Welcome to the course on

Basics of Geocomputation and Geoweb Services

Duration: October 19, 2020 to October 29, 2020

Course Contents

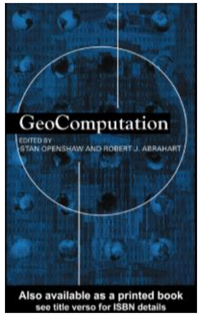
- Introduction to Geocomputation- Technology and applications
- Introduction to online GIS and Geo-web services
- Open Geodata Repositories & ISRO Geoweb Services for thematic applications
- Programming concepts for Geo-computation- Introduction to Python and R
- Overview on concept of DBMS, RDBMS and SDBMS for geo-data handling
- Basics of cyber infrastructure requirement for GIS, computing, storage, archive, network and firewall infrastructure in GIS lab
- Crowdsourcing and participatory GIS using online tools & technologies
- Practical Demonstration on: Virtual reality, 2D and 3D geo-visualizations in web platforms
- Practical Demonstration on: Introduction to Cloud GIS

Introduction to Geocomputation- Technology and Applications

Dr. Harish Chandra Karnatak
Head & Scientist SG
Geoweb Services, IT & Distance Learning (GIT&DL) Department
Geospatial Technology an Outreach Programme Group

Email- harish@iirs.gov.in

Geo-computation



Few Definitions by the Researchers:

- The art and science of solving complex spatial problems with computers (www.geocomputation.org);
- The process of applying computing technology to geographical problems (Rees and Turton, 1998)
- The computational theory of complex socio-temporal processes (Couclelis, 1998);
- Geo-computation is the ways in which new computational tools and methods are used. also fundamentally about the depiction of spatial process (Longley, 1998).

We can say- Geocomputation is a method to use computations for processing and analysing geographical data for information extraction and decision making

Geo-computation

- Stan Openshaw is generally recognized as the father of Geo-Computation. He state that "Geocomputation is about using the various different types of geo-data and about developing relevant geo-tools within the overall context of a 'scientific' approach

- Stan identified three aspects that make Geo-Computation special .
    - The first is emphasis on "geo" subjects, i .e .GeoComputation is concerned with geographical or spatial information .
    - Second, the intensity of the computation required is distinctive .It allows new or better solutions to be found for existing problems, and also lets us solve problems heretofore insoluble.
    - Finally, GeoComputation requires a unique mind set, because it is based on "...replacing vast amounts of computation as a substitute for missing knowledge or theory and even to augment intelligence."

Geo-computation in context

Following four questions to be most important with respect to the definition given by Stan Openshaw et al :

- is there a focus on high performance computing?
- is Geo-computation simply a grab bag of tools?
- What is the relationship between GIS and Geo-computation?
- what are the key concepts of Geo-computation?

What is Geo-data

- A spatial database is a database that is optimized to store and query data that is related to objects in space, including points, lines and polygons.
- While typical databases can understand various numeric and character types of data, additional functionality needs to be added for databases to process spatial data types. *These are typically called geometry or feature.*

What is Geo-data (Cont.)

In addition to typical SQL queries such as SELECT, CREATE statements, spatial databases can perform a wide variety of spatial operations like:

- *Spatial Measurements*: Finds the distance between points, polygon area, etc.
- *Spatial Functions*: Modify existing features to create new ones, for example by providing a buffer around them, intersecting features, etc.
- *Spatial Predicates*: Allows true/false queries such as 'is there a residence located within a mile of the area we are planning to build the landfill?`

What is Geo-data (Cont.)

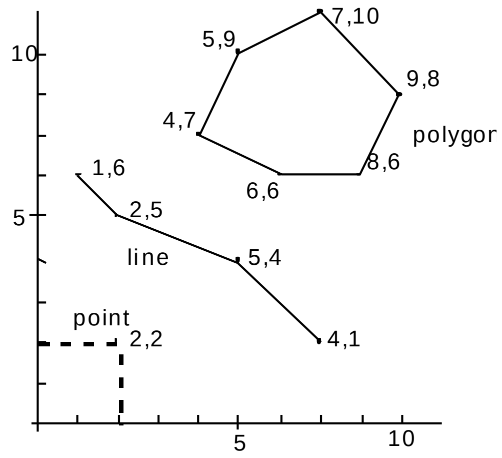
In addition to typical SQL queries such as SELECT, CREATE statements, spatial databases can perform a wide variety of spatial operations like:

