Big Data Spatial Analytics – An Introduction

Marwa Mabrouk
Mansour Raad
Agenda

- What is Big Data?
- What is Hadoop?
- How does Spatial integrate with Big Data and Hadoop?
- How do I get started?
New Challenges for Organizations

- Better Decision making
- Intelligence
- Insight/ Foresight

- Social Data Analysis
- Log files analysis
- Fraud Detection
Collect Data!!!
What is Big Data???

Is it a large network of a million points???

Is it Satellite & Aerial Imagery???

Is it a simple Geodatabase with Billions of records???

Is it this technology called hadoop?
Big Data, a new Data Type

• Volume
  - Ever growing data, petabytes / zetabytes
  - Use 350 billion annual meter readings to better predict power consumption

• Velocity
  - Time-sensitive, two mins can be too late
  - Turn 12 TBs of tweets into updates about people’s conditions during a storm for emergency response

• Variety
  - Any type of data, text / audio/ video / click streams / other
  - Exploit documents, images, voice recordings, videos in customer experience analysis programs to improve customer satisfaction
What is hadoop?
- Parallel Framework
- Executes Map Reduce Task
- Reads HDFS
- Java/Python/Awk
$> \text{cat input | Map | sort | Reduce} > \text{out}$
Apache Hive
“SQL”

MapReduce Job
hive> select * from cities
where country= ‘lebanon’ ;
What About Spatial?
Big Data – A new data type for Geospatial
Geospatial in Big Data

• Big Data with location in Enterprise Big Data warehouse

• Run spatial queries on data where it resides

• Pull Results in ArcGIS: Visualize results as a map; Include in a report; Publish in a web or mobile app
GIS Tools for Hadoop


- Support Running
  geometry-based spatial queries inside Hadoop

- Open Source
  - Apache 2.0 license

- Two types of users
  - Developers
  - ArcGIS users
GIS tools for Hadoop libraries

- **ArcGIS**
  - Geoprocessing Tools
  - Connect From ArcGIS to Hadoop using GP

GIS Tools for Hadoop
(Use samples with Hadoop & Hive to get started)

- **Hive**
  - Spatial Framework
- **Hadoop**
  - Esri Geometry API

Run Hive Queries with spatial operators

Build Map/Reduce Spatial Apps in Java
GIS Tools for Hadoop

Github repos

http://esri.github.com/gis-tools-for-hadoop/
Spatial Storage

- CSV, TSV Lat, Lon
- Esri JSON format
  - `{geometry:{x:-123,y:45},attributes:{}}`
- Custom
User Defined Functions

- select `tolower("ESRI")`;
- select * from mytable where `cos(rad) < 0.1`;
Spatial UDF !
select * from cities where near (x, y, -84.2, 39.4);
select * from cities
where contains (x,y,’#mypolys’);
How to get started?

Technology

• Download GIS Tools for Hadoop
  - Sample (jars + sample data/ query)

• Hadoop cluster
  - Amazon Web Services EMR
  - Cloudera Manager and VM, CDH

• Run sample
How to get started?

Data and Query

• Use Case, problem definition

• Dataset
  - Formats
  - Conversion
  - Location
  - Loading

• Query
  - Hive Script
AWS Elastic Map Reduce
GP Tools
Cloudera

Map Reduce
Big Data

The Data warehouse
Unlocking Data-Driven White Spaces through Collaboration
The Power of Contextual Data

Contact jolimpio@esri.com to discuss starting a joint Esri – IBM Research project at ADLAB

Esri + IBM Partnership
Big Data Partners

IBM

Cloudera

Ask Bigger Questions

Amazon Web Services

THINKBIG

TERRACOTTA

Go Big. Go Fast.
Thank you…

Marwa Mabrouk - mmabrouk@esri.com
Mansour Raad - mraad@esri.com, http://thunderheadxpler.blogspot.com/

Please fill out the session evaluation

First Offering ID: 2333

Online – www.esri.com/ucsessionssurveys
Paper – pick up and put in drop box