SPATIAL DATA MODELLING, SPATIAL DATABASE AND QUERY EXECUTION

Soumya K. Ghosh
Shreya Ghosh
Indian Institute of Technology Kharagpur
The Pennsylvania State University, USA

OGC Stack Winter School, Hyderabad | Dec 30, 2022

OVERVIEW

PennState

Data collected across both space and time & describe a phenomenon

Converting data from external sources by transferring it to your database in files

Include spatial data, represents object defined in a geometric space

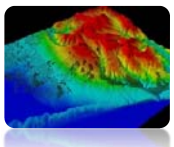
Retrieving a data subset from map layer by working directly with map features.

SPATIAL DATA MODELLING

EXPORTING SPATIAL DATA

SPATIAL DATABASE and TOOL

Spatial Query Execution



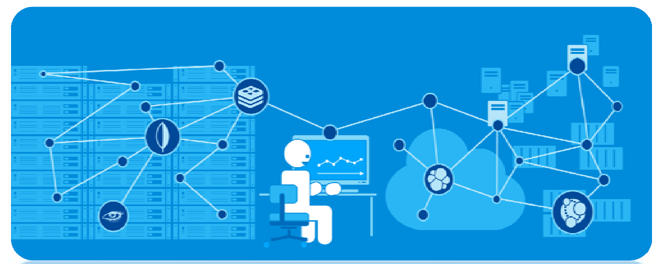
GEOSPATIAL APPLICATIONS…

Identify

- ❖ Identifying the source information
- ❖ Picking the data points that need to be analysed
- ❖ Extracting the relevant information from the data

Analyze

- ❖ Identifying the key values from the extracted data set
- ❖ Learning techniques



Inference

- ❖ Interpreting and reporting the results

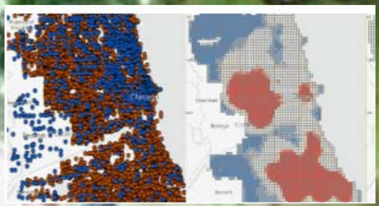
PennState

FAIR – Findable, Accessible, Interoperable and Reusable

“Spatial”: Power of ***WHERE***

- Have you ever looked at a map of crime in your city and tried to figure out what areas have high crime rates?
- Have you explored other types of information, such as school locations, parks, and demographics to try to determine the best location to buy a new home?

Whenever we look at a map, we inherently start turning that map into information by analyzing its contents—*finding patterns, assessing trends, or making decisions*.



Crime Studies

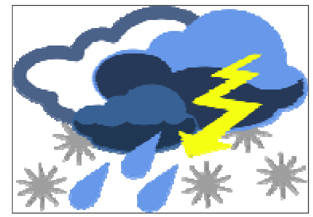
Drought Analysis

Finding optimal paths

Predictions

Spatial Data Modeling

- ❑ Spatial data is comprised of objects in multi-dimensional space



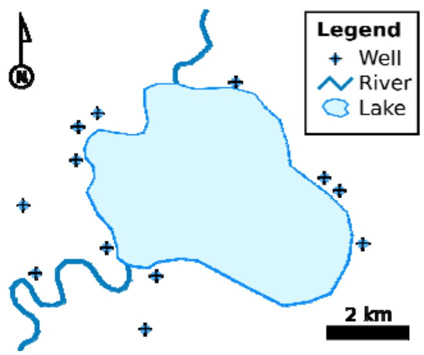
  - ➢ Transportation
  - ➢ Weather prediction
  - ➢ City Planning
  - ➢ Wild life migration patterns
  - ➢ Insurance risk considering location risk profiles
  - ➢ Emergency response determining quickest route to victim
  - ➢ Mobile phone companies tracking phone usage

- ❑ Need a structural representation of spatial data sets – *easy to share, access and analyze!*

What is .shp/ .geom?

A shapefile is an Esri vector data storage format for storing the location, shape, and attributes of geographic features.

It is stored as a set of related files and contains one feature class.




```
Distance(geometry, geometry) : number
Equals(geometry, geometry) : boolean
Disjoint(geometry, geometry) : boolean
Intersects(geometry, geometry) : boolean
Touches(geometry, geometry) : boolean
Crosses(geometry, geometry) : boolean
Overlaps(geometry, geometry) : boolean
Contains(geometry, geometry) : boolean
Length(geometry) : number
Area(geometry) : number
Centroid(geometry) : geometry
```

Spatial Database and Tools

1. https://www.postgresql.org/download/
2. https://postgis.net/install/
3. https://www.pgadmin.org/download/
4. https://qgis.org/en/site/forusers/download.html

Configuration Steps

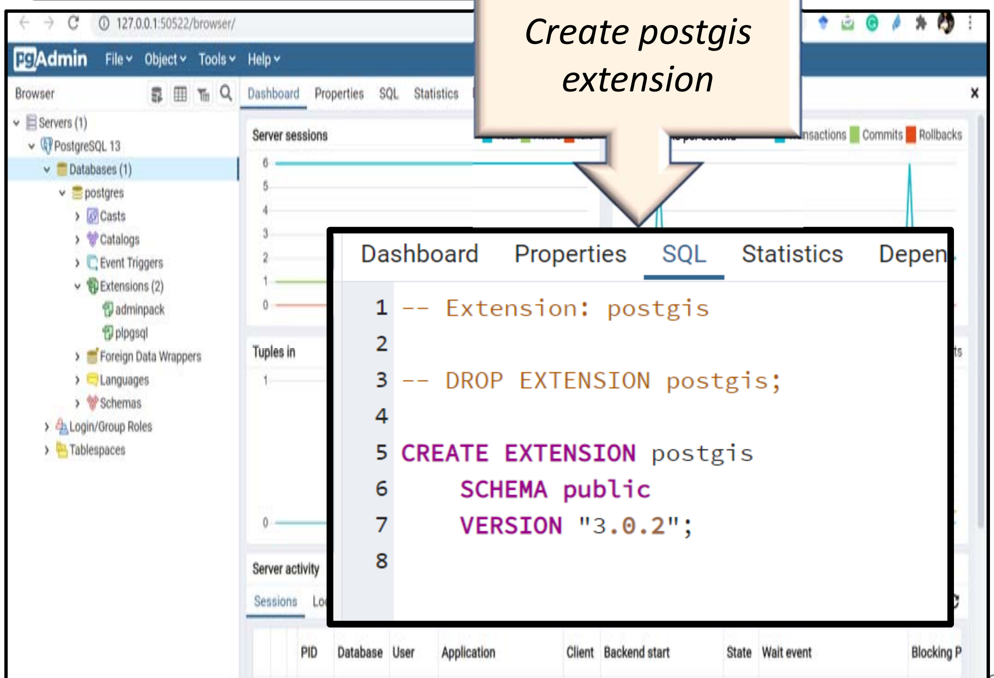
(I) Download and Install PostgreSQL

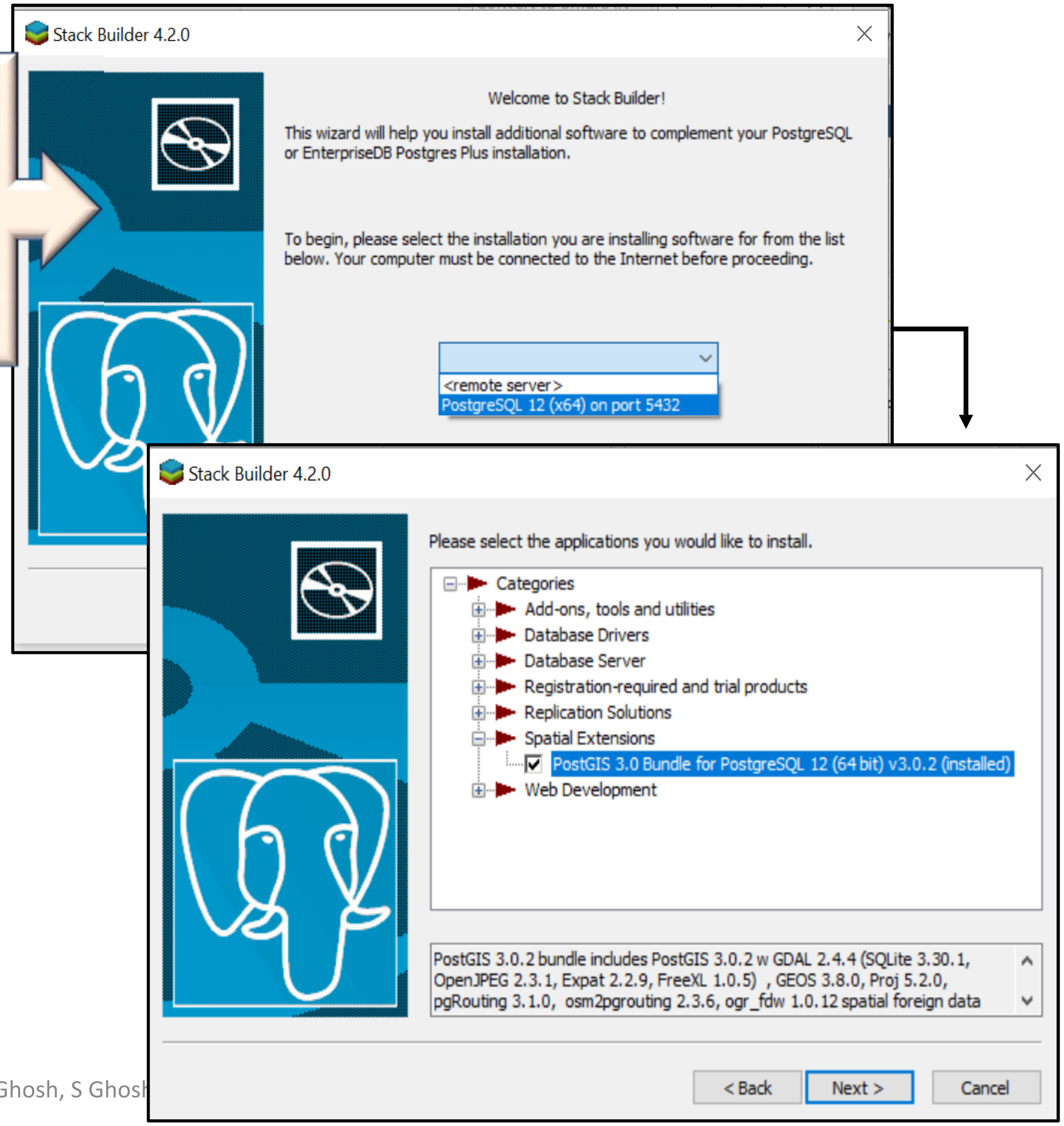
(II) PostGIS Extension

(III) Download and Install pgadmin4

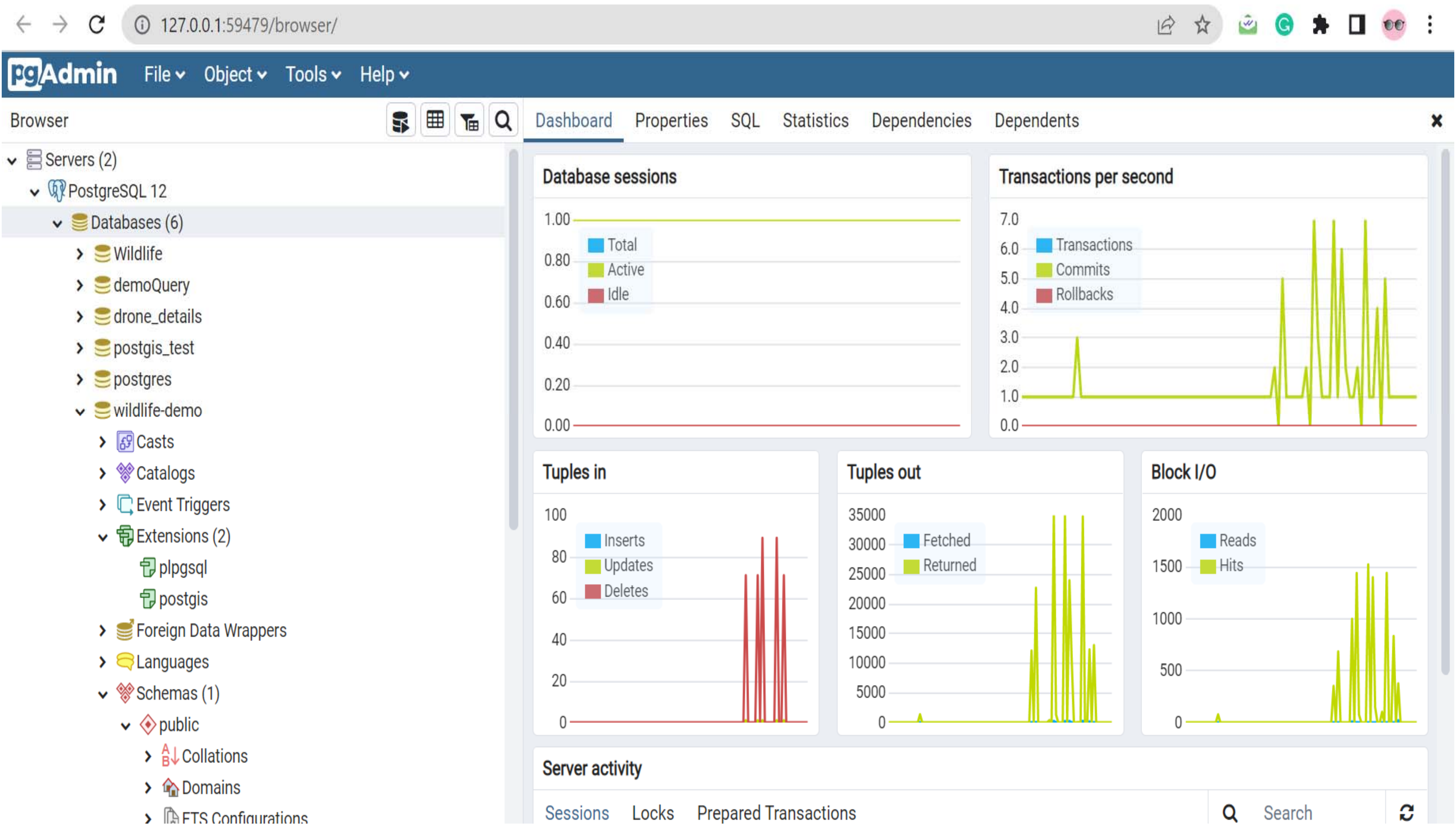
Stackbuilder to add spatial extensions

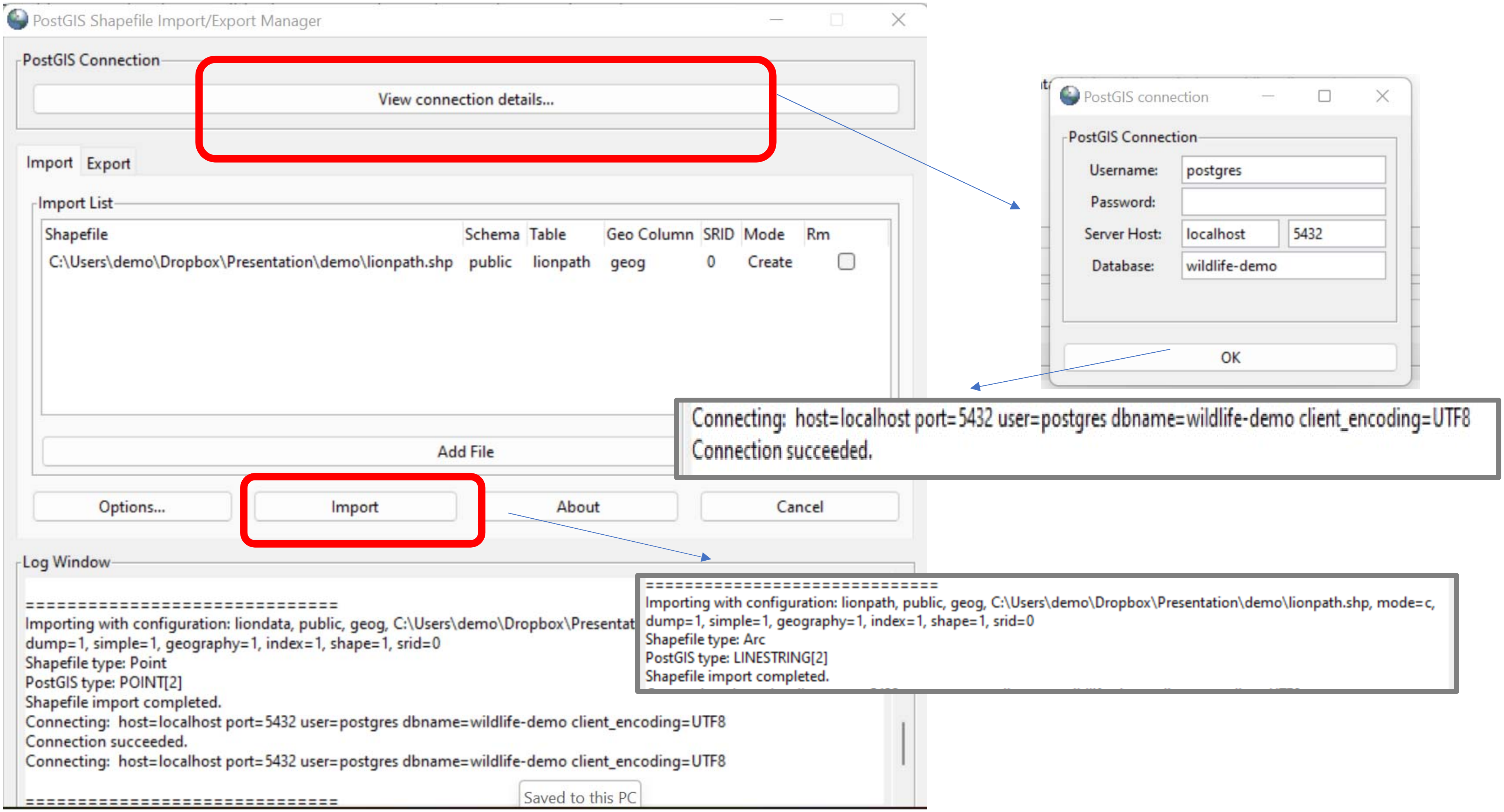
Create postgis extension





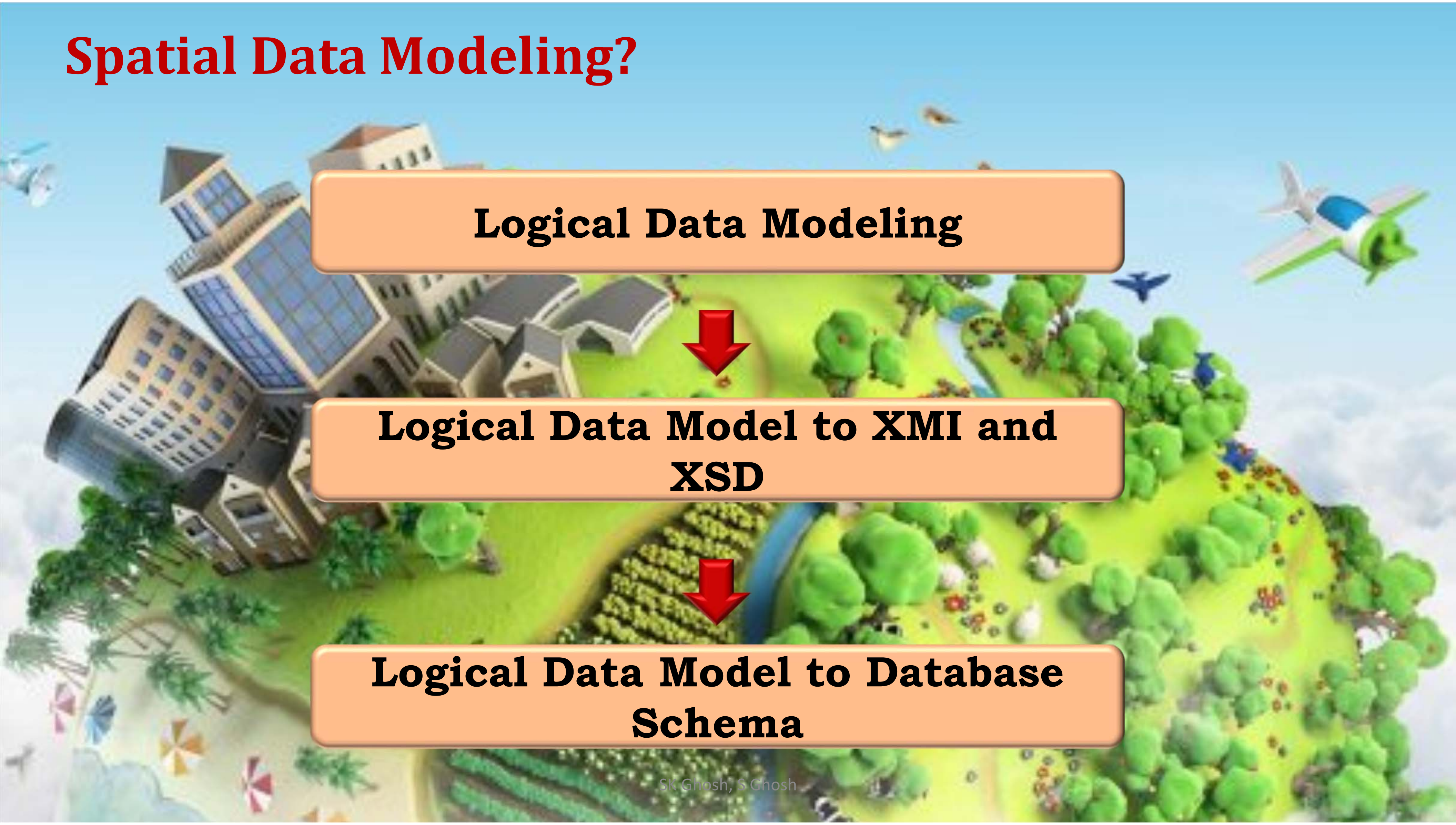
pgAdmin 4
127.0.0.1:60420/browser/
pgAdmin
File
Object
Tools
Help
Browser
Servers (2)
PostgreSQL 12
PostgreSQL 12
Dashboard
Properties
SQL
Statistics
Dependencies
Dependents
Welcome
pgAdmin
Management Tools for PostgreSQL
Feature rich | Maximises PostgreSQL | Open Source
pgAdmin is an Open Source administration and management tool for the PostgreSQL database. It includes a graphical administration interface, an SQL query tool, a procedural code debugger and much more. The tool is designed to answer the needs of developers, DBAs and system administrators alike.
Quick Links
Add New Server
Configure pgAdmin
Getting Started
PostgreSQL Documentation
pgAdmin Website
Planet PostgreSQL
Community Support


127.0.0.1:59479/browser/
pgAdmin
File
Object
Tools
Help
Browser
Servers (2)
PostgreSQL 12
Databases (6)
Wildlife
demoQuery
drone_details
postgis_test
postgres
wildlife-demo
Casts
Catalogs
Event Triggers
Extensions (2)
plpgsql
postgis
Foreign Data Wrappers
Languages
Schemas (1)
public
Collations
Domains
FTS Configurations
Dashboard
Properties
SQL
Statistics
Dependencies
Dependents
Database sessions
Total
Active
Idle
Transactions per second
Transactions
Commits
Rollbacks
Tuples in
Inserts
Updates
Deletes
Tuples out
Fetched
Returned
Block I/O
Reads
Hits
Server activity
Sessions
Locks
Prepared Transactions
Search



PostGIS Shapefile Import/Export Manager
PostGIS Connection
View connection details...
Import
Export
Import List
Shapefile
Schema
Table
Geo Column
SRID
Mode
Rm
C:\Users\demo\Dropbox\Presentation\demo\lionpath.shp
public
lionpath
geog
0
Create
Add File
Options...
Import
About
Cancel
Log Window
==============================
Importing with configuration: liondata, public, geog, C:\Users\demo\Dropbox\Presentat
dump=1, simple=1, geography=1, index=1, shape=1, srid=0
Shapefile type: Point
PostGIS type: POINT[2]
Shapefile import completed.
Connecting: host=localhost port=5432 user=postgres dbname=wildlife-demo client_encoding=UTF8
Connection succeeded.
Connecting: host=localhost port=5432 user=postgres dbname=wildlife-demo client_encoding=UTF8
==============================
Saved to this PC
PostGIS connection
PostGIS Connection
Username:
postgres
Password:
Server Host:
localhost
5432
Database:
wildlife-demo
OK
Connecting: host=localhost port=5432 user=postgres dbname=wildlife-demo client_encoding=UTF8
Connection succeeded.
==============================
Importing with configuration: lionpath, public, geog, C:\Users\demo\Dropbox\Presentation\demo\lionpath.shp, mode=c,
dump=1, simple=1, geography=1, index=1, shape=1, srid=0
Shapefile type: Arc
PostGIS type: LINESTRING[2]
Shapefile import completed.

Spatial Data Modeling?



Logical Data Modeling

↓

Logical Data Model to XMI and XSD

↓

Logical Data Model to Database Schema

Spatial Data Modeling

- ❑ **Logical Data Modeling**
- ❑ Logical Data Model to XMI and XSD
- ❑ Logical Data Model to Database Schema

❑ A logical data model or logical schema is a data model of a specific problem domain expressed independently of a particular database management product or storage technology (physical data model) but in terms of data structures such as relational tables and columns, object-oriented classes, or XML tags.

