

Efficient Storage and Management of Environmental Information

Vijay Atluri

atluri@andromeda.rutgers.edu

Rutgers University

joint work with

Nabil Adam, Songmei Yu (Rutgers)

Yelena Yesha (UMBC)

Outline

- Environmental Data warehouse
- Traditional Data warehouse
- Challenges and Proposed Solutions
- System Architecture and preliminary results
- Summary and Ongoing Work

Environmental Data Warehouse

- Assimilation, cataloging and dissemination of a vast array of environmental data
- Various data sources and data types
 - various types of satellite images with different resolutions captured by different sensors
 - AVHRR: direct downloads from polar orbiting satellites(NOAA 12, NOAA 14 and NOAA 15)
 - LANDSAT and RADAR: obtained from NASA archives
 - Hyper-spectral images: from AISA (Airborne Imaging Spectrometer for Applications) sensor
 - Aerial ortho-photographs
 - Spot images

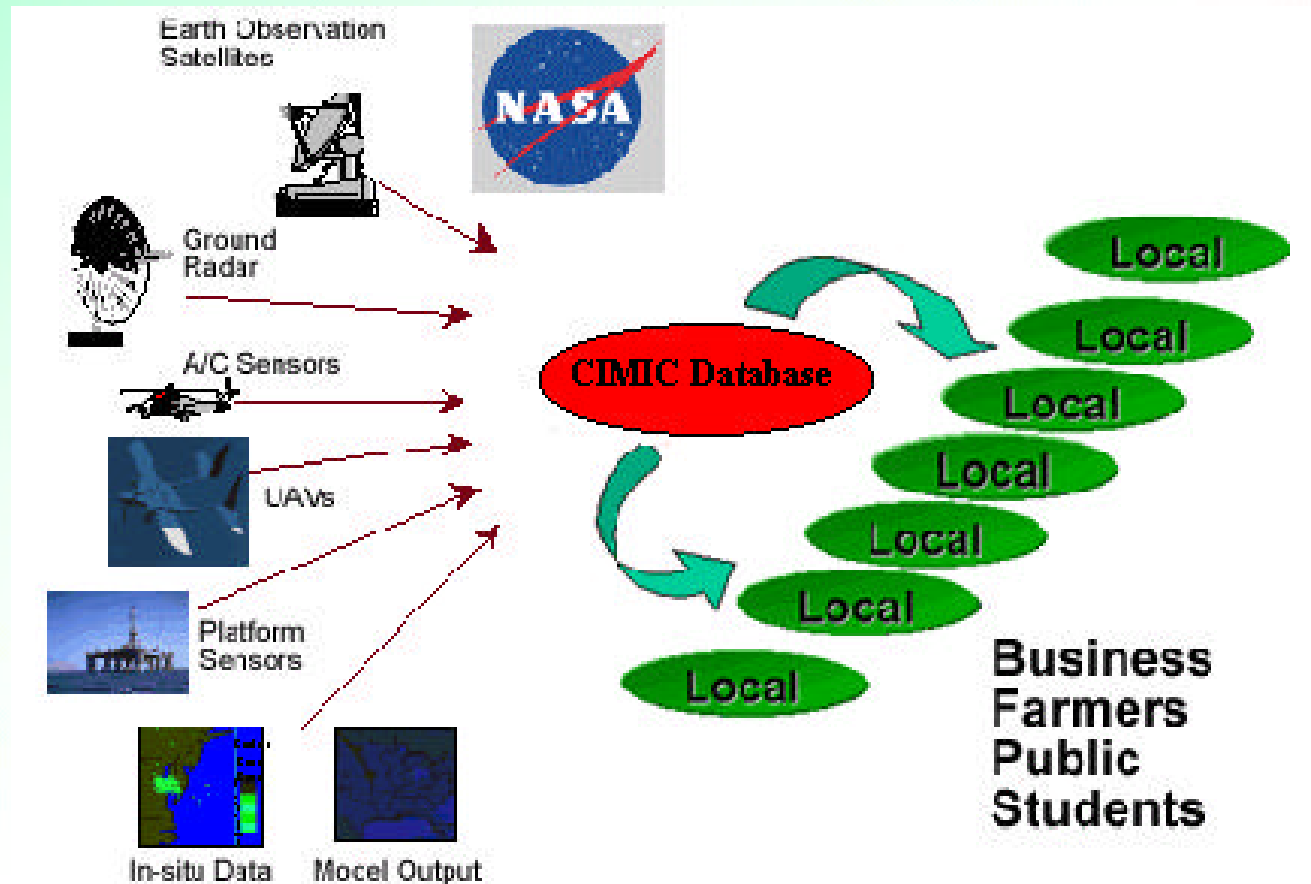
Environmental Data Warehouse

- Value-added products:
 - vegetation
 - water
 - temperature
 - true colors (composites)
- models of the topography and spatial attributes of the landscape
 - roads, rivers, parcels, schools, zip code areas, city streets and administrative boundaries
 - Maps, reports, data sets from government agencies
- census information that describes the socio-economic and health characteristics of the population
- real-time data from ground monitoring stations
-

At CIMIC

Images

- AVHRR
- Landsat
- Radar
- Orthophoto
- Etc.

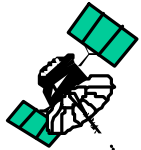


Users

- HMDC scientists and urban planners
- Rutgers scientists and students
- General public