- *Constructor Functions*: Creates new features with an SQL query specifying the vertices (points of nodes) which can make up lines. If the first and last vertex of a line are identical the feature can also be of the type polygon (a closed line).
- *Observer Functions*: Queries which return specific information about a feature such as the location of the center of a circle

Due to these special characteristics and specific query requirement, a standard database management system may not be sufficient for storage and management of spatial data. However many industry standard RDBMS packages are now supporting spatial objects inside RDBMS.

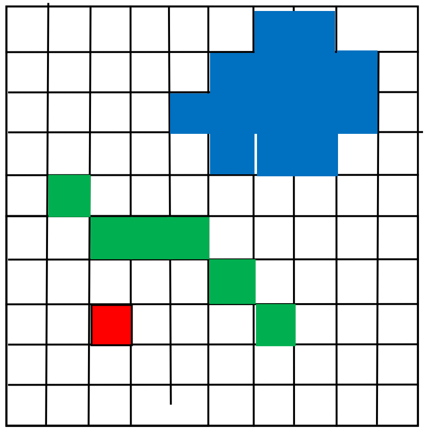
Spatial data storage

- Vector model



As geometric objects: points, lines, polygons

- Raster model



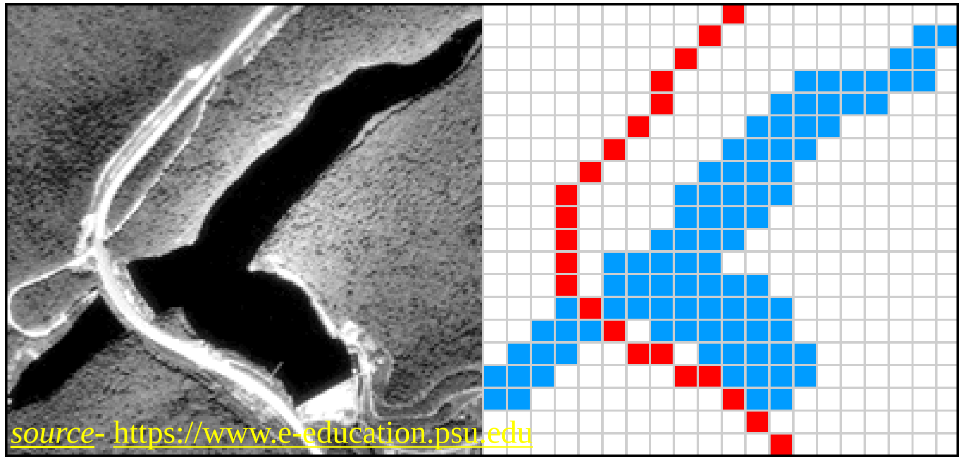
As image files composed of grid-cells (pixels)

Vector data model

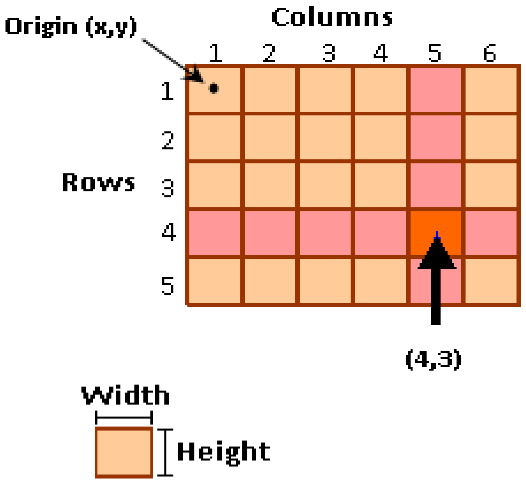
- Advantage of the vector data format: allows precise representation of points, boundaries, and linear features.
  - useful for analysis tasks that require accurate positioning,
  - for defining spatial relationship (i.e. the connectivity and adjacency) between coverage features (topology), important for such purposes as network analysis (for example to find an optimal path between two nodes in a complex transport network)
- Main disadvantage of vector data is that the boundaries of the resulting map polygons are discrete (enclosed by well-defined boundary lines), whereas in reality the map polygons may represent continuous gradation or gradual change, as in soil maps.