❑ A logical data model includes

- ❖ entities (tables)
- ❖ attributes (columns/fields) and
- ❖ relationships (keys)

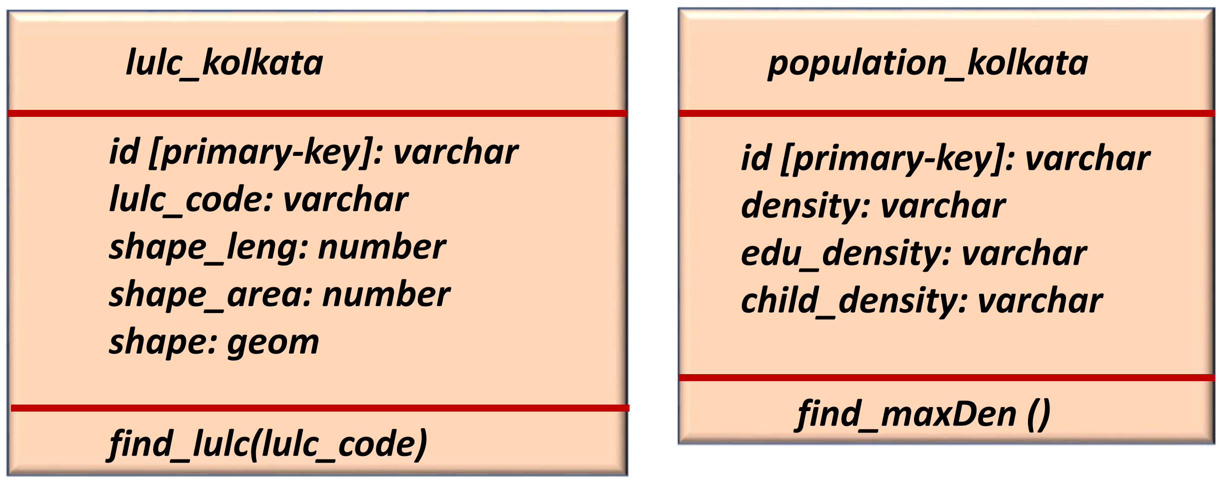
Spatial Data Modeling

- ❑ **Logical Data Modeling**
- ❑ Logical Data Model to XMI and XSD
- ❑ Logical Data Model to Database Schema

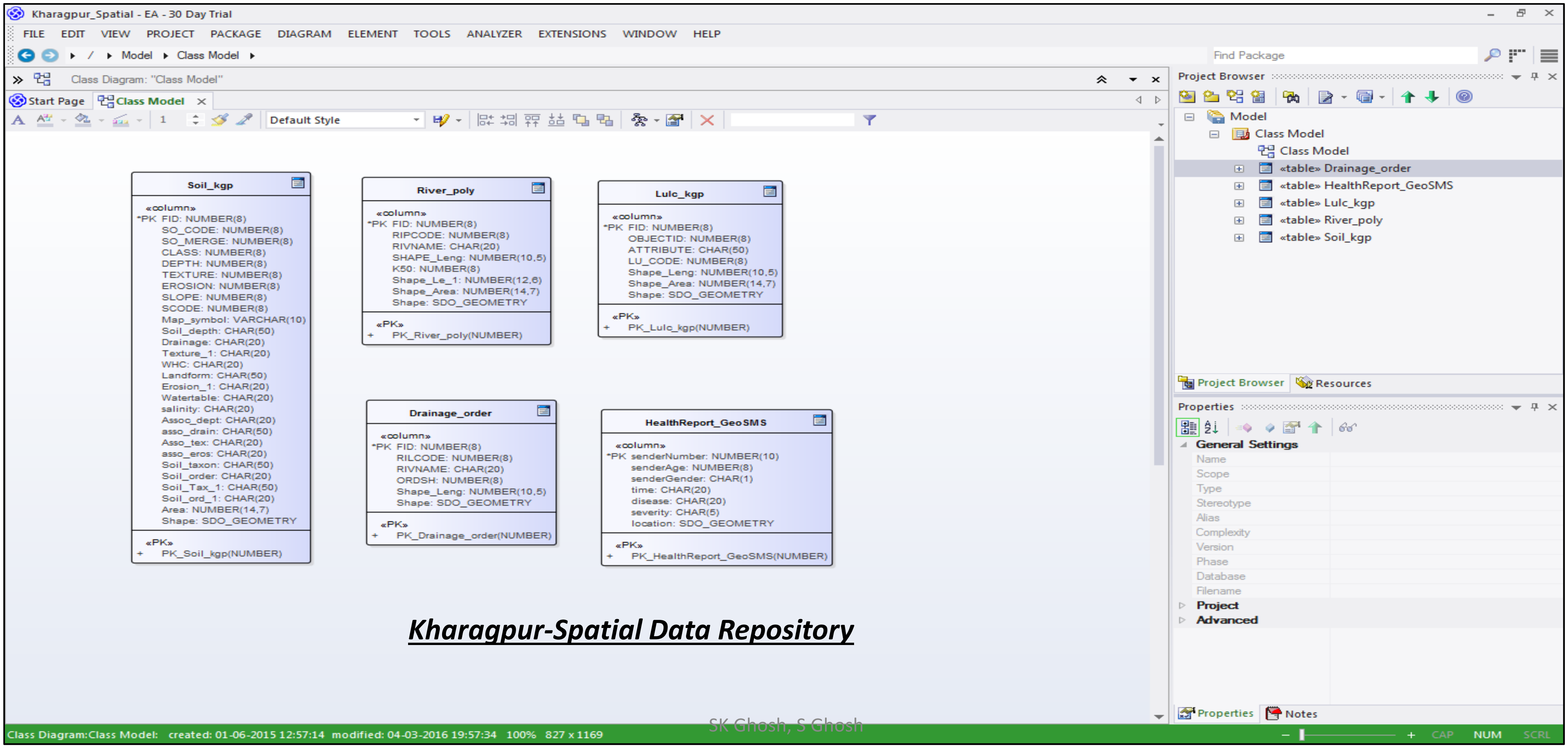
A logical data model (class diagram) includes

- **entities** (tables)
- **attributes** (columns/fields) and
- **relationships** (keys)

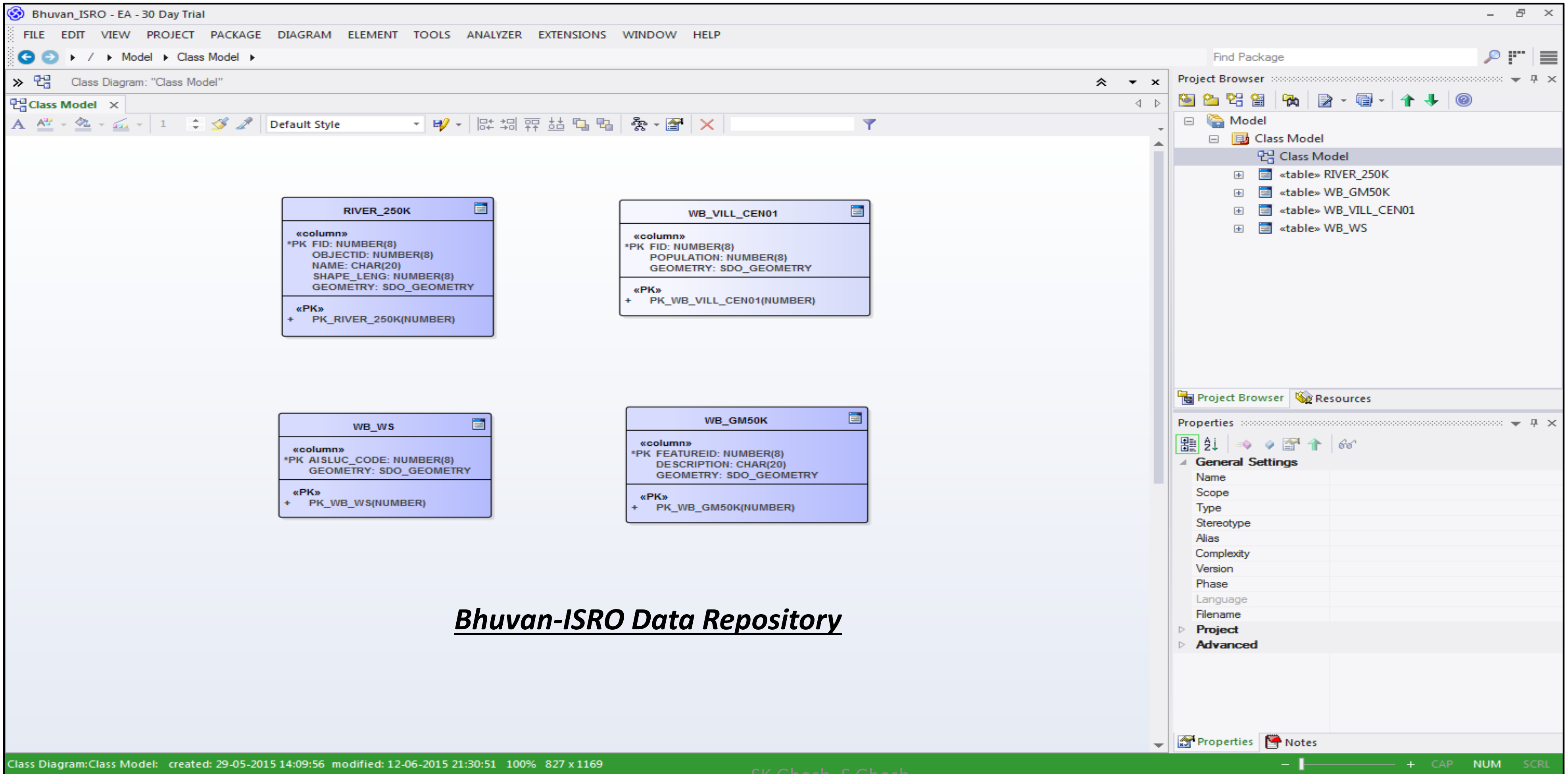
Logical data model of a ROI



Logical Data Model (Class Diagram) of Spatial Data Repositories



Class Diagram of Spatial Data Repositories



Spatial Data Modeling

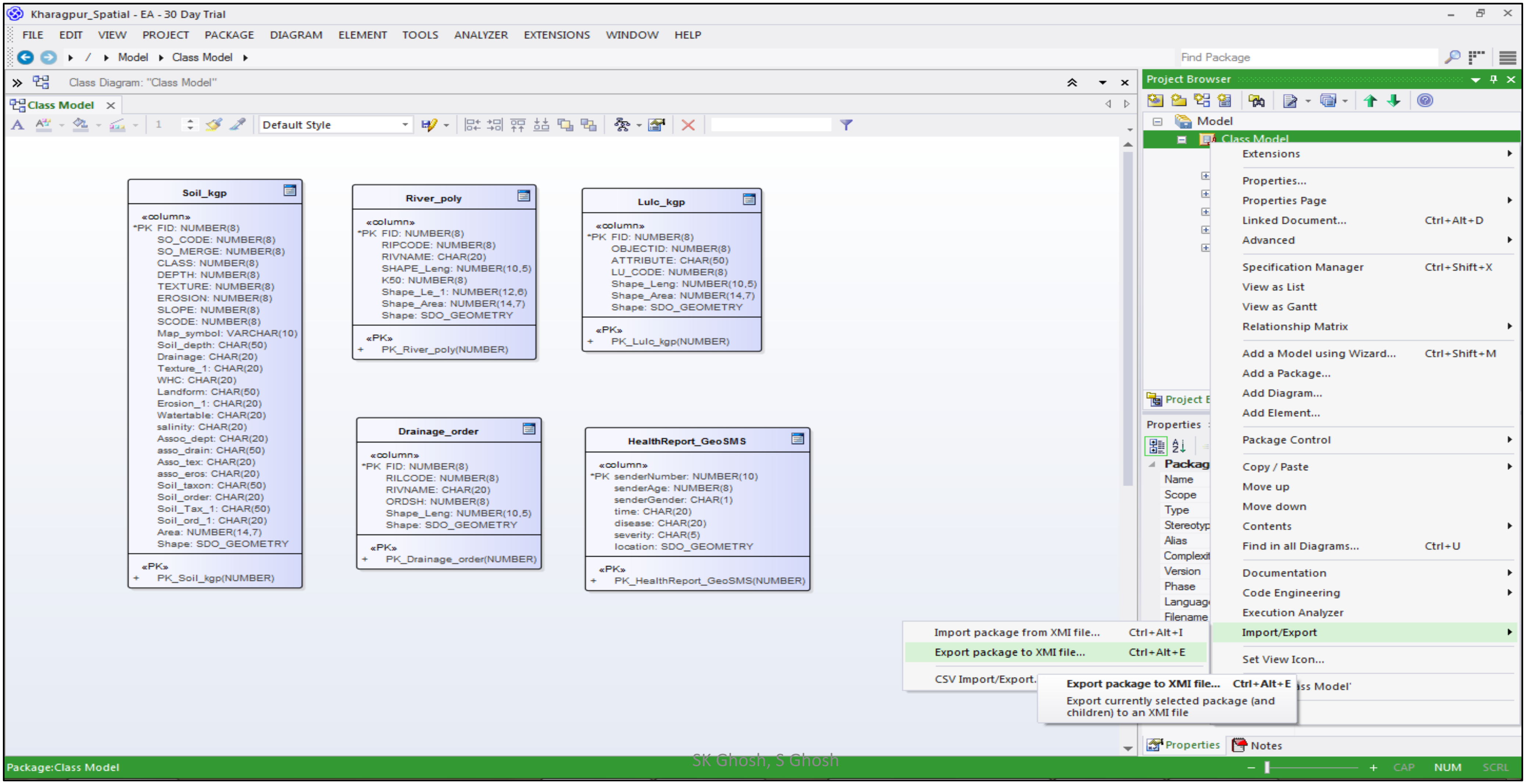
- Logical Data Modeling
- **<u>Logical Data Model to XMI and XSD</u>**
- Logical Data Model to Database Schema

- **<u>XML</u>** is a ***markup language*** that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable

- Main purposes behind the proposal of XMI were as follows:
  - To help programmers, using the Unified Modeling Language (UML), with different languages and development tools to exchange their data models with each other
  - To facilitate in exchanging information about data warehouses

- **<u>XSD</u>** is an XML schema definition language which can be used to express a set of rules to which an XML document must conform in order to be considered 'valid' according to that schema
- It is designed with the intent that determination of a document's validity would produce a collection of information, adhering to specific data types

Exporting Logical Data Model to XMI



XMI of logical data model of *Kharagpur-Spatial Data Repository*

This XML file does not appear to have any style information associated with it. The document tree is shown below.

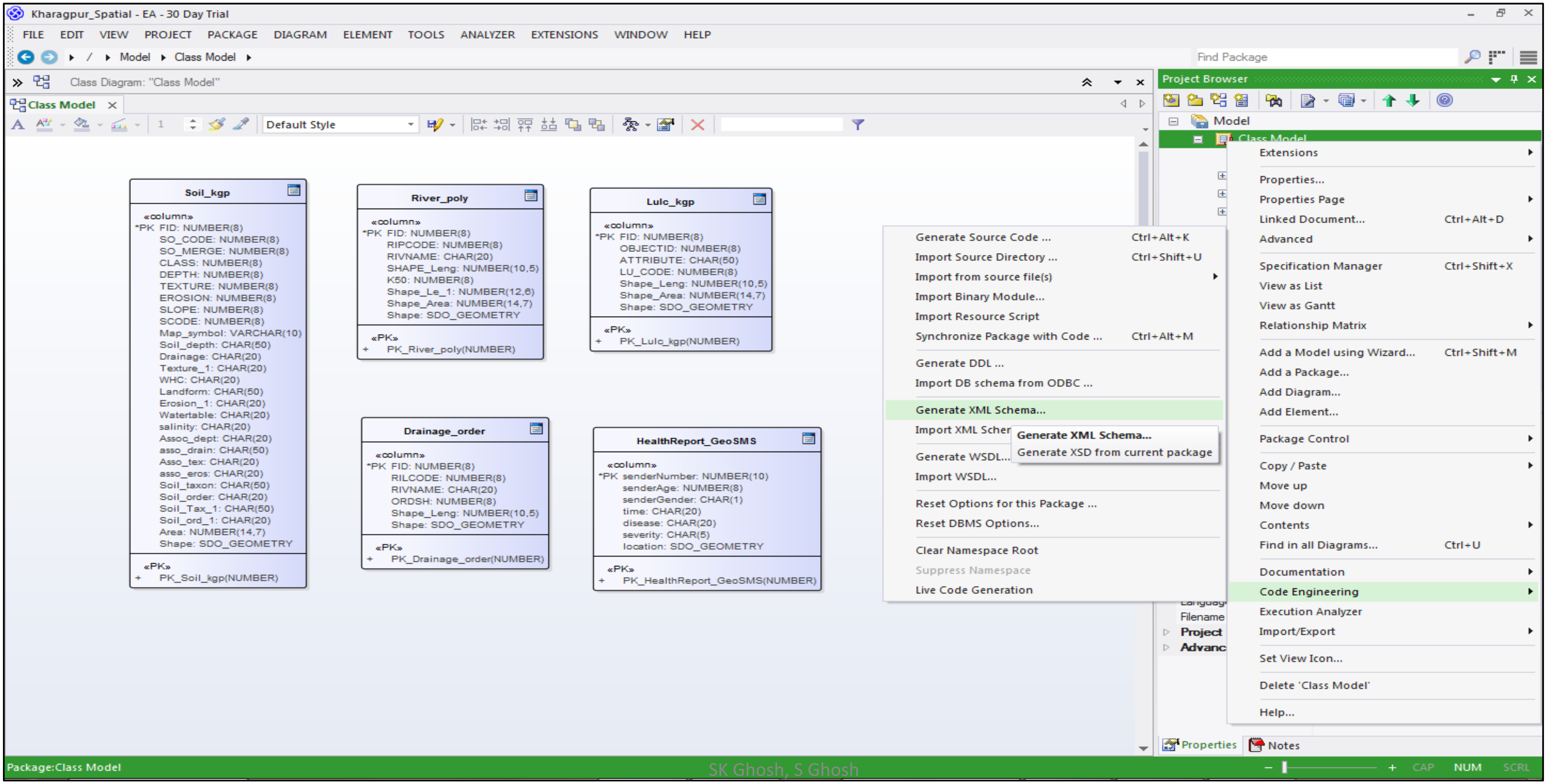

```
▼<XMI xmlns:UML="omg.org/UML1.3" xmi.version="1.1" timestamp="2015-06-12 21:33:08">
  ▼<XMI.header>
    ▼<XMI.documentation>
        <XMI.exporter>Enterprise Architect</XMI.exporter>
        <XMI.exporterVersion>2.5</XMI.exporterVersion>
      </XMI.documentation>
    </XMI.header>
  ▼<XMI.content>
    ▼<UML:Model name="EA Model" xmi.id="MX_EAID_E3D8CCE6_DFE8_4f9e_B206_3C5C57D9A064">
      ▼<UML:Namespace.ownedElement>
          <UML:Class name="EARootClass" xmi.id="EAID_11111111_5487_4080_A7F4_41526CB0AA00" isRoot="true" isLeaf="false" isAbstract="false"/>
        ▼<UML:Package name="Class Model" xmi.id="EAPK_E3D8CCE6_DFE8_4f9e_B206_3C5C57D9A064" isRoot="false" isLeaf="false" isAbstract="false" visibility="public">
          ▼<UML:ModelElement.taggedValue>
              <UML:TaggedValue tag="parent" value="EAPK_A87F0067_DFDC_48e9_B634_E48448AB9A68"/>
              <UML:TaggedValue tag="ea_package_id" value="2"/>
              <UML:TaggedValue tag="created" value="2015-06-01 12:57:14"/>
              <UML:TaggedValue tag="modified" value="2015-06-01 12:57:14"/>
              <UML:TaggedValue tag="iscontrolled" value="FALSE"/>
              <UML:TaggedValue tag="isnamespace" value="1"/>
              <UML:TaggedValue tag="lastloaddate" value="2015-06-01 12:57:14"/>
              <UML:TaggedValue tag="lastsavedate" value="2015-06-01 12:57:14"/>
              <UML:TaggedValue tag="isprotected" value="FALSE"/>
              <UML:TaggedValue tag="usedtd" value="FALSE"/>
              <UML:TaggedValue tag="logxml" value="FALSE"/>
              <UML:TaggedValue tag="tpos" value="6"/>
              <UML:TaggedValue tag="packageFlags" value="isModel=1;VICON=3;CRC=0;"/>
              <UML:TaggedValue tag="batchsave" value="0"/>
              <UML:TaggedValue tag="batchload" value="0"/>
              <UML:TaggedValue tag="phase" value="1.0"/>
              <UML:TaggedValue tag="status" value="Proposed"/>
              <UML:TaggedValue tag="complexity" value="1"/>
              <UML:TaggedValue tag="ea_stype" value="Public"/>
              <UML:TaggedValue tag="tpos" value="6"/>
              <UML:TaggedValue tag="genfile" value="C:\Users\Moni\Desktop\Kharagpur_Spatial.sql"/>
            </UML:ModelElement.taggedValue>
          ▼<UML:Namespace.ownedElement>
            ▼<UML:Class name="Lulc_kgp" xmi.id="EAID_2F20EFB5_E6AF_4b51_8D84_3FAA93670C08" visibility="public" namespace="EAPK_E3D8CCE6_DFE8_4f9e_B206_3C5C57D9A064" isRoot="false"
              isLeaf="false" isAbstract="false" isActive="false">
              ▼<UML:ModelElement.stereotype>
```

XMI of logical data model of *Bhuvan-ISRO Spatial Data Repository*

This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
▼<XMI xmlns:UML="omg.org/UML1.3" xmi.version="1.1" timestamp="2015-06-12 21:34:18">
  ▼<XMI.header>
    ▼<XMI.documentation>
        <XMI.exporter>Enterprise Architect</XMI.exporter>
        <XMI.exporterVersion>2.5</XMI.exporterVersion>
      </XMI.documentation>
    </XMI.header>
  ▼<XMI.content>
    ▼<UML:Model name="EA Model" xmi.id="MX_EAID_020C96A0_4321_4f0b_9E1D_9454A780AE92">
      ▼<UML:Namespace.ownedElement>
          <UML:Class name="EARootClass" xmi.id="EAID_11111111_5487_4080_A7F4_41526CB0AA00" isRoot="true" isLeaf="false" isAbstract="false"/>
        ▼<UML:Package name="Class Model" xmi.id="EAPK_020C96A0_4321_4f0b_9E1D_9454A780AE92" isRoot="false" isLeaf="false" isAbstract="false" visibility="public">
          ▼<UML:ModelElement.taggedValue>
              <UML:TaggedValue tag="parent" value="EAPK_99A749A0_6B1A_4b84_8BB2_6A2EE6804F47"/>
              <UML:TaggedValue tag="ea_package_id" value="2"/>
              <UML:TaggedValue tag="created" value="2015-05-29 14:09:56"/>
              <UML:TaggedValue tag="modified" value="2015-05-29 14:09:56"/>
              <UML:TaggedValue tag="iscontrolled" value="FALSE"/>
              <UML:TaggedValue tag="isnamespace" value="1"/>
              <UML:TaggedValue tag="lastloaddate" value="2015-05-29 14:09:56"/>
              <UML:TaggedValue tag="lastsavedate" value="2015-05-29 14:09:56"/>
              <UML:TaggedValue tag="isprotected" value="FALSE"/>
              <UML:TaggedValue tag="usedtd" value="FALSE"/>
              <UML:TaggedValue tag="logxml" value="FALSE"/>
              <UML:TaggedValue tag="tpos" value="6"/>
              <UML:TaggedValue tag="packageFlags" value="isModel=1;VICON=3;CRC=0;"/>
              <UML:TaggedValue tag="batchsave" value="0"/>
              <UML:TaggedValue tag="batchload" value="0"/>
              <UML:TaggedValue tag="phase" value="1.0"/>
              <UML:TaggedValue tag="status" value="Proposed"/>
              <UML:TaggedValue tag="complexity" value="1"/>
              <UML:TaggedValue tag="ea_stype" value="Public"/>
              <UML:TaggedValue tag="tpos" value="6"/>
              <UML:TaggedValue tag="genfile" value="C:\Users\Moni\Desktop\Bhuvan_ISRO.sql"/>
            </UML:ModelElement.taggedValue>
          ▼<UML:Namespace.ownedElement>
            ▼<UML:Class name="RIVER_250K" xmi.id="EAID_3EAA15EC_E878_4a99_A838_0CB534DED917" visibility="public" namespace="EAPK_020C96A0_4321_4f0b_9E1D_9454A780AE92" isRoot="false"
              isLeaf="false" isAbstract="false" isActive="false">
              ▼<UML:ModelElement.stereotype>
```

Exporting Logical Data Model to XSD



XSD of logical data model of *Kharagpur-Spatial Data Repository*

Getting Started | Imported From Firef... | Other bookmarks

This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
▼<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
   <xs:element name="Soil_kgp" type="Soil_kgp"/>
 ▼<xs:complexType name="Soil_kgp">
   ▼<xs:sequence>
      <xs:element name="FID" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="SO_CODE" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="SO_MERGE" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="CLASS" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="DEPTH" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="TEXTURE" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="EROSION" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="SLOPE" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="SCODE" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Map_symbol" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Soil_depth" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Drainage" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Texture_1" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="WHC" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Landform" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Erosion_1" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Watertable" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="salinity" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Assoc_dept" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="asso_drain" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Asso_tex" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="asso_eros" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Soil_taxon" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Soil_order" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Soil_Tax_1" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Soil_ord_1" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Area" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="Shape" type="xs:string" minOccurs="1" maxOccurs="1"/>
    </xs:sequence>
  </xs:complexType>
  <xs:element name="River_poly" type="River_poly"/>
 ▼<xs:complexType name="River_poly">
   ▼<xs:sequence>
      <xs:element name="FID" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="RIPCODE" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="RIVNAME" type="xs:string" minOccurs="1" maxOccurs="1"/>
```

XSD of logical data model of *Bhuvan-ISRO Spatial Data Repository*

This XML file does not appear to have any style information associated with it. The document tree is shown below.

