

FOSS4G

Prizren, 2023

FOSS4G 2023 – 26 June – 2 July 2023, Prizren – Kosovo

GEOSPATIAL BIG DATA ANALYTICS FOR SUSTAINABLE SMART CITIES

Muhammed Oguzhan METE

Department of Geomatics Engineering, Istanbul Technical University, Türkiye



İTÜ

ABOUT ME



Muhammed Oguzhan METE

PhD, Asst. Prof.

Istanbul Technical University
Geomatics Engineering
Department

metemu@itu.edu.tr



Biography

Dr. Muhammed Oguzhan METE is currently working as Assistant Professor at Istanbul Technical University, Geomatics Engineering Department. He is also a Community Builder at Amazon Web Services for two years. His research interests include Land Management, Real Estate Management, Cadastre, Geographic Information Systems, Machine Learning, Deep Learning, Big Data Analytics and Cloud Computing.

Interests

- Land Management
- Real Estate Valuation
- Cadastre
- Geographic Information Systems
- Machine Learning / Deep Learning
- Big Data Analytics
- Cloud Computing

Education

- PhD in Geomatics Engineering, 2019 - 2022
Istanbul Technical University
- MSc in Geomatics Engineering, 2017 - 2019
Istanbul Technical University
- BSc in Geomatics Engineering, 2012 - 2017
Istanbul Technical University



İTÜ

TABLE OF CONTENTS

1) Sustainability

- a) Sustainable Development Goals
- b) European Green Deal
- c) Net Zero Target

2) Smart City

- a) Definition
- b) Components (Smart Energy, Smart Infra. Smart Environment)

3) Geospatial Big Data

- a) How to handle it?
- b) Parallel & Cluster Processing
- c) Dask-Geopandas & Apache Sedona
- d) Data File Formats (GeoParquet vs others)
- e) Benchmark Test: Apache Sedona and DaskGeopandas



3
iTÜ

TABLE OF CONTENTS

4) Sustainable Smart Energy Management: Achieving the UK Net Zero Carbon Emission Target

- a) Building scale analytics
- b) Spatial Join, Clustering, Visualization (Datashader)



