

# SKYSHARK: A Benchmark with Real-world Data for Line-rate Stream Processing with FPGAs

Maximilian Langohr, Tim Vogler, Klaus Meyer-Wegener

## OUR MISSION STATEMENT

- A benchmark for stream processing systems enabling the measurement of modern hardware (FPGA)
- Using real-time aircraft tracking data combined with community sourced relational data
- Queries based on real problem and applications in air traffic control



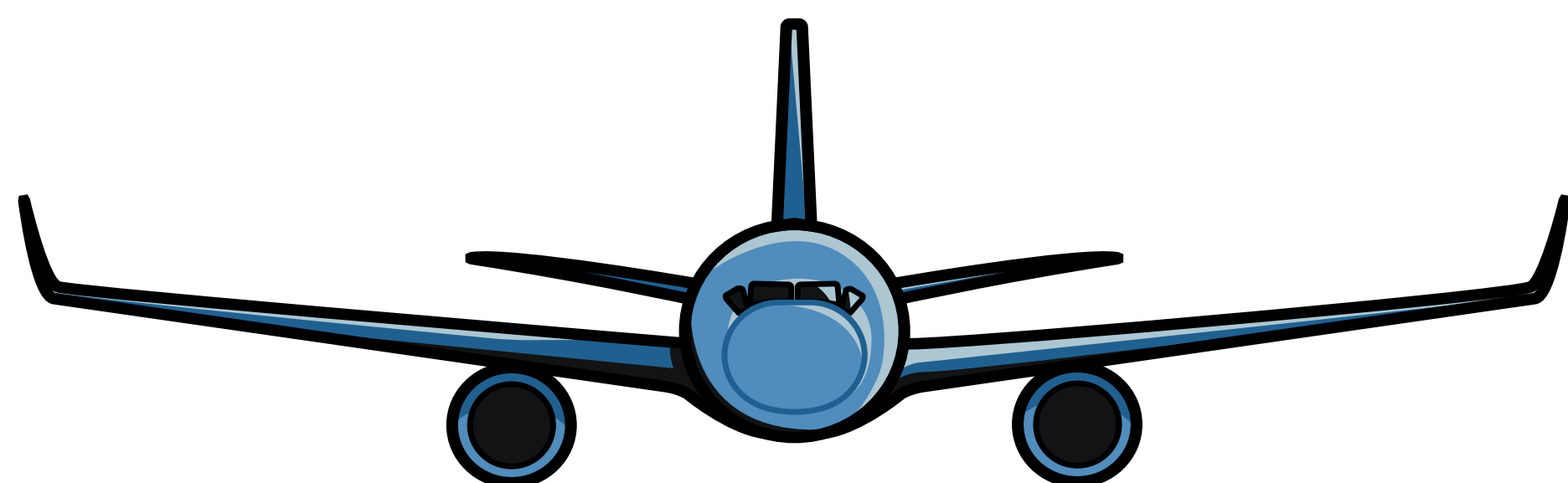
### STREAMING DATA

#### LIVE TRACKING DATA (States) ~13GB

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

#### OPENSKY NETWORK

- Collects tracking data (ADS-B) using a large community
- Provides free access to the collected data using REST API and Trini Shell
- Developed by researchers for researchers



### RELATIONAL DATA

#### AIRCRAFTS (387,183 tuples)

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

#### AIRPORTS (73,736 tuples)

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

#### FLIGHT SCHEDULE (128,091 tuples)

- Created from states
- Only approximate flight schedule
- Used for static joins against real-time tracking data



### METRICS

#### THROUGHPUT

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

#### LATENCY

- Can be measured per tuple
- Only for queries with no blocking operation

## THE BENCHMARK

- 14 queries for stream processing systems
- Varying focus and complexity
- Queries 1-9 allow tuple wise latency measurement
- Queries 10-14 use blocking operations (e.g. windows, aggregations)

## TOOLING

### BENCHMARKING TOOL

- Offers multiple adapters to connect your system (TCP, UDP, ZeroMQ, Kafka)

### CREATING A FLIGHT SCHEDULE

- Creates a flight schedule based on the collected states
- Estimates starting and landing time of each aircraft