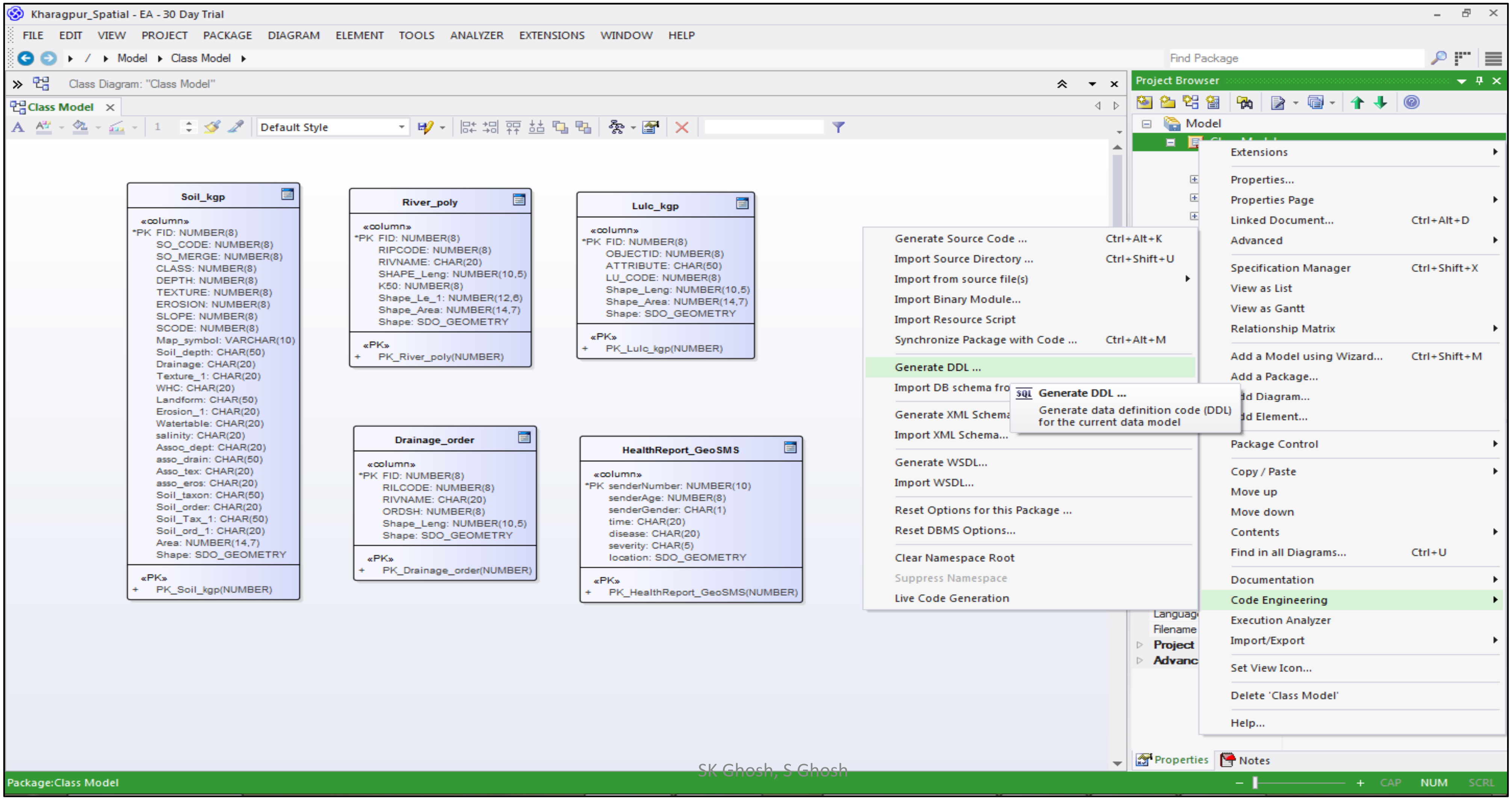

```
▼<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
   <xs:element name="RIVER_250K" type="RIVER_250K"/>
  ▼<xs:complexType name="RIVER_250K">
    ▼<xs:sequence>
       <xs:element name="FID" type="xs:string" minOccurs="1" maxOccurs="1"/>
       <xs:element name="OBJECTID" type="xs:string" minOccurs="1" maxOccurs="1"/>
       <xs:element name="NAME" type="xs:string" minOccurs="1" maxOccurs="1"/>
       <xs:element name="SHAPE_LENG" type="xs:string" minOccurs="1" maxOccurs="1"/>
       <xs:element name="GEOMETRY" type="xs:string" minOccurs="1" maxOccurs="1"/>
     </xs:sequence>
   </xs:complexType>
   <xs:element name="WB_VILL_CEN01" type="WB_VILL_CEN01"/>
  ▼<xs:complexType name="WB_VILL_CEN01">
    ▼<xs:sequence>
       <xs:element name="FID" type="xs:string" minOccurs="1" maxOccurs="1"/>
       <xs:element name="POPULATION" type="xs:string" minOccurs="1" maxOccurs="1"/>
       <xs:element name="GEOMETRY" type="xs:string" minOccurs="1" maxOccurs="1"/>
     </xs:sequence>
   </xs:complexType>
   <xs:element name="WB_WS" type="WB_WS"/>
  ▼<xs:complexType name="WB_WS">
    ▼<xs:sequence>
       <xs:element name="AISLUC_CODE" type="xs:string" minOccurs="1" maxOccurs="1"/>
       <xs:element name="GEOMETRY" type="xs:string" minOccurs="1" maxOccurs="1"/>
     </xs:sequence>
   </xs:complexType>
   <xs:element name="WB_GM50K" type="WB_GM50K"/>
  ▼<xs:complexType name="WB_GM50K">
    ▼<xs:sequence>
       <xs:element name="FEATUREID" type="xs:string" minOccurs="1" maxOccurs="1"/>
       <xs:element name="DESCRIPTION" type="xs:string" minOccurs="1" maxOccurs="1"/>
       <xs:element name="GEOMETRY" type="xs:string" minOccurs="1" maxOccurs="1"/>
     </xs:sequence>
   </xs:complexType>
 </xs:schema>
```

Spatial Data Modeling

- ❑ Logical Data Modeling
- ❑ Logical Data Model to XMI and XSD
- ❑ **Logical Data Model to Database Schema**

- ❑ A database schema of a database system is *its structure described in a formal language supported by the database management system* (DBMS)
- ❑ It refers to the organization of data as a blueprint of how a database is constructed (*divided into database tables in case of Relational Databases*)
- ❑ In the context of Oracle databases, a *schema object is a logical data storage structure*
- ❑ In general, the *language by which the database schema is described is called Data Definition language* (DDL)

Exporting Logical Data Model to DDL



SK Ghosh, S Ghosh

DDL of *Kharagpur-Spatial Data Repository*

```sql
--  --------------------------------------------------
--   Generated by Enterprise Architect Version 11.0.1106
--   Created On : Thursday, 11 June, 2015
--   DBMS       : Oracle
--  --------------------------------------------------
--  Drop Tables, Stored Procedures and Views
DROP TABLE Drainage_order CASCADE CONSTRAINTS;
DROP TABLE HealthReport_GeoSMS CASCADE CONSTRAINTS;
DROP TABLE Lulc_kgp CASCADE CONSTRAINTS;
DROP TABLE River_poly CASCADE CONSTRAINTS;
DROP TABLE Soil_kgp CASCADE CONSTRAINTS;
--  Create Tables
CREATE TABLE Drainage_order
(
    FID          NUMBER(8) NOT NULL,
    RILCODE      NUMBER(8),
    RIVNAME      CHAR(20),
    ORDSH        NUMBER(8),
    Shape_Leng   NUMBER(10,5),
    Shape        SDO_GEOMETRY
);
CREATE TABLE HealthReport_GeoSMS
(
    senderNumber   NUMBER(10) NOT NULL,
    senderAge      NUMBER(8),
    senderGender   CHAR(1),
    time           CHAR(20),
    disease        CHAR(20),
    severity       CHAR(5),
    location       SDO_GEOMETRY
);
CREATE TABLE Lulc_kgp
(
    FID          NUMBER(8) NOT NULL,
    OBJECTID     NUMBER(8),
    ATTRIBUTE    CHAR(50),
    LU_CODE      NUMBER(8),
    Shape_Leng   NUMBER(10,5),
```


```sql
--  --------------------------------------------------
--   Generated by Enterprise Architect Version 11.0.1106
--   Created On : Thursday, 11 June, 2015
--   DBMS       : Oracle
--  --------------------------------------------------
--  Drop Tables, Stored Procedures and Views
DROP TABLE Drainage_order CASCADE CONSTRAINTS;
DROP TABLE HealthReport_GeoSMS CASCADE CONSTRAINTS;
DROP TABLE Lulc_kgp CASCADE CONSTRAINTS;
DROP TABLE River_poly CASCADE CONSTRAINTS;
DROP TABLE Soil_kgp CASCADE CONSTRAINTS;
--  Create Tables
CREATE TABLE Drainage_order
(
    FID          NUMBER(8) NOT NULL,
    RILCODE      NUMBER(8),
    RIVNAME      CHAR(20),
    ORDSH        NUMBER(8),
    Shape_Leng   NUMBER(10,5),
    Shape        SDO_GEOMETRY
);
CREATE TABLE HealthReport_GeoSMS
(
    senderNumber   NUMBER(10) NOT NULL,
    senderAge      NUMBER(8),
    senderGender   CHAR(1),
    time           CHAR(20),
    disease        CHAR(20),
    severity       CHAR(5),
    location       SDO_GEOMETRY
);
CREATE TABLE Lulc_kgp
(
    FID          NUMBER(8) NOT NULL,
    OBJECTID     NUMBER(8),
    ATTRIBUTE    CHAR(50),
    LU_CODE      NUMBER(8),
    Shape_Leng   NUMBER(10,5),
```

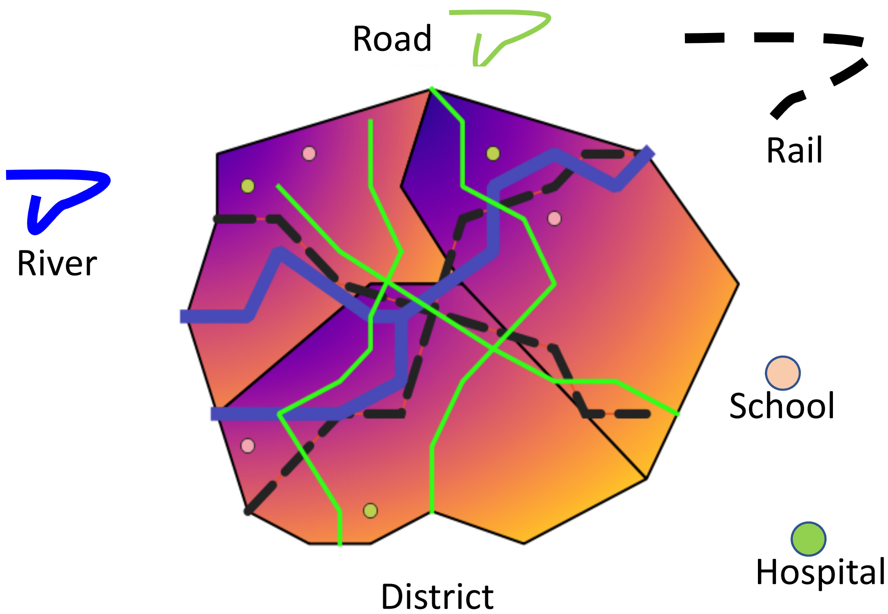

```sql
    SCODE        NUMBER(8),
    Map_symbol   VARCHAR(10),
    Soil_depth   CHAR(50),
    Drainage     CHAR(20),
    Texture_1    CHAR(20),
    WHC          CHAR(20),
    Landform     CHAR(50),
    Erosion_1    CHAR(20),
    Watertable   CHAR(20),
    salinity     CHAR(20),
    Assoc_dept   CHAR(20),
    asso_drain   CHAR(50),
    Asso_tex     CHAR(20),
    asso_eros    CHAR(20),
    Soil_taxon   CHAR(50),
    Soil_order   CHAR(20),
    Soil_Tax_1   CHAR(50),
    Soil_ord_1   CHAR(20),
    Area         NUMBER(14,7),
    Shape        SDO_GEOMETRY
);
--  Create Primary Key Constraints
ALTER TABLE Drainage_order ADD CONSTRAINT PK_Drai
    PRIMARY KEY (FID)
 USING INDEX ;
ALTER TABLE HealthReport_GeoSMS ADD CONSTRAINT PK_
    PRIMARY KEY (senderNumber)
 USING INDEX ;
ALTER TABLE Lulc_kgp ADD CONSTRAINT PK_Lulc_kgp
    PRIMARY KEY (FID)
 USING INDEX ;
ALTER TABLE River_poly ADD CONSTRAINT PK_River_po
    PRIMARY KEY (FID)
 USING INDEX ;
ALTER TABLE Soil_kgp ADD CONSTRAINT PK_Soil_kgp
    PRIMARY KEY (FID)
 USING INDEX ;
```

DDL of *Bhuvan-ISRO Spatial Data Repository*

```sql
--  ----------------------------------------------------
--  Generated by Enterprise Architect Version 11.0.1106
--  Created On : Thursday, 11 June, 2015
--  DBMS       : Oracle
--  ----------------------------------------------------
-- Drop Tables, Stored Procedures and Views
DROP TABLE RIVER_250K CASCADE CONSTRAINTS;
DROP TABLE WB_GM50K CASCADE CONSTRAINTS;
DROP TABLE WB_VILL_CEN01 CASCADE CONSTRAINTS;
DROP TABLE WB_WS CASCADE CONSTRAINTS;
-- Create Tables
CREATE TABLE RIVER_250K
(
    FID          NUMBER(8) NOT NULL,
    OBJECTID     NUMBER(8),
    NAME         CHAR(20),
    SHAPE_LENG   NUMBER(8),
    GEOMETRY     SDO_GEOMETRY
);
CREATE TABLE WB_GM50K
(
    FEATUREID    NUMBER(8) NOT NULL,
    DESCRIPTION  CHAR(20),
    GEOMETRY     SDO_GEOMETRY
);
CREATE TABLE WB_VILL_CEN01
(
    FID          NUMBER(8) NOT NULL,
    POPULATION   NUMBER(8),
    GEOMETRY     SDO_GEOMETRY
);
CREATE TABLE WB_WS
(
    AISLUC_CODE  NUMBER(8) NOT NULL,
    GEOMETRY     SDO_GEOMETRY
);
-- Create Primary Key Constraints
ALTER TABLE RIVER_250K ADD CONSTRAINT PK_RIVER_250K
    PRIMARY KEY (FID)
 USING INDEX ;

ALTER TABLE WB_GM50K ADD CONSTRAINT PK_WB_GM50K
    PRIMARY KEY (FEATUREID)
 USING INDEX ;

ALTER TABLE WB_VILL_CEN01 ADD CONSTRAINT PK_WB_VILL_CEN01
    PRIMARY KEY (FID)
 USING INDEX ;

ALTER TABLE WB_WS ADD CONSTRAINT PK_WB_WS
    PRIMARY KEY (AISLUC_CODE)
 USING INDEX ;
```

Demo I *(Synthetic Dataset)*

- **Road** [gid, road_name, road_km, road_type, geom]
- **Rail** [gid, track_type, track_km, geom]
- **Hospital** [gid, hospital_name, hospital_type, geom]
- **School** [gid, school_name, school_type, geom]
- **District** [gid, dist_name, dist_perimeter, dist_area, geom]
- **River** [gid, river_name, river_km, geom]



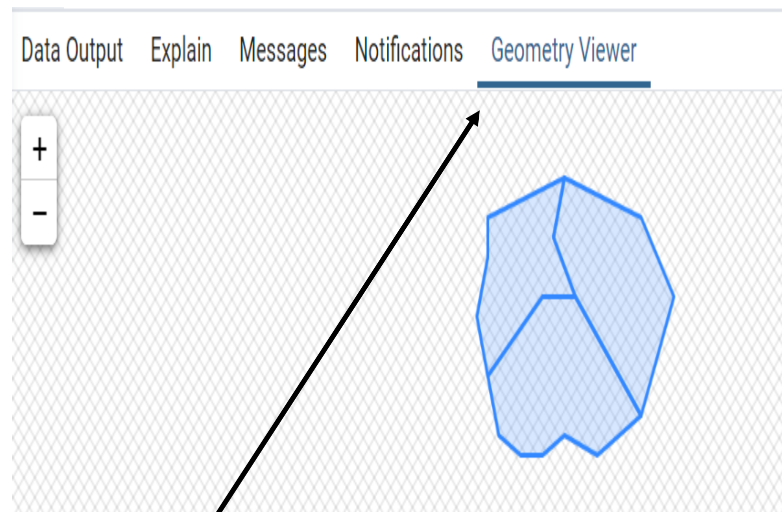
Sample SQL Queries

Extract data from tables

SELECT * FROM hospital;

Data Output | Explain | Messages | Notifications | Geometry Viewer

	gid [PK] integer	hospital_n character varying (2)	hospital_t character varying (16)	geom geometry
1	1	SH	State	010100000000000...
2	2	CH	Central	010100000000000...
3	3	PV	Private	010100000000000...



SELECT * FROM district;

Data Output | Explain | Messages | Notifications | Geometry Viewer

	gid [PK] integer	dist_name character varying (6)	dist_pari smallint	dist_area smallint	geom geometry
1	1	DIST_1	[null]	[null]	010600000001000...
2	2	DIST_2	[null]	[null]	010600000001000...
3	3	DIST_3	[null]	[null]	010600000001000...

Spatial Query : Buffer

- **Point Buffer Query**

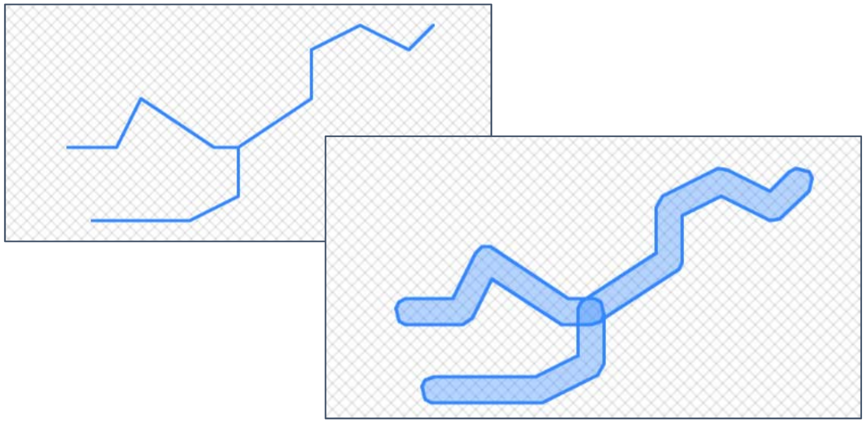
```
SELECT hospital_n, ST_Buffer(hospital.geom, 4)
FROM hospital;
```

- **Polyline Buffer Query**

```
SELECT river_name, geom, ST_Buffer(river.geom, 0.5)
FROM river;
```

Data Output | Explain | Messages | Notifications | Geometry Viewer

	river_name character varying (2)	geom geometry	st_buffer geometry
1	R1	010500000001000...	01030000000100...
2	R2	010500000001000...	01030000000100...



SK Ghosh, S Ghosh

Spatial Query : Cross and Touch Operation

- **Cross Operation**

```
SELECT road.road_name, rail.track_type
FROM road, rail
WHERE ST_Crosses(road.geom, rail.geom);
```

- **Touch Operation**

```
SELECT b.dist_name
FROM district a, district b
WHERE ST_Touches(a.geom, b.geom)
AND a.dist_name = DIST_1';
```

Spatial Query : Intersect Operation and Area–Perimeter Calculation

- **Intersect Operation**

```
SELECT d.dist_name, r.road_name
FROM district d, road r
WHERE ST_Intersects(d.geom, r.geom)
ORDER BY (dist_name);
```

- **Area and Perimeter calculation of districts**

```
SELECT dist_name, ST_Area(geom) AS Area,
ST_PERIMETER(district.geom) AS Perimeter
FROM district
ORDER BY ST_Area(geom) DESC;
```

Spatial Query : Nearest Neighbor

- **Nearest Neighbor Query**

```
SELECT school.gid, school.school_nam,
ST_Distance(school.geom, hospital.geom) AS distance
FROM school, hospital
WHERE hospital.hospital_n = 'SH'
ORDER BY ST_Distance(school.geom, hospital.geom) ASC;
```

Demo II *(IIT KGP Dataset)*

	gid [PK] integer	r_name character varying (50)	start_pnt character varying (50)	end_pnt character varying (50)	const_type character varying (50)	no_of_lane smallint	type character varying (50)	speed_lim smallint	direction character varying (50)	geom geometry
1	110	IIT Main Road	Puri Gate	Prembazzar	Pakka	2	[null]	30	N-S	010500000001000...
2	226	IIT Main Road	Puri Gate	Prembazzar	Pakka	2	[null]	30	N-S	010500000001000...
3	375	IIT Main Road	Puri Gate	Prembazzar	Pakka	2	[null]	30	N-S	010500000001000...
4	383	Nursery Road	Dreamland	Super Duper	Pakka	2	[null]	30	E-W	010500000001000...
5	386	Nursery Road	Dreamland	Super Duper	Pakka	2	[null]	30	E-W	010500000001000...

road_network

Data Output | Explain | Messages | Notifications | Geometry Viewer

	gid [PK] integer	osm_id character varying (254)	name character varying (254)	amenity character varying (254)	geom geometry
1	6	[null]	Indian Institute of Technology ...	university	010600000001000...
2	3	7255545	TSG Lake	Lake/Pool	010600000001000...
3	7	[null]	New Wagon Shop	Shop	010600000001000...
4	13	[null]	Police Quarters	[null]	010600000001000...
5	14	[null]	Jubilee Park	[null]	010600000001000...
6	22	[null]	VIKRAMSHILA SCHOOL OF M...	[null]	010600000001000...

kgp_poi

Spatial Query : Nearest Neighbor

- **Nearest Neighbor Query**

```
SELECT school.gid, school.school_nam,
ST_Distance(school.geom, hospital.geom) AS distance
FROM school, hospital
WHERE hospital.hospital_n = 'SH'
ORDER BY ST_Distance(school.geom, hospital.geom) ASC;
```

Spatial Query : Areas of the Halls

Select name,geom, st_area(geom) as area from kgp_poi where amenity='Hall of Residence' order by area;

Data Output | Explain | Messages | Notifications | Geometry Viewer

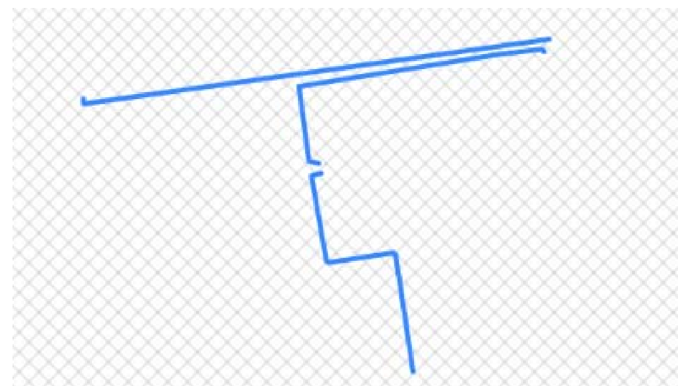
	name character varying (254)	geom geometry	area double precision
1	Sister Nivedita Hall of Residen...	010600000001000...	1793.7335926305623
2	Indira Gandhi Hall of Residence	010600000001000...	2570.7360038824245
3	Mother Terese Hall of Residen...	010600000001000...	5195.289427014087
4	SAM	010600000001000...	7054.38867677244
5	Lal Bahadur Shastri Hall Of Re...	010600000001000...	9494.568795897754
6	Gokhale	010600000001000...	9540.358845181045
7	HJB Hall of Residence	010600000001000	10764.193362758739

Spatial Query : Find the road segments within 50meter of Takshila [st_intersects and st_buffer]

```
select r.r_name,r.geom,kgp.geom from road_NETWORK r, kgp_poi
kgp where st_intersects(r.geom, st_buffer(kgp.geom,50)) and
kgp.name='Takshila Complex'
```

Data Output Explain Messages Notifications Geometry Viewer

	r_name character varying (50)	geom geometry	geom geometry
1	[null]	010500002011OF0...	010600002011OF0...
2	[null]	010500002011OF0...	010600002011OF0...
3	Ardeshir Dalal Avenue	010500002011OF0...	010600002011OF0...
4	Ardeshir Dalal Avenue	010500002011OF0...	010600002011OF0...



Spatial Query : Find the distance between CSE and student halls [st_distance]

```
SELECT B.name, A.name, st_distance(B.geom, A.geom) dis

FROM kgp_poi B, kgp_poi A WHERE B.name='DEPT OF COMPUTER

SCIENCE AND ENGINEERING' and A.amenity='Hall of Residence'

order by dis;
```

Data Output | Explain | Messages | Notifications | Geometry Viewer

	name character varying (254)	name character varying (254)	dis double precision
1	DEPT OF COMPUTER SCIENC...	Indira Gandhi Hall of Residence	167.1834011325025
2	DEPT OF COMPUTER SCIENC...	Mother Terese Hall of Residen...	244.47008446460518
3	DEPT OF COMPUTER SCIENC...	Bidhan Chandra Roy Hall of R...	255.22709207325013
4	DEPT OF COMPUTER SCIENC...	SAM	294.1172468864576
5	DEPT OF COMPUTER SCIENC...	Sister Nivedita Hall of Residen...	324.7456225175059
6	DEPT OF COMPUTER SCIENC...	Gokhale	364.56701643152473
7	DEPT OF COMPUTER SCIENC...	Rajendra Prasad Hall of Resid...	385.0800371694558

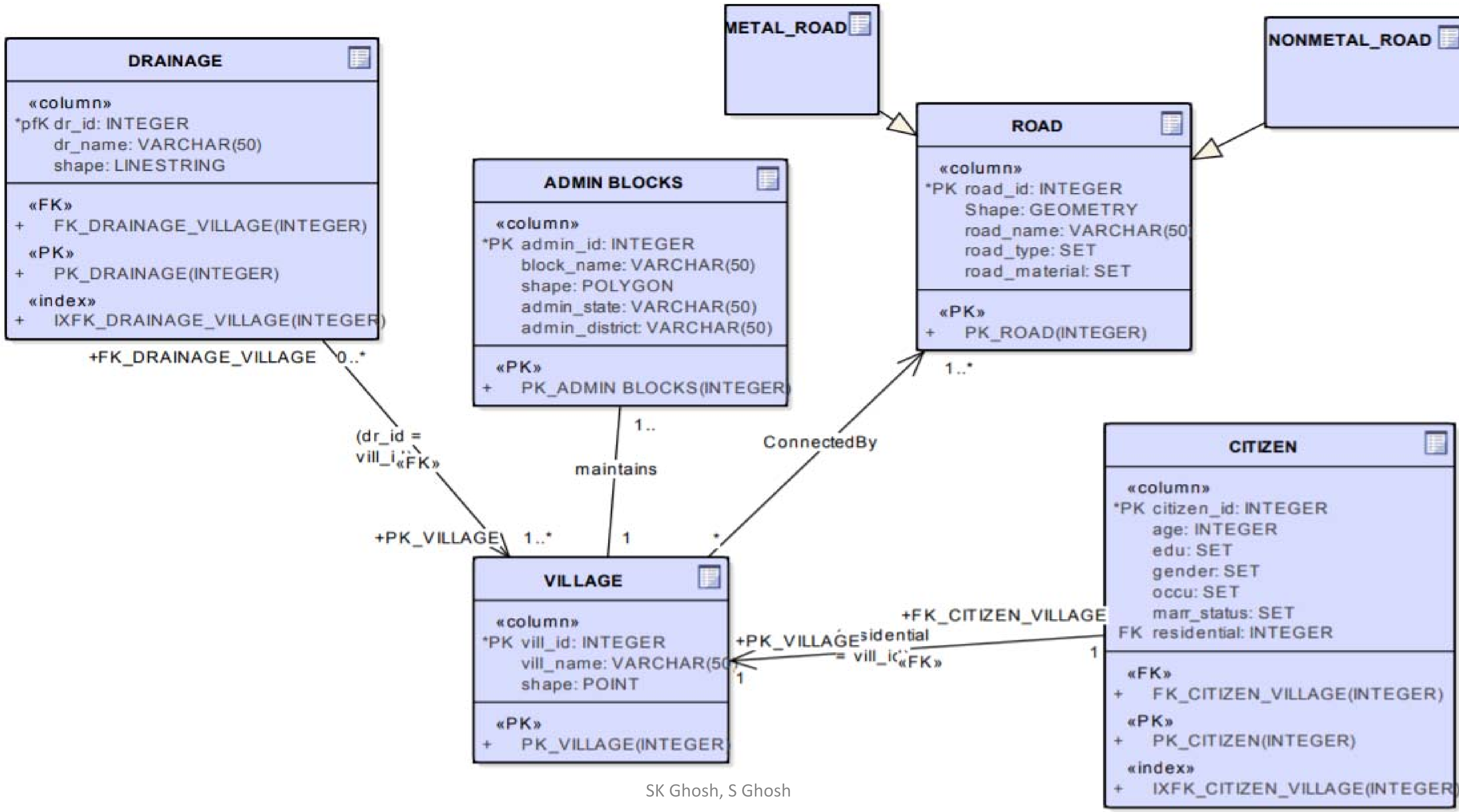
Example Scenario

Consider 4 independent repositories of a region **P**, namely, ***ROAD (R), DRAINAGE (D), VILLAGE (V)*** and ***ADMIN BLOCKS (A)***.
[Road: Polyline; Drainage: Polyline; Village: Point; Admin: Polygon]

Spatial Query 1: Find the villages which are likely to be affected during flood.
[Flood: Areas within 1km of a drainage network are inundated]

Spatial Query 2: Find the Roads likely to be affected if River R1 is flooded?