4
iTÜ



PROBLEM DEFINITION AND AIM OF THE STUDY

Increasing Urbanization



<https://www.quora.com/What-effects-does-urbanization-have-on-the-environment>



social



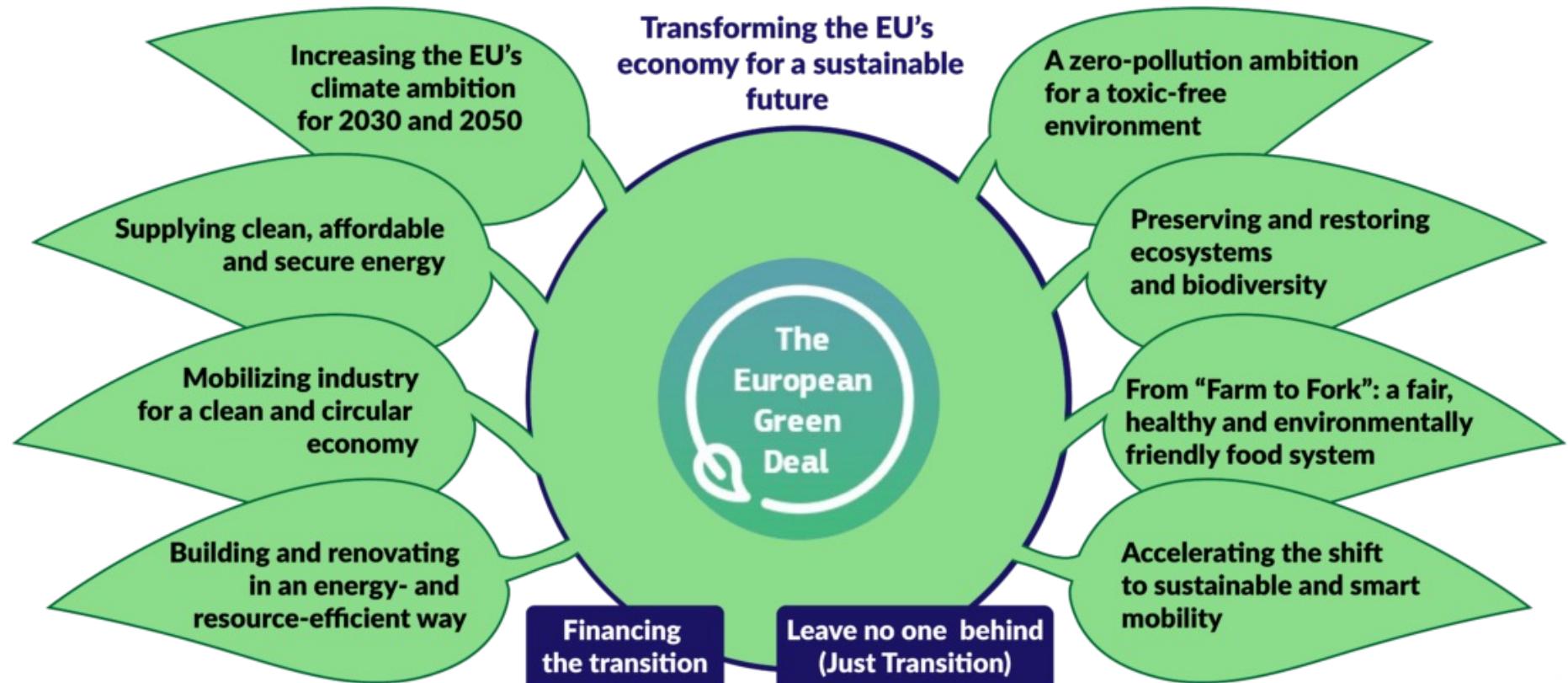
environmental

economic



İTÜ

SUSTAINABLE SMART CITIES



<https://www.gisreportsonline.com/r/european-green-deal/>

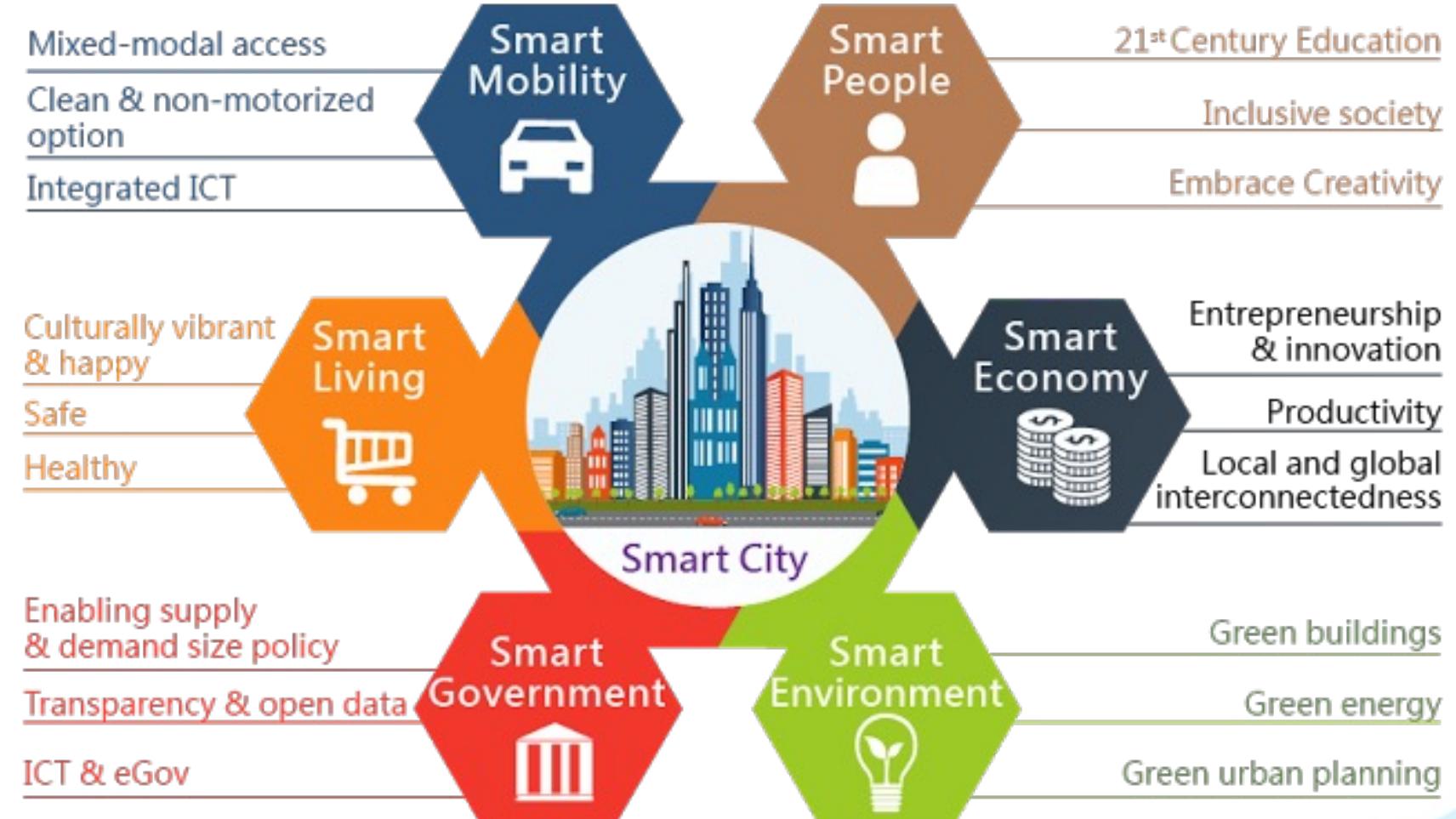


İTÜ

SUSTAINABLE SMART CITIES



SUSTAINABLE SMART CITIES



Mode, V., Varin, V., & Wanchai, R. (2018).



GEOSPATIAL BIG DATA

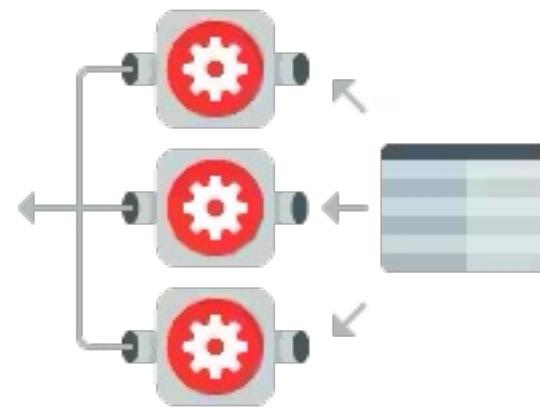
How to handle Geospatial Big Data?



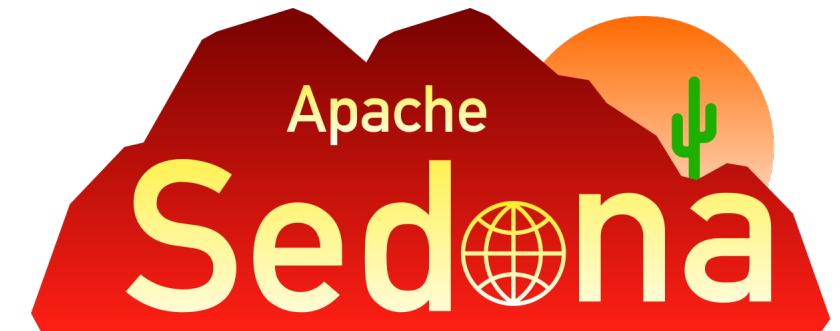
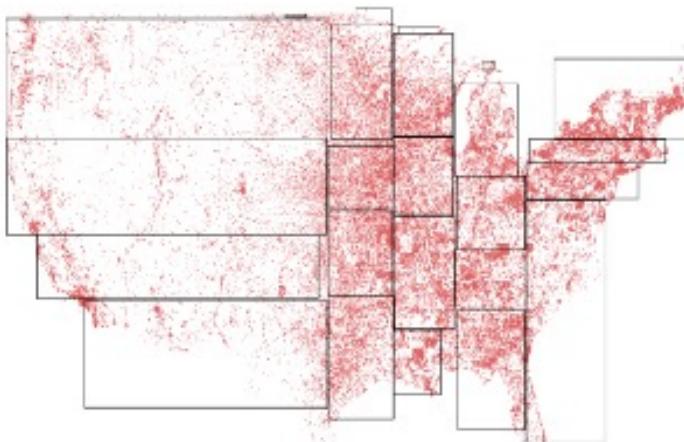
<https://www.sap.com/uk/products/technology-platform/what-is-big-data.html>

