Cloud-based Geospatial open systems for mitigating climate change: research directions, challenges, and future perspectives

Isam Al Jawarneh, PhD.

Postdoc researcher

Department of Computer Science and Engineering - DISI
University of Bologna, Italy
isam.aljawarneh3@unibo.it

FOSS4G BUENOS AIRES 2021

30 Sept 2021 on Buenos Aires room

Outline

- Overview & motivating scenario
- Challenges in exploitation open-source geospatial systems climate change
- Novel Cloud-based architectures for spatial-meteorological integrity
- Future research frontiers and recommendations

FOSS4G BUENOS AIRES 2021

Background & Motivating Scenario

Motivating Application Scenario

Promoting the health of lightweight dwellers:

- Relationships between mobility and meteorological data
  - How vehicle emissions affect health of people
- Interactive visualization
  - Interactive heatmaps showing pollution and mobility
- Green pathfinder
  - Origin-destination health-aware paths

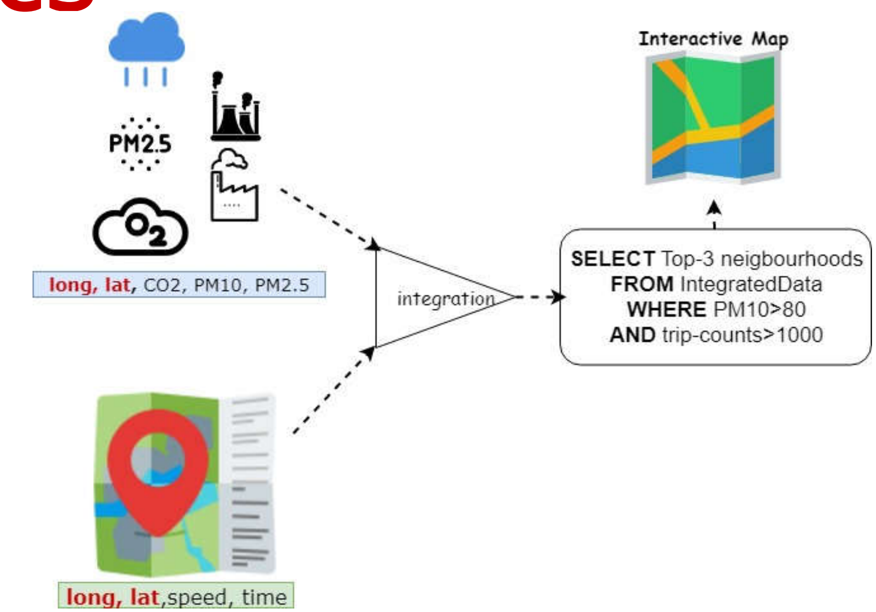
➢ Primitive geospatial queries
- Proximity queries
- Spatial join
- Spatial clustering
- Spatial geo-statistics (e.g., sampling)
- k-Nearest Neighborhoods

QoS goals include
- ✓ Low latency
- ✓ High throughput
- ✓ Maximum resource utilization
- ✓ Maximum accuracy

FOSS4G BUENOS AIRES 2021

Shared meteorological and mobility analytics

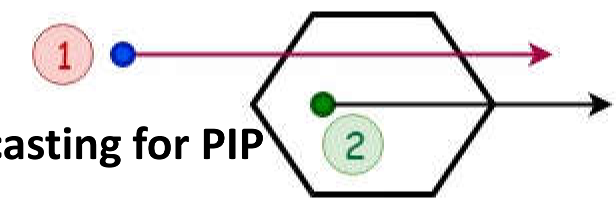
- Spatial join between **mobility & meteorological data**
- Enables new interesting interactive queries



Geospatial Queries

- Proximity query:
  - Data sorted on distance to a POI criteria (optionally specifying a range of distance)
  - Can be solved by applying Point-in-Polygon (PIP) test, by transforming the problem into an inclusion (**Containment)** query