Satellite Imagery

NOAA 12, 14, 15



Direct Readout
AVHRR

elements.dat
Downloaded daily
from US Navy

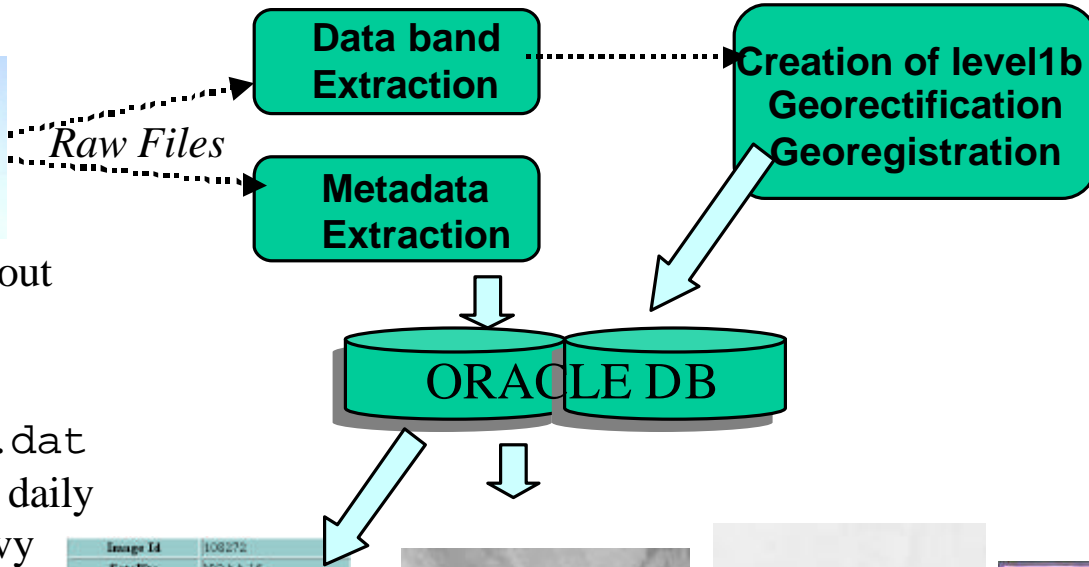
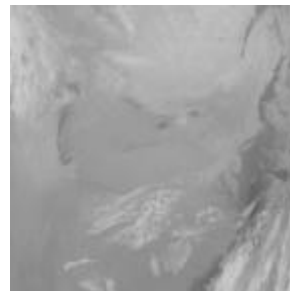


Image Id	108272
Satellite	NOAA 15
Datatype	BANDS
Image Date	26-FEB-01
Time	h
Original File	h27feb015
Thumbnail File	2001m27feb015.1.gif
Bandfile Root	2001m27feb015
Compression Type	gzip
Storage Status	ONLINE
Image Rows	2268
Image Columns	2048
Resolution Rows	1024
Resolution Columns	1024

