Integration of Mobility Data with Weather Information

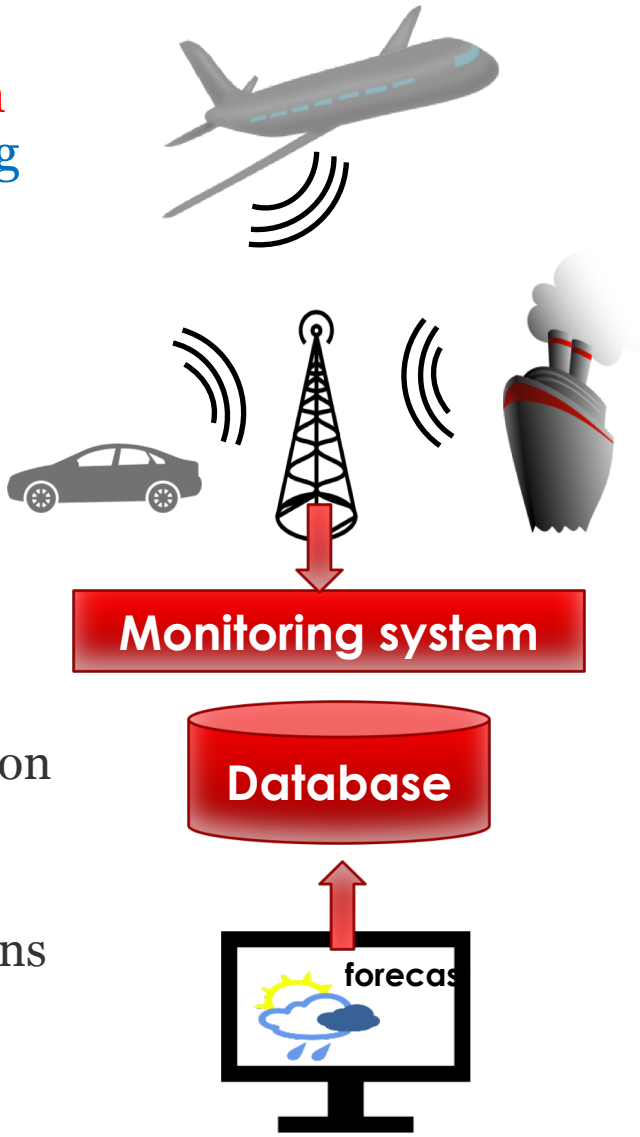
Nikolaos Koutroumanis, Georgios M. Santipantakis, Apostolos Glenis, ***Christos Doulkeridis*** *and George A. Vouros*

Department of Digital Systems
School of Information and Communication Technologies
University of Piraeus, Greece

This work is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 687591 (datAcron), No 780754 (Trqck&Know), No 779747 (BigDataStack) and No 777695 (MASTER)

Motivation

- In several applications, surveillance data from moving objects are collected at ever-increasing rates
  - Fleet management applications
  - Maritime monitoring systems
  - Aircraft surveillance systems and radars
- Challenge:
  - Integrate such dynamic data with weather data
- Why?
  - To facilitate analysis of mobility data in association with weather
  - To discover hidden patterns
  - To improve the understanding of mobility patterns

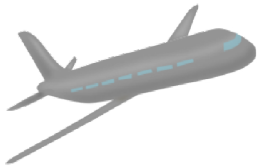


Applications

- Urban domain – fleet management
  - More accurate **fuel estimation**
  - Project Track&Know: https://trackandknowproject.eu/

- Maritime domain
  - More accurate **trajectory prediction**
  - Project BigDataStack: https://bigdatastack.eu/

- Air-traffic management domain
  - Improved **forecasting** of **flight regulations** and **delays**
  - Project datAcron: http://www.datacron-project.eu/



Contributions

- The design and implementation of a **weather integration system**, with several salient features:
  - it works as a standalone and re-usable tool for data integration of mobility data with weather,
  - it is efficient in terms of processing performance, thus making it suitable for application in online scenarios (stream processing)
  - it supports enrichment of complex geometries (e.g., polylines, polygons) with weather data, which is not straightforward

