**Air Traffic scaling, delay- analysis and prediction using PySpark on Hadoop Cluster**

Scaling, Simulation, delay analysis and delay prediction of 2019 US domestic air flights

Mohammed Zoher Guniem Megh Raj Upreti

University of Stavanger, Norway University of Stavanger, Norway

m.guniem@stud.uis.no mr.upreti@stud.uis.no

# ABSTRACT

As more commercial airplanes take off to the sky, data generated from air traffic is rapidly increasing in both size and complexity. According to the International Air Transport Association (IATA), Air travel industry is anticipated to annually grow by 3.5% in the next two decades. This growth raises big questions about how to gather, store, analyze and use the data generated from the those flying cities in the sky.

# KEYWORDS

Hadoop, PySpark, Power BI, Python, Aviation, Scaling, delay, Prediction, Air traffic.

# 1 INTRODUCTION