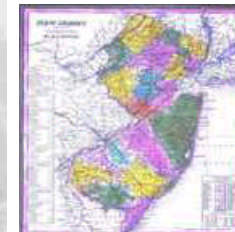
metadata



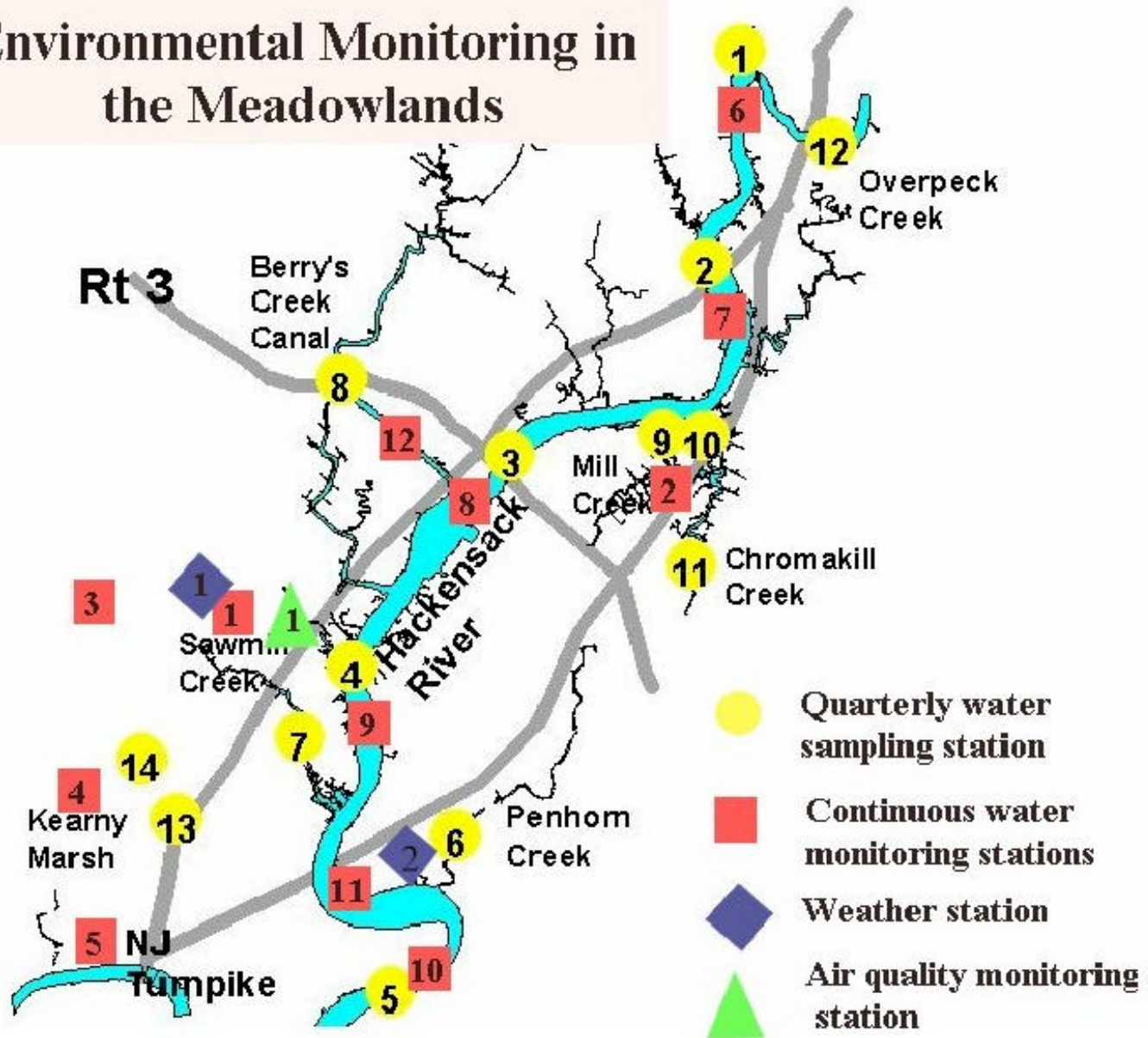
thumbnails



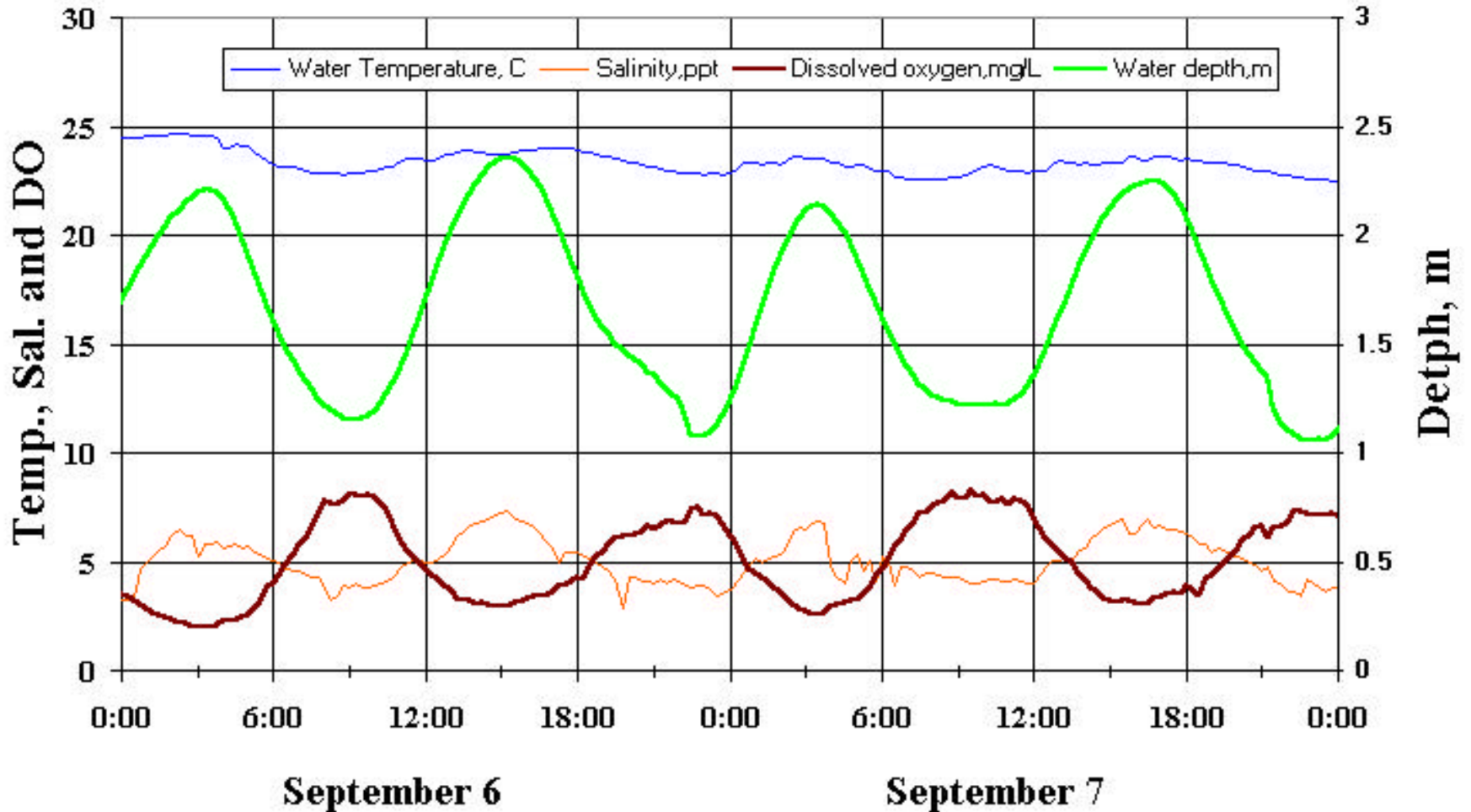
Region Of Interest



Environmental Monitoring in the Meadowlands



Water Quality in Hackensack River at the Meadowlands Marina



Diverse User Groups with specific data requirements

- **Policy Makers:**

- query various critical parameters such as ambient air and water quality and visualize the results in a graphical form
- gain help in the evaluation and formulation of environmental policies

- **Scientists:**

- make scientific observations such as the changes in vegetation pattern and its effect on temperature over the years, ..

- **Citizens:**

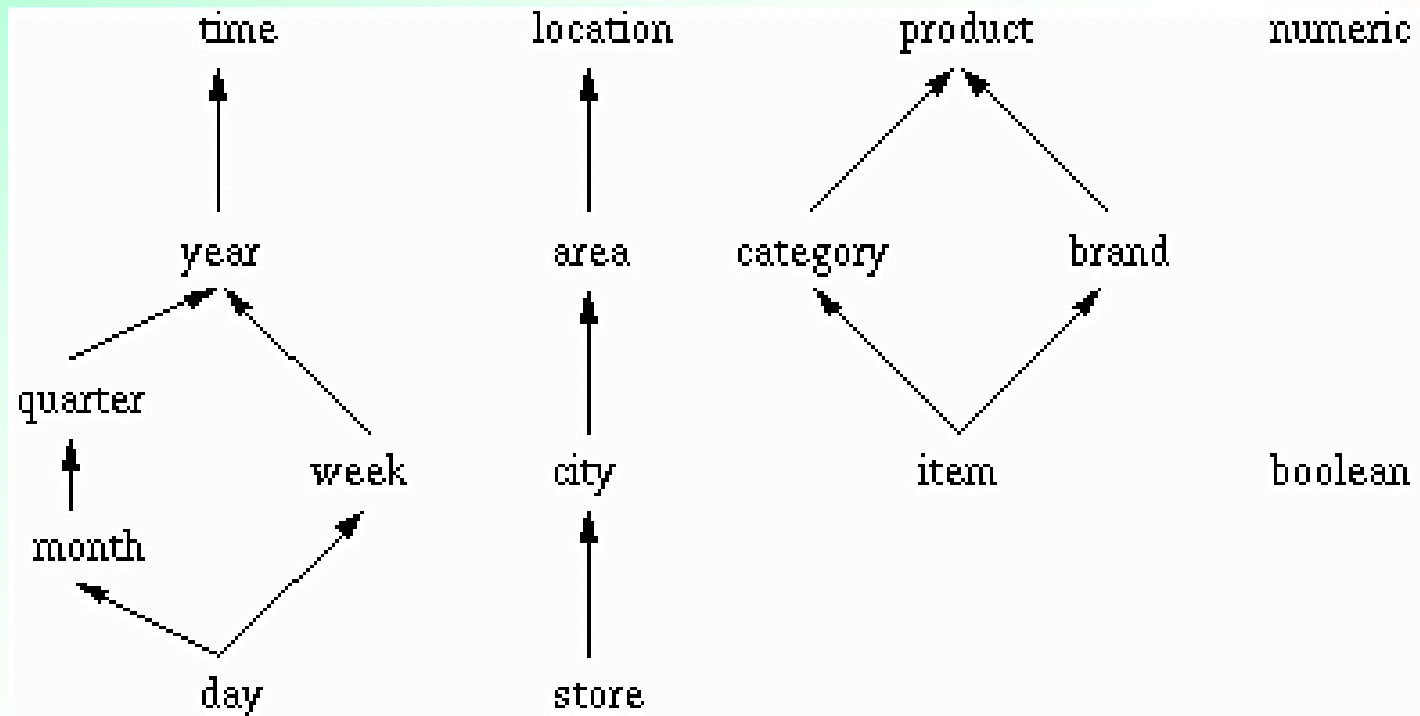
- learn information about their county, community, home on such issues as environment, health, and infrastructure