Available for download:
https://github.com/nkoutroumanis/Weather-Integrator

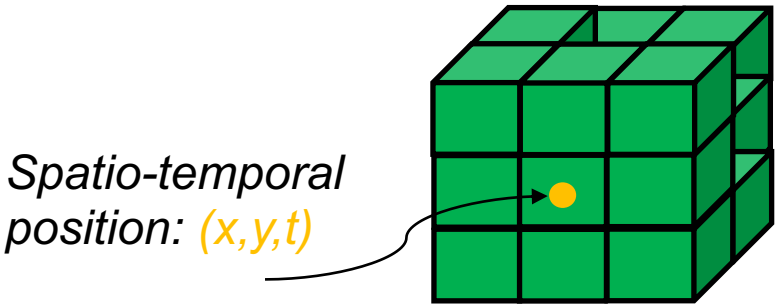
Weather data

- GRIB: a file format for storage and transport of gridded meteorological data
  - The GRIB standard was designed to be self-describing, compact and portable
    - Maintained by the World Meteorological Organization
    - All National Met Services (NMS) use GRIB files to store and exchange forecast data
- Data is provided from Numerical Weather Prediction models (computer-generated)
  - Provided by NOAA (National Oceanic and Atmospheric Administration)

Source: http://weather.mailasail.com/Franks-Weather/Grib-Files-Explained

Weather data

- Internal file representation
  - Values are on a 3-D grid of points
  - Contains dozens of weather attributes (more than 100)
  - Each GRIB file contains weather data for a timespan of 6 hours
    - One day is composed of 4 GRIB files