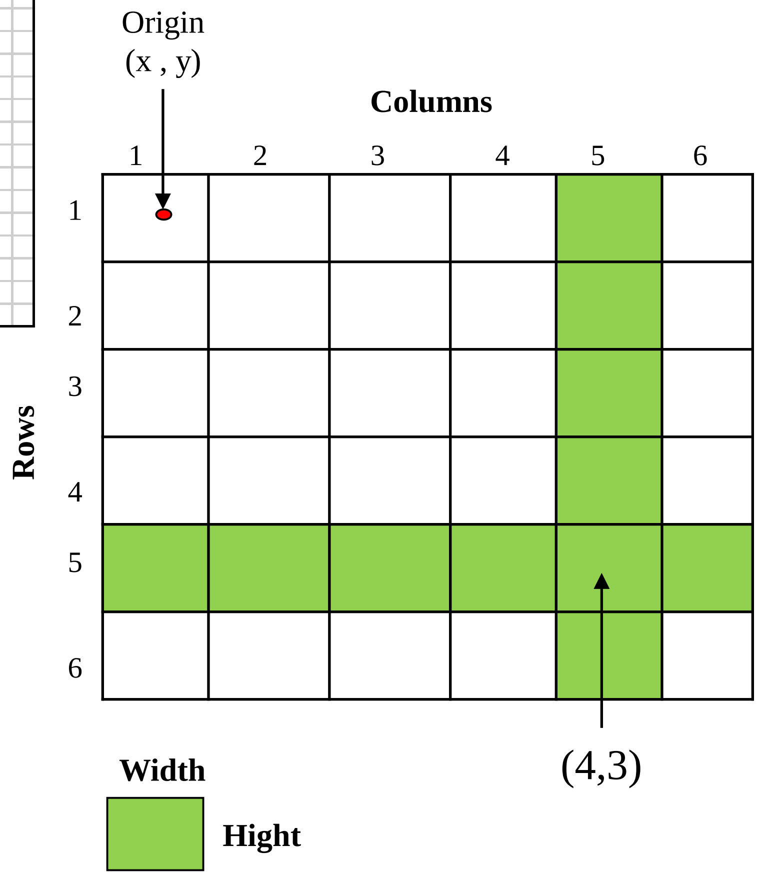
Raster data model

- Good for representing indistinct boundaries thematic information on soil types, soil moisture, vegetation, ground temperatures
- As reconnaissance satellites and aerial surveys use raster-based scanners, the information (i.e. scanned images) can be directly incorporated into GIS
- The higher the grid resolution, the larger the data file is going to be



source- https://www.e-education.psu.edu





Projection, Scale, Accuracy and Resolution

the key properties of spatial data

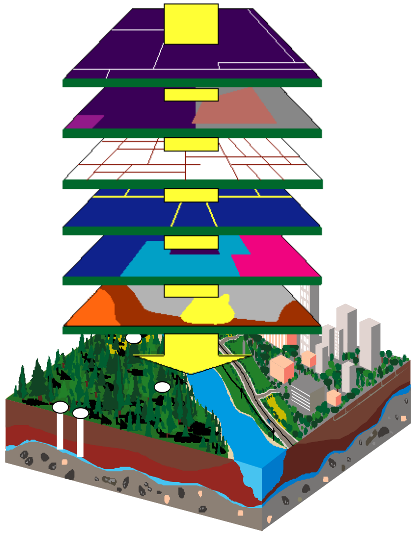
- **Projection:** the method by which the curved 3-D surface of the earth is represented by X,Y coordinates on a 2-D flat map/screen
  - distortion is inevitable
- **Scale:** the ratio of distance on a map to the equivalent distance on the ground
  - in theory GIS is scale independent but in practice there is an implicit range of scales for data output in any project
- **Accuracy:** how well does the database info match the real world
  - *Positional*: how close are features to their real world location?
  - *Consistency*: do feature characteristics in database match those in real world
    - is a road in the database a road in the real world?
  - *Completeness*: are all real world instances of features present in the database?
    - Are all roads included.
- **Resolution:** the size of the smallest feature able to be recognized
  - for raster data, it is the *pixel* size

The tighter the specification, the higher the cost.