Spatial Query 3: To setup a new industry the requirement is: It should be in Admin Blocks A2 or A7, 2km from NH, no Drainage within 1km, within 5kmof villages with working population (20-50yrs) greater than 100


DRAINAGE
«column»
*pfK dr_id: INTEGER
dr_name: VARCHAR(50)
shape: LINESTRING
«FK»
+ FK_DRAINAGE_VILLAGE(INTEGER)
«PK»
+ PK_DRAINAGE(INTEGER)
«index»
+ IXFK_DRAINAGE_VILLAGE(INTEGER)
+FK_DRAINAGE_VILLAGE
0..*
(dr_id =
vill_i
«FK»
+PK_VILLAGE
1..*
ADMIN BLOCKS
«column»
*PK admin_id: INTEGER
block_name: VARCHAR(50)
shape: POLYGON
admin_state: VARCHAR(50)
admin_district: VARCHAR(50)
«PK»
+ PK_ADMIN BLOCKS(INTEGER)
1..
maintains
1
METAL_ROAD
NONMETAL_ROAD
ROAD
«column»
*PK road_id: INTEGER
Shape: GEOMETRY
road_name: VARCHAR(50)
road_type: SET
road_material: SET
«PK»
+ PK_ROAD(INTEGER)
1..*
ConnectedBy
*
VILLAGE
«column»
*PK vill_id: INTEGER
vill_name: VARCHAR(50)
shape: POINT
«PK»
+ PK_VILLAGE(INTEGER)
+PK_VILLAGE
1
+FK_CITIZEN_VILLAGE
esidential
= vill_i
«FK»
1
CITIZEN
«column»
*PK citizen_id: INTEGER
age: INTEGER
edu: SET
gender: SET
occu: SET
marr_status: SET
FK residential: INTEGER
«FK»
+ FK_CITIZEN_VILLAGE(INTEGER)
«PK»
+ PK_CITIZEN(INTEGER)
«index»
+ IXFK_CITIZEN_VILLAGE(INTEGER)

Consider 4 independent repositories of a region **P**, namely, ***ROAD (R), DRAINAGE (D), VILLAGE (V)*** and ***ADMIN BLOCKS (A)***. [Road: Polyline; Drainage: Polyline; Village: Point; Admin: Polygon]

Spatial Query 1: Find the villages which are likely to be affected during flood.
[Flood: Areas within 1km of a drainage network are inundated]

```
SELECT V.vill_id, V.vill_name FROM VILLAGE V, DRAINAGE D
WHERE OVERLAP(V.shape, BUFFER(D.shape,1000))=1;
```

Spatial Query 2: Find the Roads likely to be affected if River R1 is flooded?

```
SELECT R.road_id, V.vill_name FROM VILLAGE V, DRAINAGE D
WHERE OVERLAP(V.shape, BUFFER(D.shape,1000))=1
AND D.dr_name="R1";
```

Consider 4 independent repositories of a region **P**, namely, ***ROAD (R), DRAINAGE (D), VILLAGE (V)*** and ***ADMIN BLOCKS (A)***. [Road: Polyline; Drainage: Polyline; Village: Point; Admin: Polygon]

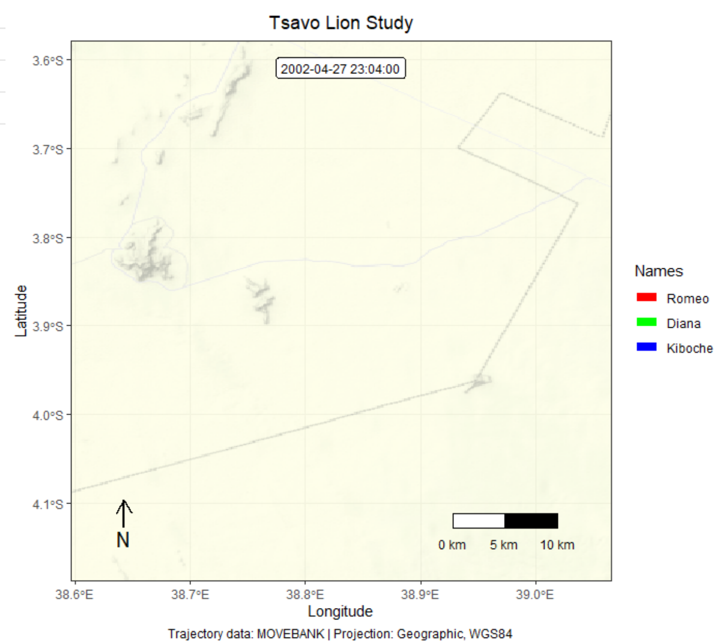
<u>Spatial Query 3:</u> To setup a new industry the requirement is: It should be in Admin Blocks A2 or A7, 2km from NH, no Drainage within 1km, within 5kmof villages with working population (20-50yrs) greater than 100

```
Create VIEW REG AS(
SELECT INTERSECT(V.shape,A.shape) AS REG_SHAPE FROM ROAD R, DRAINAGE D, VILLAGE V,
CITIZEN C WHERE
OVERLAP(V.shape, BUFFER(D.shape,1000))=0 AND
OVERLAP(V.shape, BUFFER(R.shape,2000))=1 AND COUNT(C.citizen_id)>=100 WHERE
C.age>20 AND C.age<50 AND C.residential==V.vill_id)


SELECT INTERSECT (REG_SHAPE,A.shape) FROM REG, ADMIN_BLOCKS A
WHERE
A.block_name IN (“A2”,”A7”) AND
OVERLAP(A.shape, BUFFER(REG_SHAPE,5000))==1
```

EXAMPLE WITH REAL DATA

A	B	C	D	E	F	G	H	I	J	K	L	M
event-id	visible	timestamp	location-lc	location-la	comments	eobs:t	sensor-typ	individual-taxon-canonical-name	tag-local-i	individual-	study-name	
1442760	TRUE	04:00.0	38.87065	-3.78818	24.06		gps	Panthera leo	#346658	Romeo	Tsavo Lion Study	
1442761	TRUE	03:00.0	38.92993	-3.80081	32.25		gps	Panthera leo	#346658	Romeo	Tsavo Lion Study	
1442762	TRUE	03:00.0	38.92914	-3.80073	35.56		gps	Panthera leo	#346658	Romeo	Tsavo Lion Study	
1442763	TRUE	03:00.0	38.93245	-3.80076	32.25		gps	Panthera leo	#346658	Romeo	Tsavo Lion Study	
1442764	TRUE	03:00.0	38.93511	-3.80006	30.69		gps	Panthera leo	#346658	Romeo	Tsavo Lion Study	
1442765	TRUE	04:00.0	38.93576	-3.79968	36.13		gps	Panthera leo	#346658	Romeo	Tsavo Lion Study	
1442766	TRUE	03:00.0	38.93766	-3.79954	29.56		gps	Panthera leo	#346658	Romeo	Tsavo Lion Study	
1442767	TRUE	03:00.0	38.94172	-3.79895	29.56		gps	Panthera leo	#346658	Romeo	Tsavo Lion Study	
1442768	TRUE	03:00.0	38.94081	-3.7843	34.13		gps	Panthera leo	#346658	Romeo	Tsavo Lion Study	
1442769	TRUE	03:00.0	38.92703	-3.80612	31.56		gps	Panthera leo				
1442770	TRUE	04:00.0	38.96854	-3.83329	33.75		gps	Panthera leo				
1442771	TRUE	03:00.0	38.9685	-3.83394	34.69		gps	Panthera leo				
1442772	TRUE	03:00.0	38.96831	-3.83457	33.5		gps	Panthera leo				
1442773	TRUE	02:00.0	38.96401	-3.83185	33.69		gps	Panthera leo				
1442774	TRUE	04:00.0	38.94022	-3.78442	32.31		gps	Panthera leo				

Lion trajectory data (source: MOVEBANK)

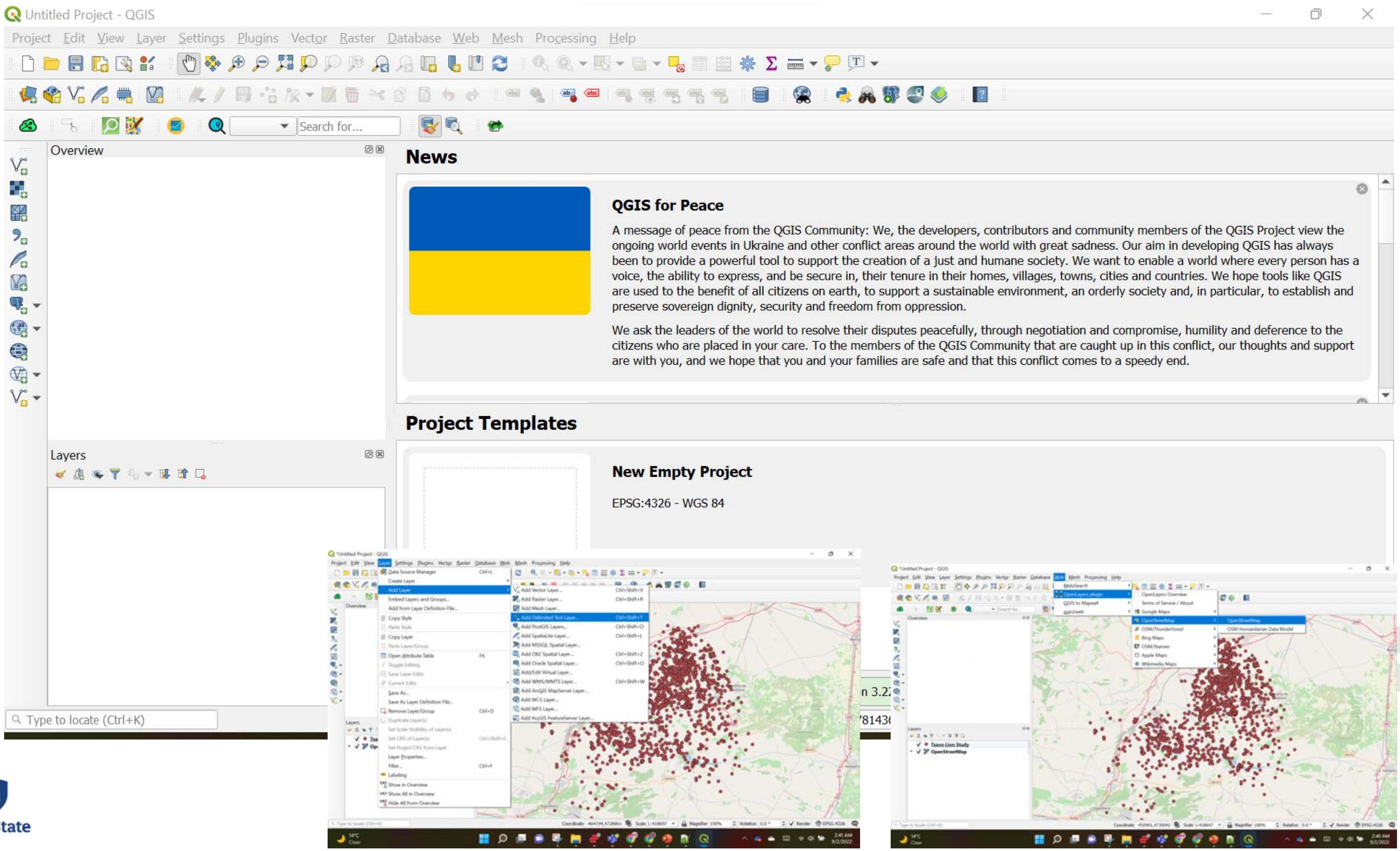


SPATIAL DATA → DATABASE

- **Method 1**: Create UML → DDL → Execute (pgadmin4) → Tables/ schema create → Insert into … (data source)
  - ❑ ALTER TABLE your_table ADD COLUMN geom geometry(Point, 4326);
  - ❑ UPDATE your_table SET geom = ST_SetSRID(ST_MakePoint(longitude, latitude), 4326);

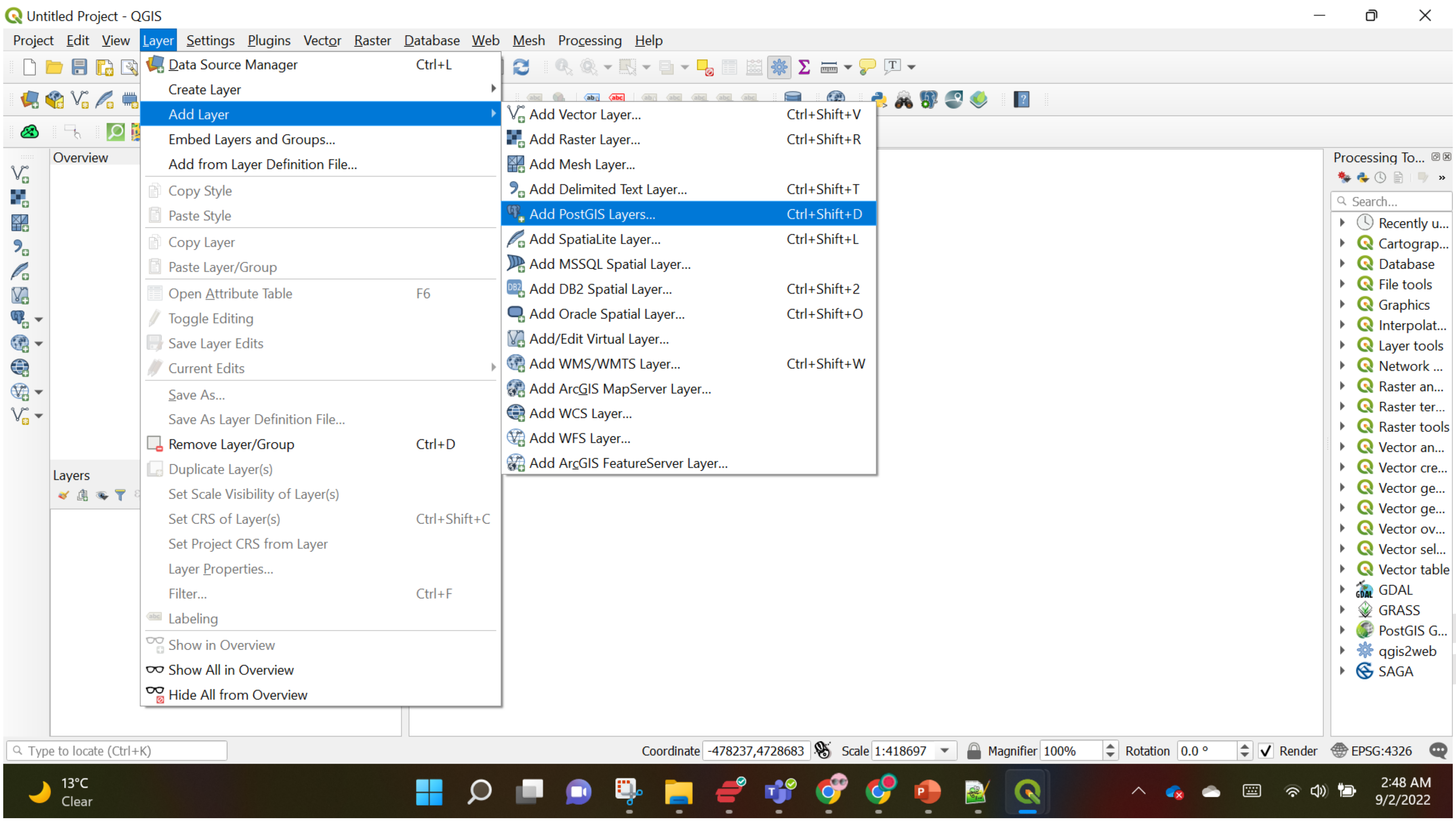
- **Method 2**: .shp file (PostGIS shape file import/ export manager) → PostGIS [You can also connect PostGIS from QGIS portal]

PennState

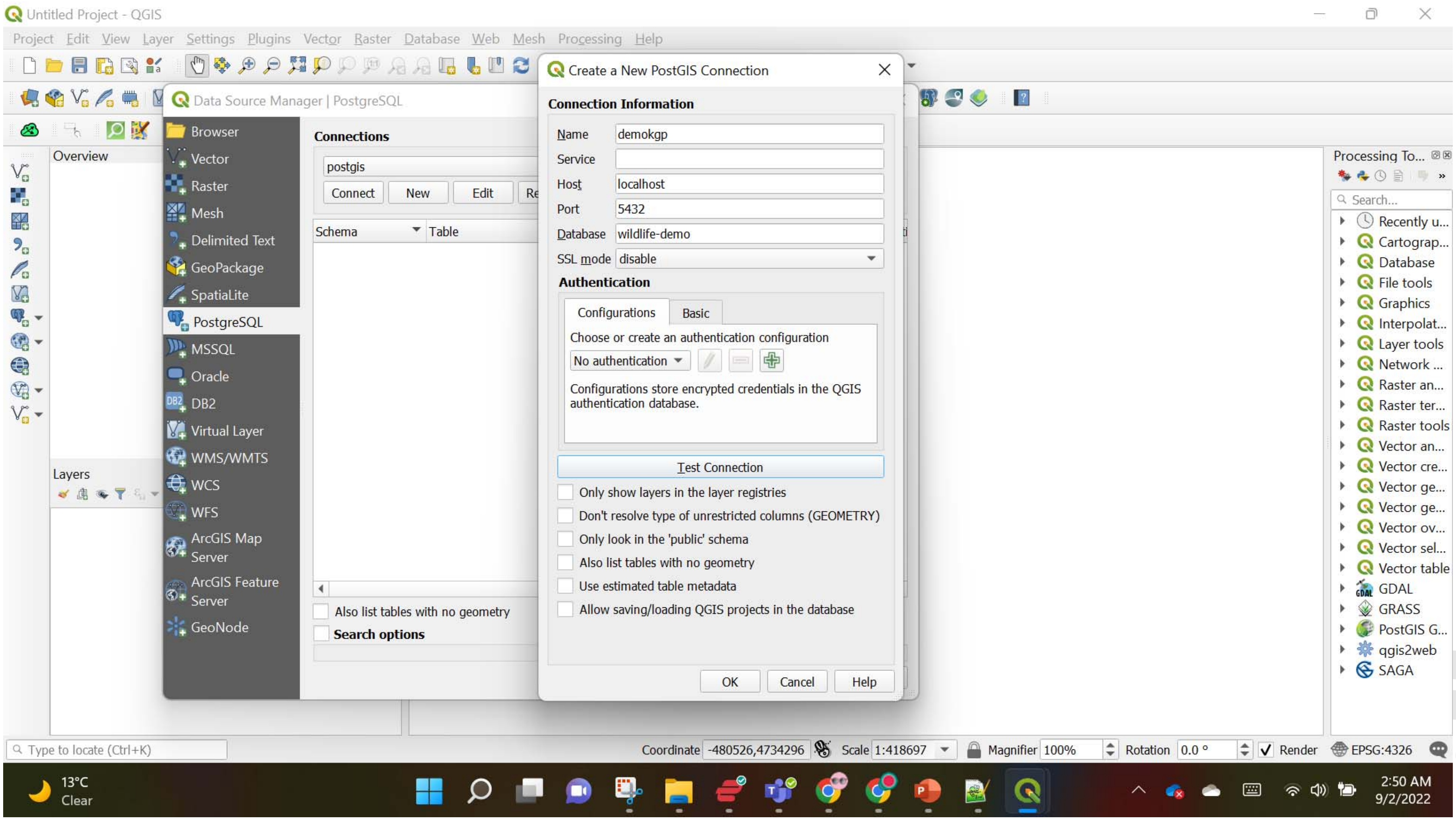

Untitled Project - QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help
Search for...
Overview
News
QGIS for Peace
A message of peace from the QGIS Community: We, the developers, contributors and community members of the QGIS Project view the ongoing world events in Ukraine and other conflict areas around the world with great sadness. Our aim in developing QGIS has always been to provide a powerful tool to support the creation of a just and humane society. We want to enable a world where every person has a voice, the ability to express, and be secure in, their tenure in their homes, villages, towns, cities and countries. We hope tools like QGIS are used to the benefit of all citizens on earth, to support a sustainable environment, an orderly society and, in particular, to establish and preserve sovereign dignity, security and freedom from oppression.
We ask the leaders of the world to resolve their disputes peacefully, through negotiation and compromise, humility and deference to the citizens who are placed in your care. To the members of the QGIS Community that are caught up in this conflict, our thoughts and support are with you, and we hope that you and your families are safe and that this conflict comes to a speedy end.
Project Templates
New Empty Project
EPSG:4326 - WGS 84
Layers
Type to locate (Ctrl+K)