Weather data

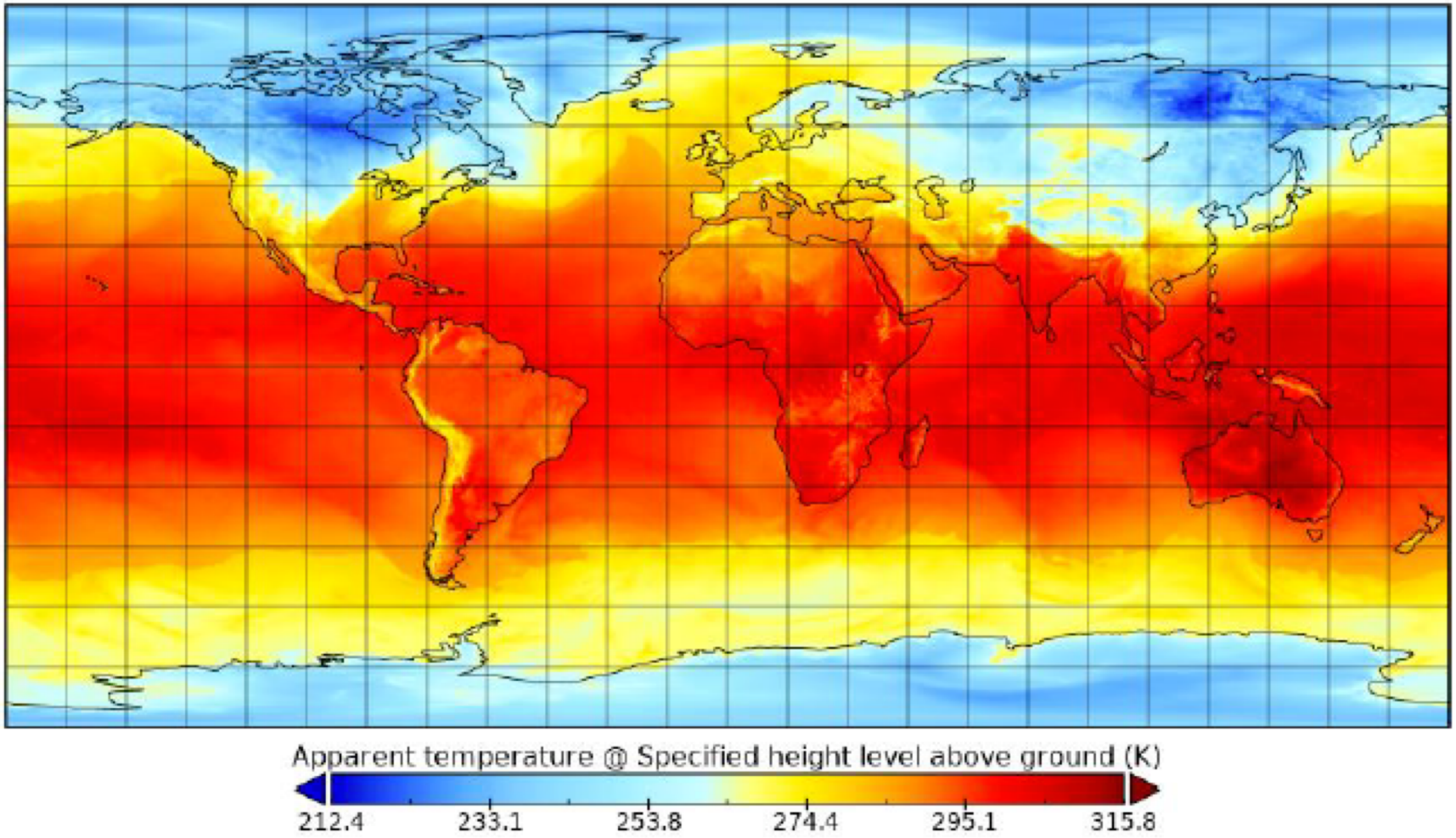