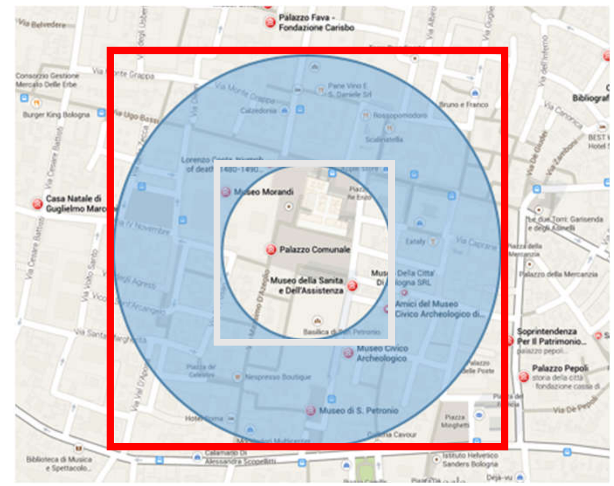
- Containment query:
  - Requests all points contained within the premises of a polygonal shape (regularly or arbitrarily shaped)
  - Requires PIP test



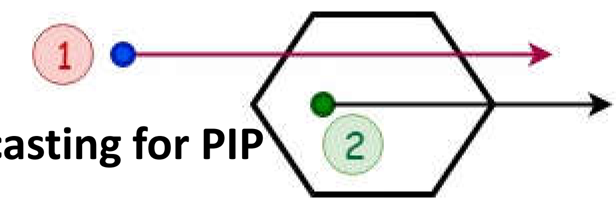
Raw casting for PIP

FOSS4G BUENOS AIRES 2021

Solving proximity queries by spatial join



- Transforming proximity query into a containment:
  - Calculate the embedding area (rectangular areas)
  - Apply a PIP algorithm to retrieve points within the embedding space. (join op., filter)
  - Calculate distance between all points from the previous list and retrieve points that fall within the range specified. (refine)



Raw casting for PIP

FOSS4G BUENOS AIRES 2021

Partial landscape of Cloud-based geospatial systems

Desired features include:

- Filter-and-refinement approach. For spatial join.
- SQL-alike support. For spatial queries
- Spatial Approximate Query Processing. e.g., Spatial sampling

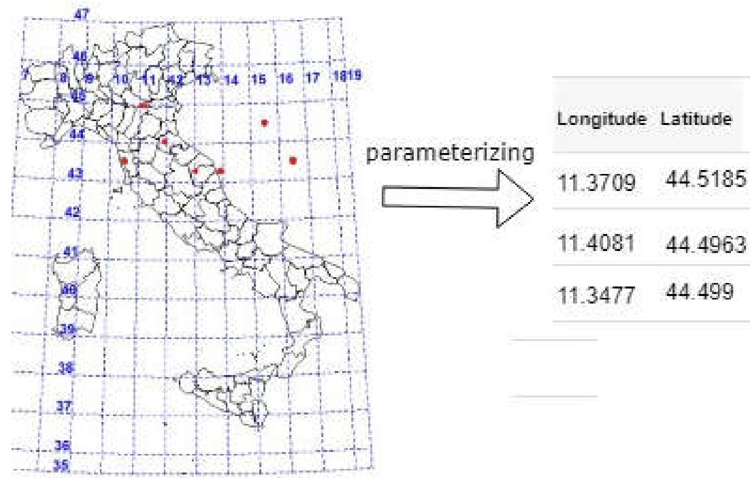
spatial join	queries	SQL	
others	join, range, containment, clustering (MR-DBSCAN), KNN	×	Stark
Broadcast , partition	join, range	×	SpatialSpark
others	join, range, KNN	×	locaionSpark
distance join	join, range, KNN	✓	simba
filter-and-refine	join, range	✓	Spark Magellan
filter-and-refine	join, range, KNN, sampling	✓	Sedona
filter-and-refine	join, range, KNN	✗	SpatialHadoop
filter-and-refine	join, range, KNN, ...	✓	geomesa
filter-and-refine	join, range, KNN	✗	GeoFlink
			Spark, Flink, Hadoop