Geo-Modelling

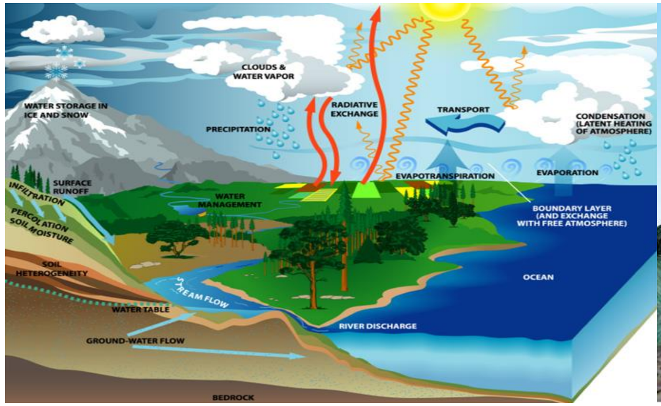
What is Model?

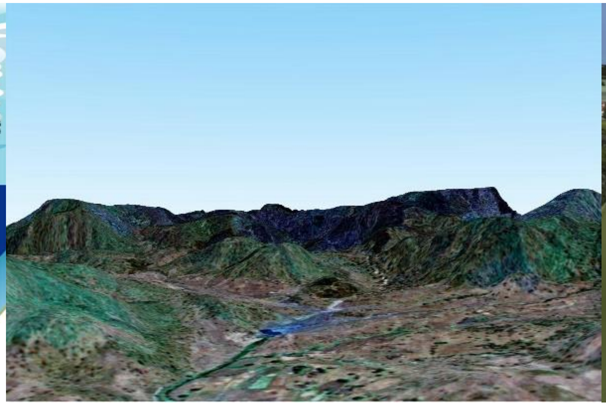
"Abstract Representation of Reality."



Spatial Model

- A model of some process operating in space (and time)
  - there is variation across the space (and through time)
  - location is important
    - the results of modeling change when locations change
    - locations must be known





Scale and digital models

- Digital models don't have a scale factor
  - but they operate at limited spatial resolution
- Spatial resolution is a critical factor
  - it determines:
    - what is left out of the model
    - the cost of collecting data and running the model
  - it contributes to the model's accuracy
    - the degree of uncertainty about the real world created by the model
- Temporal resolution is important for the same reasons

Analyze or model?

- Analysis:
  - static, one point in time
  - searching for patterns, anomalies
  - generating ideas and hypotheses
  - evaluating
- Modeling:
  - may be dynamic, multiple points in time
  - implementing ideas and hypotheses
    - to compare to the real world
  - experimenting with scenarios

Modelling & Simulation

- Modeling is the act of building a **model**. A simulation **is the process** of **using a model** to study the behavior and performance of an actual or theoretical system. While a model aims to be true to the system it represents, a simulation can use a model to explore states that would not be possible in the original system.

- Models: Three types
  - Physical
  - Process
  - Mathematical
- Simulations: Three Types
  - Live
  - Virtual
  - Constructive

