped GENERATE group,COUNT(filtered);
dump count;
```

Total flights to calculate proportion. Filtered to remove 10 header rows from results using ! = DayofWeek in the code.

```
flight = LOAD 'all_years.csv' USING PigStorage(',');
flight_limit = FOREACH flight GENERATE $3 AS day, $0 AS year;
filtered = FILTER flight_limit BY day != 'DayofWeek';
total_grouped = GROUP filtered BY day;
total_flights = FOREACH total_grouped GENERATE group,COUNT(filtered);
dump total_flights;
```

Best Days of the Week to Fly

Saturday reigns supreme in this category. According to data analysis, it's clear that there are fewer delays on Saturday than any other day of the week.

This may seem counterintuitive to some people. After all, which day is likely going to be the most crowded at a museum, at a park, or on the beach? Saturday typically takes that dubious honor. As a result, many people have avoided booking Saturday flights because they assume the crowds at the airport would be the same.

In reality, **Friday is the worst day to travel with 29.47% of flights being delayed on average.** It's also the day with the most flights scheduled, meaning there is more opportunity for delays to begin cascading throughout the airport, until one finally affects your flight.

In order from least busy to most busy, here are the typical delay percentages in airports:

1. Saturday - 18.11%
2. Thursday - 21.69%
3. Wednesday - 23.59%
4. Sunday - 24.13%
5. Monday - 25.69%
6. Tuesday - 26.60%
7. Friday - 29.75%

(Medina, 2019)

Question 9

Hive was not working on the University Cloudera Distribution, so a Virtual Box Sandbox was loaded onto my laptop, in order to set up my own working account. HUE could not be used due to a file loading problem on my machine, so the command line was used instead.

Loading file into HDFS

```
[cloudera@quickstart ~]$ hdfs dfs -put all_years.csv /user/cloudera
```

Creating table in Hive

```
CREATE TABLE IF NOT EXISTS flightdata2 (
years INT,
months INT,
daymonth INT,
dayweek INT,
deptime BIGINT,
crsdepttime BIGINT,
arrtime BIGINT,
crsarrtime BIGINT,
carrier STRING,
flightno BIGINT,
tailno STRING,
actelap BIGINT,
crselap BIGINT,
airtime BIGINT,
arrdelay BIGINT,
depdelay BIGINT,
origin STRING,
dest STRING,
dist BIGINT,
taxin BIGINT,
taxout BIGINT,
canc BIGINT,
cancode STRING,
divert BIGINT,
carrdelay BIGINT,
weathdelay BIGINT,
nasdelay BIGINT,
securdelay BIGINT,
aircftdelay BIGINT)

ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;
```

A further line had to be added to remove the header row as it kept overwriting the new schema. Code used was:

```
tblproperties ("skip.header.line.count"="1");
```

Loading data into the table

LOAD DATA INPATH '/user/cloudera/all_years.csv' INTO TABLE flightdata2;

SELECT months, SUM(nasdelay)
FROM flightdata11
WHERE year > 2004
GROUP BY months;

Question 10

Any other areas of interest

The total minutes and count were determined for each of the delay categories in order to compare using the script above but putting in each category in turn.

SELECT months, SUM(*carrdelay*)
FROM flightdata11
WHERE years >2004 AND *carrdelay* IS NOT NULL AND *carrdelay* != 0
GROUP BY months;

SELECT months, COUNT(*carrdelay*)
FROM flightdata11
WHERE years >2004 AND *carrdelay* IS NOT NULL AND *carrdelay* != 0
GROUP BY months;

Appendix 4 contains the detail results from running these queries.

Security Delays Taking A Closer Look

SELECT years, SUM(securdelay)
FROM flightdata11
WHERE years >2004
GROUP BY years;

```
2005    141045
2006    221371
2007    176906
2008    114316
```

SELECT months, SUM(securdelay)
FROM flightdata11
WHERE years = 2006
GROUP BY months;

```
1     8056
2     22340
3     18280
4     23017
5     7559
6     11849
7     11401
8     64327
9     13413
10    11375
11    10087
12    19667
```

SELECT daymonth, SUM(securdelay)
FROM flightdata11
WHERE years = 2006, months = 8
GROUP by daymonth;

```
25      538
1       464
26      422
2       107
27      739
3       217
28      1000
4       228
29      323
5       403
30      141
6       396
31      244
7       512
8       97
9       366
10      43906
11      3926
12      1793
13      1644
14      1255
15      498
16      483
17      593
18      689
19      874
20      816
21      391
22      398
23      259
24      605
```

Question 11

How many flights left in 2007 on each day of the week respectively?