FOSS4G BUENOS AIRES 2021

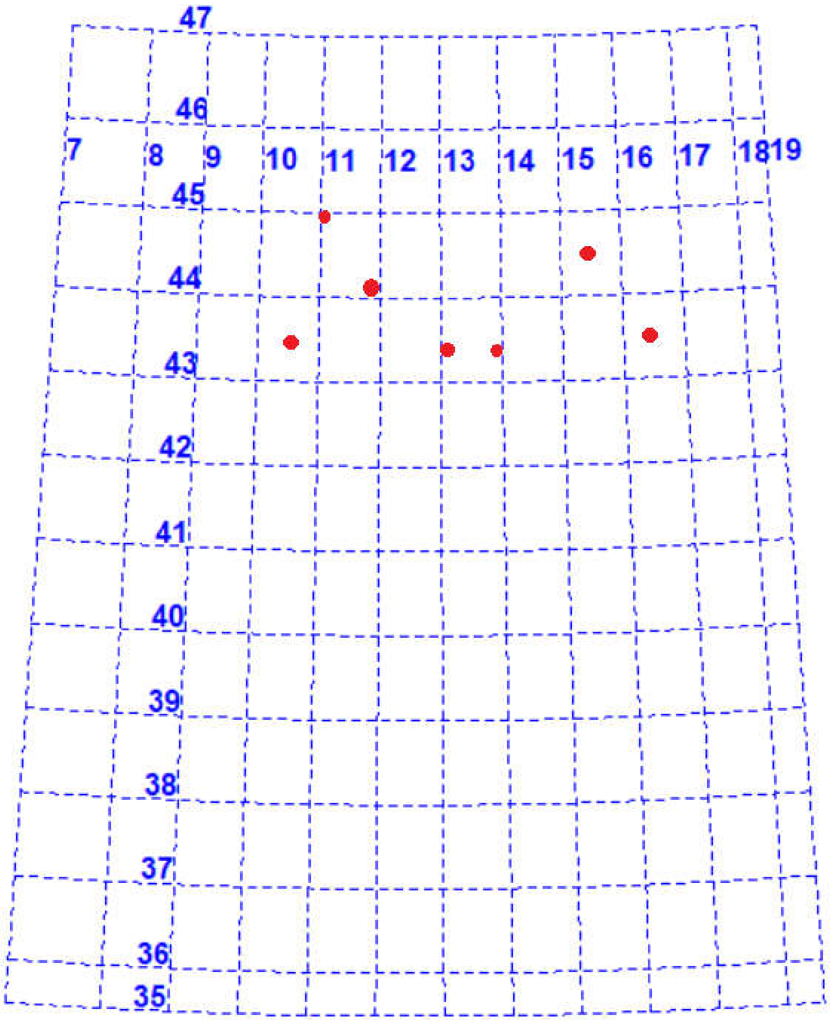
challenges

Geospatial data transformation

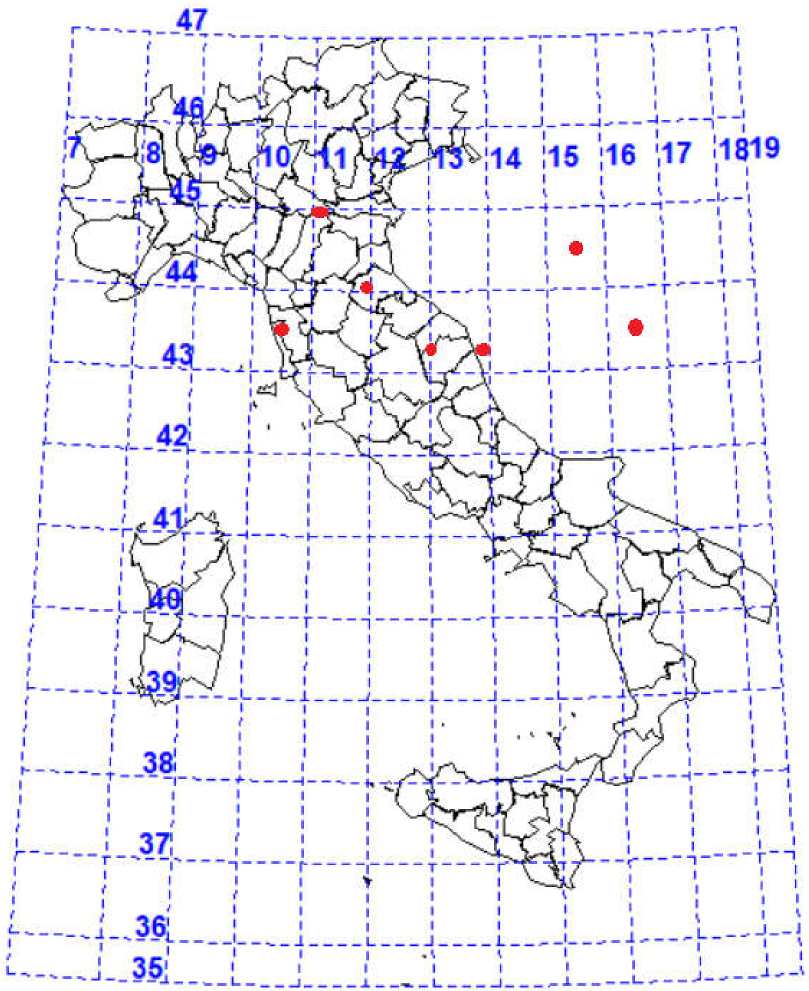
- Big geospatial data management is challenging
  - geospatial records are parametrized (longitudes and latitudes)
  - GPS is rarely 100% accurate,
    - susceptible to acceptable error-bounds
  - Real geometries are lost by this kind of transformation
  - Reconstructing into real geometries is **expensive**



