Web Scrapping

Web scrapping to get airline schedule and status of following airlines:

a) Yeti Airlines

b) Buddha Airlines

c) Tara Airlines

d) Simrik Airlines

e) Shree Airlines

f) Shree Airlines

g) Sita Airlines

h) Saurya Airlines

i) NAC Airlines

Server: 10.13.189.55

Database:

**airlines\_schedule\_status(temp)** for temporary storage of airline schedule one for each airline.

**airlines\_schedule\_status\_info** for combined storage of airline schedule and status.

Discussion 3/5/2020:

1. Ensure schedule run for airlines deemed possible for table extraction and dump in database on daily basis.
2. Explore data extraction for airlines currently not possible to extract using traditional webscrapping techniques

Web scrapping has been completed for status and schedule of following airlines:

a) Flight Schedule (Yeti, Tara, Simrik, Sita, NAC,TIA)

b) Flight Status (Yeti, Buddha)

Web scrapping for flight schedule and flight status of remaining airlines were not performed because of unavailability of corresponding information in official website of these airlines.

 For Flight Schedule of Buddha, though information is available in its official website, web scrapping could not be performed because of different approach being used(soap). Web scrapping for this needs to be explored.

Tables included in airlines\_schedule\_status are intermediate tables that contains schedule and status of different airlines (one for each airline).

Tables included in airlines\_schedule\_status\_info are final tables that contains airline schedule and status information in a single table (one for schedule, one for status).

Need for creating intermediate table was  format of information obtained from different airlines is different. Final table contains all information in same same format.

Python script is ready.

I tried to run this script thrice a day by using crontab in 55 server.

This could not be completed as respective sites were blocked in this server.

Script will run as per schedule once these sites are whitelisted.

Web Scrapping for flight schedule for Buddha airline is successfully completed.