Can we adopt traditional Data warehouse models?

Traditional Warehouse Models

- **Star schema**
 - A single object (fact table) in the middle connected to a number of objects (dimension tables)
- **Snowflake schema**
 - A refinement of star schema where the dimensional hierarchy is represented explicitly by normalizing the dimension tables.
- **Fact constellation schema**
 - Multiple fact tables share dimension tables.
- **Are best suited for traditional data such as banking, insurance, retail sales**

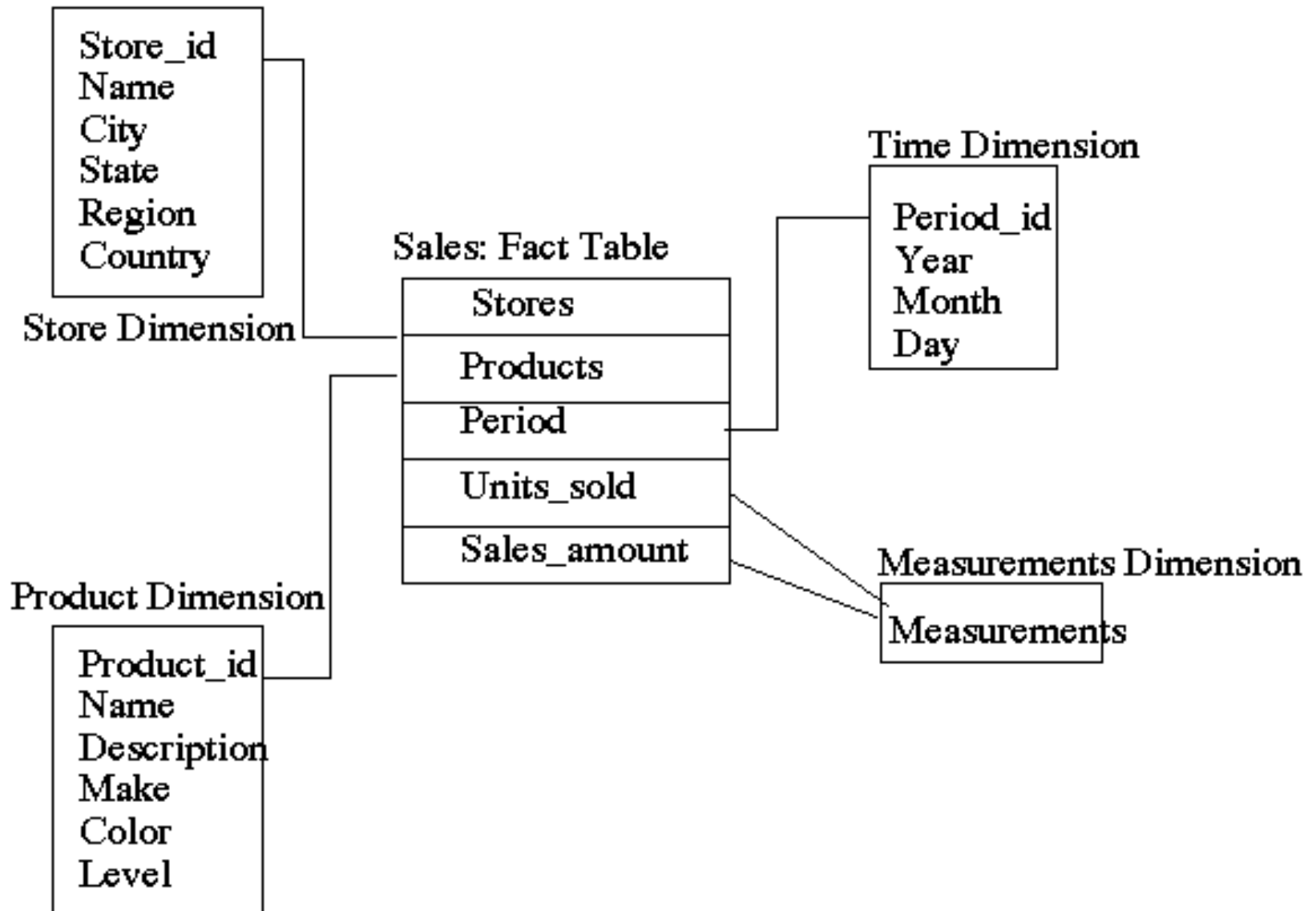
A Multidimensional fact table scheme