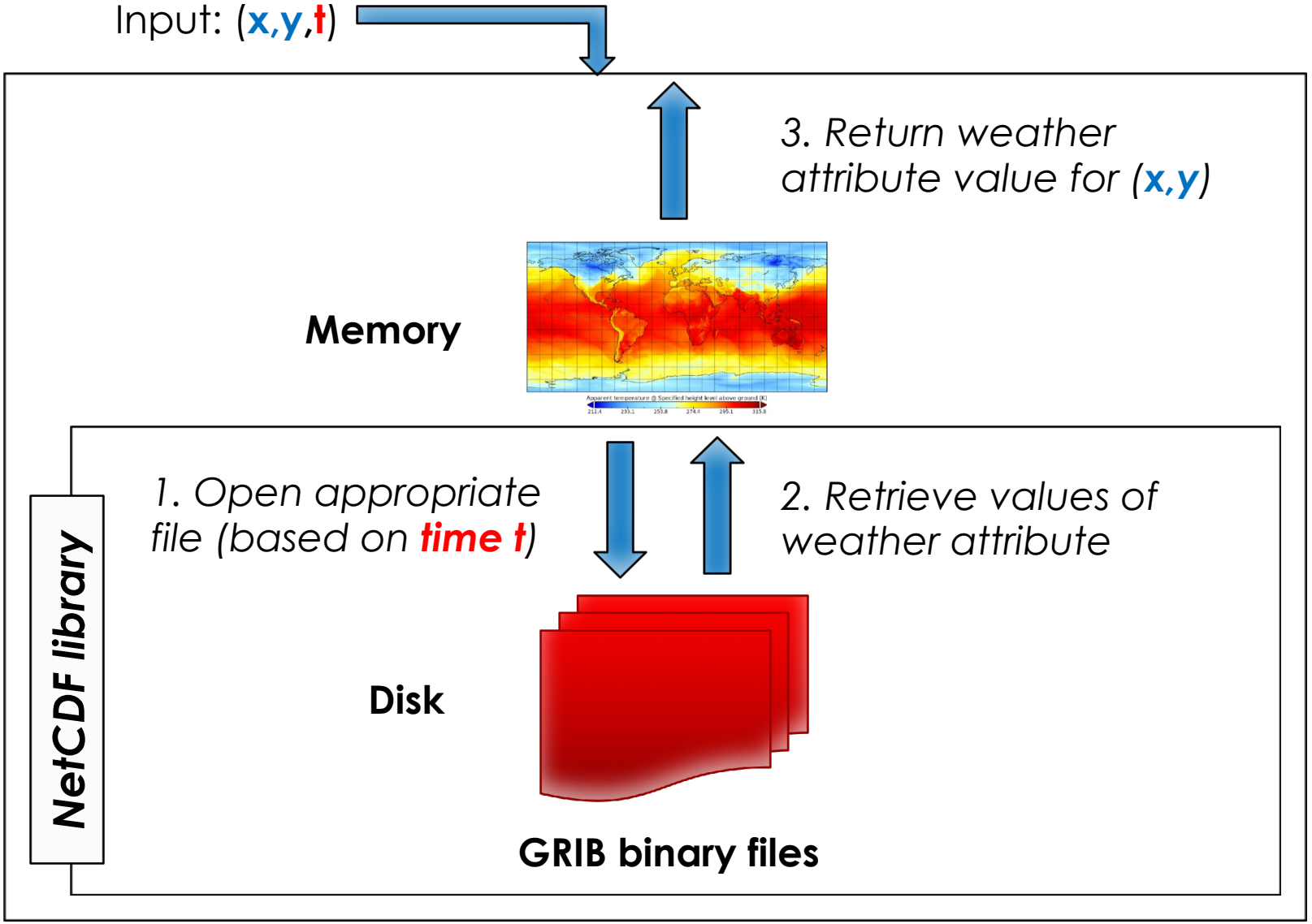