PennState

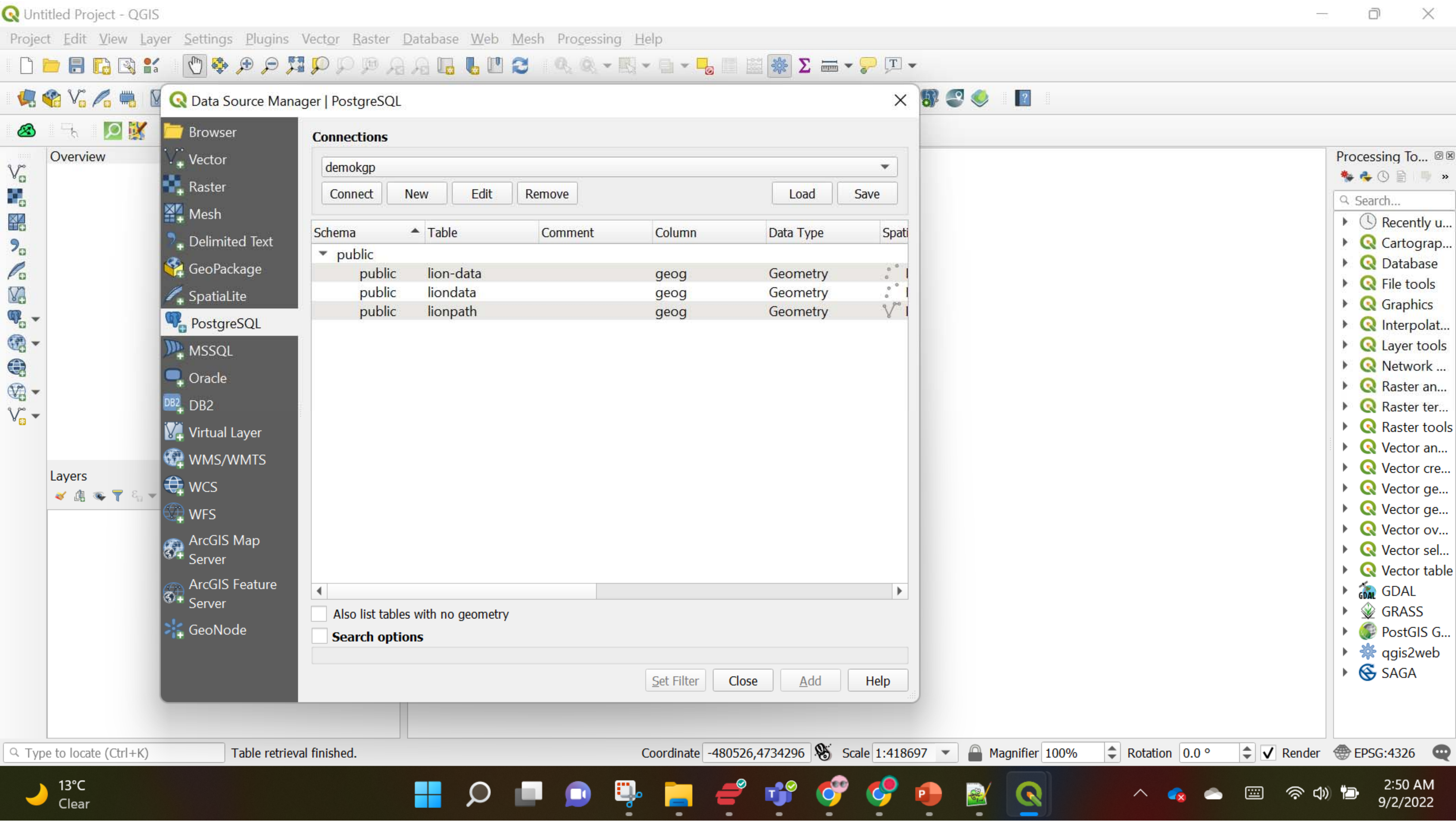
CONNECT POSTGIS FROM QGIS

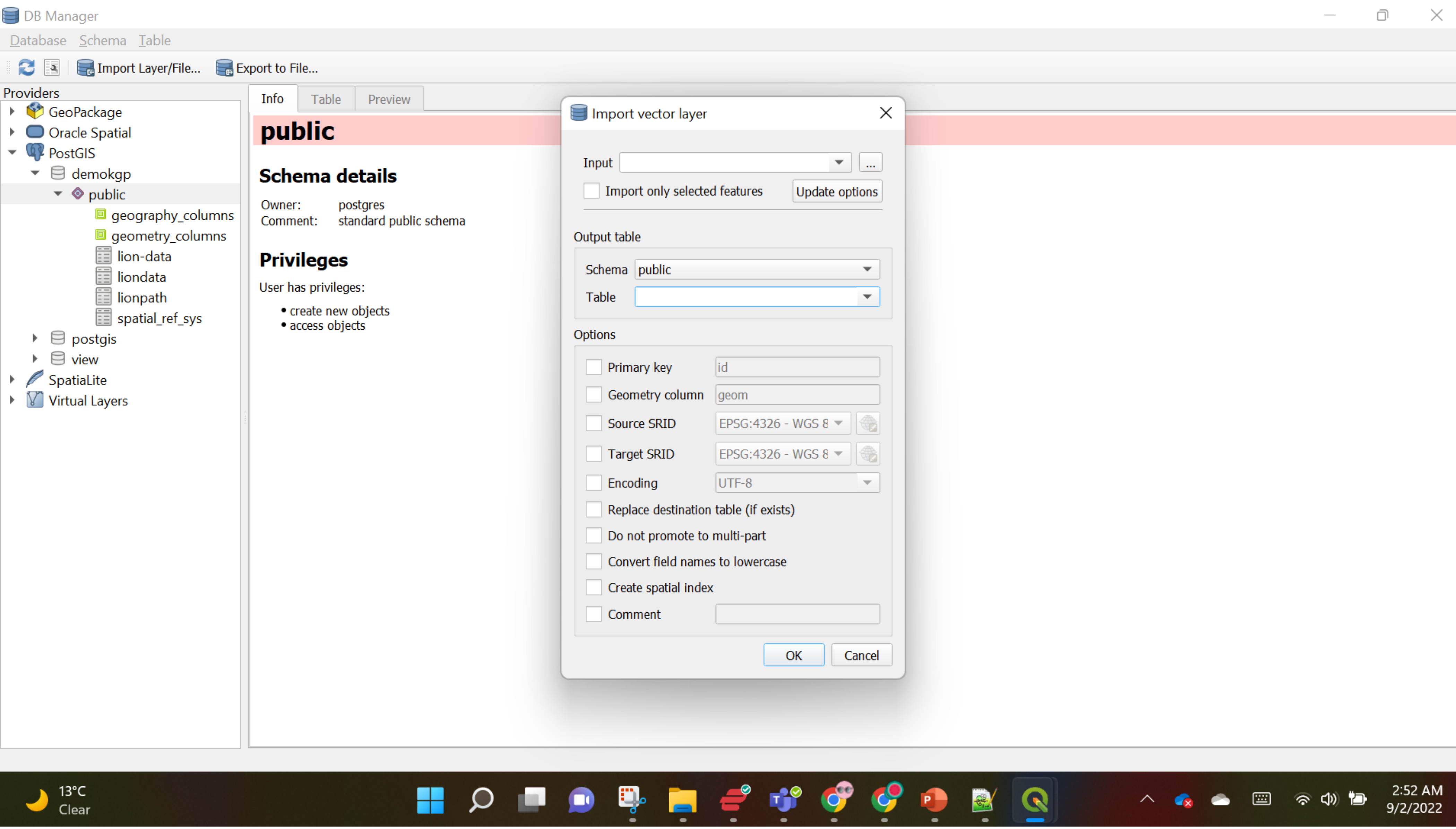


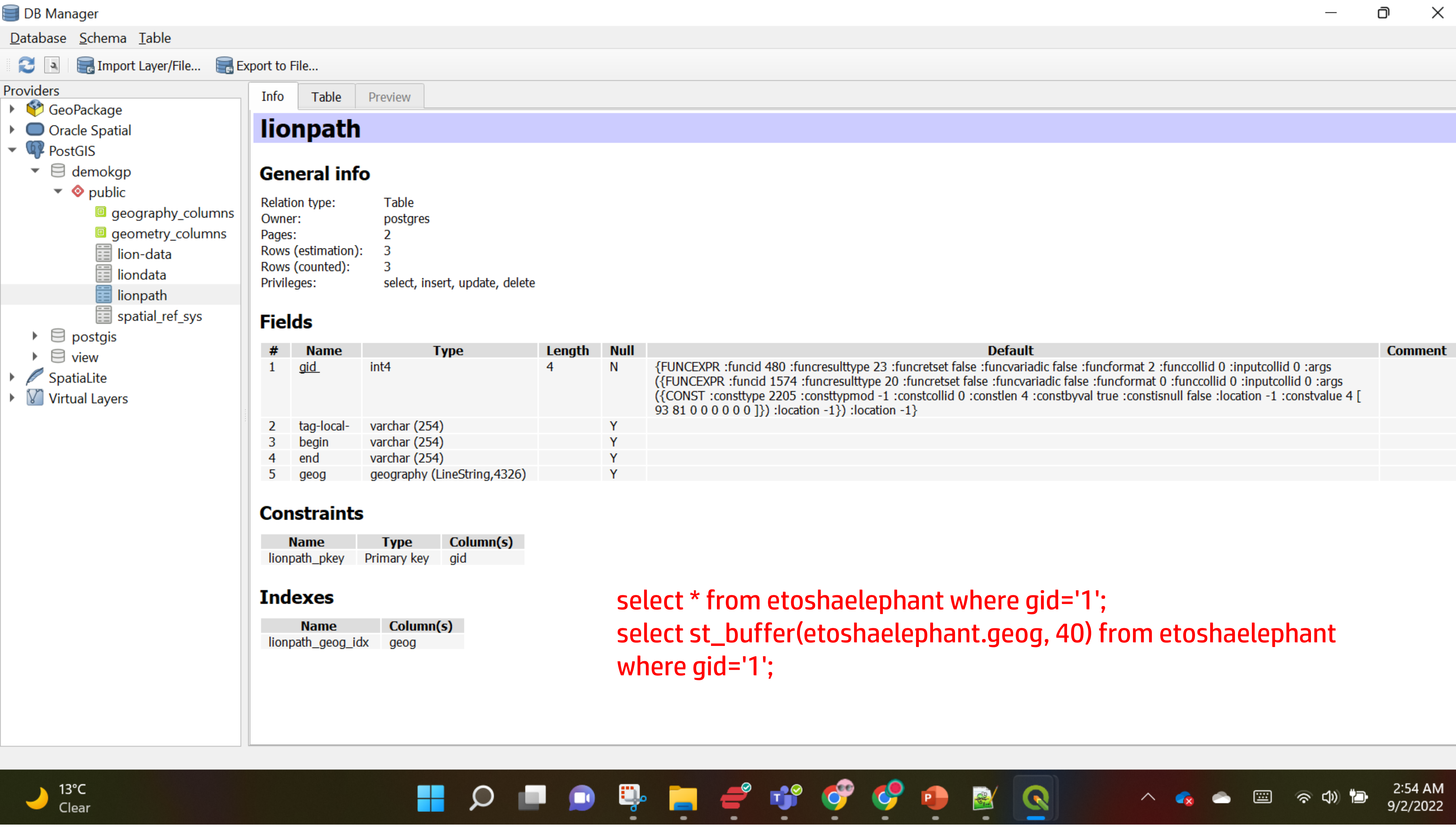
PennState


Untitled Project - QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help
Data Source Manager | PostgreSQL
Browser
Vector
Raster
Mesh
Delimited Text
GeoPackage
SpatiaLite
PostgreSQL
MSSQL
Oracle
DB2
Virtual Layer
WMS/WMTS
WCS
WFS
ArcGIS Map Server
ArcGIS Feature Server
GeoNode
Connections
postgis
Connect
New
Edit
Schema
Table
Also list tables with no geometry
Search options
Create a New PostGIS Connection
Connection Information
Name demokgp
Service
Host localhost
Port 5432
Database wildlife-demo
SSL mode disable
Authentication
Configurations
Basic
Choose or create an authentication configuration
No authentication
Configurations store encrypted credentials in the QGIS authentication database.
Test Connection
Only show layers in the layer registries
Don't resolve type of unrestricted columns (GEOMETRY)
Only look in the 'public' schema
Also list tables with no geometry
Use estimated table metadata
Allow saving/loading QGIS projects in the database
OK
Cancel
Help
Overview
Layers
Processing To...
Search...
Recently u...
Cartograp...
Database
File tools
Graphics
Interpolat...
Layer tools
Network ...
Raster an...
Raster ter...
Raster tools
Vector an...
Vector cre...
Vector ge...
Vector ge...
Vector ov...
Vector sel...
Vector table
GDAL
GRASS
PostGIS G...
qgis2web
SAGA
Type to locate (Ctrl+K)
Coordinate -480526,4734296
Scale 1:418697
Magnifier 100%
Rotation 0.0 °
Render
EPSG:4326
13°C
Clear
2:50 AM
9/2/2022

PennState


Untitled Project - QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help
Data Source Manager | PostgreSQL
Browser
Vector
Raster
Mesh
Delimited Text
GeoPackage
SpatiaLite
PostgreSQL
MSSQL
Oracle
DB2
Virtual Layer
WMS/WMTS
WCS
WFS
ArcGIS Map Server
ArcGIS Feature Server
GeoNode
Connections
demokgp
Connect
New
Edit
Remove
Load
Save
Schema
Table
Comment
Column
Data Type
Spati
public
public lion-data geog Geometry
public liondata geog Geometry
public lionpath geog Geometry
Also list tables with no geometry
Search options
Set Filter
Close
Add
Help
Overview
Layers
Processing To...
Search...
Recently u...
Cartograp...
Database
File tools
Graphics
Interpolat...
Layer tools
Network ...
Raster an...
Raster ter...
Raster tools
Vector an...
Vector cre...
Vector ge...
Vector ge...
Vector ov...
Vector sel...
Vector table
GDAL
GRASS
PostGIS G...
qgis2web
SAGA
Type to locate (Ctrl+K)
Table retrieval finished.
Coordinate -480526,4734296
Scale 1:418697
Magnifier 100%
Rotation 0.0 °
Render
EPSG:4326
13°C
Clear
2:50 AM
9/2/2022


DB Manager
Database Schema Table
Import Layer/File...
Export to File...
Providers
GeoPackage
Oracle Spatial
PostGIS
demokgp
public
geography_columns
geometry_columns
lion-data
liondata
lionpath
spatial_ref_sys
postgis
view
SpatiaLite
Virtual Layers
Info
Table
Preview
public
Schema details
Owner: postgres
Comment: standard public schema
Privileges
User has privileges:
• create new objects
• access objects
Import vector layer
Input
Import only selected features
Update options
Output table
Schema public
Table
Options
Primary key id
Geometry column geom
Source SRID EPSG:4326 - WGS 8
Target SRID EPSG:4326 - WGS 8
Encoding UTF-8
Replace destination table (if exists)
Do not promote to multi-part
Convert field names to lowercase
Create spatial index
Comment
OK
Cancel
13°C
Clear
2:52 AM
9/2/2022




```
select * from etoshaelephant where gid='1';
select st_buffer(etoshaelephant.geog, 40) from etoshaelephant
where gid='1';
```

Find out best locations electric vehicle charging point

Identify pickpocket suspects!

Estimate Accurate Traffic Status: Intelligent Path Recommender

Forecasting City Crowd Behaviour

Optimized Traffic Planning

Personalized Travel-Planner/ Security enhancement

Automated Vehicle Routing

Smart-City Planning

Travel Demand Prediction: Optimized Resource Allocation

Transferring Mobility Knowledge to another Region: Better Urban planning

Effective Land-Use Planning / Business Settlement

Wildlife migration -- Sustainability

Location Based Services (nearest Restaurants/ Petrol Stations/ Emergency centers)

Urban Sensing thru' data acquisition

SPATIAL BIG DATA

SPATIAL INFORMATICS – FROM DATA PERSPECTIVE

- ✓The first law of geography
- ✓Coordinate System
- ✓Uncertainty
- ✓Modifiable Areal Unit Problem

PennState

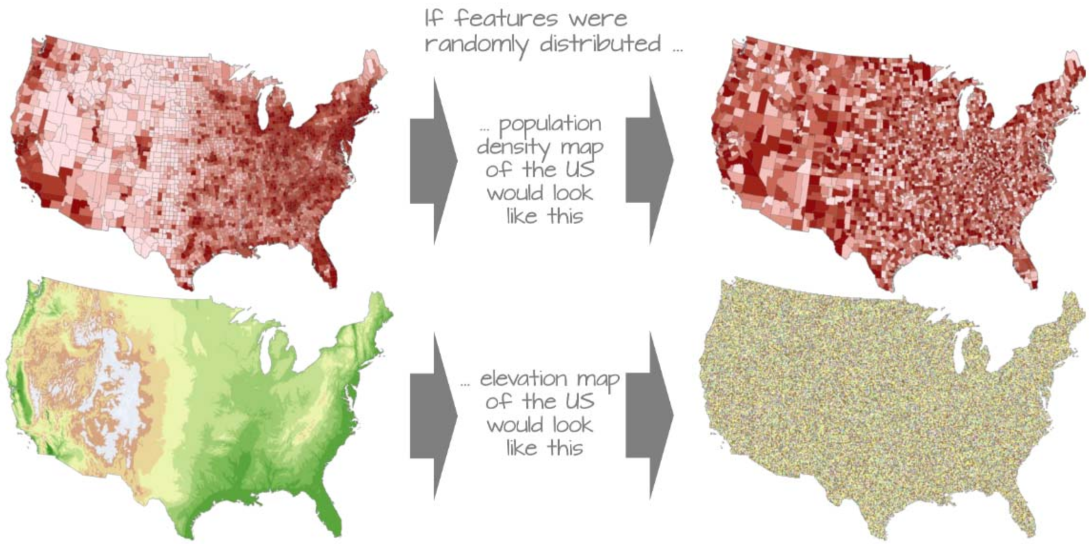
SPATIAL INFORMATICS – FROM DATA PERSPECTIVE

✓The first law of geography (Waldo R. Tobler, 1970)

Everything is related to everything else, but near things are more related than distant things

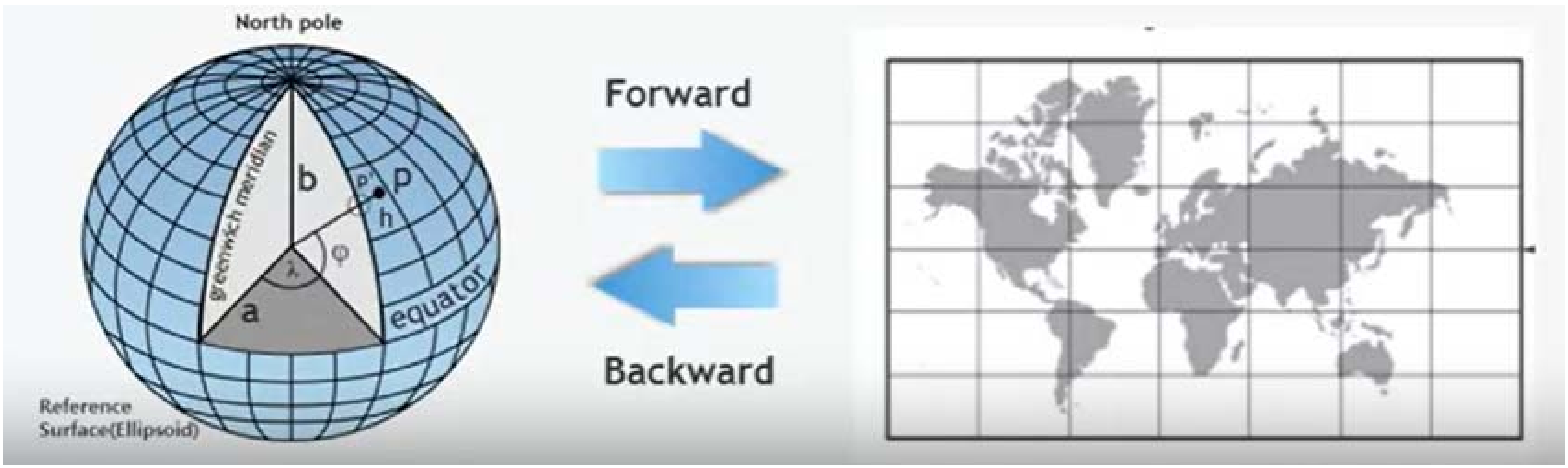
Q: whether or not features with similar values are clustered, randomly distributed or dispersed

Spatial autocorrelation can be defined as the measure of the degree to which one object is similar to other nearby objects.



SPATIAL INFORMATICS – FROM DATA PERSPECTIVE

Coordinate System



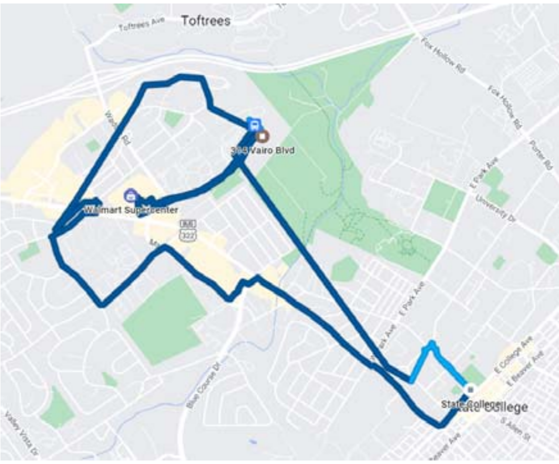
SPATIAL INFORMATICS – FROM DATA PERSPECTIVE

✓Uncertainty

- ❑ Spatial data are based on measurements
- ❑ Spatial data are inherently prone to error
- ❑ Spatial data analytics is stochastic, i.e., probabilistic

Bank Account
$ 5000

Deterministic

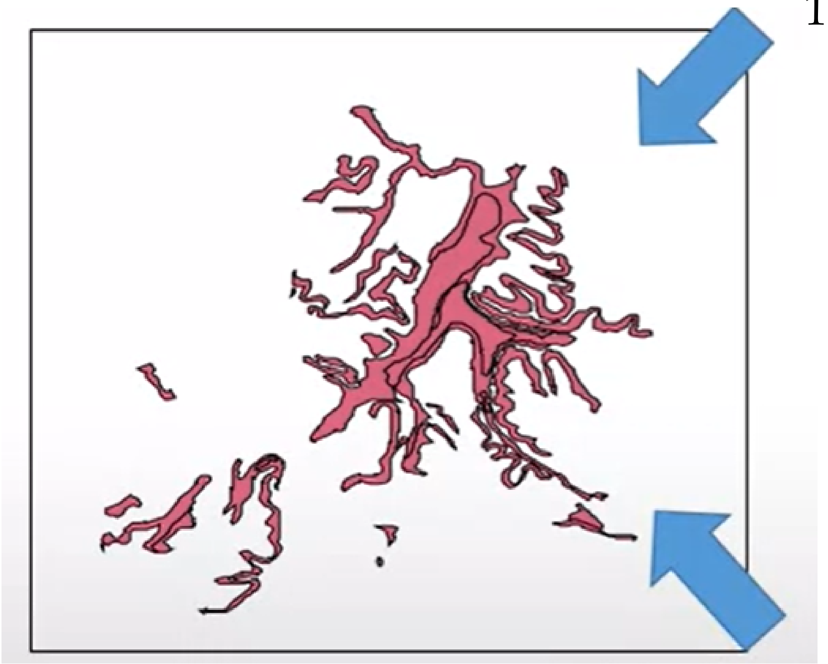


Trajectory

SPATIAL INFORMATICS – FROM DATA PERSPECTIVE

✓Uncertainty

The red region is flooded at the precipitation rate of 1 inch/hour for 3 hours



The red region is expected to be flooded at the precipitation rate of 1 inch/hour for 3 hours at the confidence level of 95%

PennState

SPATIAL INFORMATICS – FROM DATA PERSPECTIVE

✓Modifiable Areal Unit Problem

Same basic data yield different results when aggregated in different ways

- Scale Effect: Analytic difference depending on the size of units used
- Zoning Effect: Analytic difference depending on the way of aggregation

SPATIAL BIG DATA

- ❑ Spatial Big Data exceeds the capacity of commonly used spatial computing systems

  due to volume, variety and velocity

- ❑ Spatial Big Data comes from many different sources

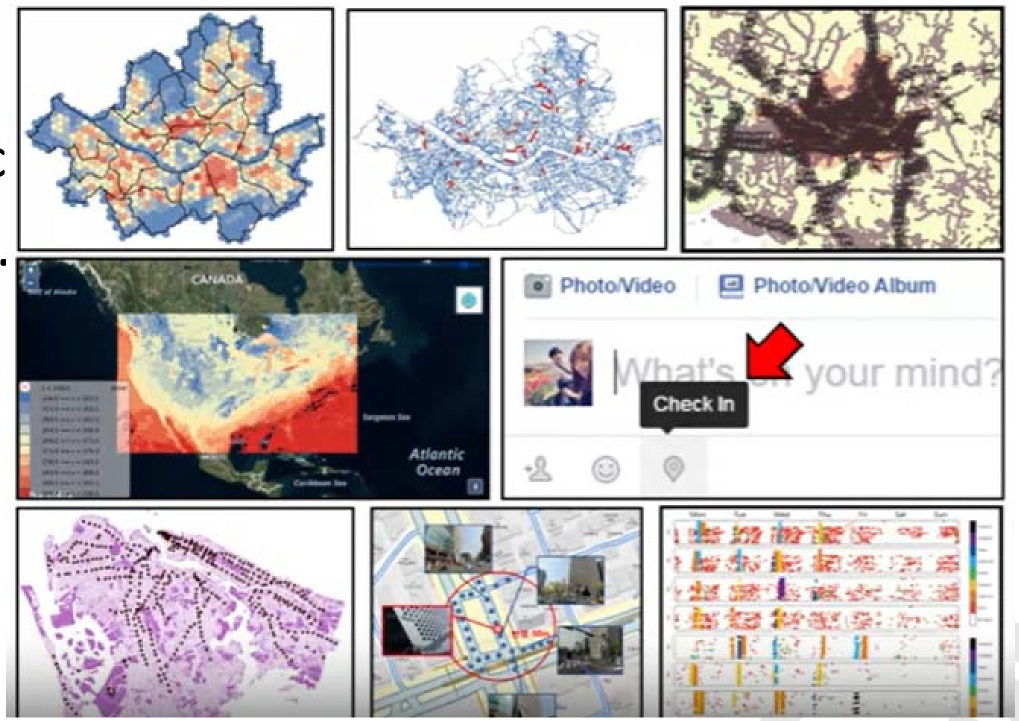
  satellites, drones, vehicles, geosocial networking services, mobile devices, cameras

- ❑ A significant portion of big data is in fact spatial big data

3 V's: **Volume, Velocity** and **Variety**

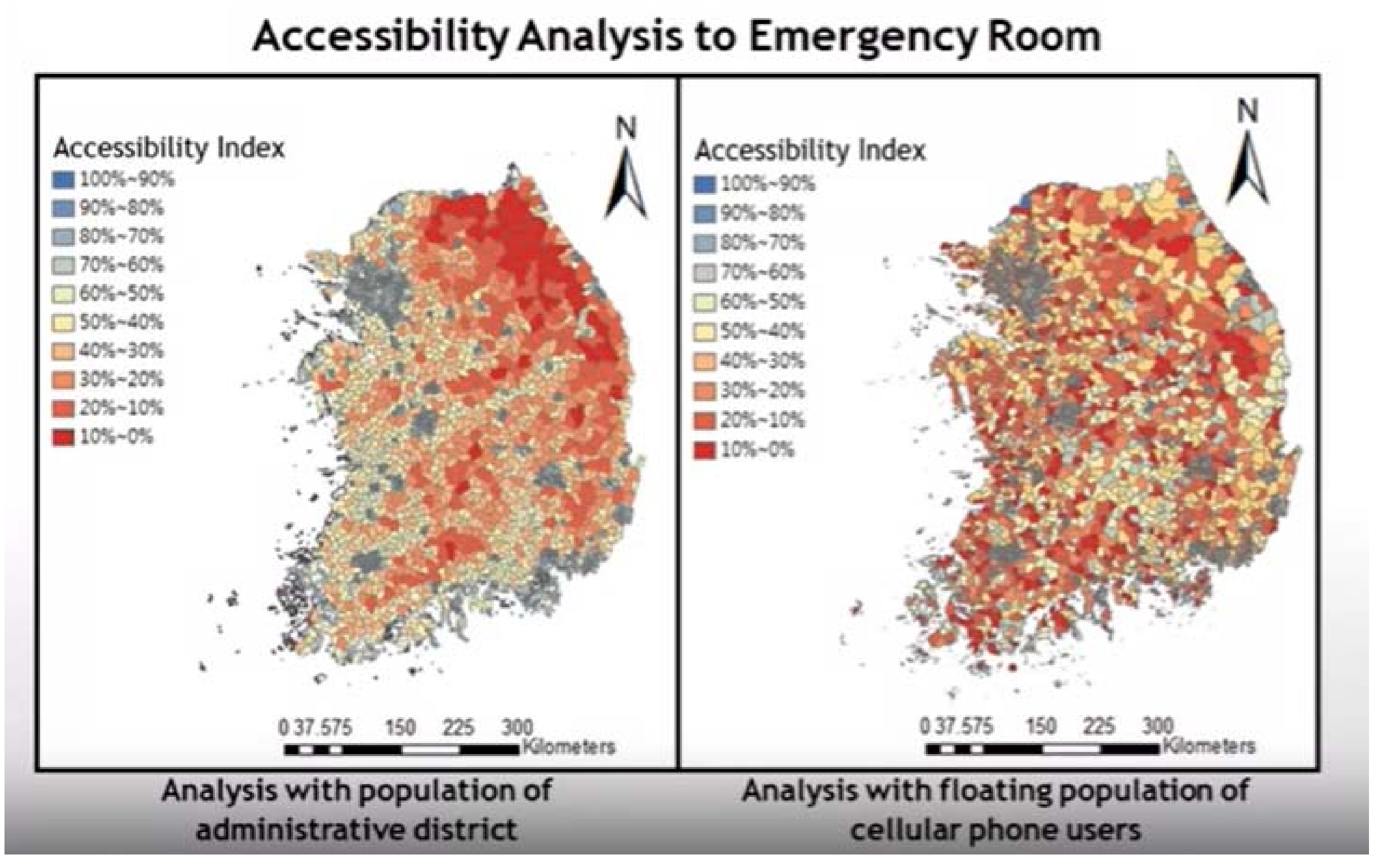
TYPES OF SPATIAL BIG DATA

- Taxi, truck – vehicle trajectory
- Public transportation card transac
- LBSN check-in (twitter, facebook…
- Geo-tagged photos
- Weather data
- Earth observing satellite images



PennState

IMPORTANCE OF SPATIAL BIG DATA



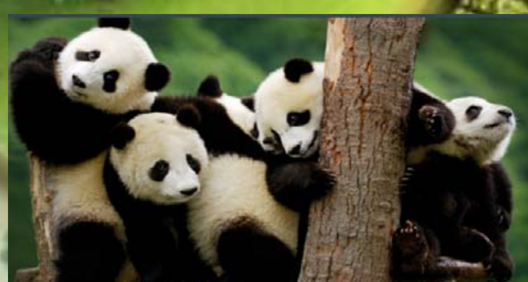
Spatial Big Data

- Spatial data that ***exceeds the capacity of commonly used spatial computing systems*** due to volume, variety and velocity….
- Sources:
  - satellites
  - drones
  - vehicles
  - social networking services
  - mobile devices
  - cameras etc.
- Challenges:
  - Retaining computational efficiency
  - Storing into the cloud
  - Repartitioning

Tracking endangered species… Wildlife analytics

Movebank: a free online database of animal tracking data

- 2.4 billion locations
- 3.1 billion other sensor records
- 5,915 studies
- 3,000 data owners



Wildlife data analysis

Computational Research Question

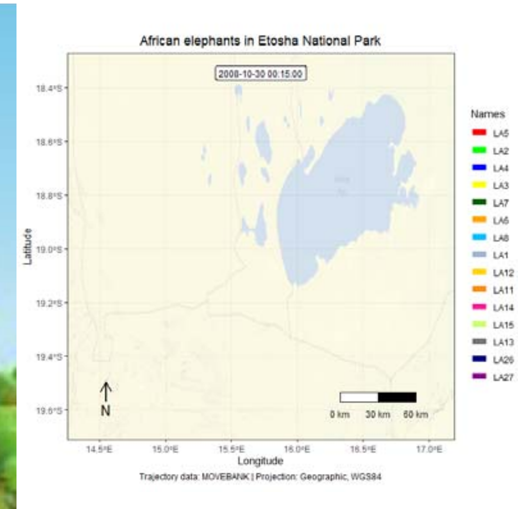
(R1): Can we identify/ classify wildlife category from input trajectory traces? What are the most contributing features and how do they vary on temporal and spatial scales?