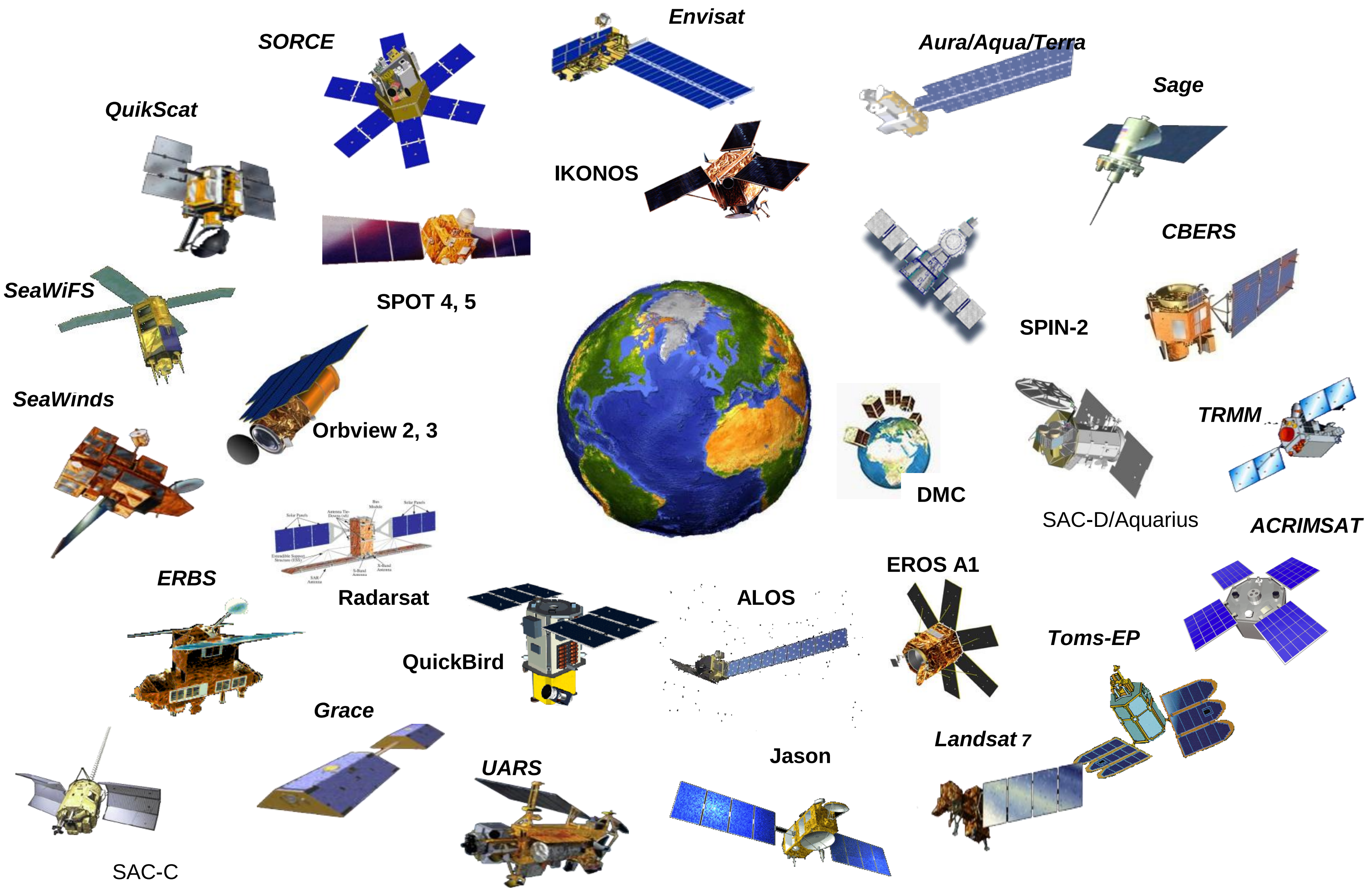
Geo-computation Methods

- Geo-simulation (CA and ABM)
- Artificial Intelligence and Machine Learning
- Artificial Neural Networks (ANNs)
- Evolutionary computing (EC)
- Automated zoning/re-districting (e.g. AZP)
- Cluster hunting (e.g. GAM/K)
- Interactive data mining tools (e.g. brushing and linking, cross-tabbed attribute mapping)
- Visualisation tools (e.g. 3D and 4D visualisation, immersive systems… some also very new!)
- Advanced raster processing (e.g. ACS/distance transforms, visibility analysis, image processing etc.)
- Heuristic and metaheuristic spatial optimisation, …. and more!

Emerging computation challenges in GIS

- Big data analytics for geo-spatial data analysis, processing and modelling;
- Cloud computing for geo-spatial domain [IaaS, PaaS, SaaS and MaaS implementations)
- Symantec Sensor web;
- 3D and 4D GIS using ORDBMS;
- Participatory GIS- Data quality and validations;
- Real time simulation and virtualizations;
- Geo-computation and spatial analysis and modelling;
- Spatial computation concepts (network, track and distance computation),
- Computational spatial statistics, Search, classification, prediction and modelling, Knowledge discovery in spatial data.
- Contents based image retrieval and applications.
- Data mining and its applications using pattern recognition.

There is plenty of data from space….