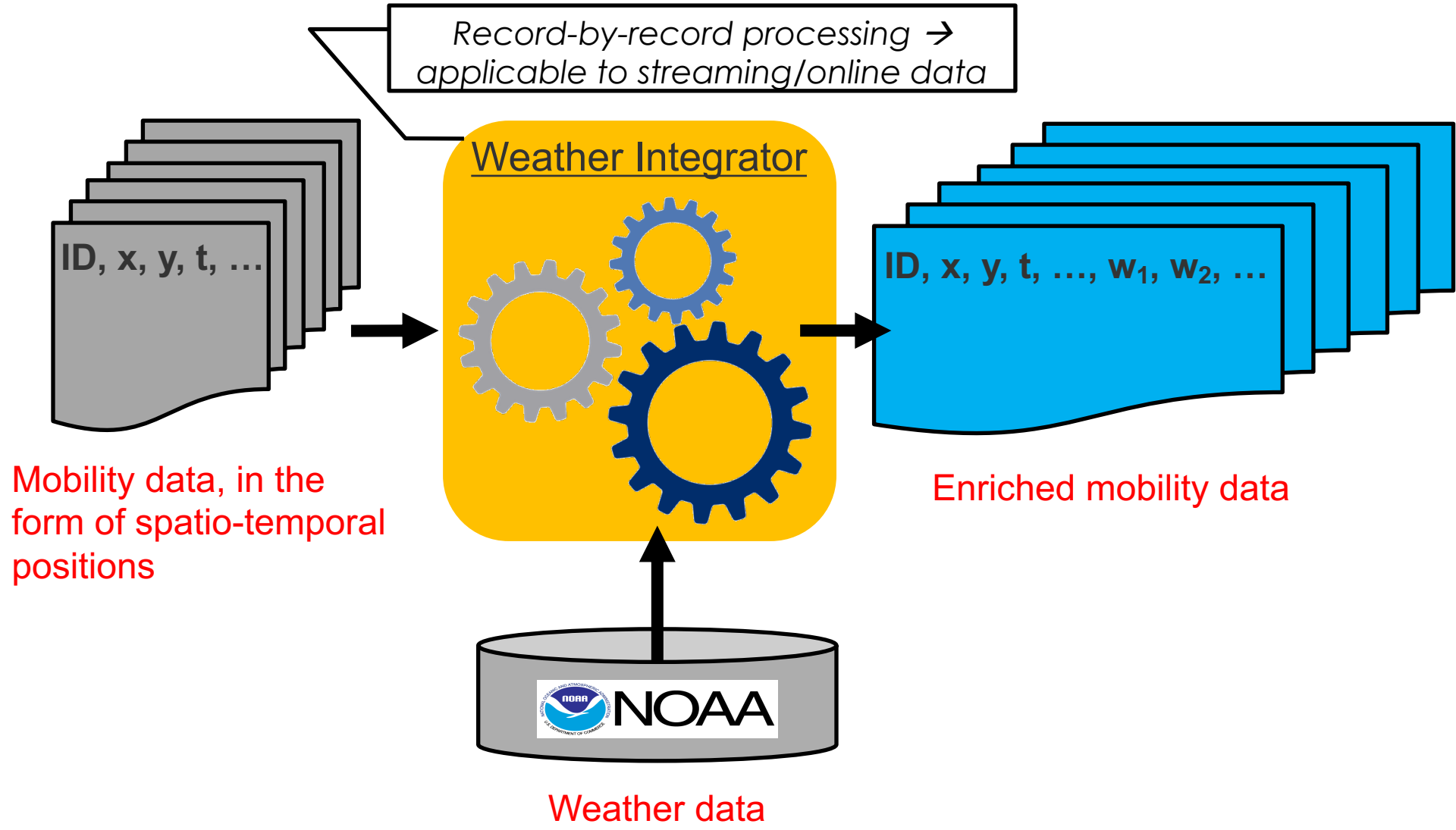