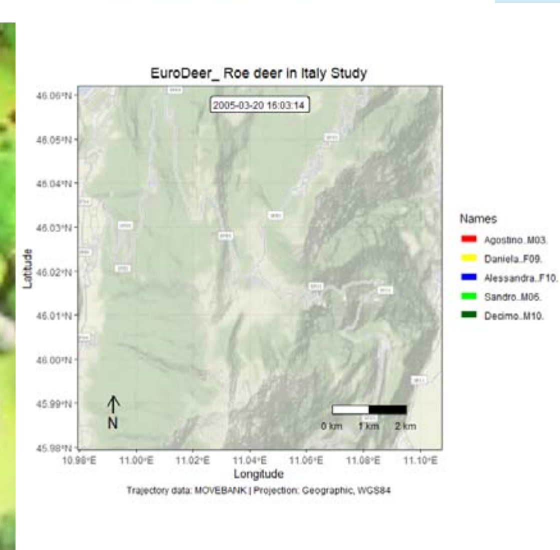
(R2): What is the most useful similarity scores/ distance metric to quantify environmental variables? How do we capture correlations between environmental change and wildlife mobility/ activity?

(R3): How do we quantify the effects of human habitat modifications on wildlife behavior?

(R4): Query extraction, Knowledge capture and representation, spatio-temporal correlation mining and visual analytics

Dataset: Lion (Tsavo) | Elephant (Etosha, Krugar, Java) | Eurodeer | Wolves (Alberta)





PennState

ECO-ROUTING

<u>Next generation routing service</u>

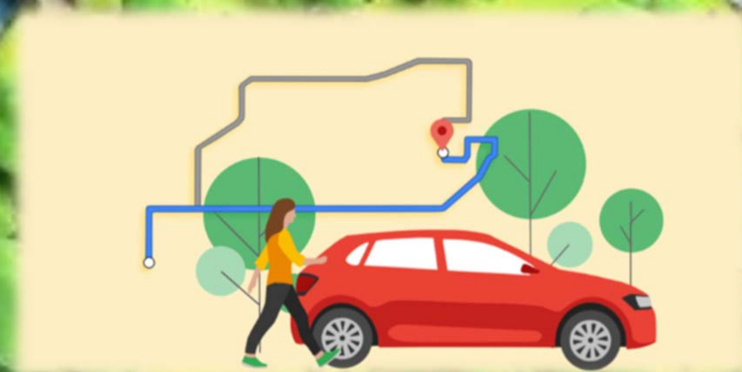
- avoids congestion
- reduces idling at red lights
- avoids left turns

<u>Takes into account various datasets</u>

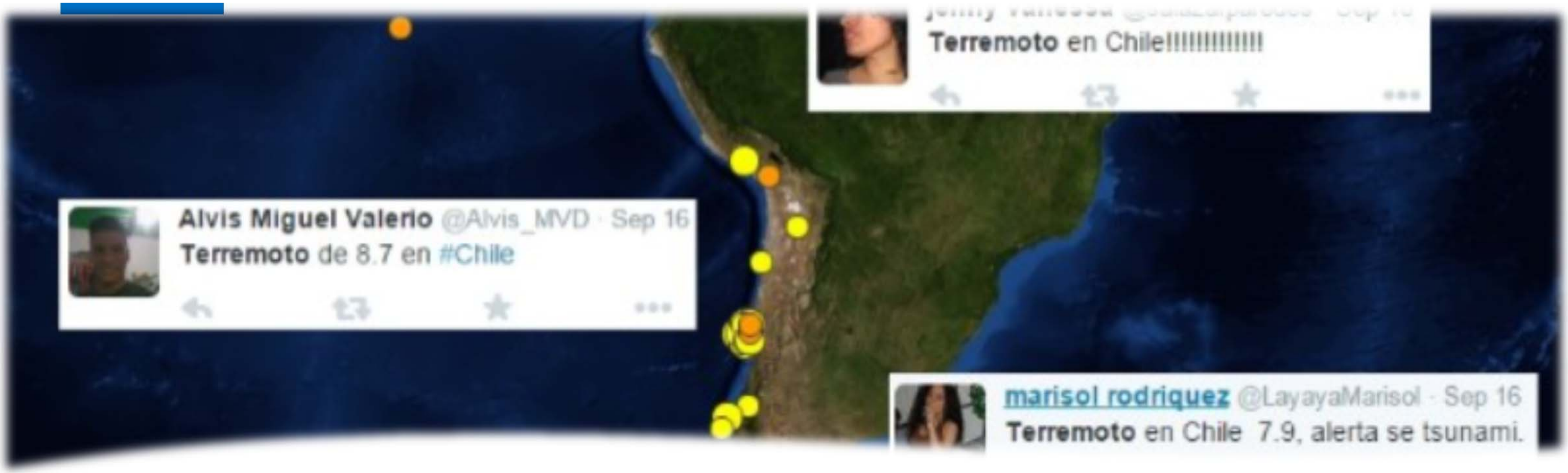
- real-time and historic traffic data of engine measurements
- speed-limits
- road types
- "rush hour vs non-rush hour"

Use eco-friendly routing on your Google Maps app

https://support.google.com/maps/answer/11470237?hl=en



PennState



DETECTING EXTREME EVENTS

- Earthquakes
- Wildfires
- Flooding
- Other calamities

How to detect

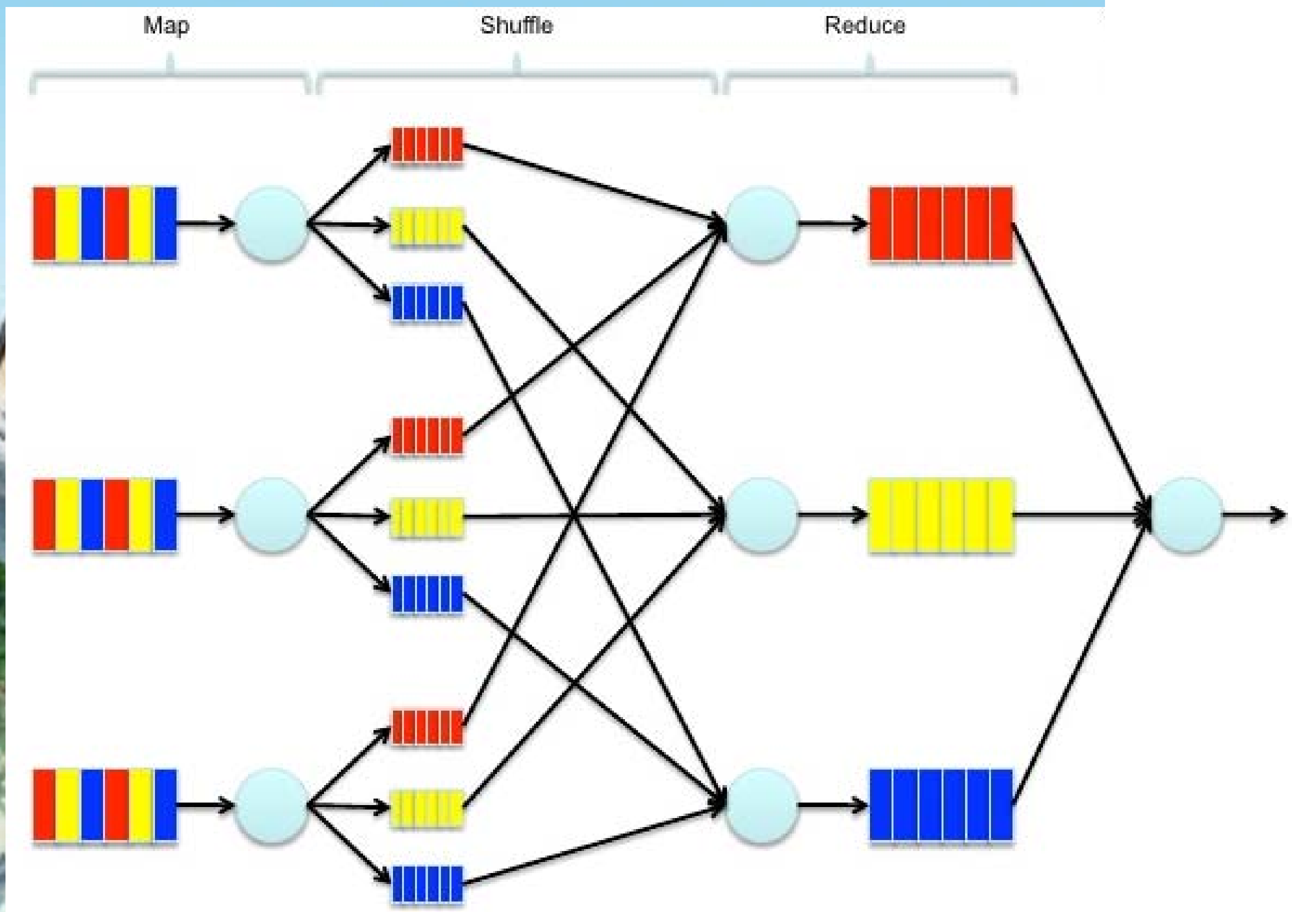
- Built-in motion detectors in mobile phones
- Using unstructured data sets can be used such as tweets

PennState

SPATIAL BIG DATA?

- New Datasets -> need to rapidly integrate new datasets and algorithms
- Computational cost increases as the diversity of Spatial Big Data grows
- Easy to collect, sensors (or sensor networks) are becoming more and more common (Internet of things)

Big data system?

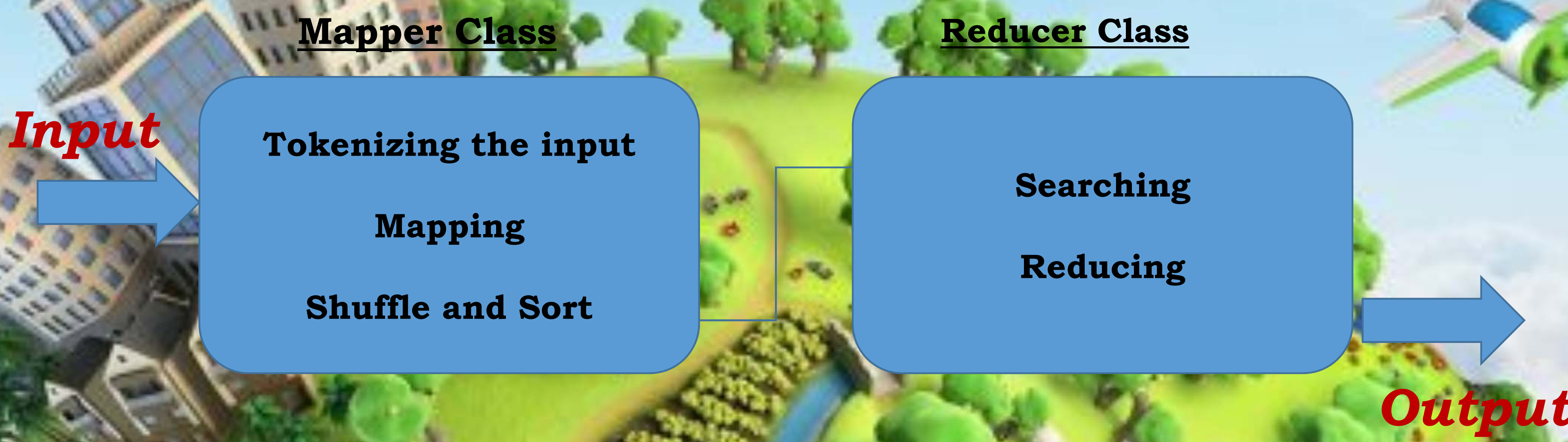


- MapReduce consists of two distinct tasks – Map and Reduce
- First is the map job, where a block of data is read and processed to produce key-value pairs as intermediate outputs.
- The output of a Mapper or map job (key-value pairs) is input to the Reducer.
- The reducer receives the key-value pair from multiple map jobs.
- Then, the reducer aggregates those intermediate data tuples (intermediate key-value pair) into a smaller set of tuples or key-value pairs which is the final output.

MapReduce

Ref: https://code.google.com/

How to start??


Mapper Class
Reducer Class
Input
Tokenizing the input
Mapping
Shuffle and Sort
Searching
Reducing
Output

Average of a set of integers

A = [10,20,30,40,50]

Assume, we are taking 3 mapper nodes and 1 reducer node

Mapper Nodes

Tokenizing the input

Mapping

Shuffle and Sort

Reducer Node

Searching

Reducing

Average of a set of integers

A = [10,20,30,40,50]

Assume, we are taking 3 mapper nodes and 1 reducer node

Average of a set of integers

Mapper Nodes — *<Average, cardinality>*

Mapper Nodes	<Average, cardinality>
10, 20	<15,2>
30, 40	<35,2>
50	<50,1>

Reducer Node

- 15*2 + 35*2 + 50*1 =150
- 2+2+1= 5
- 150/5=30

```
def map(L)
  sum=0
   for i in range(0,len(L)):
        sum=sum+L[i]
   avgM=(sum*1.0)/len(L)
   print (avgM,len(L))
```

Mapper function

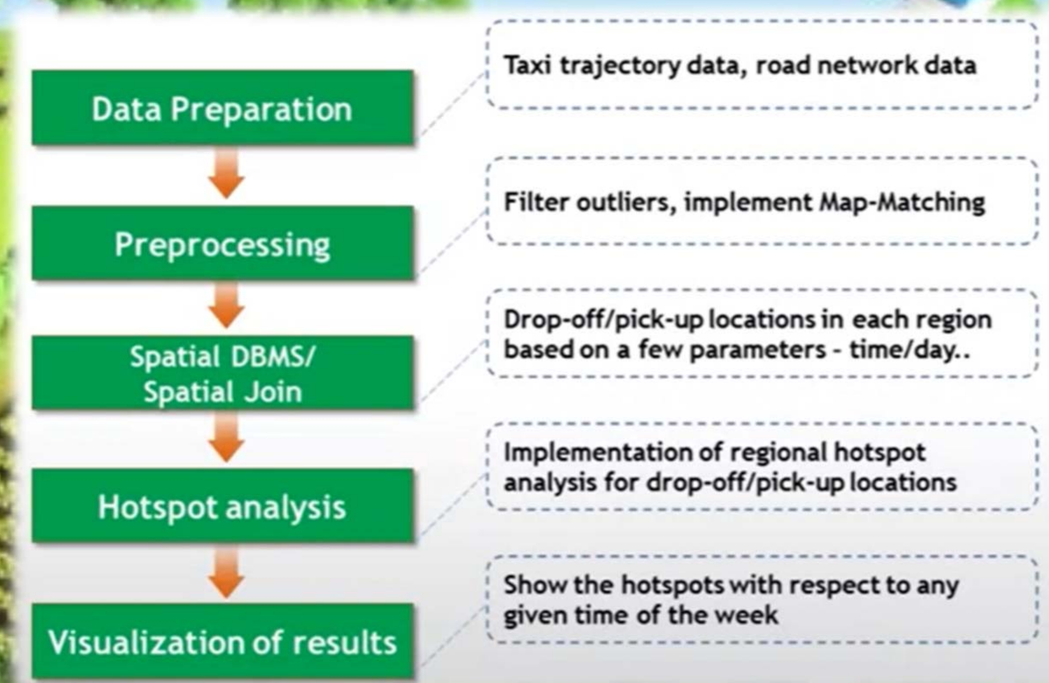
```
def reduce(avgML,leL,m)
   sum=0
    for i in range(0,m):
         sum=sum+avgML[i]*leL[i]
         count=count+leL[i]
    avg=(sum*1.0)/count
    print (avg)
```

Reducer function

TAXI TRAJECTORY ANALYSIS FOR FINDING PICK-UP HOTSPOTS

Passenger Finder: Guide to the places where more passengers are waiting for taxi cabs

Taxi data collected every day and the size is huge and noisy



PennState

TAXI TRAJECTORY ANALYSIS FOR FINDING PICK-UP HOTSPOTS

Map Class

```java
public static class MapClass extends Mapper<Object, Text, Text, Text> {

    public void map(Object key, Text value, Context context) {
        try {
            String[] str = value.toString().split(",");
            if (str.length == 11) {
                double lat = Double.valueOf(str[10]);
                double lng = Double.valueOf(str[9]);
                int status = Integer.valueOf(str[6]);
                if (lat > 37.42830 && lat < 37.70146
                        && lng > 126.76448 && lng < 127.18380
                        && status == 2) {
                    String mykey = str[0]; // CarID
                    context.write(new Text(mykey), new Text(value));
                }
            } else {
                System.out.println("Wrong format:" + value);
            }
        } catch (Exception e) {
            System.out.println(e.getMessage());
        }
    }
}
```

Check format on each record

Check the conditions before implement map tasks with key CarID

TAXI TRAJECTORY ANALYSIS FOR FINDING PICK-UP HOTSPOTS

Reduce Class

```
public static class ReduceClass extends Reducer<Text, Text, Text, Text> {

    public void reduce(Text key, Iterable<Text> values, Context context)
            throws IOException, InterruptedException {
        try (Connection myConnection = get_mySqlConnect()) {
            for (Text val : values) {
                String[] str = val.toString().split(",");

                double lat = Double.valueOf(str[10]);
                double lng = Double.valueOf(str[9]);

                String rs = find_RoadMatching(myConnection, lng, lat);
                context.write(null, new Text(val + "," + rs));
            }
        } catch (Exception ex) {
            System.out.println(ex);;
        }
    }
}
```

Implement Map-Matching and write the result to HDFS

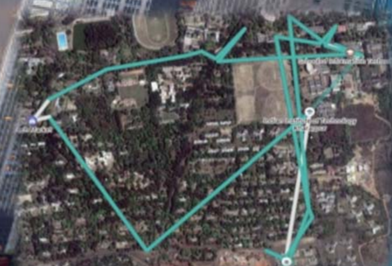
PennState

USE-CASES:

LOCATION BASED SERVICES/ MOBILITY ANALYTICS

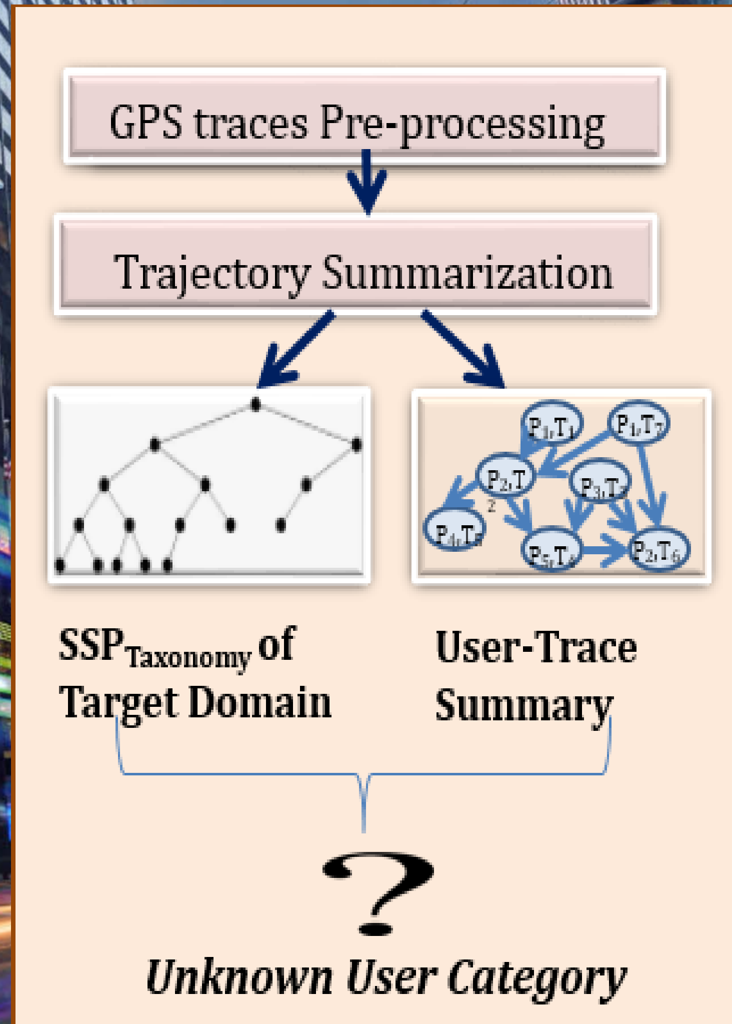
Categorization of Mobile Users and trip-purposes from GPS Traces

- Label scarcity (insufficient training test in the target region!)
- Less number of participants to carry out the learning task
- Less number of GPS traces of the participants – unable to find any pattern



Source Region:
IIT Kharagpur Campus
Labelled data Available

Target Region: Dartmouth Campus, Hanover, USA
Labelled data Unavailable



Labelled data is available in different but same type of region-of-interest!

Categorization of Mobile Users from GPS Traces

1. *Can we use the available labelled data?*
2. *If Yes, then how to map the knowledge of these two geographically dispersed region?*
3. *When and how to compare these region-of-interest?*

Source Region:
IIT Kharagpur Campus
Labelled data Available

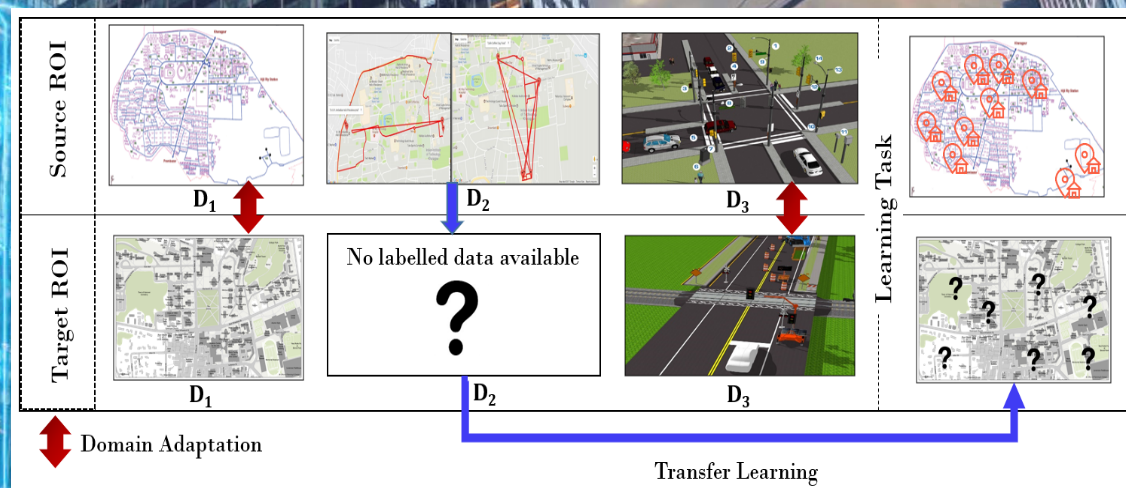
Target Region:
Dartmouth Campus, Hanover, USA
Labelled data Unavailable

- Given a semantic source region of interest S_{ROI} and user-classification task T_S; different but related semantic target region of interest T_{ROI}, how to learn the classier from S_{ROI} and T_S to perform the classification or learning task in T_T
- We define semantic source region of interest as $S_{ROI} = (SSP^S_{Taxnomy}; P(w_S))$ and semantic target region of interest as $T_{ROI} = (SSP^T_{Taxnomy}; P(w_T))$

	Kharagpur Region		*Dartmouth Region*	
User Category	Precision	Recall	Precision	Recall
Undergrad Student	0.75	0.75	0.78	0.82
Grad Students	0.846	0.95	0.9	0.82
Grad Students (Research Scholars)	0.6315	0.545	NA	NA
Employees/ Professors	0.555	0.625	0.62	0.56
Non-Residential Students	0.714	0.714	NA	NA

Trip-purpose identification: Semantic Enrichment of Trajectory Traces

- *Mobility Knowledge Graph*: Time-dependent mobility knowledge graph (MKG) based on the users' movement history and personal logs
  - MKG captures correlations among locations, temporal information and movement semantics.
- *User Mobility Semantics*: Deep learning architecture for annotating GPS (movement) traces with trip-intent and trip-purpose
- *Mobility Knowledge Transfer*: Mobilytics presents a transfer learning technique for transferring mobility knowledge from source region to target region and MKG completion



To the best of our knowledge, ours is the first work to capture the correlations between two geographically dispersed regions, and to transfer mobility knowledge

Mobility Knowledge Graph (MKG)

Information Retrieval

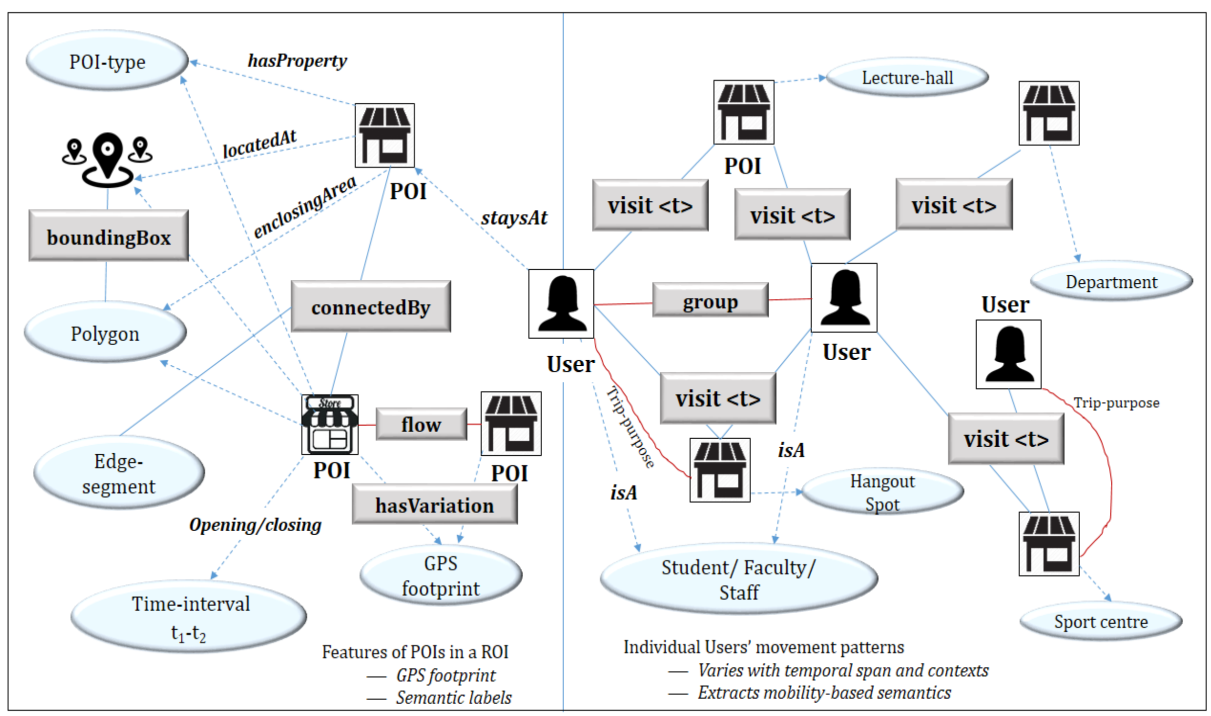
❿ *Transforming the underlying relations of mobility semantics in a machine-readable format to support information retrieval and query-processing*

Graph Storage

❿ *Spatio-temporal mobility dataset can be represented by graph structure effectively instead of other storage*