Alternate Remote Sensing Platforms are generating large data sets

Airborne Sensor platform

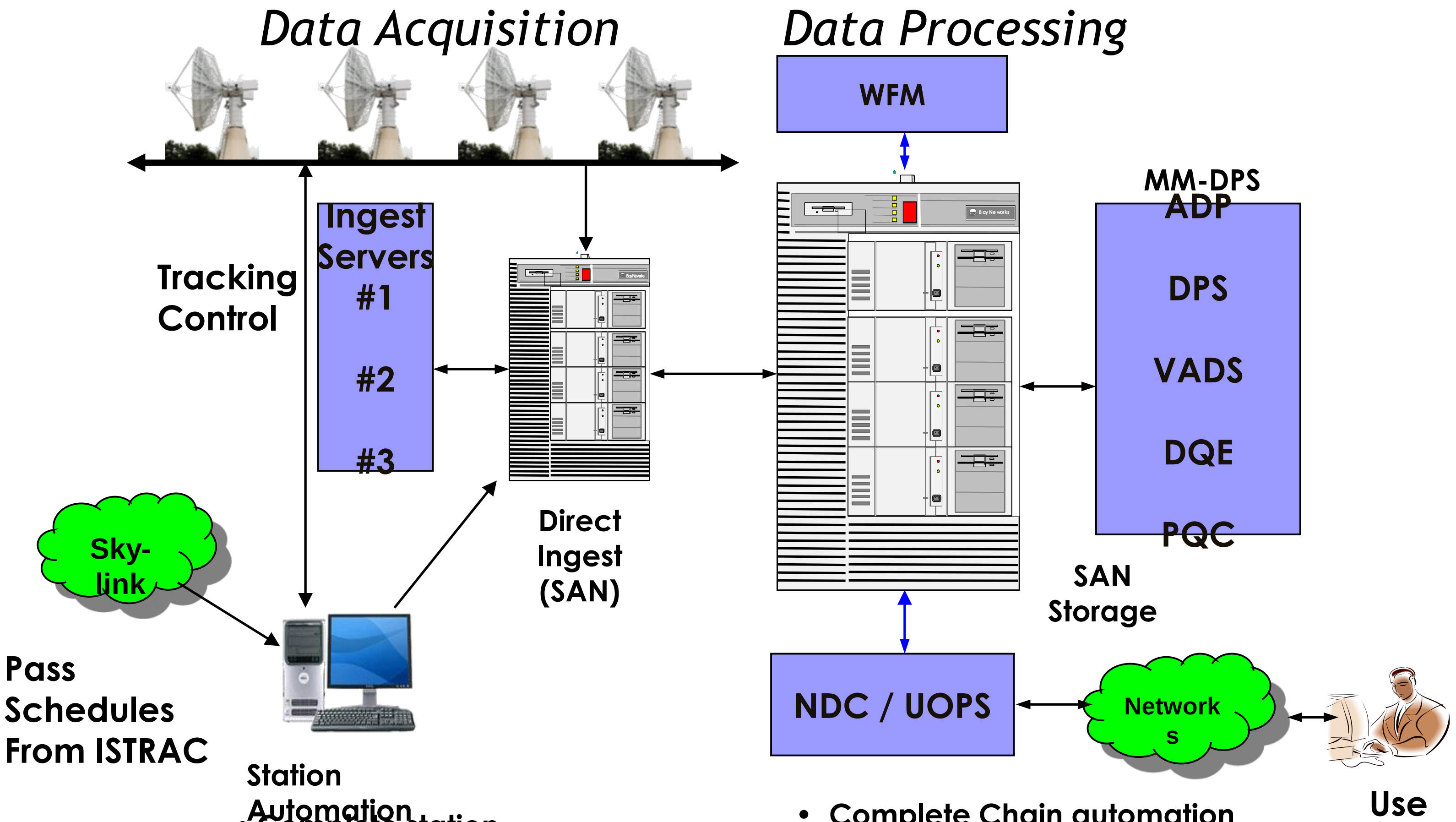
Unmanned Remote Sensing Platforms

Huge data repositories are generated in Geo-spatial communities

- NASA’s Earth Observation System (EOS) program has about 10.2 petabytes of Data;
- Similar sized collections can be expected from Other space agencies like ISRO;
- EOS contains mostly satellite data…not air photos, map or field data;
- In addition to this derived map products (mapped and modelled) are also part of the repositories;
- Data generated using “crowd sourcing’ and participatory approach – Difficult to size.