GEOSPATIAL BIG DATA

Parallel Processing



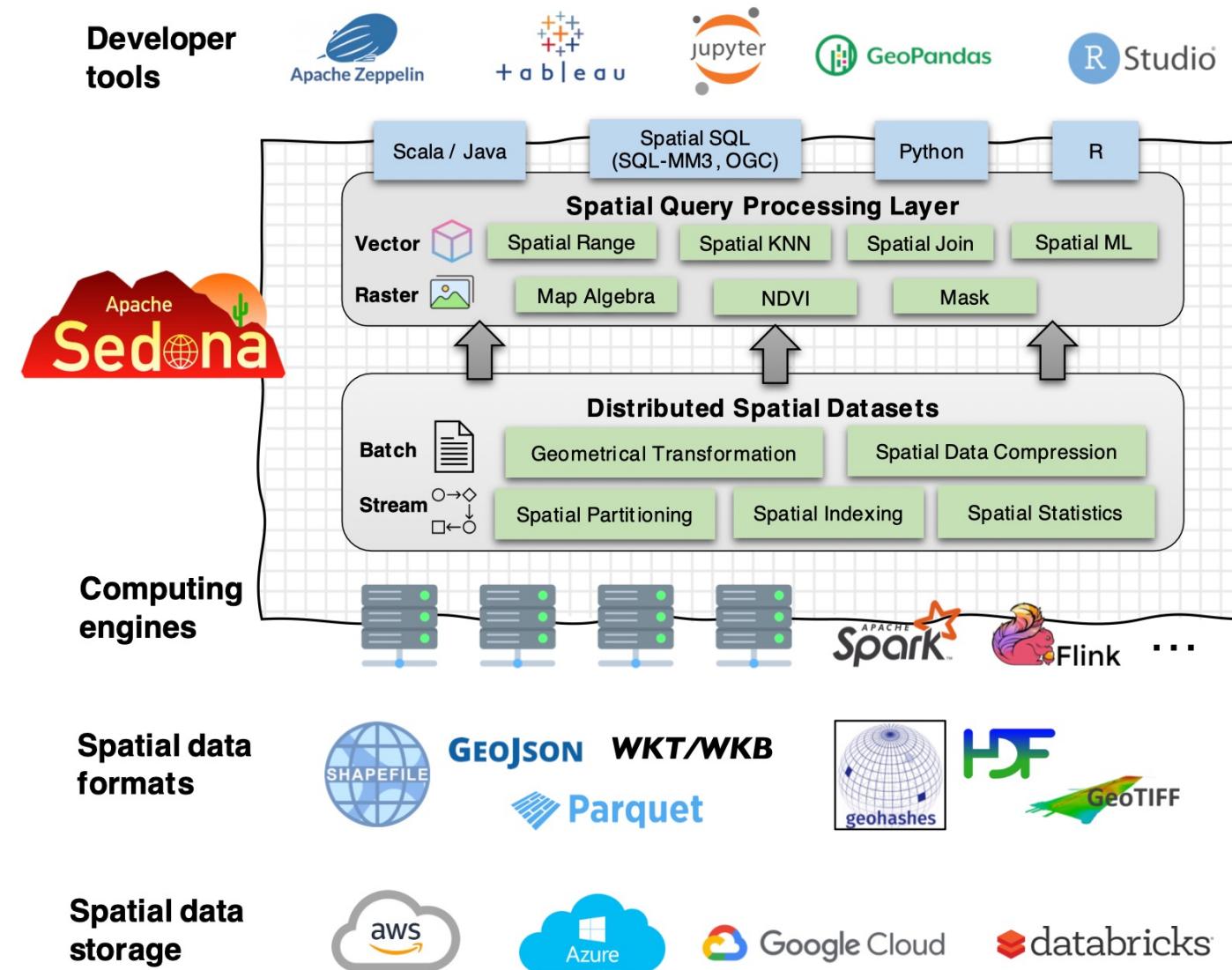
Spatial Index



10
iTÜ



GEOSPATIAL BIG DATA

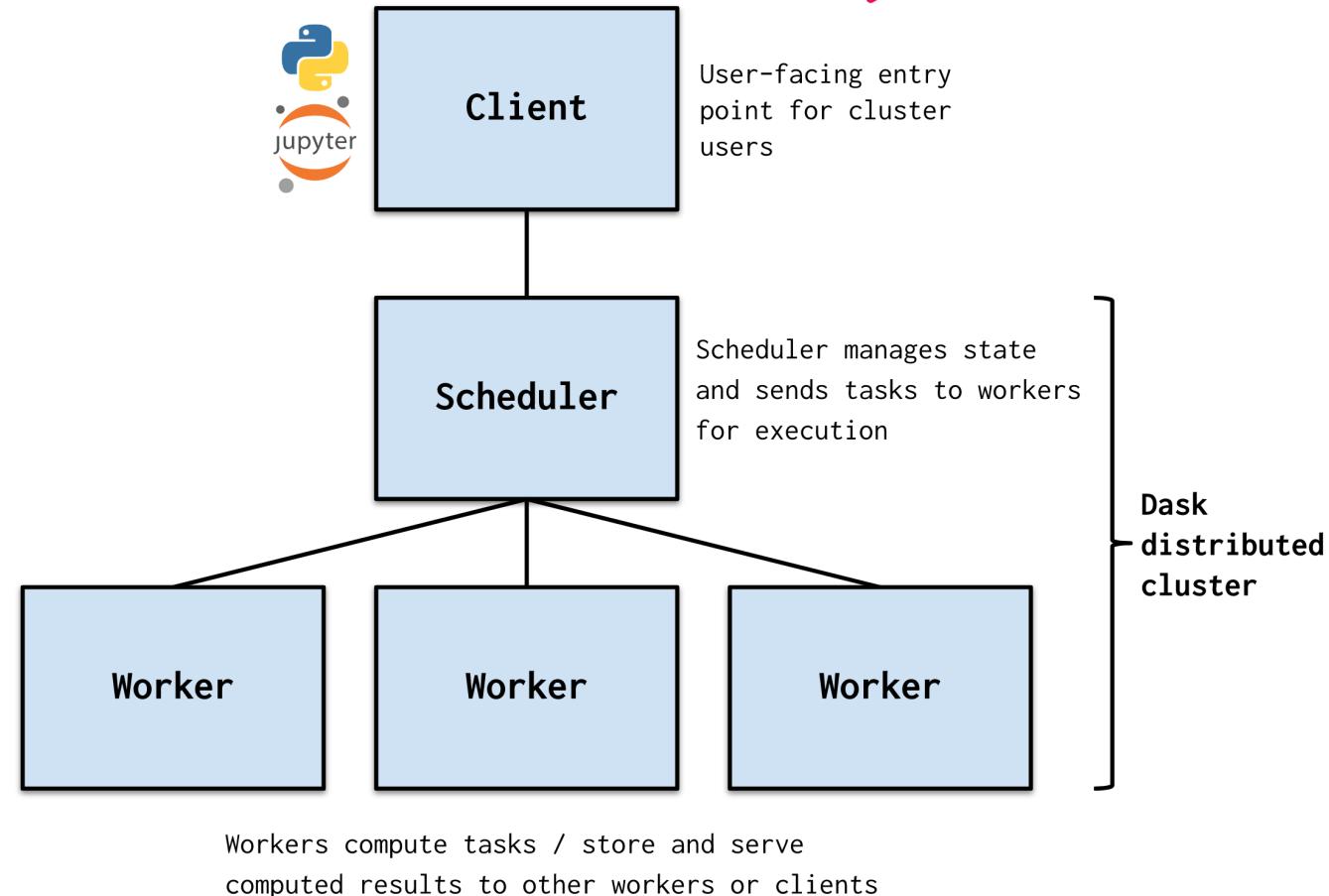