Flexible Schema

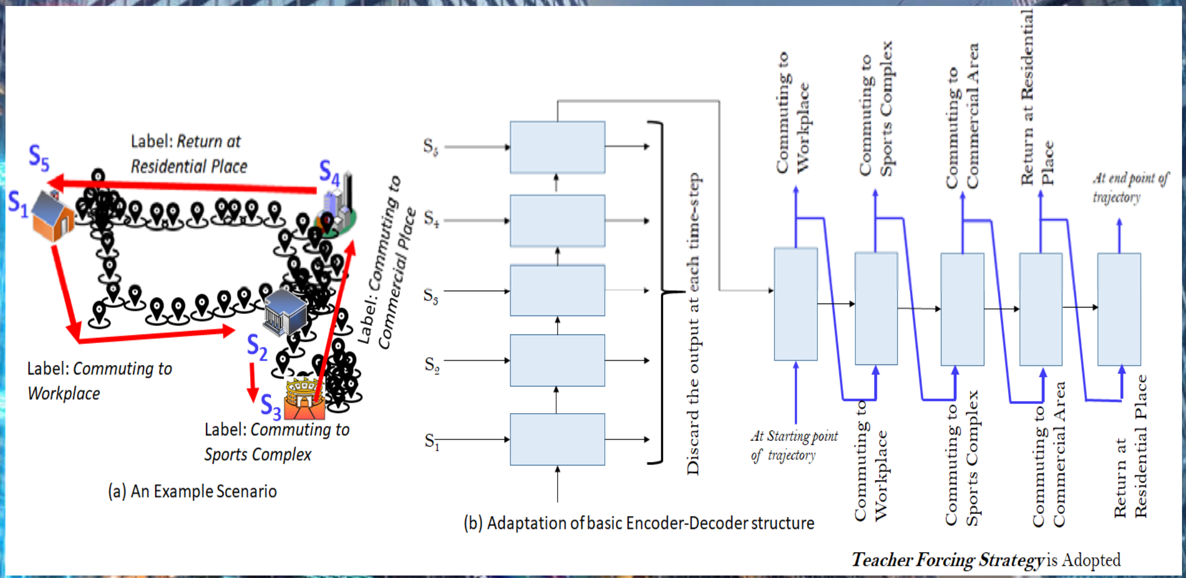
❿ *Updation of relations and facts can be easily incorporated in this structure compared to SQL-based processing*



- The mobility knowledge graph is formulated as triplet of $< s, r, o >$, where $s \subseteq \Omega$ and $o \subseteq \Omega$ are entities and $r \subseteq \varphi$ is the relation between two entities
- Each triplet (or fact) has a time-slot, when the fact is valid. The facts of MKG takes the form of $MF :< s, r, o, [t1, t2], f >$ where the $[t1, t2]$ entry denotes the time-interval when the fact is true
- f is the feature value of the relation. This f is introduced to capture the strength of the correlation

Annotation of the Trajectory segments

- *Multi-class single label classification*, where given the trajectory trace, the model outputs the corresponding labels of the stay-point transitions
- Users' movement behaviours in varied spatio-temporal contexts along with the sequential correlations efficiently
  - ❖ *LSTM architecture* can address these issues and capable to learn the long-range dependencies in sequential patterns
- Embedding Layer *(Location, time-of-visit, duration, distance_covered, time_taken)*



❖ Two trajectories with very different spatial and temporal scales may have the same moving behavior

For example, one's commute to work may take fifteen minutes and the other may take one hour

❖ The spatial and temporal scales of the trajectories representing similar moving behaviors may not be similar

❖ *Dynamic Time Warping (DTW)* or *Longest Common Subsequence (LCSS) distance* cannot be used as a robust method for extracting similar movement behaviours

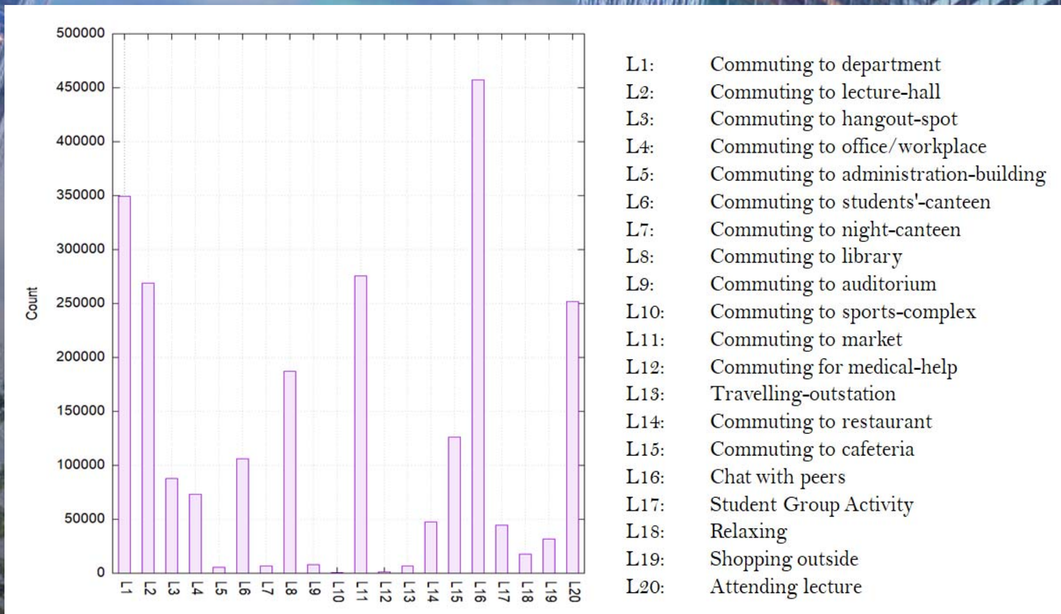
Mobility Knowledge Transfer

- Mobilytics deploys domain adaptation technique as the source instances and target data instances are from different distribution
- Utilize the labelled data instances of source ROI to classify un-labelled trajectory instances of target ROI
- Mobilytics considers ***agent*** as the **next transition predictor/ planner** of the user
  Given the input (current position) of the user and the environment, agent predicts the next transition (travel-distance to reach the next stay point) along with the time spent at the next stay-point
- ***Actions*** are defined in two-folds - (i) α = (1, *pa*,*i*): user visits the POI *pa* after traversing *i* distance and (ii) α = (0, *pa*, *t*): user stays at POI *pa* for *t* time-duration
- ***Reward*** is the weighted sum of the factors (i) *dtra*: reciprocal of the distance travelled between the real and predicted POI-visit; (ii) *durS*: reciprocal of the real and predicted time-duration spent at the POI; (iii) *act*: whether the action of staying or transition at a POI is correctly predicted

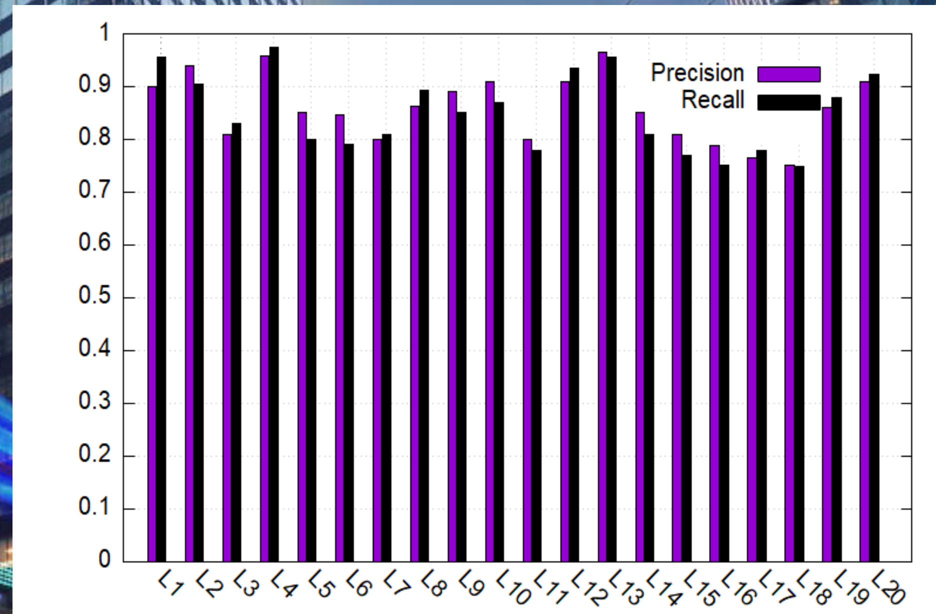
Data Modality	Source Domain (*SG*)	Target domain (*TG*)
D_1: Road Network Structure	Available	Available
G: Individuals' GPS trace (without semantics)	Available	Few Available
D_2: Labelled GPS Trace	Available	Not Available
D_3: Aggregated Movement Flow	Available	Available
Learning Target	-	Identification of POIs and Semantic labels

- Two-phase learning simultaneously in a feedback loop: (i) the agent attempts to learn the mapping from the source domain training dataset, (ii) available historical records of the mobility data of target domain is used to refine the policy
- Deep Q-network *[29]* has been deployed over the set-up

Experimental Observation



Top 20 semantic labels of mobility (trip purpose + activity) traces and their counts in the datasets



Precision and Recall measures for annotation of semantic labels

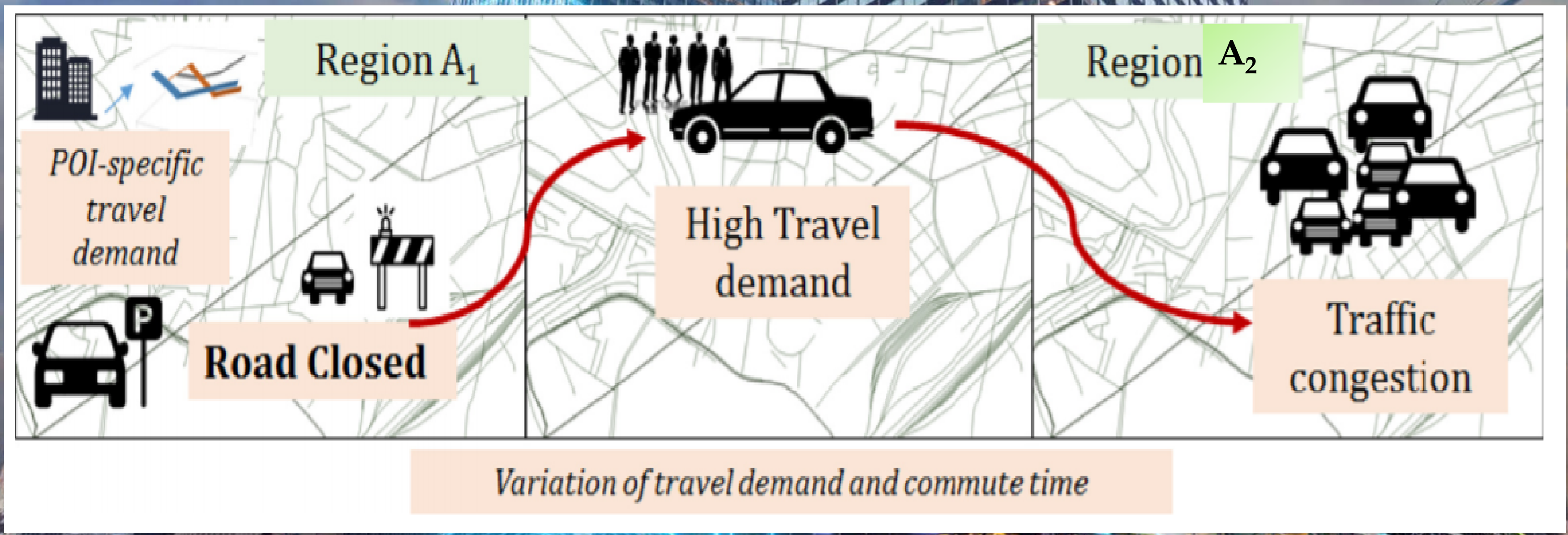
- *It is observed that the minimum precision and recall values are 0.752 and 0.748 for different semantic labels*
- *This results are the average values of precision and recall for randomly chosen user's trajectory traces*
- *Mobilytics has achieved an average of 0.852 precision value for the semantic trajectory annotation task which is quite high*

Experimental Observation

POI-type	TC	NC	POI-type	TC	NC
Academic Building	18	16	Hospital	1	1
Student Hall	23	22	ATM	3	2
Residence (Staff)	3	1	Bank	2	2
Student Canteen	6	5	Guest House	1	0
Auditorium	3	1	Library	1	1
Department	11	9	Sports Complex	2	1
Restaurant	6	5	Medical Store/ Center	3	2
Cafeteria	3	2	Post-office	1	1
Parking Area	25	23	Shopping Complex	3	3

Different identified POI-types in NITW campus and neighbouring areas after Transfer Learning (TC: Total count, NC: Number of Correctly Identified POI)

Mobility Association Rule Mining Framework (MARIO)



When the region A_1 of a city experiences traffic congestion, then with ***80%*** probability the traffic density of region A_2 will be higher after δ time-period and consequently high travel-demands in the spatial-neighborhood region of A_2

Ghosh, S., Ghosh, S. K., & Buyya, R. (2020). MARIO: A spatio-temporal data mining framework on Google Cloud to explore mobility dynamics from taxi trajectories. *Journal of Network and Computer Applications*, *164*, 102692.

Mobility Rule Template (MAR)

Rule Id	Rule template
MAR_1	$M_1(r_1, t_1, v_1) \Rightarrow M_2(r_2, t_2, v_2)$: Mobility event M_1 is followed by mobility event M_2
MAR_2	$timeVal(t) \wedge POI(p) \Rightarrow travelD(c, p, t)$: The travel demand of a region largely depends on the timestamp value and the place type information
MAR_3	$travelD(c_i, r_i, t_i) \Rightarrow footPrint(c_j, r_j, t_j) \wedge travelD(c_k, r_k, t_k)$: Travel demand in a particular region impacts footprint density and generate travel demand in other regions
MAR_4	$locTraversal(dis, S, D) \Rightarrow context \wedge timeVal(t)$: Location specific information of a taxi trip effects the context information, such as fare amount and trip time

Spatio-temporal Support ($\phi_i(r_i, t_i) \Rightarrow \phi_j(r_j, t_j)$):

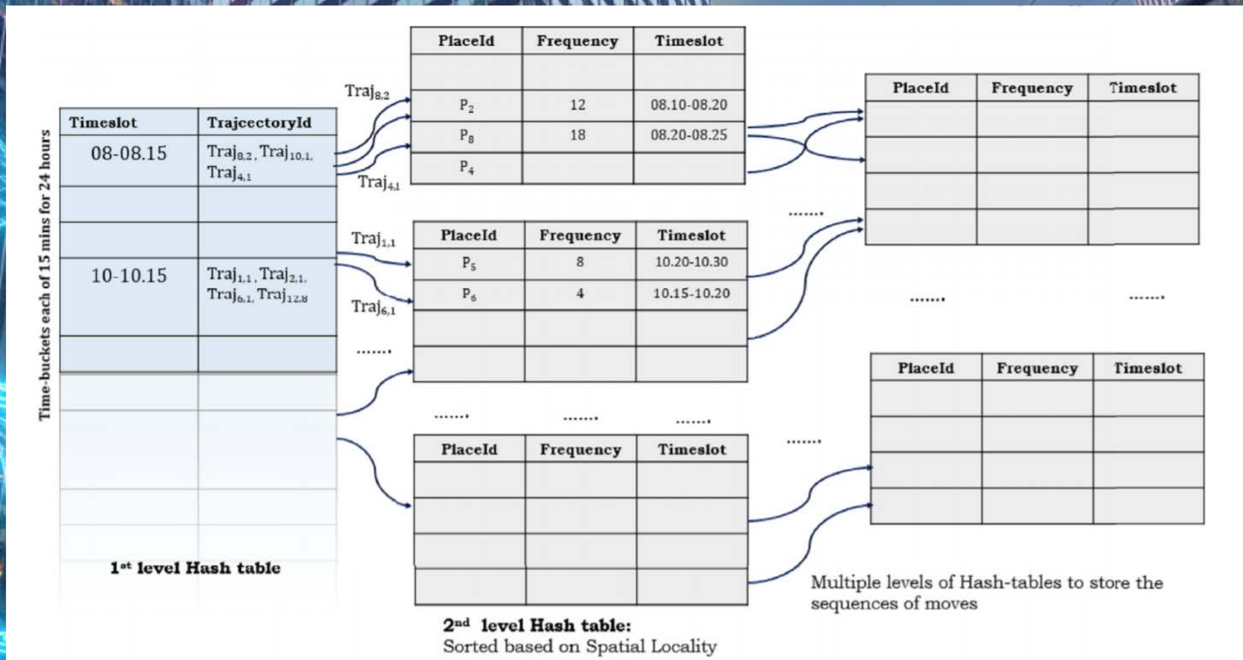
It is the scaled spatial coverage and the total length of the time-intervals in the rules

Spatio-temporal Confidence ($\phi_i(r_i, t_i) \Rightarrow \phi_j(r_j, t_j)$): It is measured as the conditional probability of the predicate ϕ_j is true given that ϕ_i is already true.

$$STconf = \frac{STsupp(\phi_i(r_i, t_i) \wedge \phi_j(r_j, t_j))}{STsupp(\phi_i(r_i, t_i))}$$

Mobility Index Construction

- MARIO deploys the k-level temporal hashing scheme to store trajectory-sequences of a region into k layers
- Beneficial for extracting movement information efficiently and in timely manner
- For example, the most followed route of a region can be discovered by analyzing the k-sequences and frequency
- On the other side, traffic states of a place can be explored by interpreting the GPS footprints of the taxi-ids in different time-slots



- *The 1st level hash table contains trajectory-ids (Traj) of different taxi-trips starting at varied time-intervals. The keys of the table are time-slots of 15 min each for a day*
- *From the next level, sequences of the trajectory-segments are maintained in different levels along with the place-id (type and location), frequency of visit and time-slots*

Cloud DataProc | Cloud Spanner

Movement Dynamics Network

- M-flow (mobility flow) follows Apriori Property: Any sub-set of infrequent spatio-temporal event-set (sequential pattern) is also infrequent

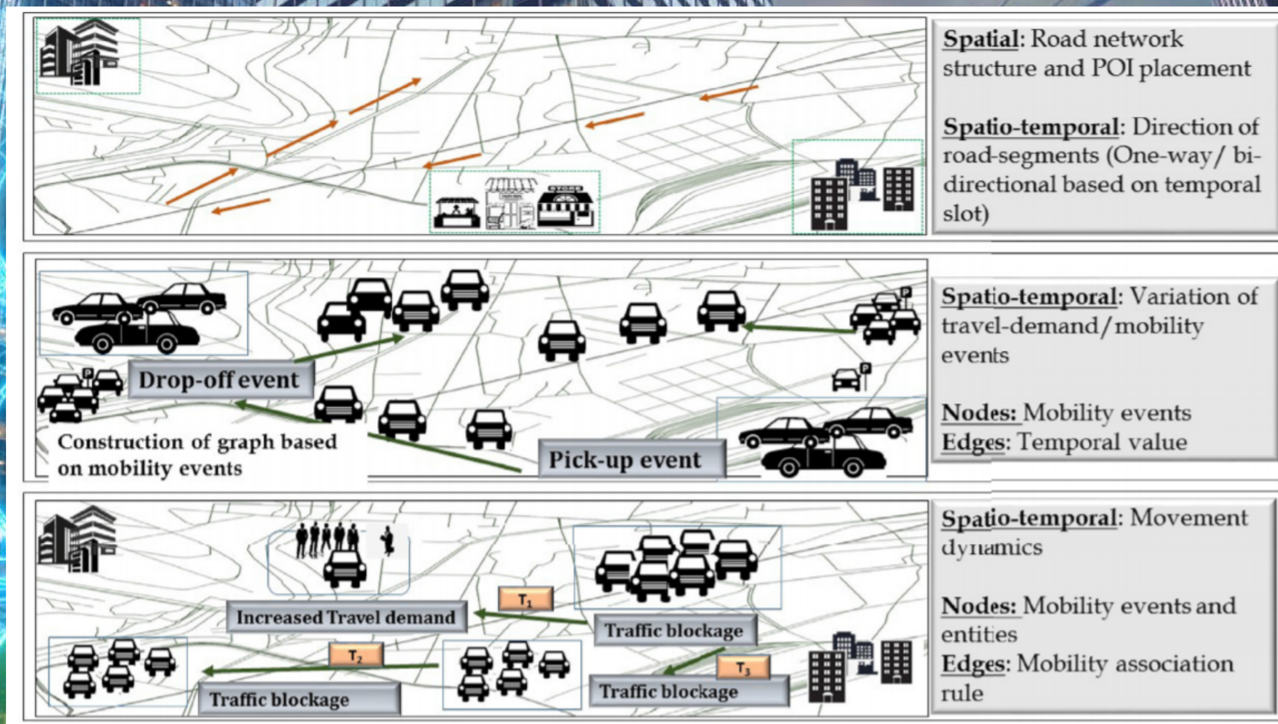
- The algorithm finds out i-itemsets (i.e., itemsets with i items having at least the minimum support) at pass i.

- It generates the set of candidate of i-itemsets and computes the count by scanning the database

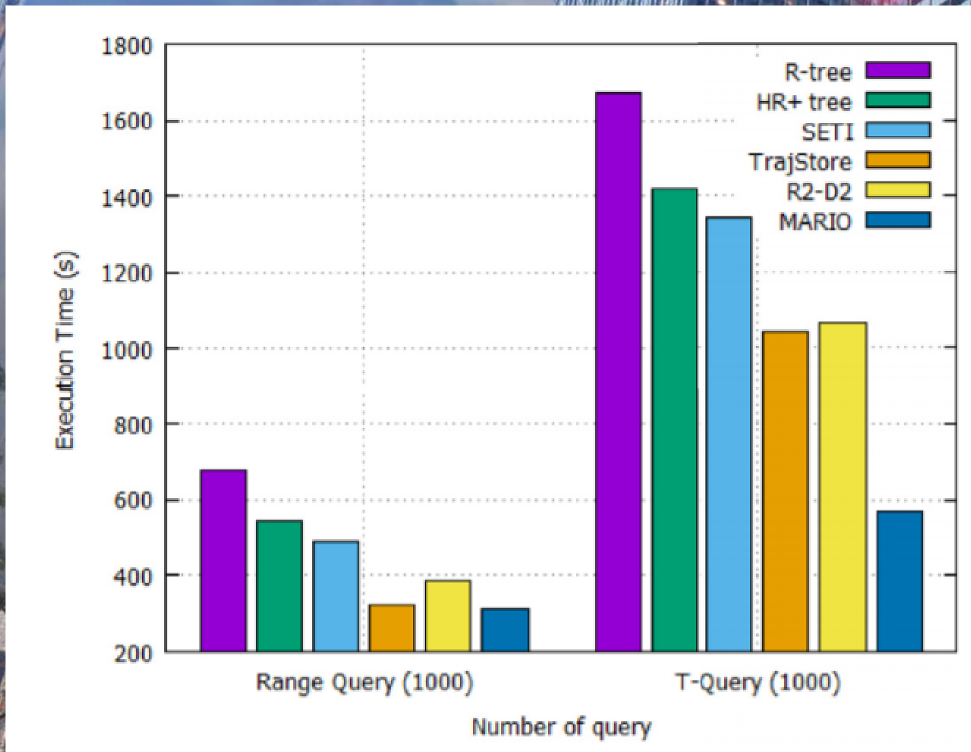
- Finally it finds out the MARs by inspecting the spatio- temporal support of all the candidate itemsets

- MAR template is used as an input

- The algorithm works in a bottom-up fashion where small temporal scale is used in the first phase followed by grouping larger time-slots given that spatio-temporal support and confidence are larger than the minimum threshold



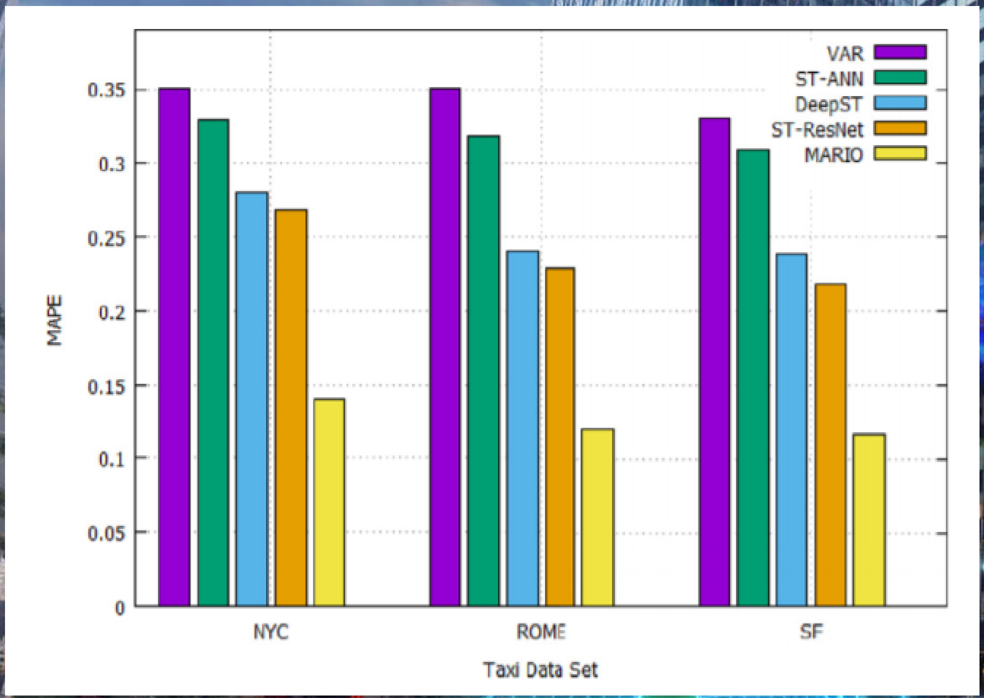
Experimental Observation



- *All experimental evaluations are carried out on VM instance of Google Cloud Platform7 having 4 vCPUs, 15 GB memory and Ubuntu 16.04, Linux as the OS*
- *The algorithms are implemented in Python, R ,and all the experiments are performed on three real datasets (Dataset III) of taxi trajectories*

- k-level temporal hashing scheme has outperformed other baselines in a huge margin *(almost 50% less execution time in average)*
- Trip-sequences of a region are stored into k-temporal levels and in consecutive buckets following the hash-function based on latitude and longitude information
- It helps to extract the range and T-query in an efficient manner compared to other methods

Experimental Observation



Average MAPE value (10 days) of travel demand in 3 cities when specific events occurred

- We select 10 days from all three datasets when specific events (accident, road-blockage, crowd due to social event) occurred and evaluate the travel demand in different places of the cities.
- It is observed that MARIO performs significantly better than other baselines in these 10 days when any event occurs.
- MARIO not only models the spatial and temporal travel demand patterns, it is also capable to model the variation of travel demand occurring due to some events