FOSS4G BUENOS AIRES 2021

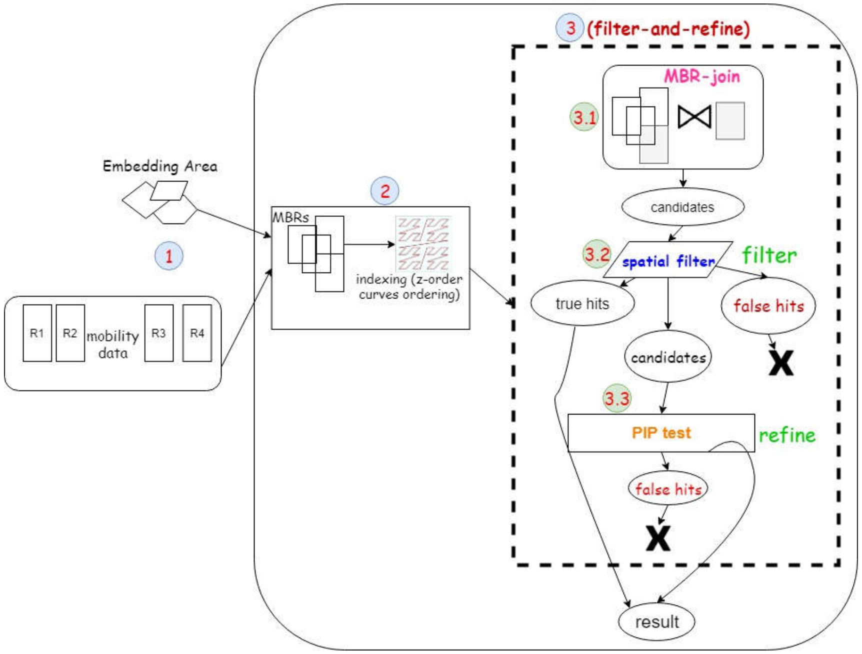


47
46
45
44
43
42
41
40
39
38
37
36
35
7
8
9
10
11
12
13
14
15
16
17
18
19


47
46
7
8
9
10
11
12
13
14
15
16
17
1819
45
44
43
42
41
40
39
38
37
36
35

Filter-and-refine

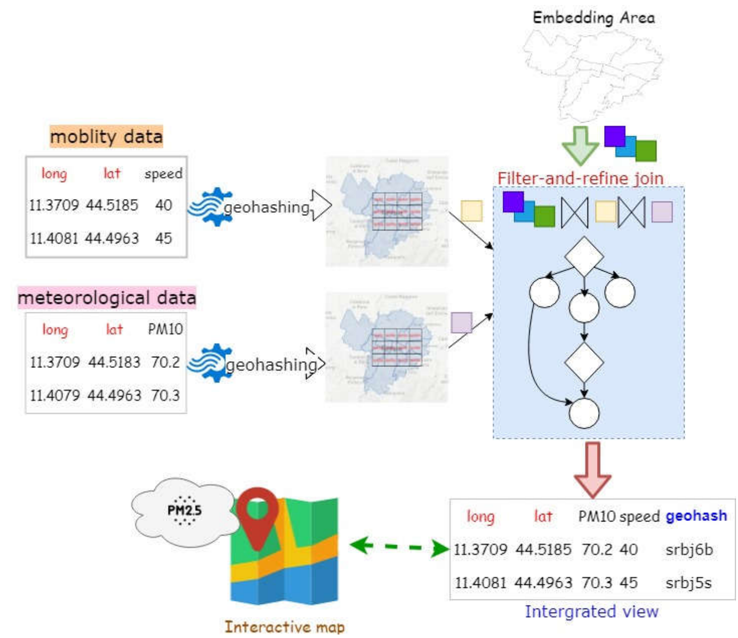
- Based on dimensionality reduction
  - Compute geohash for every point
  - Compute geohash covering of the embedding area