Image captured using Panoply: https://www.giss.nasa.gov/tools/panoply/

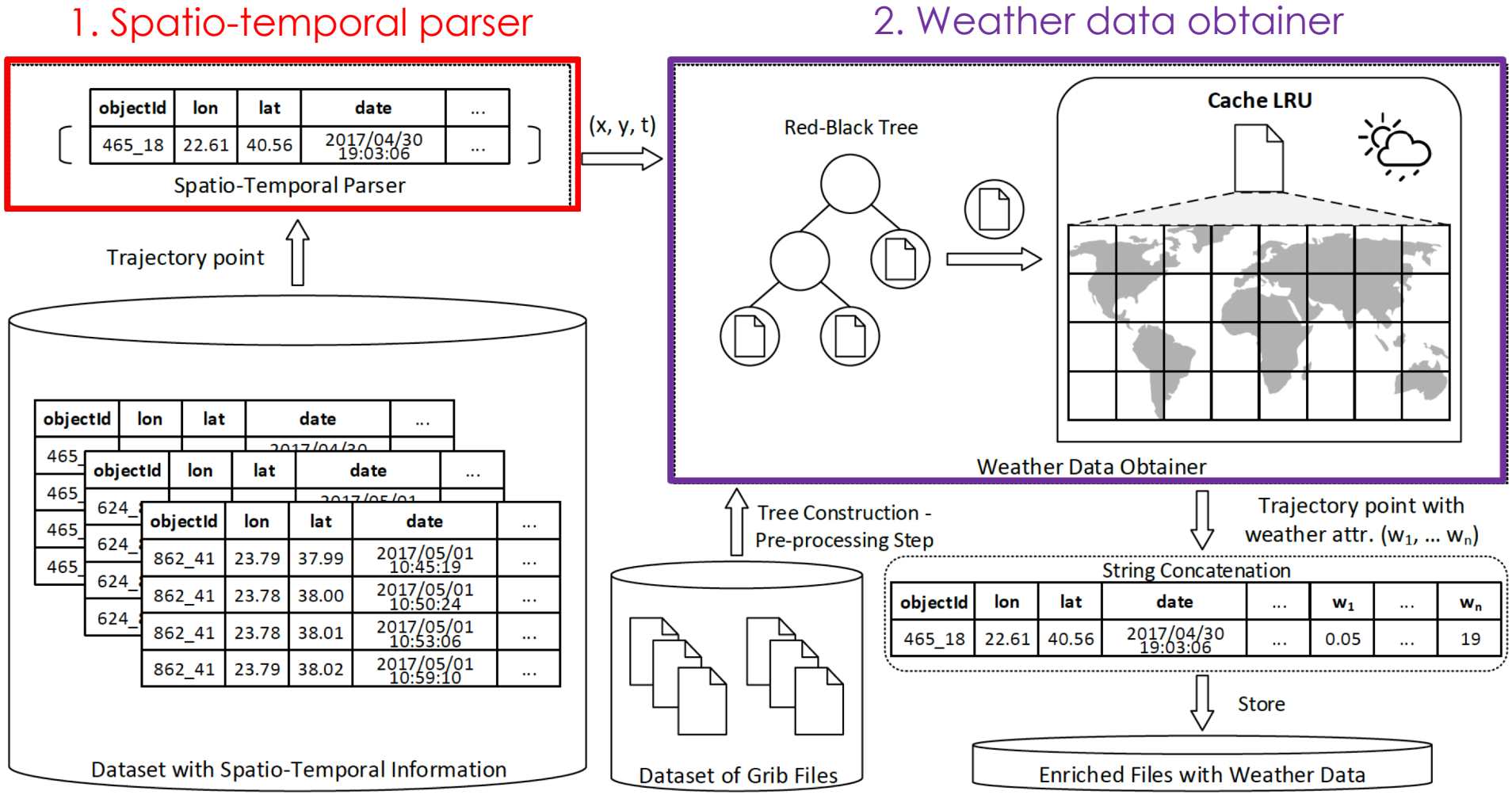
Access to weather data



Architecture (high-level)