Experimental Observation

M-Rule	S	C
$\boldsymbol{MR_1}$: TimeStamp$(T_1 \wedge T_3) \wedge Weekday \Rightarrow travelDemand(High, R_1)$	0.28	0.867
$\boldsymbol{MR_2}$: TimeStamp$(T_1 \wedge T_2) \wedge Weekday \Rightarrow travelDemand(High, R_3)$	0.21	0.785
$\boldsymbol{MR_3}$: TimeStamp$(T_3 \wedge T_4) \wedge Weekday \Rightarrow travelDemand(High, R_2)$	0.182	0.843
$\boldsymbol{MR_4}$: TimeStamp$(T_1 \wedge T_2 \wedge T_3) \wedge Weekday \Rightarrow travelDemand(High, R_5)$	0.145	0.874
$\boldsymbol{MR_5}$: TimeStamp$(T_1 \wedge T_2 \wedge T_3) \wedge Weekend \Rightarrow travelDemand(High, R_4)$	0.128	0.902
$\boldsymbol{MR_6}$: $Trip - duration(small) \wedge TimeStamp(T_1 to T_3) \Rightarrow noOfTrips(high) \wedge Region(R_1, R_3, R_5)$	0.23	0.821
$\boldsymbol{MR_7}$: $Trip - duration(small) \wedge TimeStamp(T_4 \wedge T_3) \Rightarrow noOfTrips(high) \wedge Region(R_2)$	0.24	0.876
$\boldsymbol{MR_8}$: $Trip - duration(large) \wedge TimeStamp(T_1 \wedge T_4) \Rightarrow noOfTrips(small) \wedge Region(R_5, R_6)$	0.206	0.743
$\boldsymbol{MR_9}$: $Trip - duration(large) \wedge TimeStamp(T_1 \wedge T_3) \Rightarrow noOfTrips(small) \wedge Amount(High) \wedge Region(R_6 \wedge R_4)$	0.32	0.870
$\boldsymbol{MR_{10}}$: $Trip - duration(small) \wedge TimeStamp(T_1 to T_3) \Rightarrow noOfTrips(large) \wedge Amount(Medium) \wedge Region(R_1 \wedge R_2 \wedge R_3)$	0.217	0.817

Mobility Rules and Evaluation Metrics.
S:Spatio-temporal Support, C: Spatio-temporal Confidence, T1: 0800–1000, T2: 1000–1600, T3: 1600–2100, T4: 2100 - 0800, R1: Residential area, R2: Commercial and entertainment region, R3: Academic area, R4: Areas of historic interest, R5: Railway station and Bus-stops, R6: Airport region

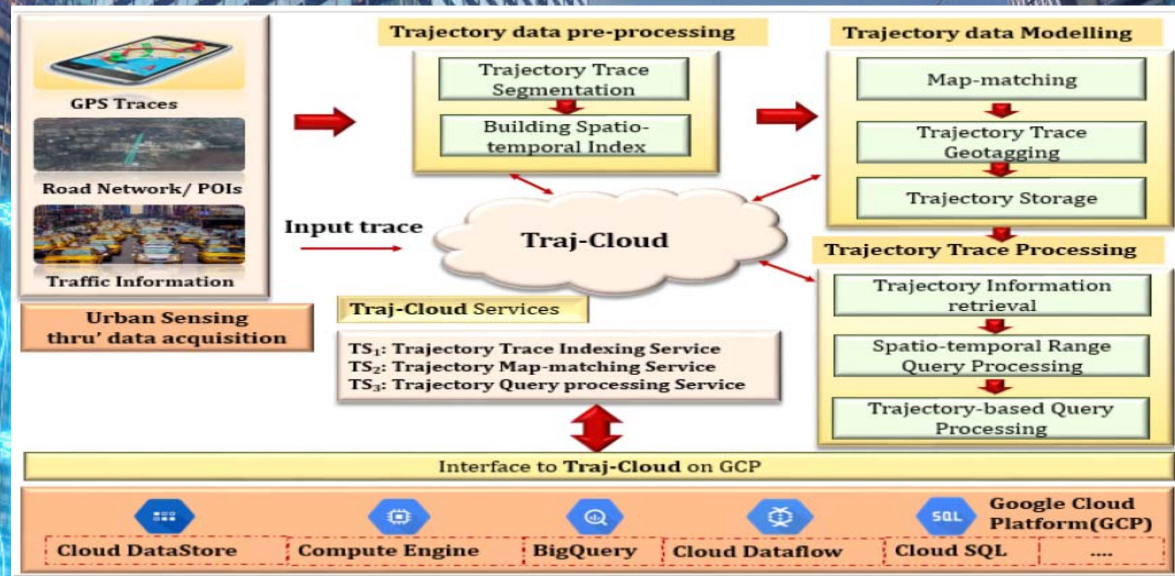
USE-CASES:

SPATIAL BIG DATA SYSTEM – CLOUD-EDGE-FOG ARCHITECTURE

PennState

Traj-Cloud (Motivations and Objectives)

- *Search, access and utilize* geospatial data
- *Configure computing infrastructure* to enable the computability of intensive simulation models
- *Adopt spatiotemporal principles* to support spatiotemporal intensive applications



The emergence of cloud computing provides a potential solution with an ***elastic, on-demand computing platform to integrate – observation systems, parameter extracting algorithms, phenomena simulations, analytical visualization and decision support****, and to provide social impact and user feedback – the essential elements of the geospatial sciences*

Traj-Cloud Services

Trajectory data Indexing Service (TS_1):
Input: GPS trajectory trace (G) and other semantic information, such as, geotagged locations or road network
Output: Spatio-temporal indices of input traces and storage of the information
GCP Component: Google BigQuery and Cloud SQL storage.

[It may be noted that the service also effectively partitions the road network structure and stores mapmatched trajectory trace effectively to ease the updation and extraction process.]

Trajectory Map-matching Service (TS_2):
Input: GPS trajectory trace (G) and road network (R)
Output: Projection of G into the corresponding R utilizing the MapReduce based platform to effectively handle huge dataload in near real-time.
GCP Component: Google Compute Engine

Ghosh, S., & Ghosh, S. K. (2019, January). Traj-cloud: a trajectory cloud for enabling efficient mobility services. In *2019 11th International Conference on Communication Systems & Networks (COMSNETS)* (pp. 765-770). IEEE.

Traj-Cloud Services

Trajectory Query Service (TS_3):
Input: GPS trajectory trace (G) log, Trajectory point and range Query Q

Output: Trajectory Trace (Point or Line shape)

GCP Component: Google Compute Engine and Cloud SQL

Point Query

— Find all the petrol stations within 500m distance of a vehicle's trajectory

Select poi_Id from POI P, Traj T where P.id="petrol_st" and overlap(P.shape,Buffer(T.shape,500))=1;

Range Query

— Find trajectory segments passing through the residential area of a city

Select traj_Id from Traj T, Region R where R.id="residential_zone" and cross(T.shape,R.shape)=1 ;

Traj-Cloud: Trajectory Map Matching Service (TS_2):

Given a set of GPS trajectory trace $T(t_1 \rightarrow \ldots t_n)$ and road network (R), TS_2 yields the map-matched trajectory trace (W).

- Generate the bounding box of the line segment (T) and calculate the geo-hashcode.
  - Based on the geo-hashcode, the geo-hashcode of the regions whose geometry overlaps or intersects with T are extracted.

- In the Map phase, the input of the mapper function is $< t_i; j >$, where t_i denotes the GPS point of the trajectory trace and j represents the corresponding grid-id where the GPS point remains.
  - The mapper function extracts all the edges (E′) within the grid j.
  - Performs point-to-curve (dist) distance calculation and extracts the nearest 3 edges.
  - Produces $< t_i; (\{e_a; v\}; \{e_b; v\}; \{e_c; v\}) >$ in ascending order of distance

- The geotagging process annotates the GPS points of the trajectory data with the nearest landmark information. For tagging landuse information, iterative reverse geocoding (IRG) is used

Traj-Cloud: Trajectory Query Processing Service (TS_3):

- The query-processing of Traj-Cloud is performed in MapReduce platform, where the mapper executes the filter step and reducer executes the refinement step of the query processing

  — In filter step, candidate spatial objects, which may satisfy the trajectory query condition are extracted
  — In refinement step, each candidate spatial object is analyzed whether they are satisfying the condition or not

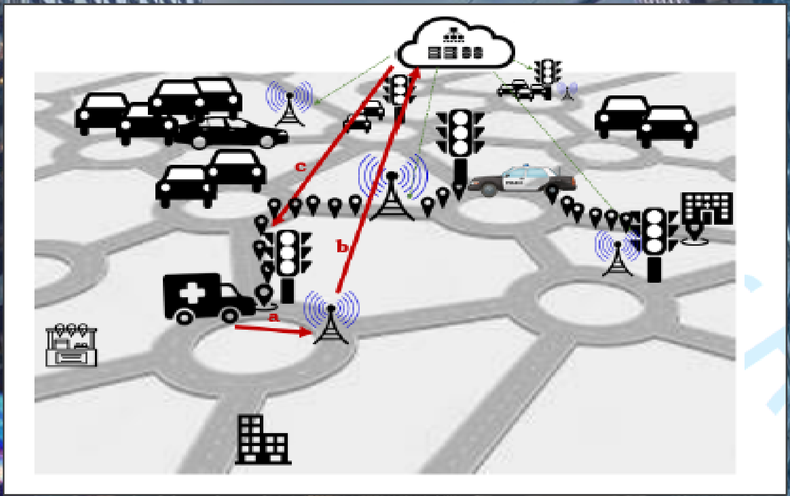
- The intersection (Q.shape) function returns all the spatial objects that intersects with the query object

  — We extract all such grids of the network, where Q.shape resides
  — The input of the mapper function for intersection is $< Q.shape_g; Grid >$ and it produces key value pair as $< Grid_{id}; Q.shape_{gj} >$
  — The mapper function segments the spatial object and extracts different grids where the segments reside
  — The combiner function finds out all objects residing in the extracted grids
  — In the reducer function, each of the objects are analyzed individually to find out whether they intersect with Q.shape or not

- The input of the mapper function for withinDistance is $< Q.shape; value >$, where Q:shape represents the geometry of the query spatial-object and distance is measured by the variable value

Mobi-IoST (Mobility-aware Internet of Spatial Things)

- ❑ To facilitate real-time applications, high-end processing and storage units are required.
- ❑ However, the cloud-only set-up is not an energy-efficient and delay-aware solution for handling such a high volume of data
- ❑ Seamless connectivity due to the mobility of IoT devices is a crucial factor to process the data in the remote cloud servers
  - *For time-critical applications such as health care, connection interruption and consequently the increase in delay in delivering the processed information, result in poor Quality of Service (QoS)*



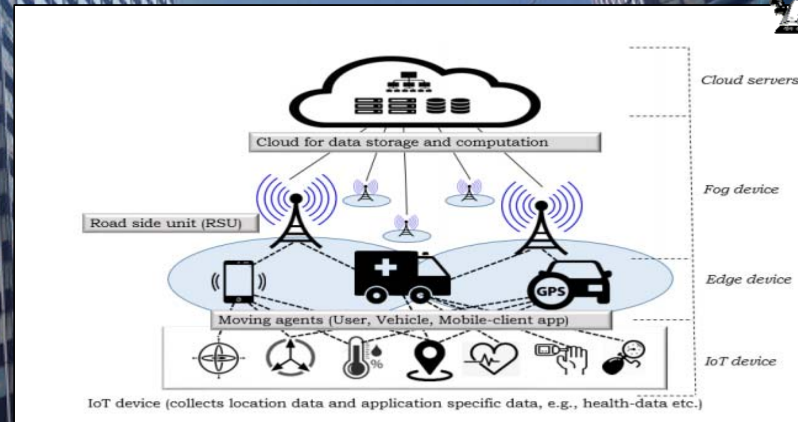
- ❑ Ambulance sends health data generated from IoT devices to RSU
- ❑ RSU sends the result to cloud along with the location information
- ❑ Cloud predicts the nearby health care centre, shortest path and helps to actuate traffic signal

Internet of Spatial Things (IoST) brings IoT in the spatial context (Eldrandaly et al. [33])

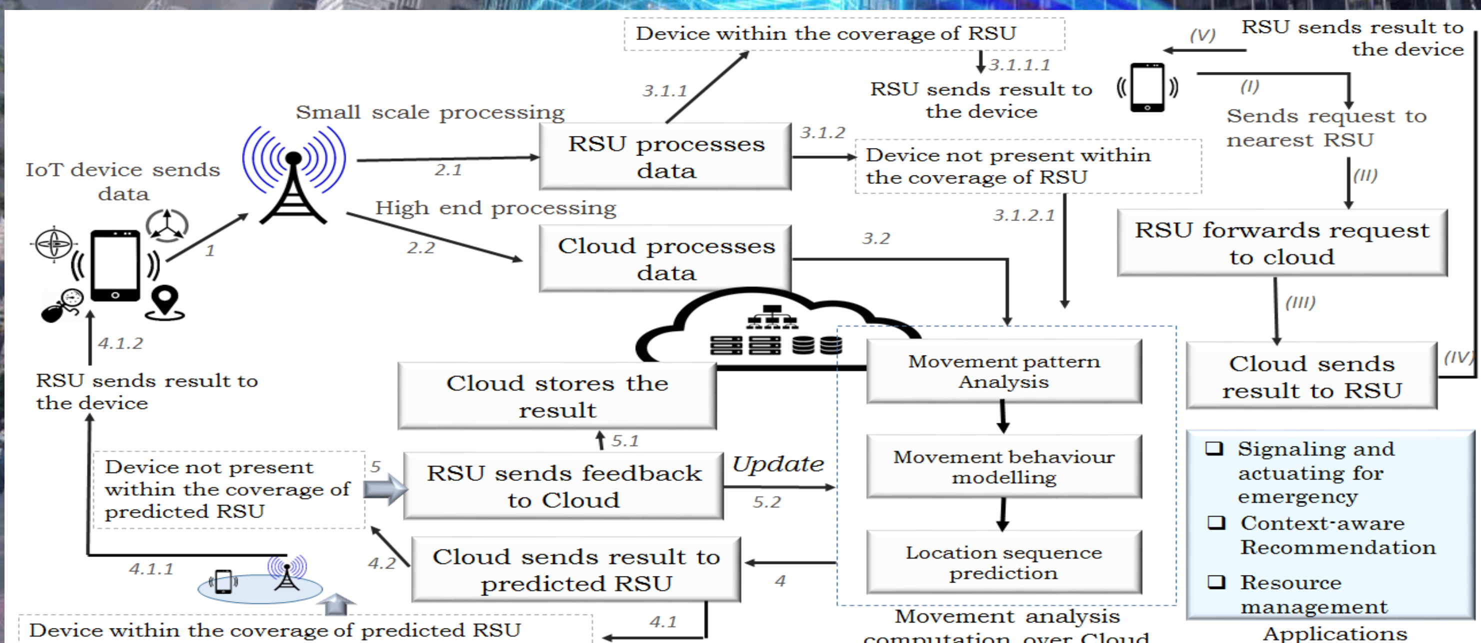
Ghosh, S., Mukherjee, A., Ghosh, S. K., & Buyya, R. (2019). Mobi-iost: mobility-aware cloud-fog-edge-iot collaborative framework for time-critical applications. *IEEE Transactions on Network Science and Engineering*, *7*(4), 2271-2285.

Mobi-IoST

- Information processing and delivering result based on the prediction of user's current location
- Exploits the mobility knowledge of the agents to predict the probable user location



Hierarchical placement of IoT, Edge, Fog devices and Cloud in Mobi-IoST framework



SK Ghosh, S Ghosh

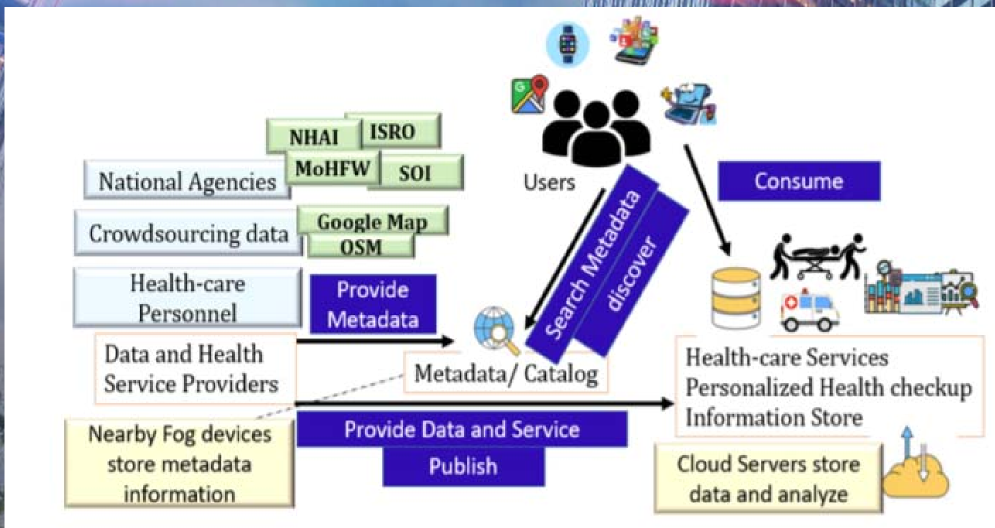
Working model of Mobi-IoST

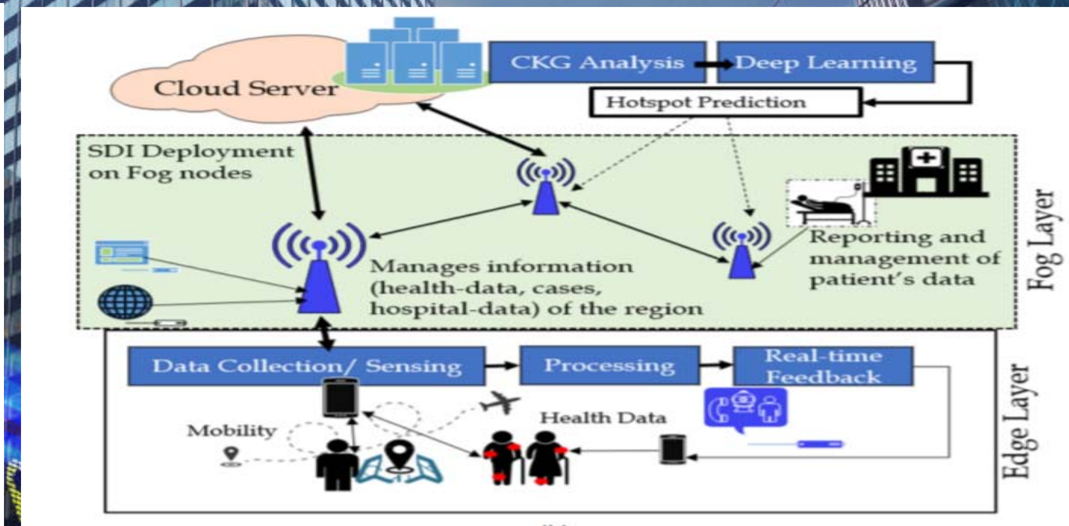
USE-CASES:

LOCATION AWARE INTERNET OF HEALTH THINGS (IOHT)

PennState

Location-based Application – IoSHT for combating COVID-19

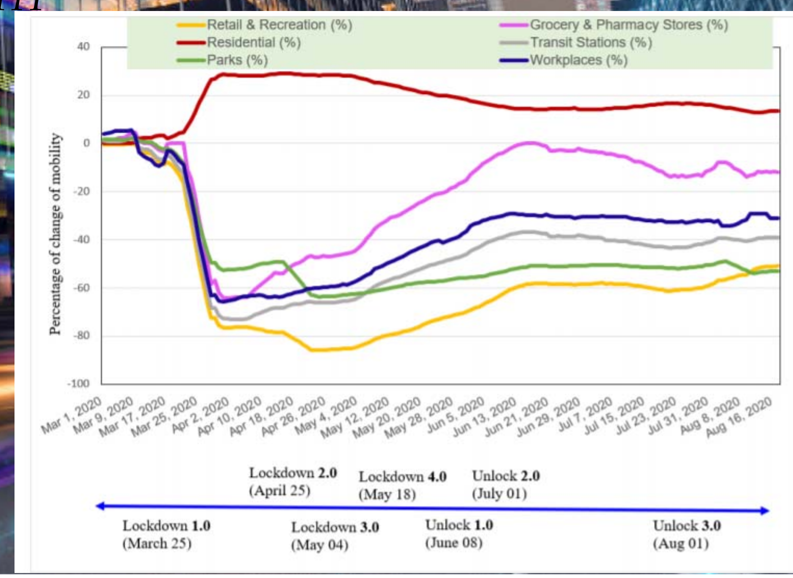




SDI (Spatial Data Infrastructure) in the context of IoSHT

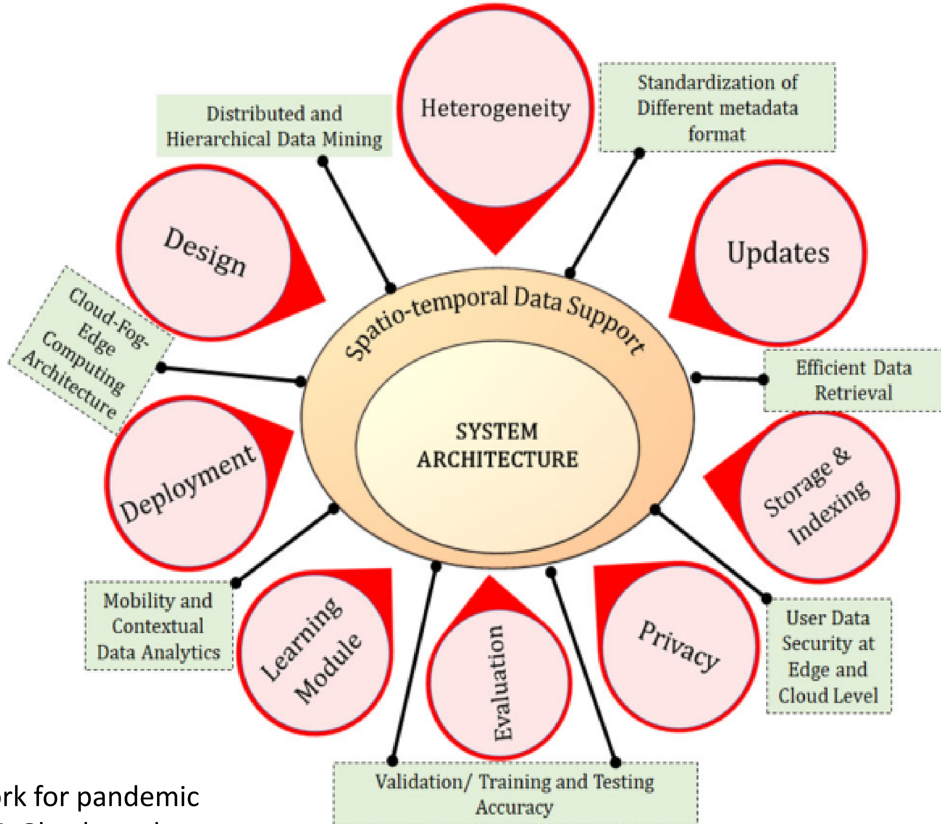
- SDI : *Ease of search and discovery of geospatial services and information • reduce data duplication of information among the national agencies (government) • seamless data sharing technique • maintaining the data integrity and privacy*

- Next hot-spot prediction: *Whether and how cascading and co-occurrence movement patterns impact the disease spread?*



SPATIAL DATA INFRASTRUCTURE

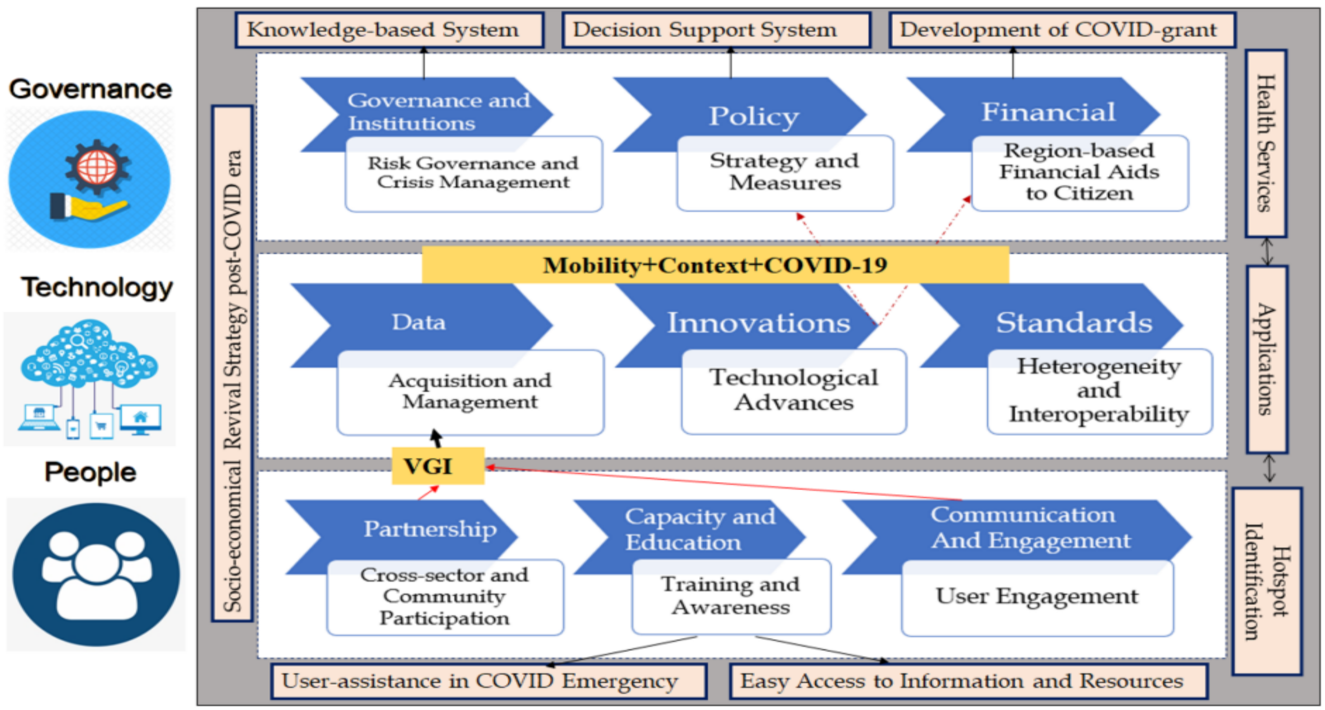
- ❑ Search and discovery of spatio-temporal services and information efficiently and with minimal manual inference
- ❑ Reducing data duplication among the national agencies (government) and eliminating false information by capturing and sharing reliable data
- ❑ Heterogeneous data sharing in seamless manner
- ❑ Data integrity and privacy conservation



STOPPAGE: Spatio-temporal data driven cloud-fog-edge computing framework for pandemic monitoring and management Shreya Ghosh, Anwesha Mukherjee, Soumya K. Ghosh, and Rajkumar Buyya. Software: Practice and Experience (2022)

PennState

STRUCTURAL MODEL OF INTEGRATED GEOSPATIAL INFORMATION FRAMEWORK IN THE CONTEXT OF COVID-19 PANDEMIC



Ghosh, S., & Mukherjee, A. (2022). STROVE: spatial data infrastructure enabled cloud–fog–edge computing framework for combating COVID-19 pandemic. *Innovations in Systems and Software Engineering*, 1-17.

PennState

THANK YOU!

Soumya K Ghosh (IIT Kharagpur)
Shreya Ghosh (Penn State, USA)

skg@cse.iitkgp.ac.in, shreya@psu.edu

PennState