



Motivation

- 👉 **We need to test and benchmark our systems**
 - 👉 Reviewers love good evaluations
- 👉 **Many benchmarks use synthetic/generated data (TPC-DS, TPC-H, Yahoo, etc.)**
 - 👉 Sometimes we want more complex data and realistic applications
- 👉 **There are not a lot of benchmarks out there for stream processing systems using modern hardware**
 - 👉 Some require the modification of your system or side-channel communication

Problem?!

- 👉 **Real and complex data is hard to find**
 - 👉 Companies don't like to share their data

Proposal

- **A benchmark for stream processing systems enabling the measurement of modern hardware (FPGA)**
- **Using real-time aircraft tracking data (ADS-B) combined with community sourced relational data**
 - OpenSky Network, OurAirports, etc.
- **Queries based on real problems and applications in air traffic control**
- **Benchmarking tool to benchmark your SPS**

"OpenSky Network" : <https://opensky-network.org>

"Our Airports" : <https://ourairports.com/>

Aircraft Tracking Data

✦ **Radar is primarily used to survey the air space (Primary Radar)**

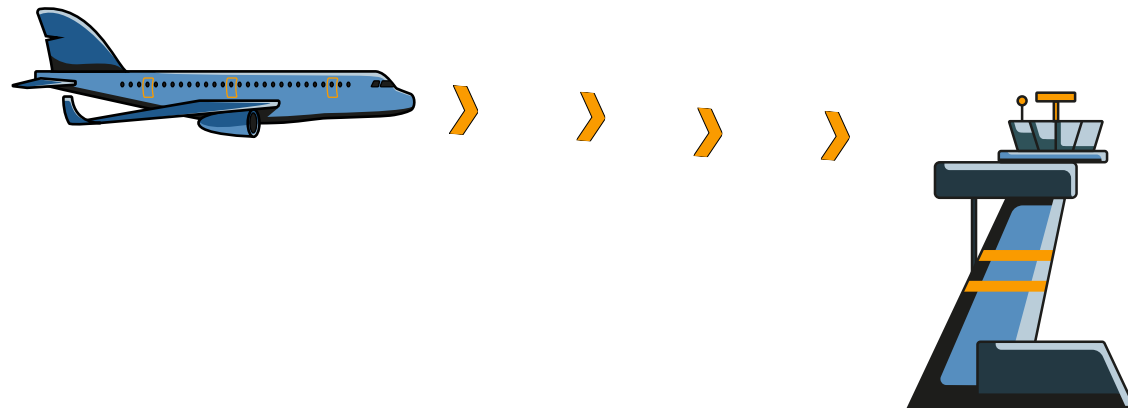
✦ Requires line of sight

✦ **Aircrafts broadcast their current identification, position etc. continuously (ADS-B)**