```
[cloudera@quickstart ~]$ cat '2007.csv'|python quest10_mapper.py|python quest10_reducer.py
(1090187, 1052983, 1062432, 1072592, 1077472, 917348, 1019453)
```

```python
#!/usr/bin/env python

import sys
for line in sys.stdin:

    line1 = line.replace('"','')
    flight = line1.split(",")

    if flight[21] == '1':
        flew = flight[21]
        if flew.isdigit:
            day = flight[3]
            print(day,flew)
```

```python
#!/usr/bin/env python
import sys

mon = 0
tues = 0
wed = 0
thurs = 0
frid = 0
sat = 0
sun = 0


for line in sys.stdin:
    line3 = line.replace("'","")
    line2 = line3.replace("(","")
    line1 = line2.replace(")","")

    flights = line1.split(',')

    if flights[0] == '1':
        mon = mon + 1
    elif flights[0] == '2':
        tues = tues + 1
    elif flights[0] == '3':
        wed = wed + 1
    elif flights[0] == '4':
        thurs = thurs + 1
    elif flights[0] == '5':
        frid = frid + 1
    elif flights[0] == '6':
        sat = sat + 1
    elif flights[0] == '7':
        sun = sun + 1


print('mon',mon,'tues', tues,'wed', wed,'thurs', thurs, 'frid', frid,'sat', sat,'sun', sun)
```

Appendix 4

AIRCRAFT DELAY STATISTICS BY TYPE OF DELAY FROM 2004 ONWARDS

Month	NAS	Av/year	Incidents *	min/delay **	Carrier Delay	Average/yr	Incidents *	min/delay **	Weather	Av.year	Incidents *	min/delay **
January	8,836,052	2,209,013	328,918	26.86	8,340,044	2,085,011	241,287	34.56	1,887,668	471,917	44,408	42.51
February	8,461,705	2,115,426	325,865	25.97	8,167,515	2,041,879	237,699	34.36	1,840,315	460,079	42,274	43.53
March	9,090,563	2,272,641	339,200	26.80	8,820,884	2,205,221	264,804	33.31	1,590,550	397,638	37,368	42.56
April	7,208,456	1,802,114	284,183	25.37	7,028,633	1,757,158	211,191	33.28	1,102,785	275,696	25,825	42.70
May	7,374,299	1,843,575	283,687	25.99	6,753,854	1,688,464	205,327	32.89	1,202,400	300,600	28,832	41.70
June	11,248,884	2,812,221	379,099	29.67	10,028,531	2,507,133	287,676	34.86	2,478,242	619,561	52,713	47.01
July	10,938,369	2,734,592	364,231	30.03	10,541,221	2,635,305	292,645	36.02	2,533,151	633,288	54,418	46.55
August	9,019,703	2,254,926	326,689	27.61	9,407,777	2,351,944	260,730	36.08	1,906,312	476,578	42,099	45.28
September	6,228,476	1,557,119	233,353	26.69	6,155,335	1,538,834	166,458	36.98	1,009,937	252,484	21,847	46.23
October	7,658,133	1,914,533	281,633	27.19	6,563,696	1,640,924	192,809	34.04	1,119,570	279,893	25,988	43.08
November	7,274,668	1,818,667	264,120	27.54	6,154,509	1,538,627	183,666	33.51	935,747	233,937	22,975	40.73
December	10,689,851	2,672,463	400,396	26.70	10,591,826	2,647,957	318,236	33.28	2,334,322	583,581	54,218	43.05
Total	104,029,159	26,007,290	3,811,374	27.29	98,553,825	24,638,456	2,862,528	34.43	19,940,999	4,985,250	452,965	44.02

Month	Aircraft	Av/year	Incidents *	min/delay **	Security	Av/year	Incidents *	min/delay **	Total	Av.year	Incidents *	min/delay **
January	10,587,031	2,646,758	243,563	43.47	46,244	11,561	2,535	18.24	29,697,039	7,424,260	860,711	34.50
February	10,781,782	2,695,446	246,101	43.81	55,134	13,784	2,679	20.58	29,306,451	7,326,613	854,618	34.29
March	11,690,466	2,922,617	265,589	44.02	69,693	17,423	3,160	22.05	31,262,156	7,815,539	910,121	34.35
April	8,730,025	2,182,506	209,277	41.72	53,767	13,442	2,302	23.36	24,123,666	6,030,917	732,778	32.92
May	8,665,673	2,166,418	209,213	41.42	34,325	8,581	1,997	17.19	24,030,551	6,007,638	729,056	32.96
June	14,143,769	3,535,942	301,897	46.85	53,167	13,292	2,808	18.93	37,952,593	9,488,148	1,024,193	37.06
July	14,690,370	3,672,593	302,168	48.62	61,520	15,380	2,964	20.76	38,764,631	9,691,158	1,016,426	38.14
August	11,944,959	2,986,240	262,838	45.45	109,970	27,493	5,281	20.82	32,388,721	8,097,180	897,637	36.08
September	6,577,677	1,644,419	158,429	41.52	35,798	8,950	1,661	21.55	20,007,223	5,001,806	581,748	34.39
October	8,122,098	2,030,525	196,056	41.43	37,432	9,358	1,929	19.40	23,500,929	5,875,232	698,415	33.65
November	7,522,609	1,880,652	181,313	41.49	28,247	7,062	1,569	18.00	21,915,780	5,478,945	653,643	33.53
December	14,622,148	3,655,537	323,841	45.15	68,341	17,085	3,886	17.59	38,306,488	9,576,622	1,100,577	34.81
Total	128,078,607	32,019,652	2,900,285	44.16	653,638	163,410	32,771	19.95	351,256,228	87,814,057	10,059,923	34.92

Source: Aircraft Delay Statistics Provided By Ali Copeland (University of Chichester)