Christopher Truong cvt7bm
Cloud PA4

a. The airline with the best on-time performance and the airline with the worst on-time performance

```
ubuntu@ip-172-31-46-175:~$ hdfs dfs -cat Airline1_out/part-00000 | sort -k2,2 -n
2019-10-31 01:49:25,088 INFO sasl.SaslDataTransferClient: SASL encryption trust chec
k: localHostTrusted = false, remoteHostTrusted = false
HA 0.16634980988593157
DL 0.3277021845351563
OO 0.34233657450015337
9E 0.34968720541605963
WN 0.35271098295659337
MQ 0.38929537937255687
AS 0.40444750486066106
YX 0.4148705038452254
YV 0.4328773936831634
G4 0.43971549845130203
AA 0.4522125071049354
UA 0.45225263682122974
OH 0.4545811724029546
EV 0.5038773352132534
NK 0.5134146341463415
F9 0.545527541352176
B6 0.6280664628889412
ubuntu@ip-172-31-46-175:~$
```

HA had the best on-time performance, with B6 having the worst on-time performance.
This is calculated as number of delays over 15 min divided by the number of total flights, with a delay on departure and arrival counting for 2 delays.

Christopher Truong cvt7bm
Cloud PA4

b.

```
ubuntu@ip-172-31-46-175:~$ hdfs dfs -cat Airline2_out/part-00000 | sort -k4,4 -n | t
ail -n 50
2019-10-31 00:08:15,922 INFO sasl.SaslDataTransferClient: SASL encryption trust chec
k: localHostTrusted = false, remoteHostTrusted = false
('F9', 'JAX', 'RDU')    54.84615384615385
('B6', 'MCO', 'PSE')    54.95348837209303
('B6', 'FLL', 'PHL')    55.193548387096776
('B6', 'PVD', 'FLL')    55.225806451612904
('B6', 'RSW', 'EWR')    55.903225806451616
('F9', 'MKE', 'PHL')    56.46153846153846
('NK', 'DEN', 'BWI')    56.96774193548387
('B6', 'BDL', 'TPA')    58.03225806451613
('B6', 'MCO', 'EWR')    58.13259668508287
('B6', 'ALB', 'FLL')    58.16129032258065
('B6', 'HOU', 'JFK')    58.58064516129032
('NK', 'FLL', 'PIT')    59.0
('B6', 'FLL', 'JFK')    59.5
('F9', 'BIL', 'DEN')    60.0
('B6', 'CLT', 'JFK')    60.29032258064516
('B6', 'ACK', 'LGA')    60.333333333333336
('B6', 'SJC', 'JFK')    60.38709677419355
('NK', 'MYR', 'EWR')    61.096774193548384
('B6', 'HPN', 'FLL')    62.95161290322581
('B6', 'MCO', 'SWF')    62.96774193548387
('B6', 'FLL', 'CLE')    63.03448275862069
('B6', 'ORD', 'JFK')    63.435483870967744
('F9', 'TTN', 'RSW')    64.38461538461539
('B6', 'SWF', 'MCO')    65.06451612903226
('NK', 'MCO', 'ORD')    66.03225806451613
('B6', 'EWR', 'MCO')    66.71270718232044
('B6', 'FLL', 'EWR')    66.89247311827957
('NK', 'BWI', 'DEN')    67.58064516129032
('NK', 'DTW', 'LGA')    68.06451612903226
```

Christopher Truong cvt7bm
Cloud PA4

```
('NK', 'DTW', 'LGA')    68.06451612903226
('F9', 'MEM', 'MCO')    69.375
('B6', 'FLL', 'RIC')    69.54838709677419
('B6', 'ORH', 'FLL')    69.96774193548387
('B6', 'JFK', 'BQN')    71.16129032258064
('F9', 'SAT', 'IAD')    71.76923076923077
('B6', 'FLL', 'BNA')    73.03225806451613
('F9', 'ONT', 'SAT')    73.846153846153840
('F9', 'JAX', 'LAS')    74.05555555555556
('B6', 'MCO', 'AUS')    76.25806451612904
('NK', 'BWI', 'DTW')    76.29032258064517
('B6', 'BNA', 'FLL')    76.35483870967742
('B6', 'BQN', 'JFK')    77.09677419354838
('F9', 'MCO', 'MEM')    78.0
('NK', 'MYR', 'ACY')    78.25806451612904
('B6', 'FLL', 'CHS')    79.16129032258064
('NK', 'LGA', 'MYR')    82.53225806451613
('F9', 'TTN', 'BNA')    88.3076923076923
('F9', 'BNA', 'TTN')    89.3076923076923
('F9', 'PNS', 'ORD')    92.33333333333333
('NK', 'AVL', 'TPA')    93.23076923076923
('F9', 'PHL', 'MCI')    96.6923076923077
ubuntu@ip-172-31-46-175:~$ ▯
```

Here we define the worst ones as the ones that are the most late on average for that particular route.

B6
1. FLL → CHS
2. BQN → JFK
3. BNA → FLL
4. MCO → AUS
5. FLL → BNA
6. JFK → BQN
7. ORH → FLL
8. FLL → RIC
9. FLL → EWR
10. EWR → MCO

F9
1. PHL → MCI
2. PNS →ORD
3. BNA → TTN
4. TTN → BNA
5. MCO → MEM
6. JAX → LAS
7. ONT → SAT
8. SAT → IAD

Christopher Truong cvt7bm
Cloud PA4

9. MEM → MCO
10. TTN → RSW

NK
1. AVL → TPA
2. LGA → MYR
3. MYR → ACY
4. BWI → DTW
5. DTW → LGA
6. BWI → DEN
7. MCO → ORD
8. MYR → EWR
9. FLL → PIT
10. DEN → BWI

c. The most common reason for cancellations was code B, for weather.

```
ubuntu@ip-172-31-46-175:~$ hdfs dfs -cat Airline3_out/part-00000 | sort -k2,2 -n
2019-10-31 01:56:28,308 INFO sasl.SaslDataTransferClient: SASL encryption trust chec
k: localHostTrusted = false, remoteHostTrusted = false
"CRSElapsedTime"        1
"D"     2
"A"     3164
"C"     3613
"B"     4519
""      647163
ubuntu@ip-172-31-46-175:~$
```