✦ Unencrypted for anyone to receive

✦ Can be used to track aircrafts across oceans or larger landmasses

✦ Regulated by international law and conventions

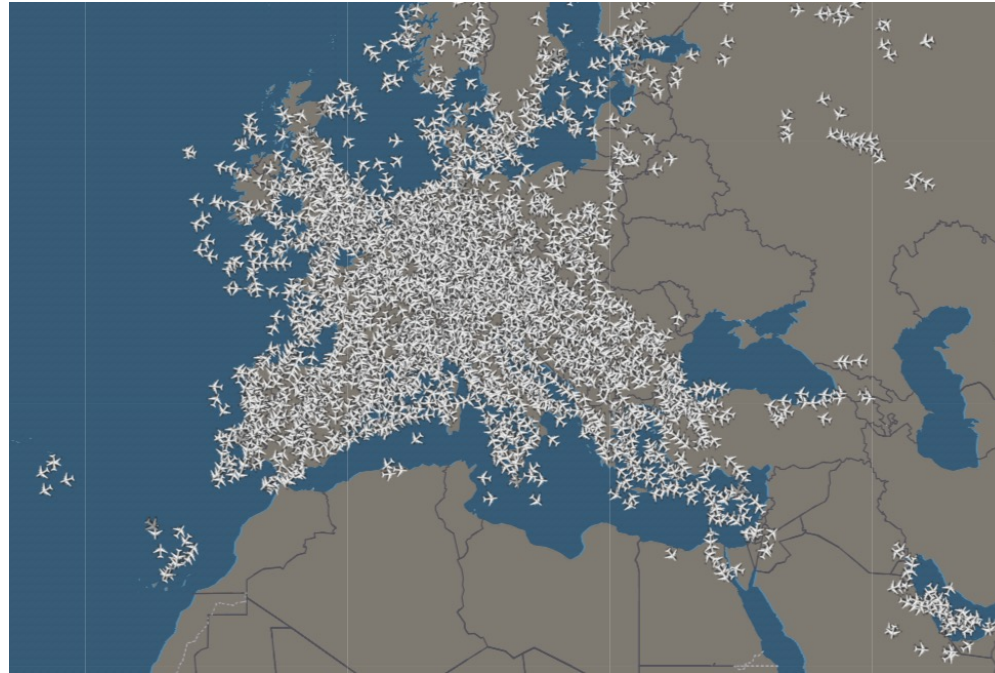


Aircraft Tracking Data

👉 **Over 30.000 flights per day in the EUROCONTROL area**

👉 **Over 45.000 flights per day in the US**

👉 Millions of datapoints that can be collected and used



"OpenSky Network" : <https://opensky-network.org/network/explorer>

Networks



Commercial Services

- ✎ Commercial providers sell ADS-B data for substantial fees
- ✎ Prominent Example: Flightradar24
- ✎ Utilize official data and data from their own receiver networks

Community Projects

- ✎ Huge communities of hobby pilots
- ✎ Platforms like **OpenSky** provide this data **for free**
- ✎ Other platforms provide data on aircrafts, airports and more (e.g. OurAirports)

"Flightradar24": <https://www.flightradar24.com/>

"Bringing Up OpenSky: A Large-scale ADS-B Sensor Network for Research".

Matthias Schäfer, Martin Strohmeier, Vincent Lenders, Ivan Martinovic, Matthias Wilhelm

In Proceedings of the 13th IEEE/ACM International Symposium on Information Processing in Sensor Networks (IPSN), pages 83-94, April 2014.

OpenSky Network

- ✎ Collects and provides ADS-B Messages
- ✎ Developed by researchers for researchers
- ✎ 5000 Receivers around the world
- ✎ Hosts conferences/symposium on air traffic related topics

Accessing the Data

✎ REST API

- ✎ Can be accessed without an account
- ✎ API call limitations

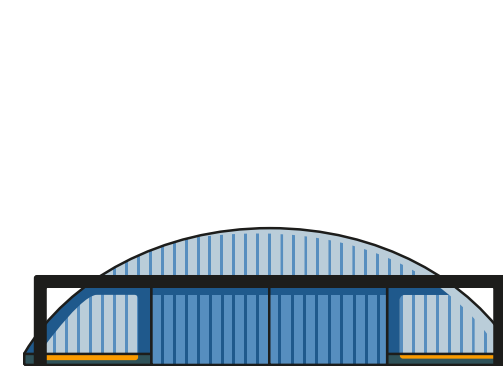
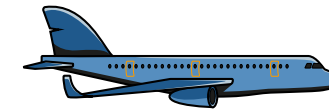
✎ Java, Python packages

✎ Impala (Trini) Shell

- ✎ Requires special permission from OpenSky

✎ Specific Data Sets

- ✎ Just contact OpenSky directly



The Data Set

Streaming Data

Live Tracking Data (States) (ca. 13 GB)

- Identifier (e.g. icao24, callsign)
- Current position (e.g. longitude, latitude)
- Course and speed (e.g. vertical rate)

```
{
  "icao24": "acdfa0",
  "callsign": "DPJ929",
  "origin_country": "United States",
  "time_position": 1687875562,
  "last_contact": 1687875562,
  "longitude": -80.8825,
  ...
}
```

Relational Data

Aircrafts (387,183 Tuples)