SALES[day:day, item:item, store:store]:numeric

PRICE-LIST[item:item, month:month]:numeric

The Star Schema



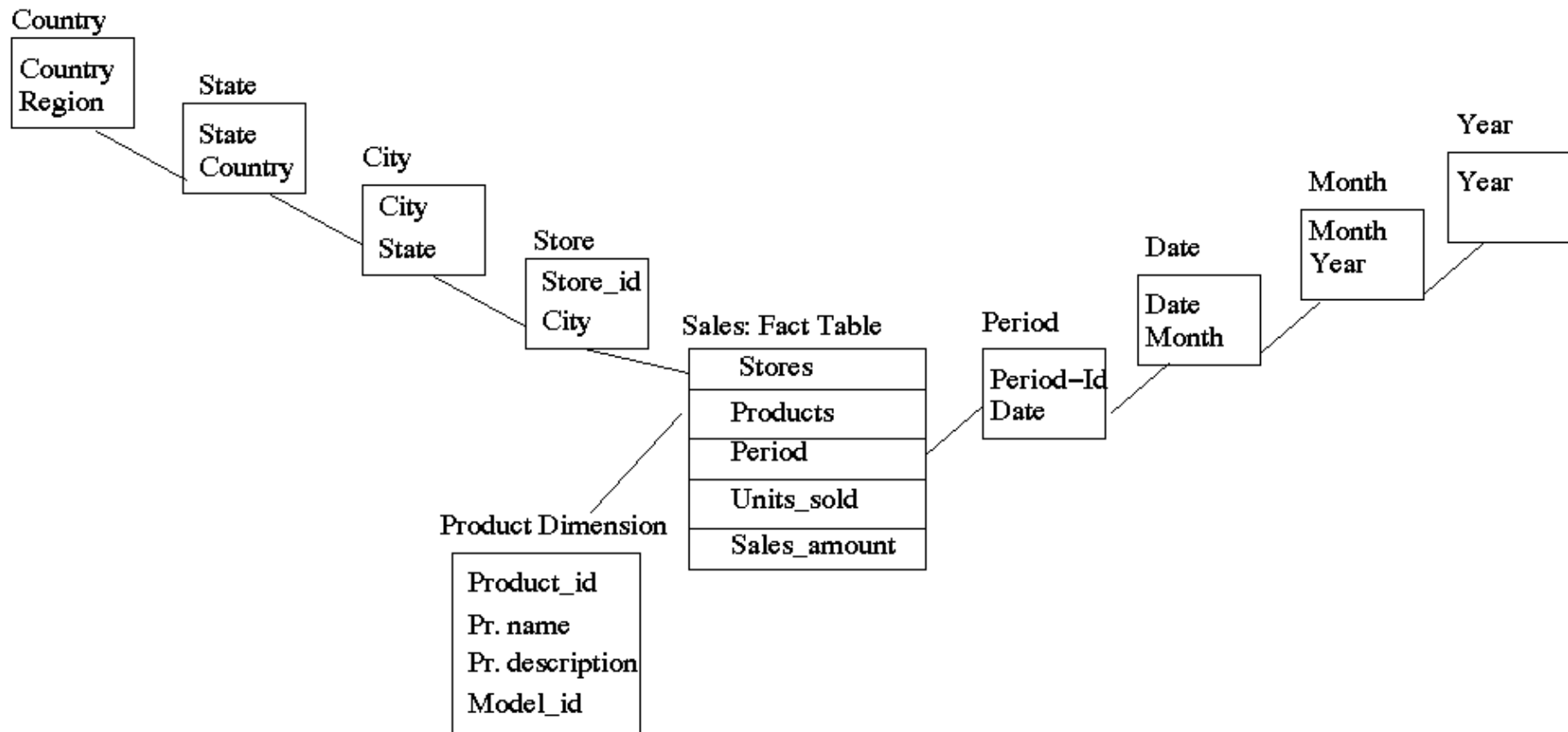
Sales

Store_id	Product_id	Period_id	Unit_sold	Sales_amount
str123	pro001	per02	10	900
str123	pro024	per02	15	1000
sta404	pro001	per02	300	45000
....
....

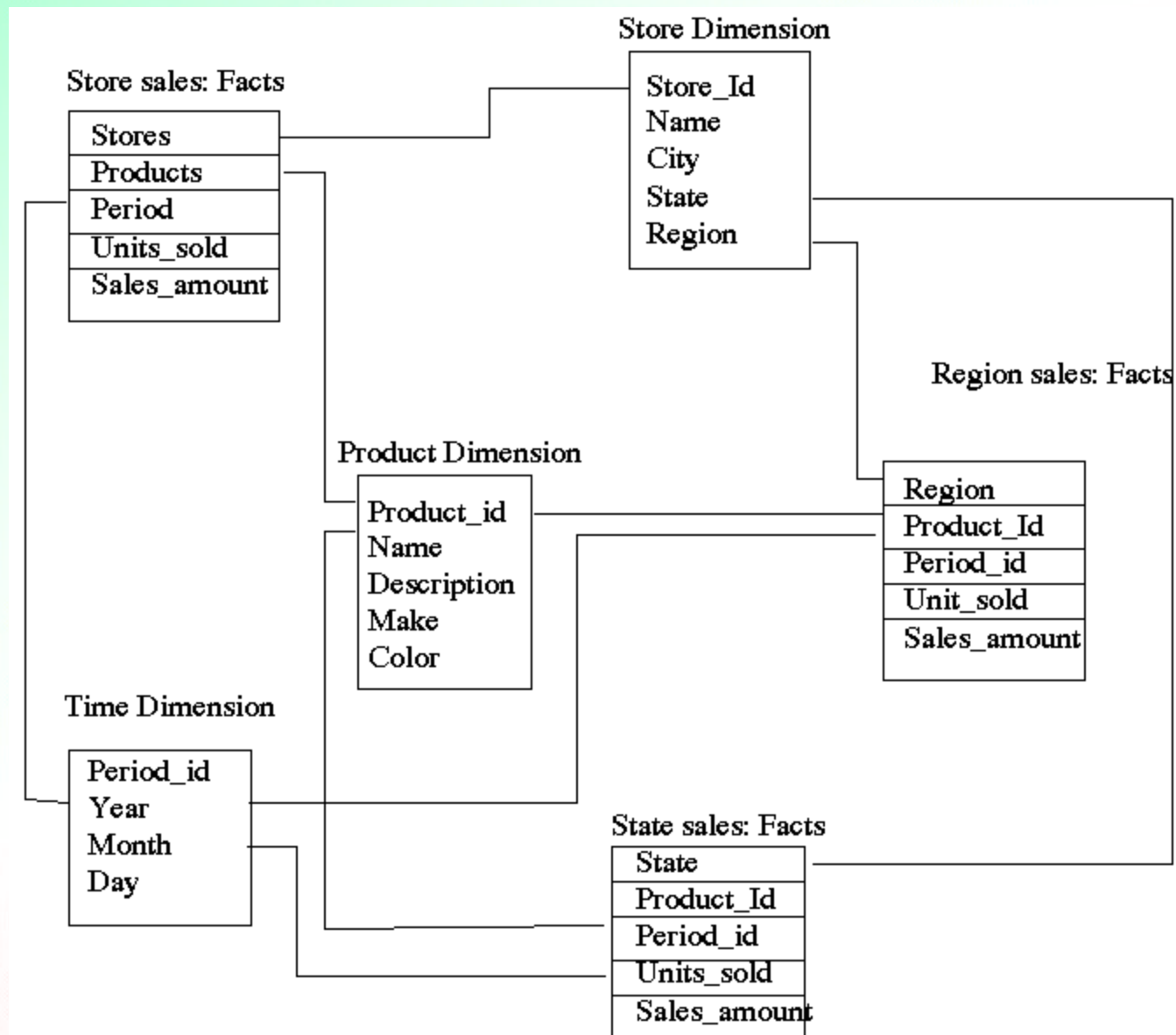
Store dimension

Store_id	Name	City	State	Region	Level
str123	John's store	Newark	New Jersey	East	1
str130	River store	Montclair	New Jersey	East	1
str234	Market Center	Piscataway	New Jersey	East	1
sta404	ALL	ALL	New Jersey	East	2
....

