Flight Data Analysis Using Big Data Tools

ANVESH NAYAN LINGAMPALLI[1,*]

[1] *School of Informatics and Computing, Bloomington, IN 47408, U.S.A.*
[*] *Corresponding authors:anveling@indiana.edu*

project-S17-IR-2016, April 16, 2017

Analysis of flight data provides insights on the United States of America's Airline data. The On-time performance of flights operated by large air carriers are tracked and made as a report, Air Travel Consumer Report, which is a big data set. Hive component of Hadoop ecosystem, is utilized to process the big data in distributed environment. Efficient accessing and processing of the user queries is acheived by this analysis on flight data.

Keywords: Apache, Hive, Ansible, Pig

https://github.com/cloudmesh/classes/blob/master/project/S17-IR-2016/report/report.pdf

1. INTRODUCTION

Aviation industry manages enormous amount of data, which consists of the information regarding the delayed, cancelled, diverted or on-time flights by large air-carriers[1]. This statistics is publicly available in the Air Travel Consumer Report. Big Data analysis of this data will provide a consistent understanding and importance of the given data. With 35 million flight departures per year, data is critically important for any planning decision made by airlines and airports. The results of analysis has benefits which can help airline operations to predict and reduce redundance[2].

2. MILESTONES

- Performing Analysis on local VM
  - Loading Data into HDFS
  - Pre-processing of the data using Pig
  - This data into Hive tables
- Analysis on the distributed cloud environment
- Visualizing the results using Tableau
- Benchmarking
- Final update with report

3. TECHNOLOGIES

- Distributed Computation and Storage:- HDFS, Hive and Pig
- Development:- Python and Java
- Deployment:- Ansible

4. DEPLOYMENT

Ansible Playbook is used as the application and configuration deployment tool. Deploying the Hive and Pig framework into the cluster environment.

5. BENCHMARKING

TBD

REFERENCES

[1] "Aviation analysis," Web Page. [Online]. Available: https://aviationanalytics.com/airport-analytics/data-analysis/

[2] "Big data in aviation," Web page. [Online]. Available: http://apex.aero/2016/11/30/big-data-aviation-industry-case-becoming-data-driven