- Identifiers (e.g. icao24)
- Type
- Manufacturer

Airports (73,736 Tuples)

- Identifiers (e.g. ident, type, name)
- Position, elevation, size

Flight Schedule (138,091 Tuples)

- icao24, callsign
- Departure airport, destination airport
- Time to destination
- Flight phase

Benchmark Metrics



Latency

- ↳ Can be measured per tuple
- ↳ Only queries with non blocking operations work (filter, projection etc.)
- ↳ We measure end-to-end

Throughput

- ↳ **Input Throughput**
 - ↳ Tuples per second sent to the SPS
- ↳ **Output Throughput**
 - ↳ Tuples per second received from the SPS

Benchmark Queries

- 👉 **14 Queries based on real problems and applications**

- 👉 Based on a master thesis (hobby pilot)

- 👉 **Varying focus and complexity**

- 👉 From simple filter expressions to multi-join

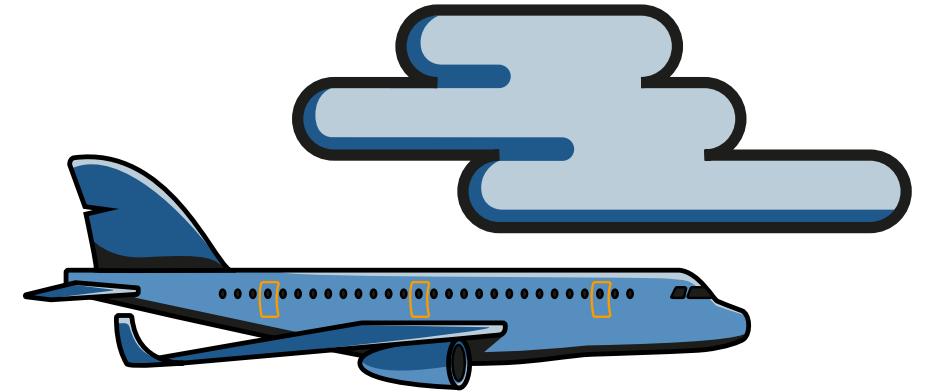
- 👉 **Queries 1-9 allow tuple wise latency measurement**

- 👉 Simple/Complex filter expressions on numerical values and strings

- 👉 Projections containing complex algorithmic expressions (suitable for FPGAs)