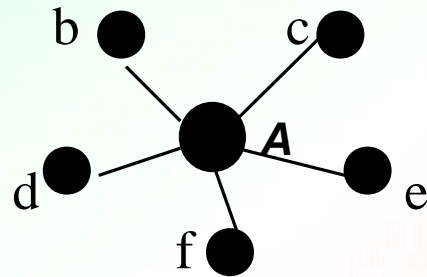
The Snowflake Schema



The Fact Constellation Schema



The Star Schema



Challenges

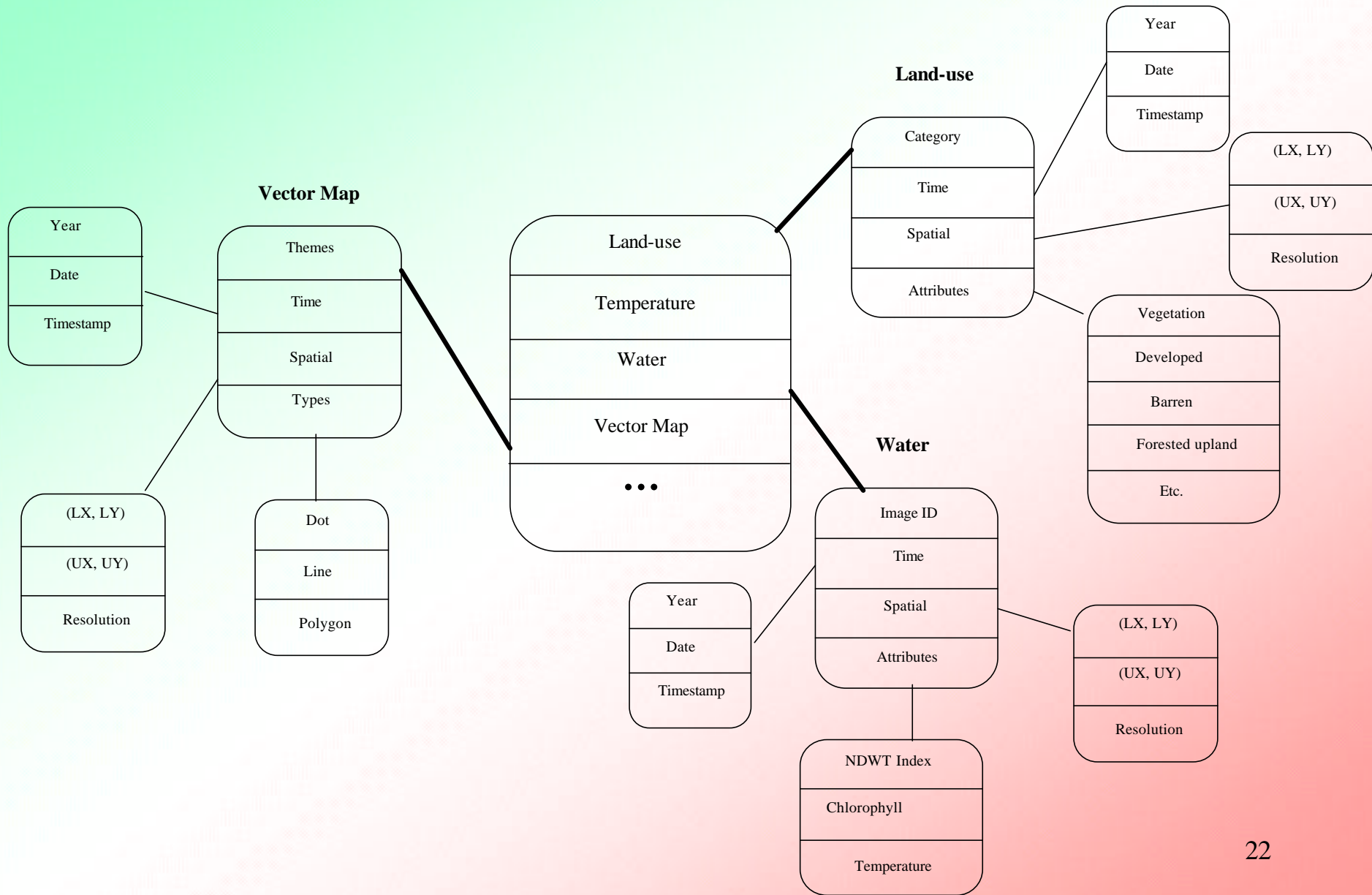
1. Nature of the Environmental data

- Each dimension in itself is multi-dimensional in nature
- raster images such as satellite downloads
 - used to generate various images of different types including land-use, water, temperature, NDVI
 - each of them have multiple dimensions
 - the geographic extent and coordinates
 - the time and date of its capture
 - resolution, ...
- similar with aerial photographs

Challenges

- regional maps represented as vector data
 - temporal and spatial
- streaming data collected from various sensors
 - temperature
 - air quality
 - atmospheric pressure
 - water quality: dissolved oxygen, mineral contents, salinity
 - geographic location (spatial dimension)
 - temporal dimension