Present EO Data Repository at ISRO

Integrated Multi-mission Environment- IMGEOS Architecture



- Complete station automation form payload programming to ADP
- One machine for each Antenna & truly multi-mission.

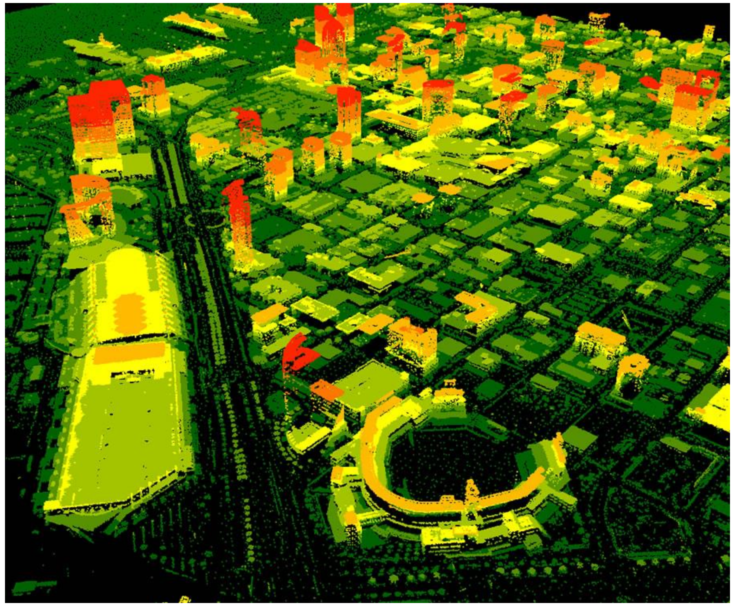
- Complete Chain automation
- Multi-mission DPS
- Data movement through Net
- High Throughput, Improved TAT

Challenges in Handling this Data

- Handling large volumes, Velocity and variety of data;
- Coordination and collaborations in data creation and collections;
- Data sharing and distribution;
- Data Calibration and corrections;
- Analysis, modelling and Simulation of data;
- …..

Big Data Challenges

- Storing unprecedented volumes of data ;
- Describing what we have (Metadata);
- Finding what user need;
- Identify un-wanted data or data which can be easily created on demand;
- Data and information quality strategies;
- Create successful tools and languages to describe and find data, so that reuse is actively encouraged
- Enable real time analysis of data;
- Historical data for change analysis studies;
- Data and Information Security.



Cluster Computing analysis challenges of massive data in Earth Science

Re-express (spatial) analysis algorithms so that they scale across HPC hardware AND Big Data:

- **Geometry**: Point / line / region / volume—algebra, selection, transformation, projection
- **Topology**: connectivity, route-finding, friction
- **(Spatial) statistics**: classification, interpolation
- **Point pattern analysis** / discovery: cluster detection

The challenge is to be SYSTEMATIC, not piecemeal

What's limiting the task?

- Memory?
  - 1TB on a single compute node now
  - 2-8TB on some equipment (e.g. SGI UV)
- CPU?
  - Tightly bound—needs a lot of inter-process communication
  - Embarrassingly parallel—can be perfectly decomposed
- Data?
  - Random? Linear? Blocky? (Degree of locality of reference)
  - Replication?
- Communications?
  - Data channels, infiniband, metadata
- Nothing?
  - Not everything needs to be parallelized…just the limiting segments