  - Perform a cheap equi-join to find which points fall within the embedding area (filter)
  - Use the ray casting algorithm to exclude false positives (refine)
- Adopted by Spark's Magellan and Geomesa



Architectures

Proposed Architecture

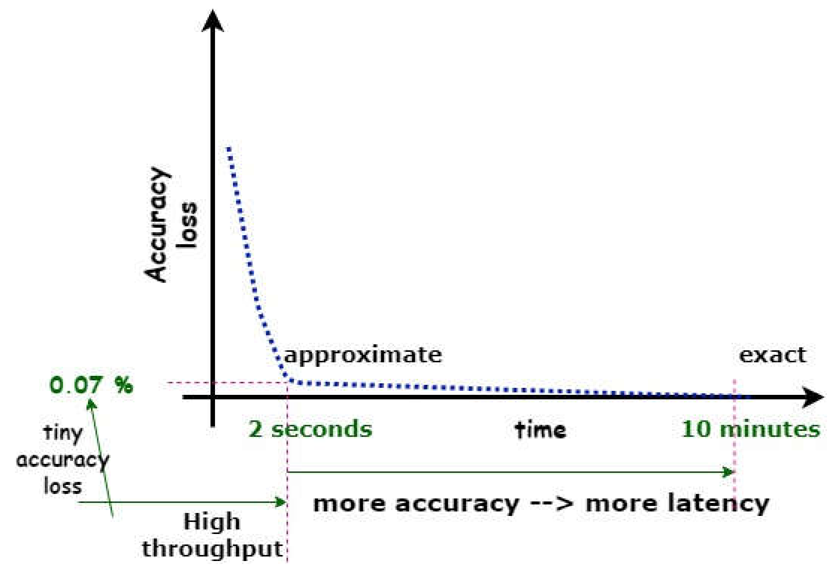
- Generate geohashes of mobility and meteorological
- Apply the filter-and-refine spatial join
  - The results contain a unified view
- Generate interactive maps (e.g., heatmaps)