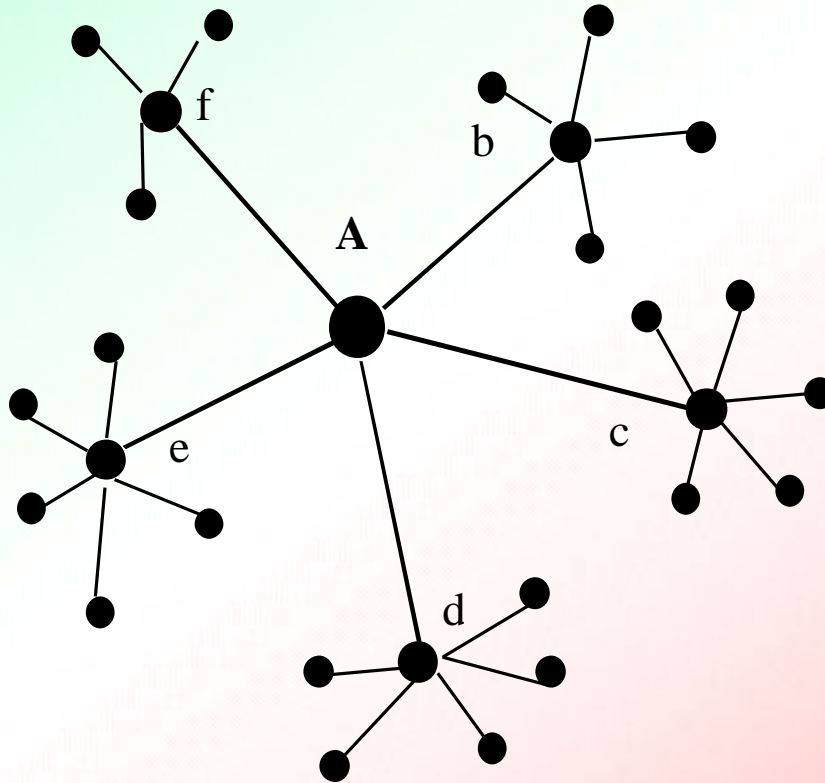
Nature of the Environmental data



Nature of the Environmental data

- each dimensional table is itself multi-dimensional by nature
- Traditional data warehouse models are not suitable for a spatial data warehouse
- Our Proposal: *cascaded star schema*

The Cascaded Star Schema



A is the fact table, and b, c, d, e and f are dimensions that are also multi-dimensional themselves

Challenges

2. Complex Nature of the queries

- (a) Retrieve the changes in the vegetation pattern over a certain region during last 10 years, and their effect on the regional maps over that time period
 - requires
 - layering of the images representing the vegetation patterns with those of the maps whose time intervals of validity overlap
 - traverse along this temporal dimension with the overlaid image
 - In the traditional data warehouse sense,
 - first construct two data cubes along the time dimensions for each of the vegetation images and maps
 - then fuse these two cubes into one

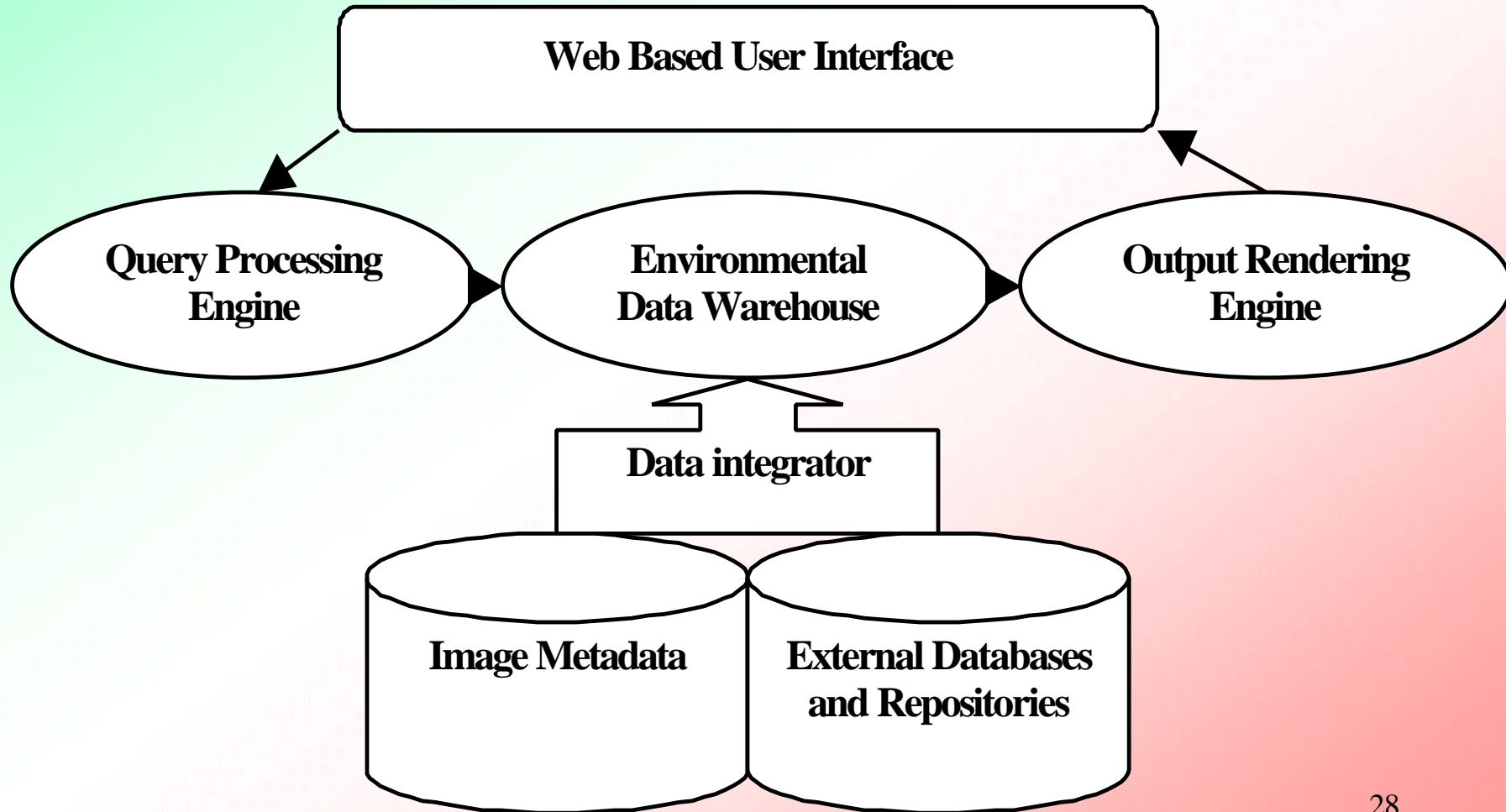
Challenges

- (b) observe the changes in the surface water and population due to the changes in the vegetation pattern
 - fusion of multiple cubes is required
- (c) simulate a fly-by over a region starting with a specific point and elevation, and traverse the region on a specific path with reducing elevation levels at a certain speed, and reaching a destination (a 3-dimensional trajectory)
 - requires
 - retrieving images that span adjacent regions that overlap the spatial trajectory, but with increasing resolution levels to simulate the effect of reduced elevation level
 - display them at a speed that matches the desired velocity of the fly-by.