<https://sedona.apache.org>



İTÜ

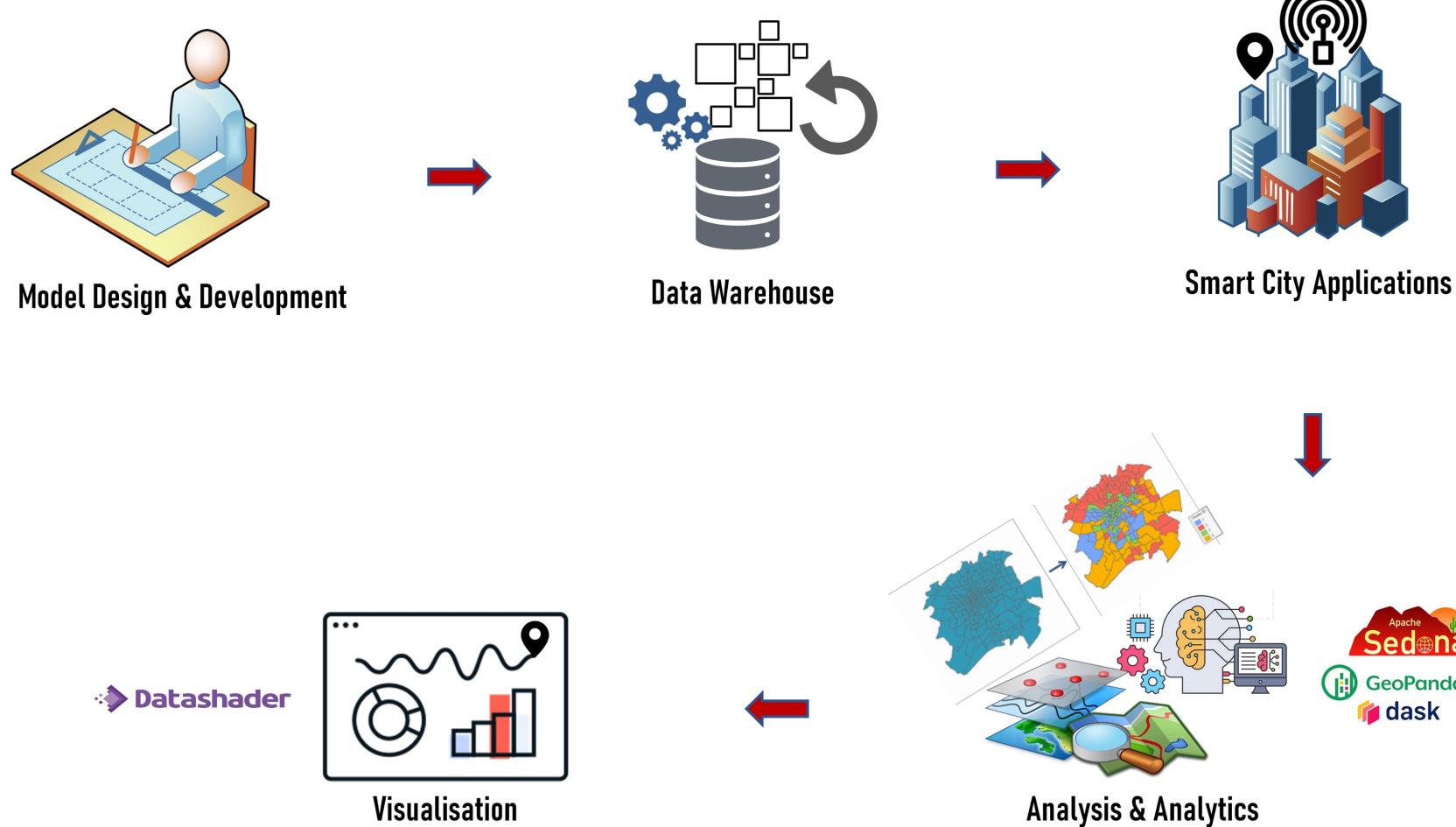
GEOSPATIAL BIG DATA



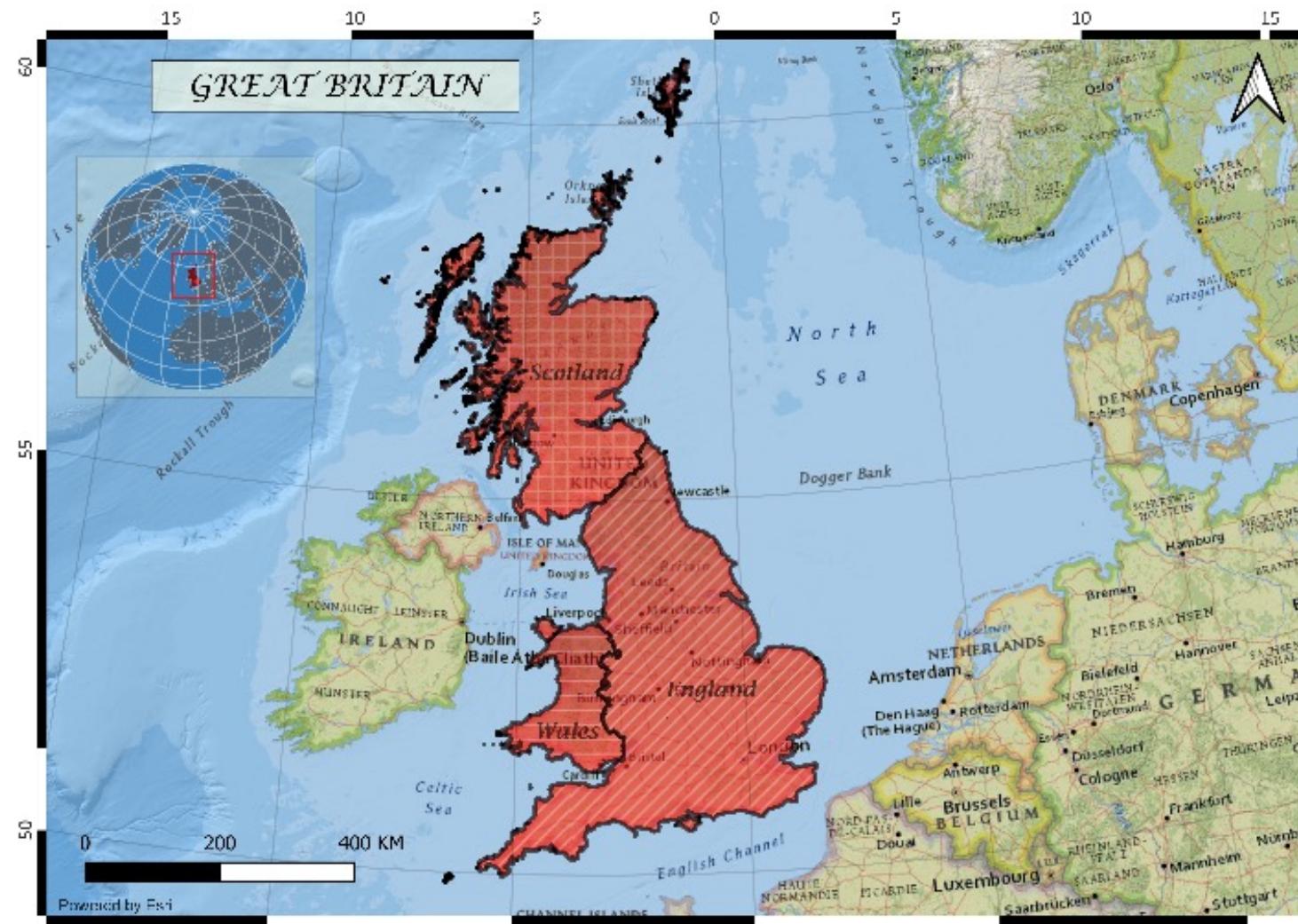
<https://github.com/dask/dask/issues/4471>