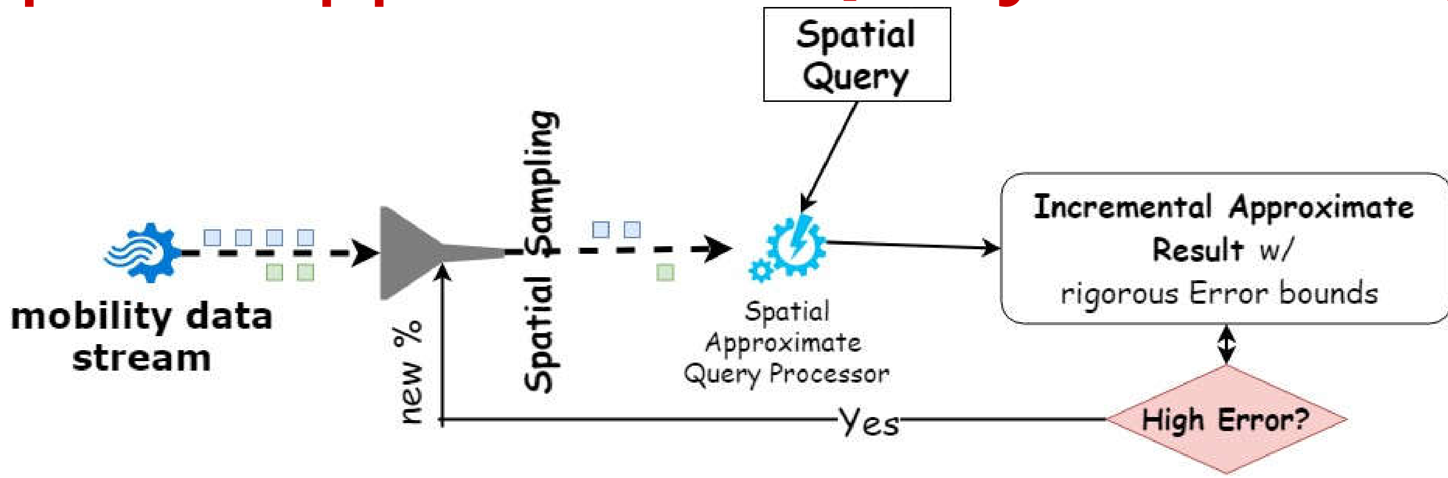
FOSS4G BUENOS AIRES 2021

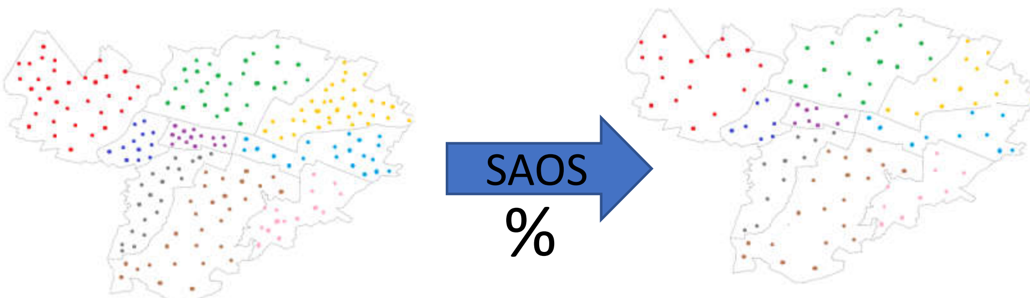
Spatial Approximate Query Processing

- Challenges
  - ➢ Data streams arrive very fast
  - ➢ Skewness and arrival rates fluctuate
- Decision makers accept tiny loss in accuracy in exchange for a throughput gain