- 👉 **Queries 10-14 use blocking operations (e.g. windows, aggregations)**

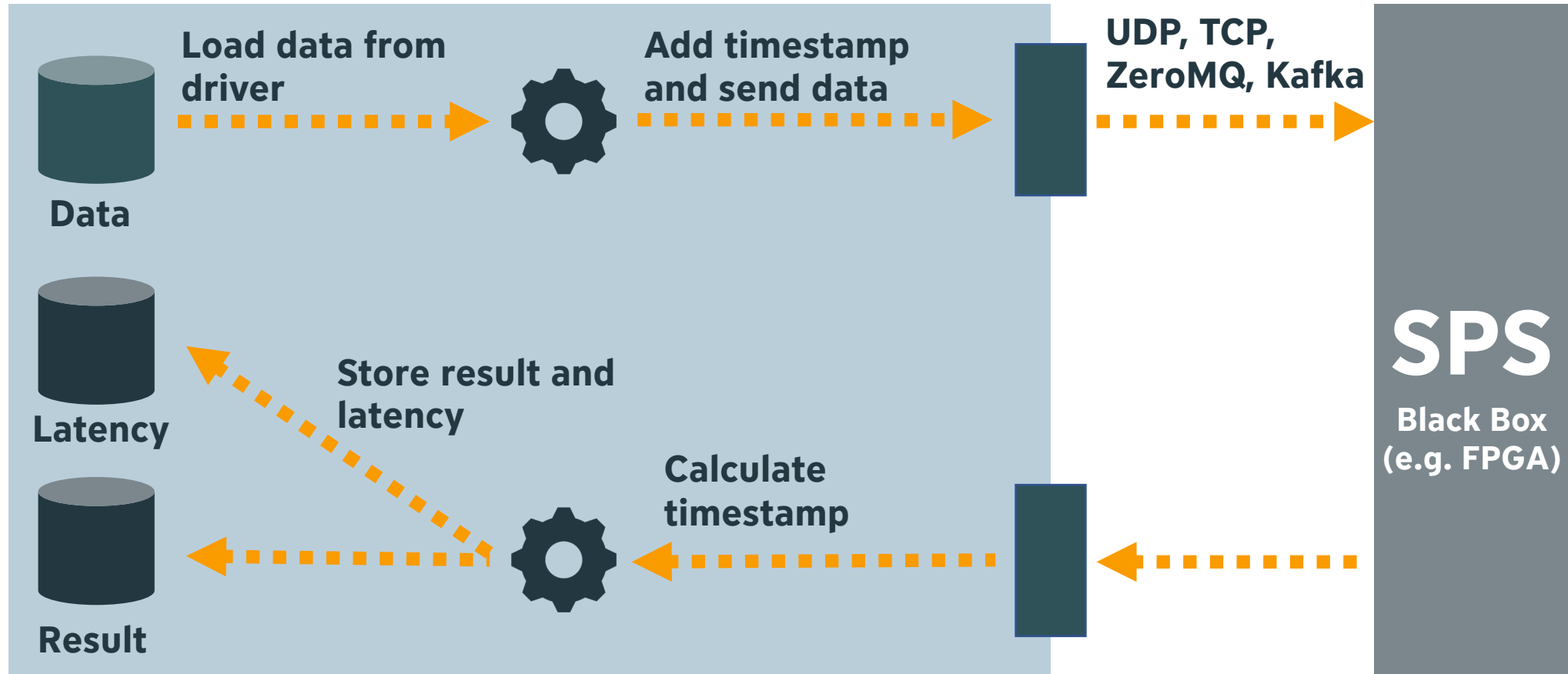
- 👉 Static joins, window expressions containing aggregations



“Development and Implementation of a Database-Benchmark Using Real-Time Flight Data (ADS-B) and Flight Schedules ”.

Tim Vogler, Master Thesis, FAU, 2023

Benchmark Tool



Future Work

✎ **Explore more data sources**

- ✎ Mainly OpenSky Network and OurAirports

✎ **Explore more use-cases for the data set**

- ✎ Relational databases
- ✎ Time-series analysis

✎ **Maintain the project and incorporate feedback from the community**

- ✎ Let us know your thoughts and ideas

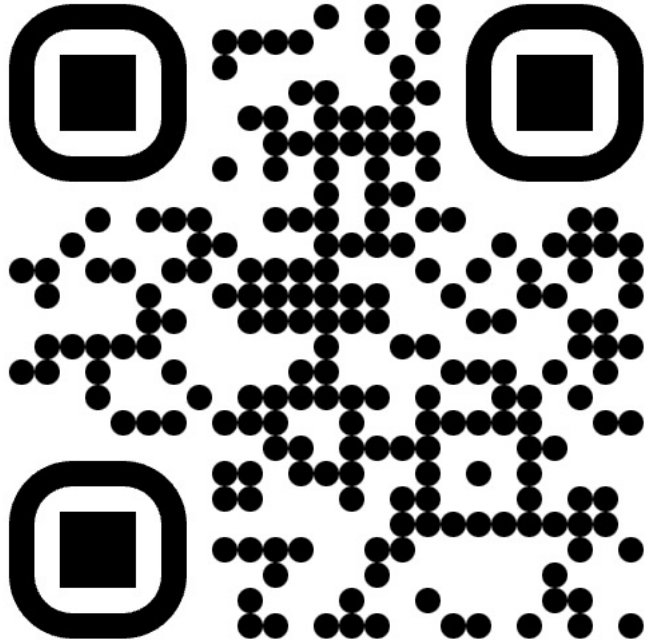


QUESTIONS ?

Follow Our Project



SKYSHARK.ORG



GITHUB



TEAM: Maximilian Langohr, Tim Vogler, Klaus Meyer-Wegener
ARTWORK: Acelya Aksu

SKYSHARK