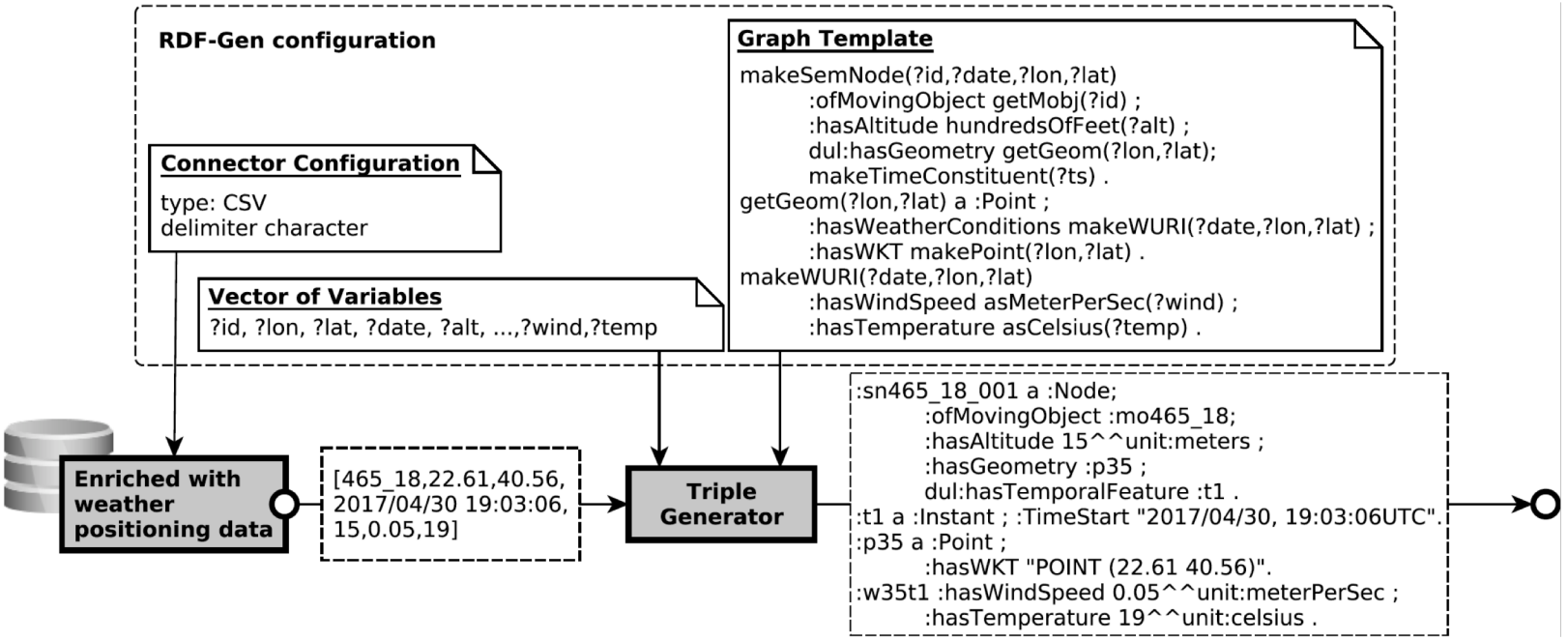
Architecture



Caching

- Aim: avoid the cost of re-opening GRIB files to access data
- This is necessary for applications where positions of moving objects are not temporally sorted
- Maintain a **cache** of references to open files
  - Cache size: user-specified
  - Cache replacement policy: LRU
- Reduces the overhead of repeated open/close operations

Extension #1 – Linked data

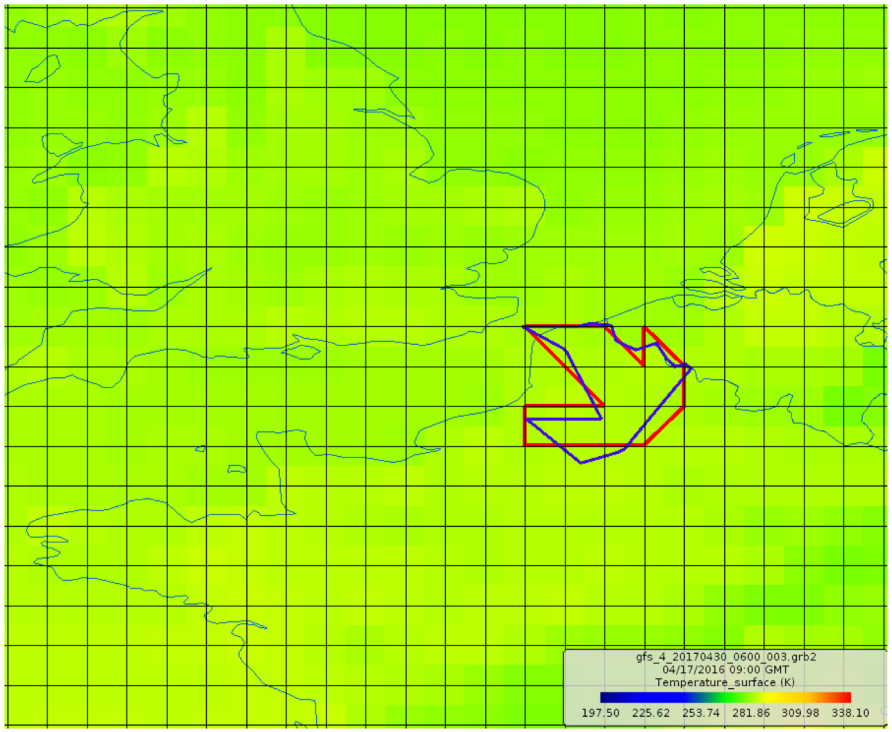


- Ability to output data in RDF format, as Linked Data
- By exploiting the RDF-Gen tool (also developed by our group)

G.M. Santipantakis, K.I. Kotis, G.A. Vouros, C. Doulkeridis: ***"RDF-Gen: Generating RDF from Streaming and Archival Data, in WIMS'18.***

Extension #2 – Complex geometries

- Beyond points
- Supporting complex geometries
  - Trajectory (polyline)
  - 3D Airblock/Sector (polygon)
- Complex geometries may span multiple cells
- Retrieve weather attribute values for points of simplified geometry
- Different kinds of aggregations are then possible (e.g., mean, median, max/min, etc.)