Spatial Approximate Query Processing



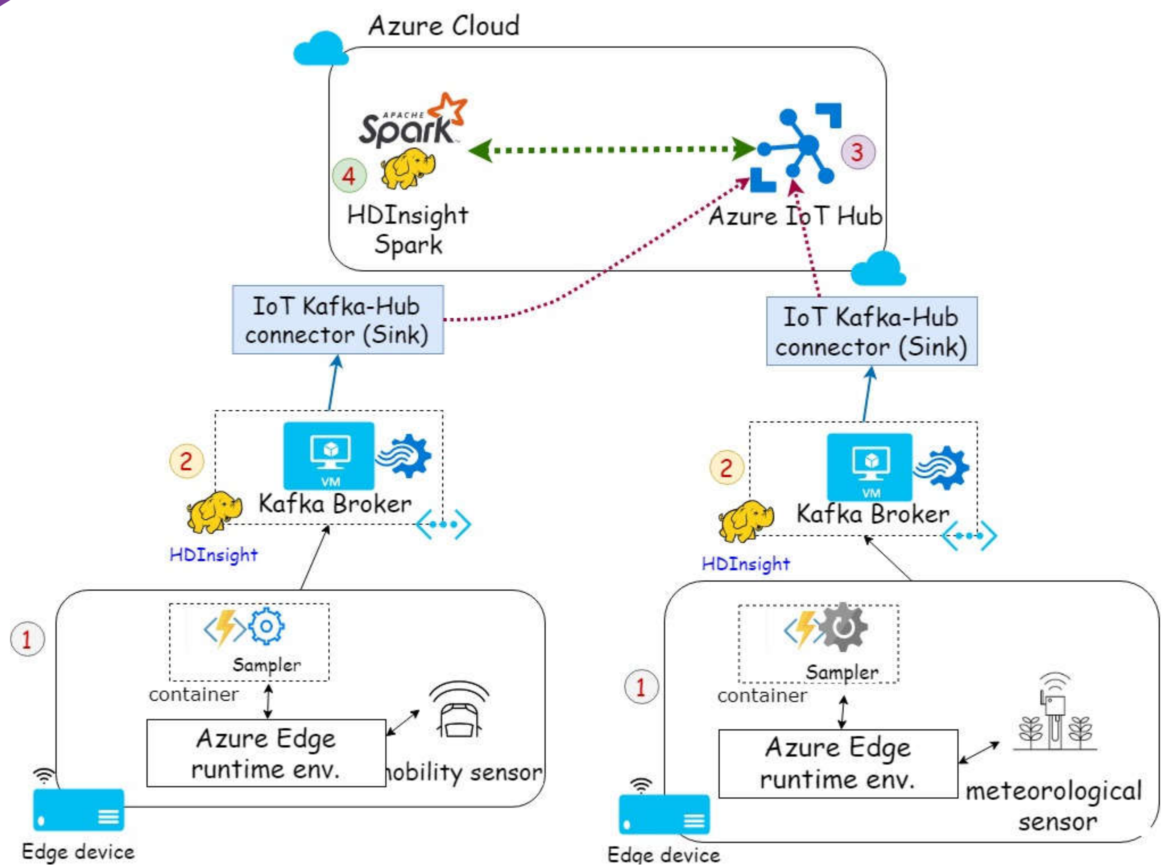


Future Research

Conclusion and Open research

- **Current** Cloud-based open-source geospatial systems are limited in their capacity
  - However, they are good springboards!
- Offloading jobs to Edge devices
  - Port some processing to Edge devices,
  - reducing time-to-insight.
    - e.g., pushing spatial sampler to Edge devices upstream (near the data sources).