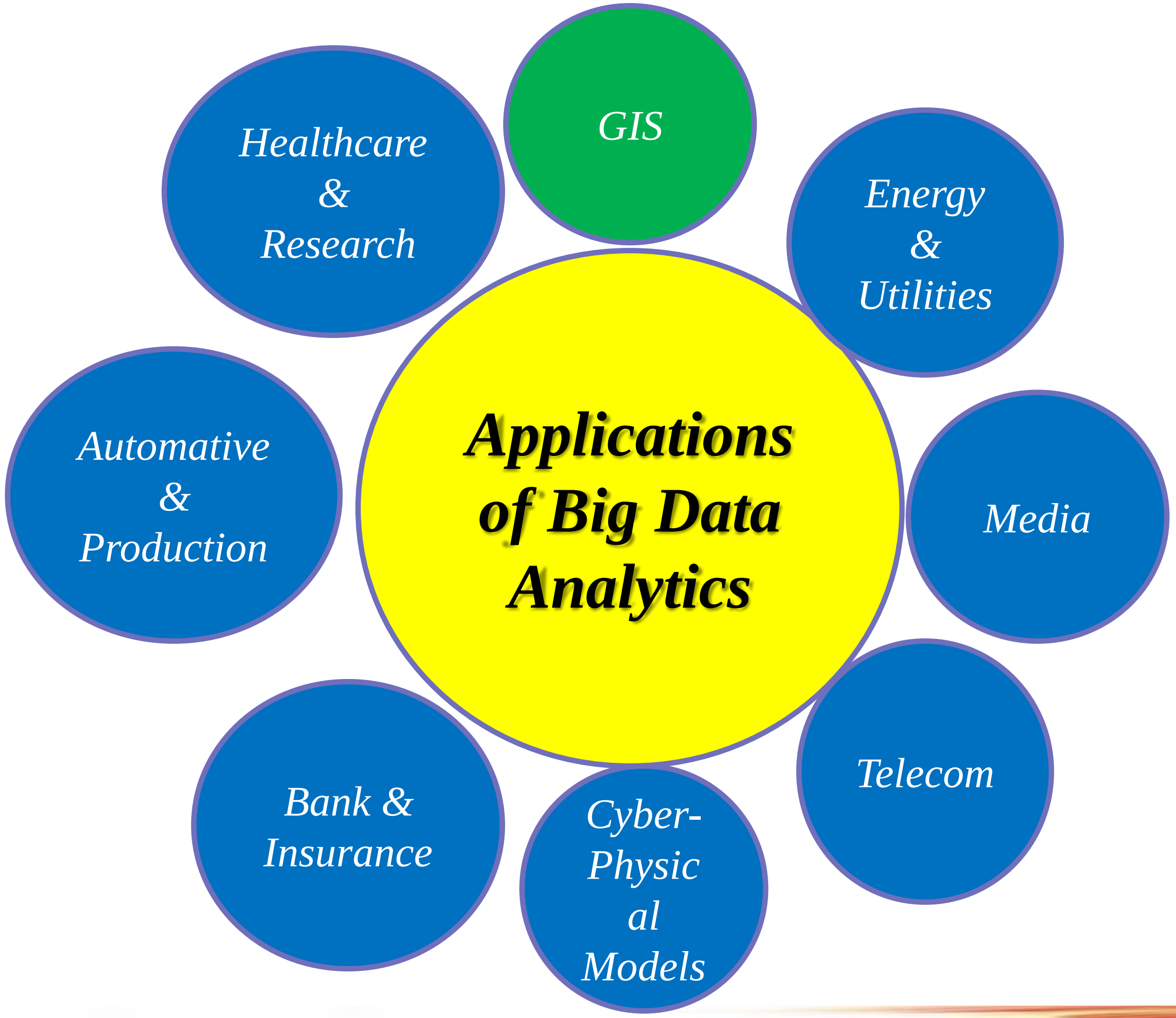
Big Data Analytics

- Big data analytics refers to the process of collecting, organizing and analyzing large sets of data ("big data") to discover patterns and other useful information.

- Big data analytics helps organizations to better understand the information contained within the data and will also help identify the data that is most important to the business and future business decisions. Big data analysts basically want the knowledge that comes from analyzing the data.

- Remember 3 Vs- Volume (amount of data), Velocity (speed of data in and out), and Variety (range of data types and sources).

GIS is one of the important user of Big Data Analytics



The Big Data approach to GIS allows analysis and decision making from huge datasets, by using algorithms, query processing and spatiotemporal data mining.

GIS Works with Big Data

- GIS tools search, sift and sieve data from multiple and disparate databases to organize it for better workflows and spatial analysis.
- GIS runs operations that aggregate terabytes and more of spatial information, run analysis, and visualize results as maps.
- All this occurs in real-time, with multiple data streaming into the existing GIS for better understanding of spatial trends and relationships.

Demonstration on Cloud Based Geo-computation System- Urban applications

Thank You