2D projection of an airblock (blue) and simplified geometry (red)

Experiments – Data sets

- Mobility traces (from industrial data provider – fleet management)
  - 25 attributes, CSV format
  - Spatial coverage: Greece
  - Small: 4.5K records, 5.1MB
    - Temporal coverage: Jan.2018
  - Large: 81.5M records, 10GB
    - Temporal coverage: Jul.2017-Jun.2018
- Weather data
  - Binary files: 124 (31*4) files, ~8GB
  - Spatial coverage: the entire world
  - Temporal coverage: Jan.2018

Experimental evaluation

- ***Experimental Setup:***
  - Task:
    - Enrich positions of moving objects with 13 rain-related weather attributes
  - Platform:
    - 3.6GHZ i7-4790 processor, 16GB DDR3 1600MHz RAM, 1TB HDD
  - Metrics:
    - Execution time
    - Throughput
    - Cache hit ratio

Performance results

Table 1: Evaluation on small data set.

Weather Integration	Execution Time	Memory Consumption	Throughtput
With Indexing	12 sec	229 MB	3,570 rows/sec
Without Indexing	1,660 sec	106 MB	26 rows/sec
Pre-processing	1 sec	57 MB	N/A

Table 2: Evaluation on large data set.

Procedure	Execution Time	Memory Consumption	Throughtput
Weather Integration	29,261 sec	176 MB	2,784 rows/sec
Pre-Processing	5 sec	59 MB	N/A

Evaluation of caching mechanism

- Randomly shuffle the input records of mobility traces
- Thus, generating a mobility data set in random temporal order
- Worst-case scenario:
  - Low probability that two consequent records will use the same GRIB file

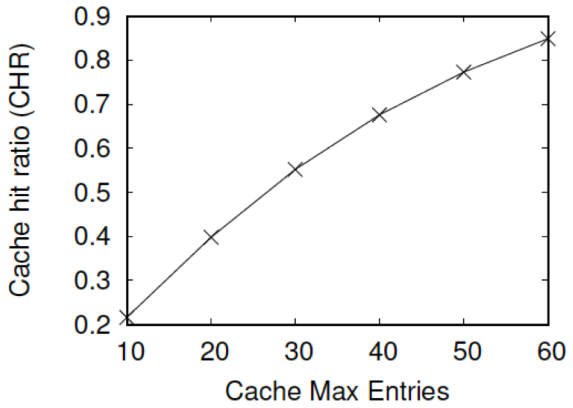


Figure 5: Cache hit ratio for increased cache size.

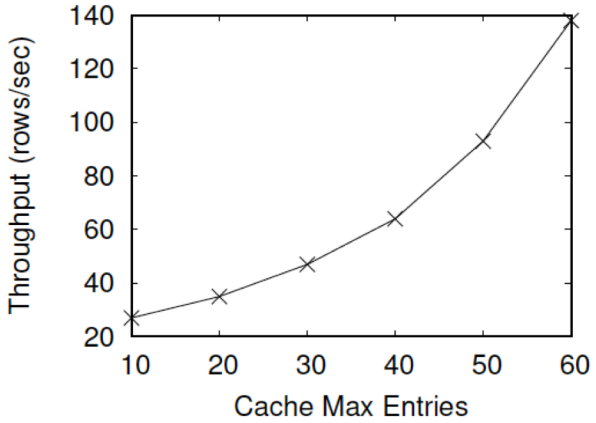


Figure 7: Throughput for increased cache size.

Conclusions & outlook

- We presented a system for enrichment of mobility data with weather
- On a conventional machine on temporally sorted mobility data
  - Achieved throughput: ~3.5K rows/sec
- Also supporting complex geometries and output of linked data
- Available for download: https://github.com/nkoutroumanis/Weather-Integrator

- In the future, we plan to extend our system:
  - Parallelize processing for improved performance
  - Improve the accuracy of weather enrichment of complex geometries

Thank you for your attention

More info:

our group: *http://www.datastories.org/*

personal page: http://www.ds.unipi.gr/en/cdoulk/

e-mail: cdoulk@unipi.gr

This work is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 687591 (datAcron), No 780754 (Trqck&Know), No 779747 (BigDataStack) and No 777695 (MASTER)