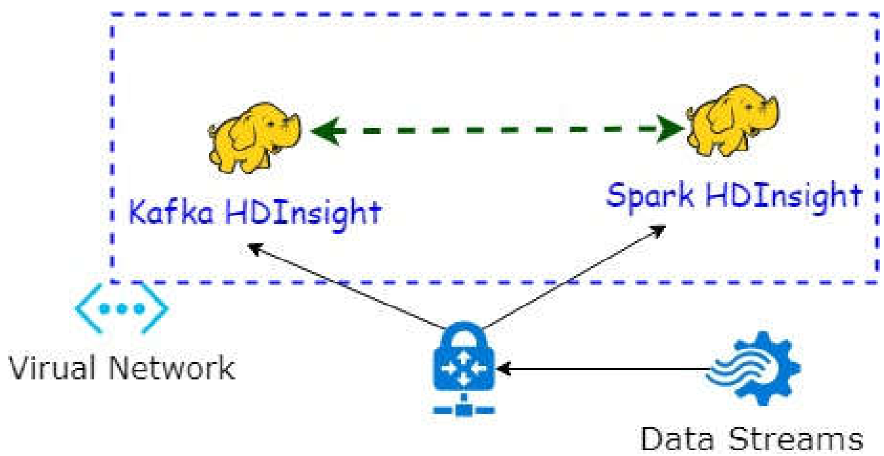
FOSS4G BUENOS AIRES 2021



Azure Cloud
Spark
4
HDInsight
Spark
3
Azure IoT Hub
IoT Kafka-Hub
connector (Sink)
IoT Kafka-Hub
connector (Sink)
2
VM
Kafka Broker
HDInsight
2
VM
Kafka Broker
HDInsight
1
Sampler
container
Azure Edge
runtime env.
mobility sensor
Edge device
1
Sampler
container
Azure Edge
runtime env.
meteorological
sensor
Edge device

Open-source code

- My code for Cloud-based Spatial Approximate Query processing with instructions to deploy on **Microsoft Azure:**
  - https://isamaljawarneh.github.io/ApproximateStream/



Open for collaboration

- Our group (**Mobile Middleware Research Group**) at **University of Bologna**:

  http://www.middleware.unibo.it/

- Feel free to contact us:
  - Dr. **Isam Al Jawarneh** : isam.aljawarneh3@unibo.it
  - Prof. **Luca Foschini**: luca.foschini@unibo.it

FOSS4G BUENOS AIRES 2021

Relevant literature

[1] **I. M. Al Jawarneh**, P. Bellavista, L. Foschini and R. Montanari, **"Spatial-aware approximate big data stream processing,"** in 2019 IEEE Global Communications Conference (GLOBECOM), 2019, pp. 1-6.

[2] **I. M. Al Jawarneh**, P. Bellavista, A. Corradi, L. Foschini, R. Montanari and A. Zanotti, **"In-memory spatial-aware framework for processing proximity-alike queries in big spatial data,"** in 2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2018, pp. 1-6.

[3] **Al Jawarneh, I. M.,** Bellavista, P., Corradi, A., Foschini, L., & Montanari, R. (2020). **Efficient QoS-Aware Spatial Join Processing for Scalable NoSQL Storage Frameworks.** IEEE Transactions on Network and Service Management, 18(2), 2437-2449.

[4] **Al Jawarneh, I. M.,** Bellavista, P., Corradi, A., Foschini, L., & Montanari, R. (2020). **Big spatial data management for the Internet of Things: a survey.** Journal of Network and Systems Management, 28(4), 990-1035.

Relevant literature

[5] **Al Jawarneh, I. M.,** Bellavista, P., Corradi, A., Foschini, L., & Montanari, R. (2020, September). **Spatially Representative Online Big Data Sampling for Smart Cities.** In 2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD) (pp. 1-6). IEEE.

[6] **Al Jawarneh, I. M.,** Bellavista, P., Foschini, L., & Montanari, R. (2019, December). **Spatial-aware approximate big data stream processing.** In 2019 IEEE global communications conference (GLOBECOM) (pp. 1-6). IEEE.

[7] **Al Jawarneh, I. M.,** Bellavista, P., Corradi, A., Foschini, L., Montanari, R., & Zanotti, A. (2018, September). **In-memory spatial-aware framework for processing proximity-alike queries in big spatial data.** In 2018 IEEE 23rd international workshop on computer aided modeling and design of communication links and networks (CAMAD) (pp. 1-6). IEEE.

[8] **Al Jawarneh, I. M.,** Bellavista, P., Casimiro, F., Corradi, A., & Foschini, L. (2018, June). **Cost-effective strategies for provisioning NoSQL storage services in support for industry 4.0.** In 2018 IEEE symposium on computers and communications (ISCC) (pp. 01227-01232). IEEE.

FOSS4G BUENOS AIRES 2021

Cloud-based Geospatial open systems for mitigating climate change: research directions, challenges, and future perspectives

Thanks a lot!

Isam Al Jawarneh, PhD.

Postdoc researcher

Department of Computer Science and Engineering - DISI
University of Bologna, Italy
isam.aljawarneh3@unibo.it

FOSS4G BUENOS AIRES 2021

30 Sept 2021 on Buenos Aires room