SUSTAINABLE SMART CITIES

Geospatial Big Data Administration Model Framework for Sustainable Smart Cities



STUDY AREA



Study Area: Great Britain, United Kingdom



14
iTÜ





ENERGY PERFORMANCE OF BUILDINGS (EPC)

Department For Communities and Local Government

Energy Performance of Buildings Data England and Wales

11,999 Domestic EPCs found

1 Oct 2016
Charlton Park Farm, Charlton Drive, Wraxall BS48 1PD

30 Sep 2016
Velen House, Hills Avenue CB1 7UY

30 Sep 2016
Tirnalia House, Hills Avenue CB1 7UY

30 Sep 2016
2, Balsam Road, West Timperley WA14 5DR

30 Sep 2016
4, Tiberius Drive, Fairfields MK11 4AX

30 Sep 2016
Height Green Barn, Ripponden HX6 4HH

30 Sep 2016
241, St. Lukes Road SR4 0AL

30 Sep 2016
243, St. Lukes Road SR4 0AL

30 Sep 2016
239, St. Lukes Road SR4 0AL

30 Sep 2016
207, Runcorn Road, Barnton CW8 4HR

30 Sep 2016
6, Tiberius Drive, Fairfields MK11 4AX

30 Sep 2016
237, St. Lukes Road SR4 0AL

30 Sep 2016
247, St. Lukes Road SR4 0AL

30 Sep 2016
Calidore House, Hills Avenue CB1 7UY

Energy Performance of Buildings Data: England and Wales

Domestic EPC

Non-domestic EPC

DEC

Help

Attribute Name	Explanation
LMK Key	Individual lodgement identifier.
Building Reference Number	Unique identifier for the property.
Current Energy Efficiency	Based on cost of energy, i.e. energy required for space heating, water heating and lighting [in kWh/year] multiplied by fuel costs. (£/m ² /year where cost is derived from kWh).
Property Type	Describes the type of property such as House, Flat, Maisonette etc.
Built Form	The building type of the Property e.g. Detached, Semi-Detached, Terrace etc.
Environment Impact Current	The Environmental Impact Rating. A measure of the property's current impact on the environment in terms of carbon dioxide (CO ₂) emissions. The higher the rating the lower the CO ₂ emissions (CO ₂ emissions in tonnes/year).
Energy Consumption Current	Current estimated total energy consumption for the property in a 12-month period (kWh/m ²).
CO ₂ Emissions Current	CO ₂ emissions per year in tonnes/year.
Lighting Cost Current	Current estimated annual energy costs for lighting the property (GBP).
Heating Cost Current	Current estimated annual energy costs for heating the property (GBP).
Hot Water Cost Current	Current estimated annual energy costs for hot water (GBP).
Total Floor Area	The total useful floor area is the total of all enclosed spaces measured to the internal face of the external walls (m ²).
Floor Level	Floor level relative to the lowest level of the property (0 for ground floor). If there is a basement, the basement is level 0 and the other floors are from 1 upwards.
Extension Count	The number of extensions added to the property.
Number Habitable Rooms	Habitable rooms include any living room, sitting room, dining room, bedroom, study and similar; and also a non-separated conservatory.
Address	Postcode, Address 1 (SAON), Address 2 (PAON), Address 3 (Street)

First 5000 results (.zip)

FAQ: why only 5000?

Display Energy Certificate

How efficiently is this building being used?

Department of Energy & Climate Change
3-8 Whitehall Place
London SW1A 2HH

HM Government

Certificate Reference Number:
0098-9592-5110-2590-0003

This certificate indicates how much energy is being used to operate this building. The operational rating is based on meter readings of all the energy actually used in the building. It is compared to a benchmark that represents performance indicative of all buildings of this type. There is more advice on how to interpret this information on the Government's website www.communities.gov.uk/epbd.

Energy Performance Operational Rating

This tells you how efficiently energy has been used in the building. The numbers do not represent actual units of energy consumed; they represent comparative energy efficiency. 100 would be typical for this kind of building.

More energy efficient

A 0-25

B 26-50

C 51-75

D 76-100

100 would be typical

E 101-125

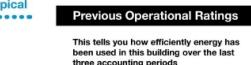
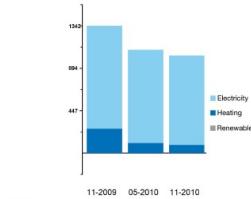
F 126-150

G Over 150

Less energy efficient

Total CO₂ Emissions

This tells you how much carbon dioxide the building emits. It shows tonnes per year of CO₂.



Technical Information

This tells you technical information about how energy is used in this building. Consumption data based on actual meter readings.

Main heating fuel: Natural Gas
Building Environment: Air Conditioning
Total useful floor area (m²): 19960
Accreditation Scheme: CIBSE Certification Limited
Employer/Trading Name: Barr Associates
Employer/Trading Address: 100 High Street, Ambleside, DY8 4BT
Issue Date: 13-11-2010
Nominated Date: 13-11-2010
Valid Until: 11-11-2011
Related Party Disclosure: Not related to the occupier

Administrative Information

This is a Display Energy Certificate as defined in SI 2007/991 as amended.