How is it done now?

- Such requests are served by first identifying the sequence of images and manually composing them
- This is time consuming and labor intensive
- cannot handle ad-hoc queries

System Architecture



Flyby Video



Demo

- <http://cimic.rutgers.edu/~songmei/dw.html>

Welcome to CIMIC RAC Spatial Data Warehouse

Specify A Satellite/Instrument Data Source:

Specify A Data Type:

Specify A Starting and Ending Date:

View all the images on one page

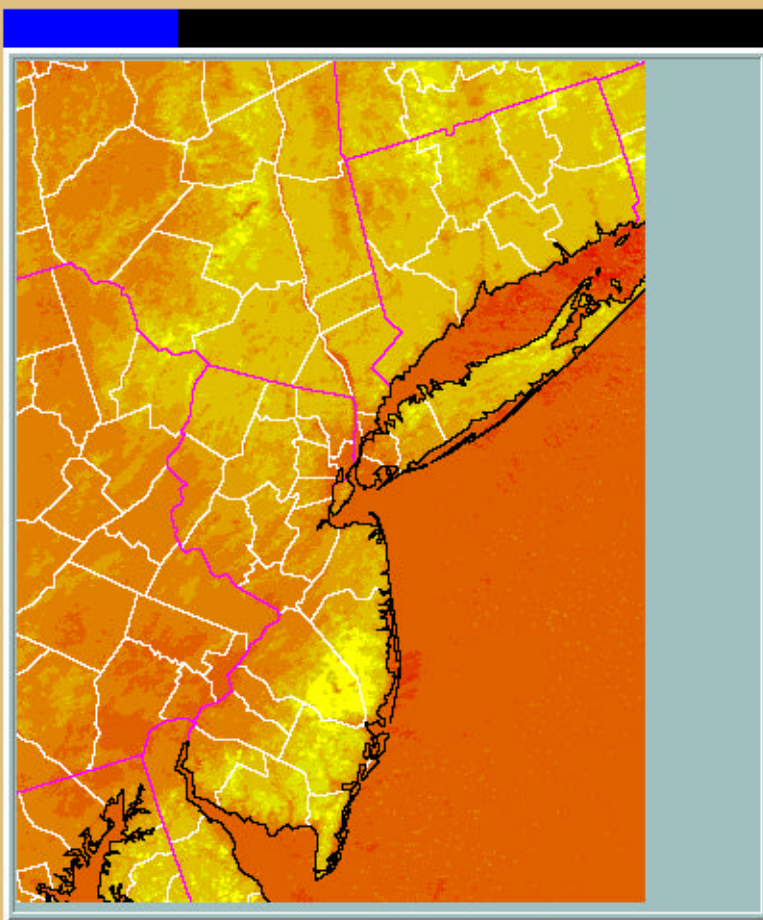
View all the images one by one

Starting Date (Year, Month and Day):

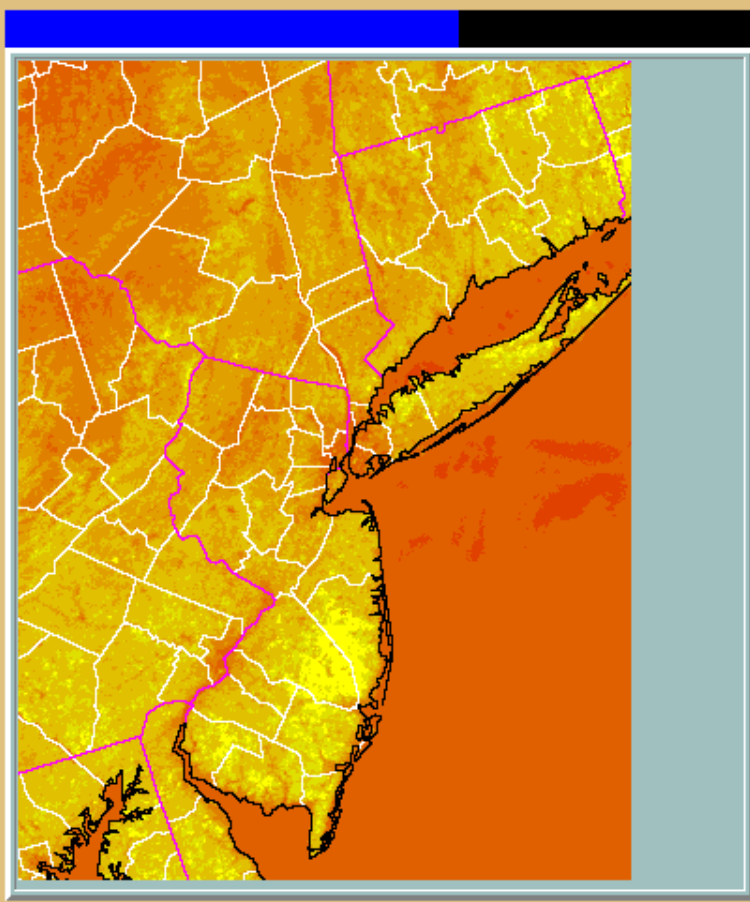
Ending Date (Year, Month and Day):

Specify One Particular Date:

Digital Meadowlands - Search Results



Digital Meadowlands - Search Results



Summary

- Recognized that traditional data warehouse models are not suitable for managing data in complex domains such as environmental and scientific
- proposed a cascaded star model

Ongoing Work

- Formulating the necessary primitives for the specification and execution of queries
- Extending the OLAP operations for the cascaded star
 - roll-up: aggregating on a specific dimension, I.e., summarize data
 - drill-down: from higher level summary to lower level detailed
 - slicing: projecting data along a subset of dimensions with an equality selection of other dimensions
 - dicing: similar to slicing except that instead of equality selection of other dimensions, a range selection is used
 - pivoting: reorient the multidimensional cube
 - zoom-in, zoom-out, aggregation of views using the above OLAP operations

Ongoing Work

- slicing
- dicing
- Pivot