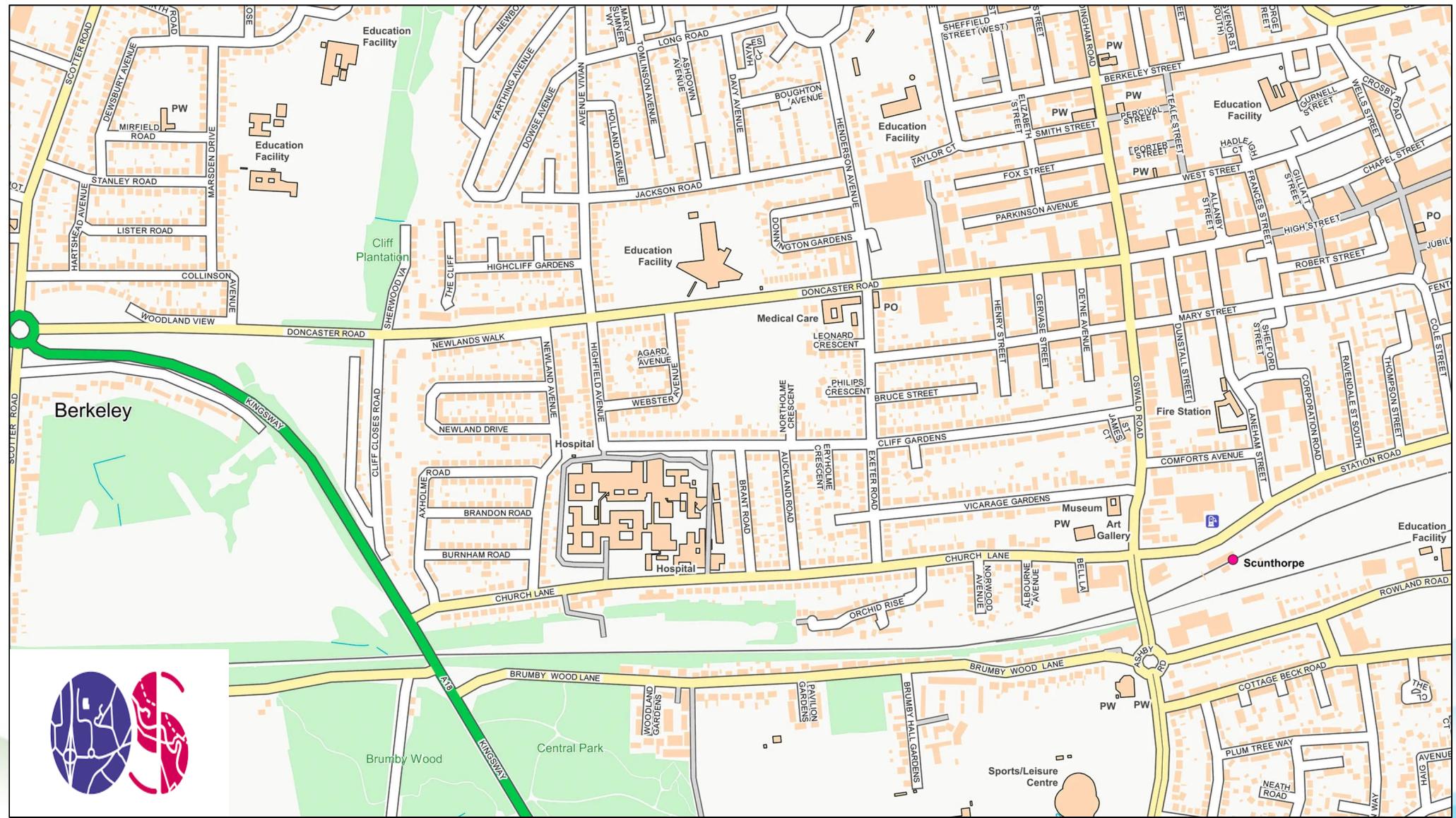
Assessor Source: GLG ORCalc v3.5.1
Property Reference: 0098-9592-5110-2590-0003
Assessor Name: Darren Myers
Assessor Number: LCEA12929
Accreditation Scheme: CIBSE Certification Limited
Employer/Trading Name: Barr Associates
Employer/Trading Address: 100 High Street, Ambleside, DY8 4BT
Issue Date: 13-11-2010
Nominated Date: 13-11-2010
Valid Until: 11-11-2011
Related Party Disclosure: Not related to the occupier

Recommendations for improving the energy efficiency of the building are contained in the accompanying Advisory Report.

CURRENT ENERGY RATING



OS OPEN MAP LOCAL

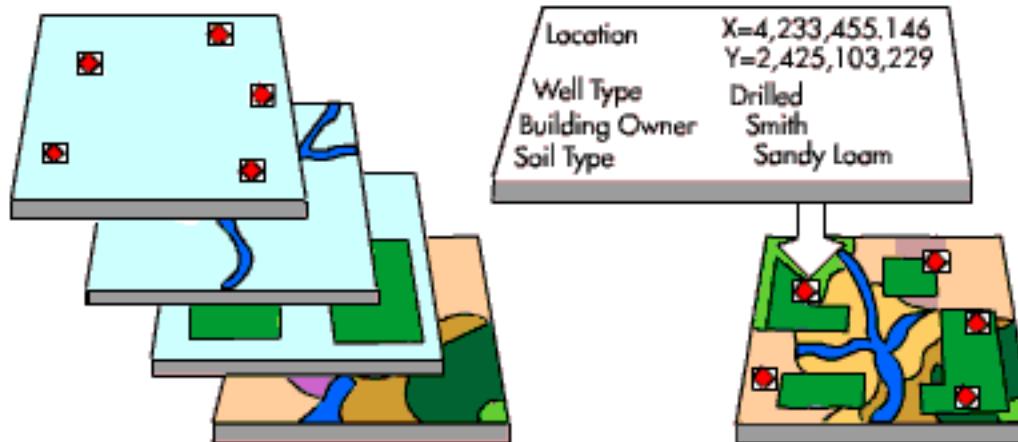


METHODOLOGY



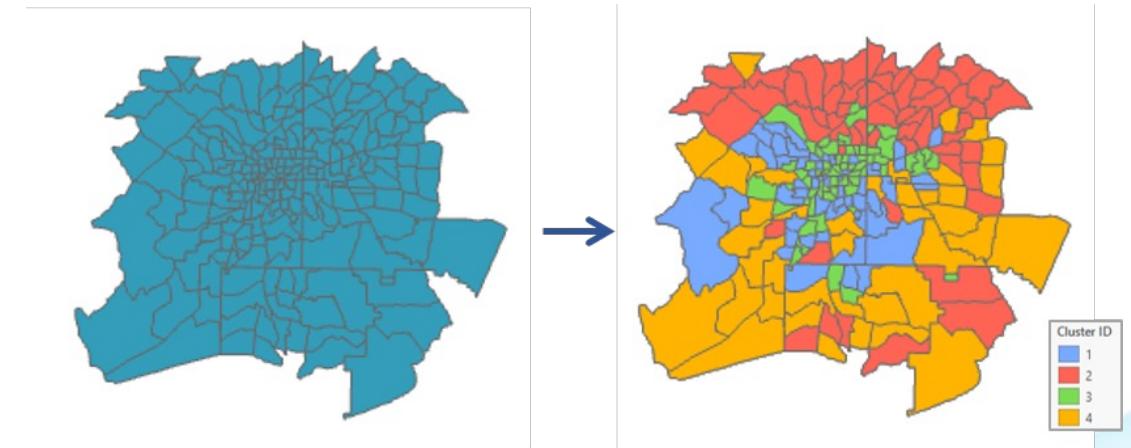
GeoPandas

Spatial Join



https://geopandas.org/en/stable/gallery/spatial_joins.html

Spatial Clustering



<https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/spatially-constrained-multivariate-clustering.htm>



İTÜ



METHODOLOGY

In [3]:

```
%%time
uprn = dask_geopandas.read_file("D:/Data/UPRN/osopenuprn_202301.gpkg", npartitions=8)
#df = gpd.read_file ("D:/Data/UPRN/osopenuprn_202301.gpkg")
#uprn = dask_geopandas.from_geopandas(df, npartitions=4)
uprn
```

CPU times: total: 18.1 s
Wall time: 18.1 s

Out [3]: Dask-GeoPandas GeoDataFrame Structure:

	UPRN	X_COORDINATE	Y_COORDINATE	LATITUDE	LONGITUDE	geometry
npartitions=8						
0	int64	float64	float64	float64	float64	geometry
5037324
...
35261268
40298589

Dask Name: read-file, 1 graph layer

```
uprn.to_parquet('D:/Data/EPC/uprn/')
ddf = dask_geopandas.read_parquet("D:/Data/EPC/uprn/", gather_spatial_partitions=False)
ddf.spatial_partitions
```



İTÜ

METHODOLOGY

```
In [20]: epc = pd.read_parquet('D:/Data/EPC/epc2.parquet')  
epc
```

Out [20]:

	LMK_KEY	ADDRESS1	ADDRESS2	ADDRESS3	POSTCODE	BUILDING_REFERENCE_NUMBER	CURRENT_E...
0	0003352408dad177e827f77c62ed7172f2f16336c44dcd...	21 HERIOT GRANGE	HARTLEPOOL	nan	TS25 3JH		10000150044
1	1514938939222017012914223649268943	9, Sea View Terrace	nan	nan	TS24 0ET		4423379478
2	1451865729102016060917325243560118	54, Chatham Road	nan	nan	TS24 8QQ		134825478
3	142065249242010020920181950000618	10, Kendal Road	nan	nan	TS25 1QY		2295640568
4	978835396212019052007505697910817	19, Wainwright Walk	nan	nan	TS25 1XA		8777671178
...

```
join_result = gpd.sjoin(buildings, uprn, how='left', predicate='intersects')  
join_result = join_result[~join_result.index.duplicated(keep='first')]  
join_result
```



İTÜ

METHODOLOGY

```
In [4]: #Read spatial data into a Sedona DataFrame  
buildings_df = spark.read.format("geoparquet").load('buildingsparquet/buildingsparquet.parquet')  
buildings_df.createOrReplaceTempView("buildings")  
buildings_df.printSchema()  
  
uprn_df = spark.read.format("geoparquet").load('uprncsv/uprncsv.parquet')  
uprn_df.createOrReplaceTempView("uprn")  
uprn_df.printSchema()  
  
root  
|-- id: long (nullable = true)  
|-- uuid: string (nullable = true)  
|-- LAD22CD: string (nullable = true)  
|-- LAD22NM: string (nullable = true)  
|-- COUNTRY: string (nullable = true)  
|-- RGN21CD: string (nullable = true)  
|-- RGN21NM: string (nullable = true)  
|-- ENVIRONMENT_IMPACT_CURRENT: long (nullable = true)  
|-- geometry: geometry (nullable = true)  
  
root  
|-- UPRN: long (nullable = true)  
|-- X_COORDINATE: double (nullable = true)  
|-- Y_COORDINATE: double (nullable = true)  
|-- LATITUDE: double (nullable = true)  
|-- LONGITUDE: double (nullable = true)  
|-- geometry: geometry (nullable = true)
```



METHODOLOGY

Dask-GeoPandas

```
join_result = dask_geopandas.sjoin(daskbuildings, daskuprn, how='inner', predicate='intersects')  
join_result.compute()
```

Apache Sedona

```
result = spark.sql("SELECT b.geometry as buildings_geom, b.ENVIRONMENT_IMPACT_CURRENT,  
u.geometry as uprn_geom, u.UPRN FROM buildings b, uprn u WHERE ST_Contains(b.geometry,  
u.geometry)")
```



İTÜ



METHODOLOGY

Dask-GeoPandas

```
kmeans = KMeans(n_clusters = 4, init = 'k-means++', random_state = 5, max_iter=400)
y_kmeans = kmeans.fit_predict(X)
k=pd.DataFrame(y_kmeans, columns=['cluster'])
gdf=gdf.join(k)
```



İTÜ

METHODOLOGY

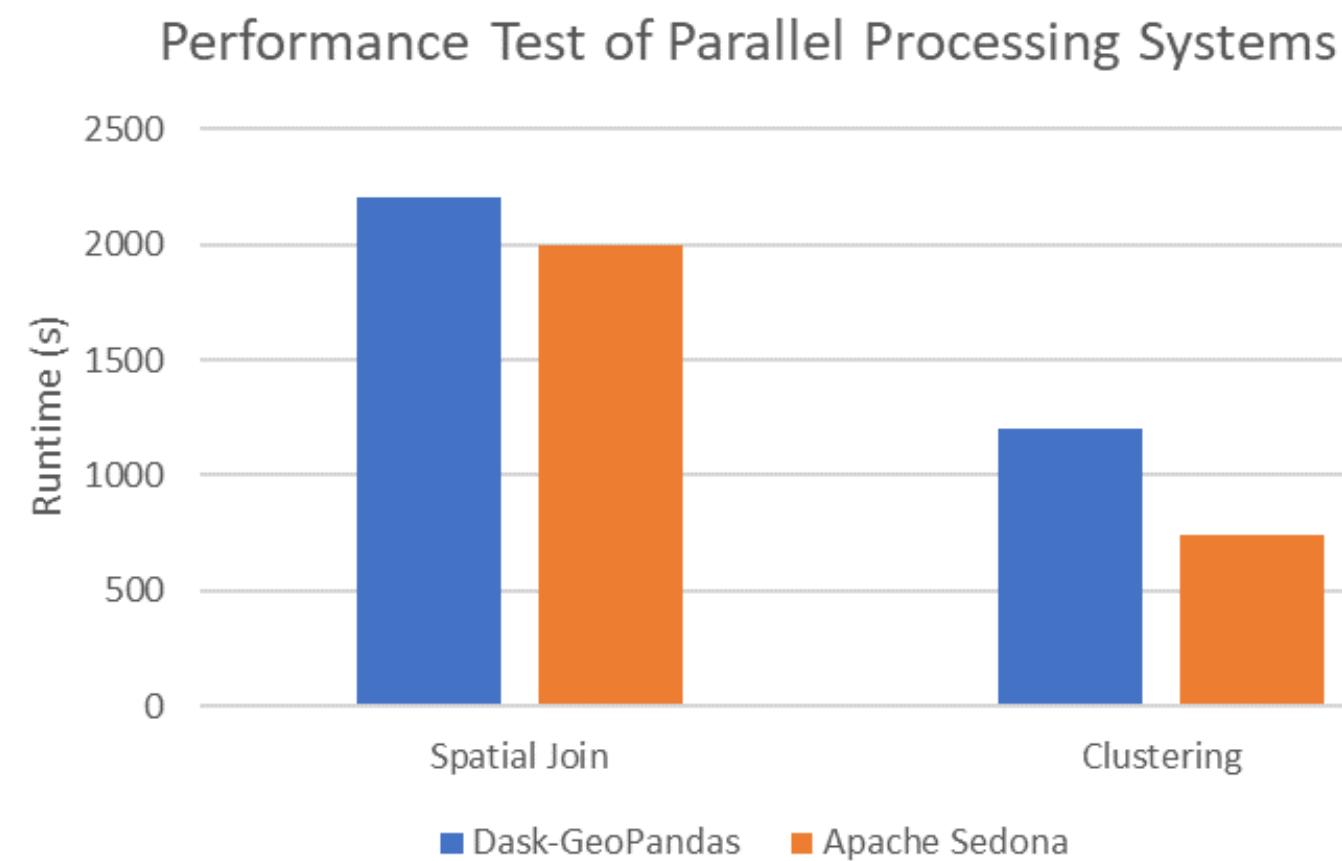
Apache Sedona

```
# Trains a k-means model.  
kmeans = KMeans().setK(2).setSeed(1)  
model = kmeans.fit(dataset)  
  
# Make predictions  
predictions = model.transform(dataset)  
  
# Evaluate clustering by computing Silhouette score  
evaluator = ClusteringEvaluator()  
  
silhouette = evaluator.evaluate(predictions)  
print("Silhouette with squared euclidean distance = " + str(silhouette))  
  
# Shows the result.  
centers = model.clusterCenters()  
print("Cluster Centers: ")  
for center in centers:  
    print(center)
```



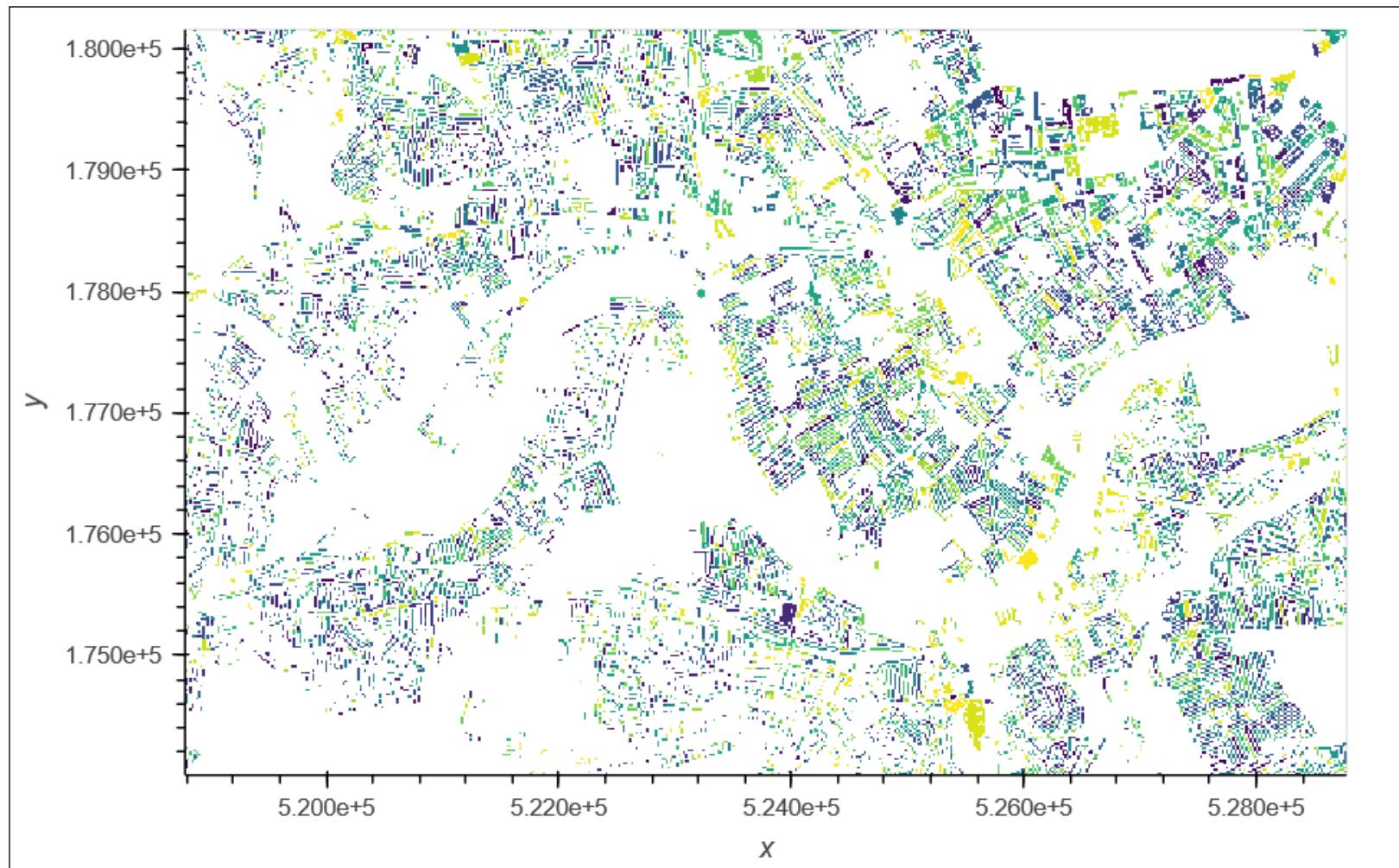
İTÜ

RESULTS AND DISCUSSION



İTÜ

RESULTS AND DISCUSSION



Visualisation of Building-scale Energy Efficiency Analytics Using Datasader



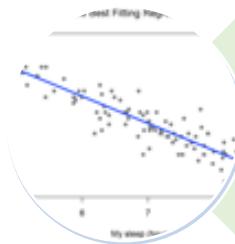
İTÜ



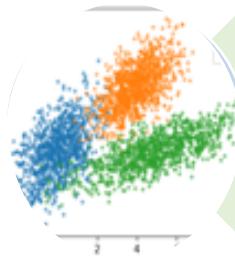
RESULTS AND DISCUSSION



Data driven tracking of smart cities for climate action



Development of a holistic Geospatial Big Data administration model

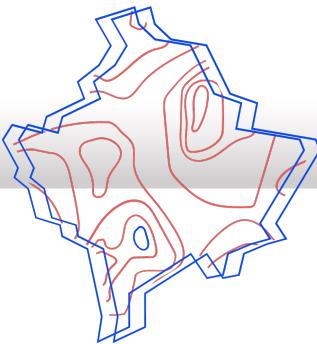


Comparison of the two Geospatial Big Data Analytics frameworks



Revealing the potential of the Geospatial Big Data Analytics

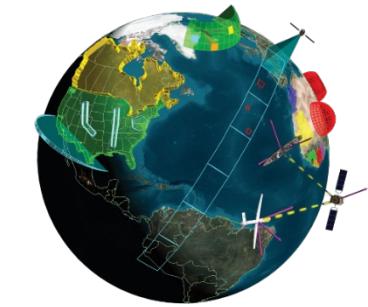




FOSS4G
Prizren, 2023

FOSS4G 2023 – 26 June – 2 July 2023, Prizren – Kosovo

THANK YOU



GEOSPATIAL BIG DATA ANALYTICS FOR SUSTAINABLE SMART CITIES



Muhammed Oguzhan METE

Department of Geomatics Engineering, Istanbul Technical University, Türkiye



İTÜ