

2013 Esri International User Conference

July 8–12, 2013 | San Diego, California



Technical Workshop

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!!!





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



**Is it Satellite & Aerial
Imagery???**

What is Big Data???



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



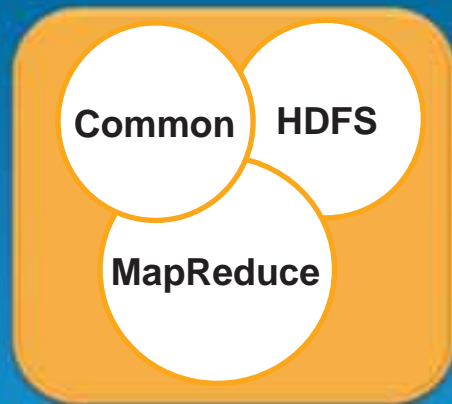
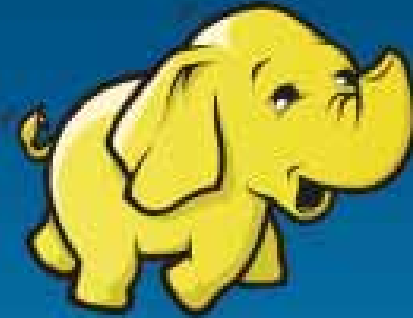
**Is it this technology
called hadoop?**

Big Data, a new Data Type

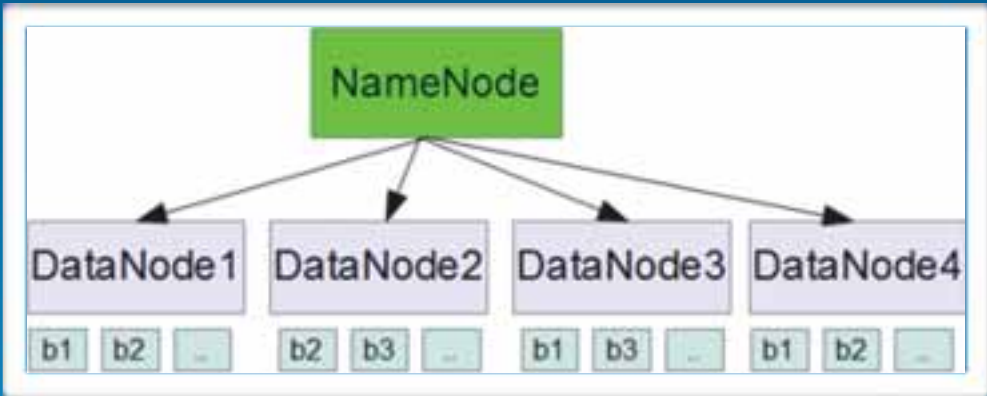


- **Volume**
 - Ever growing data, petabytes / zettabytes
 - Use 350 billion annual meter readings to better predict power consumption
- **Velocity**
 - Time-sensitive, two mins can be too late
 - Turn 12 TBs of tweets into updates about people's conditions during a storm for emergency response
- **Variety**
 - 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

```
$> cat input | Map | sort | Reduce > out
```

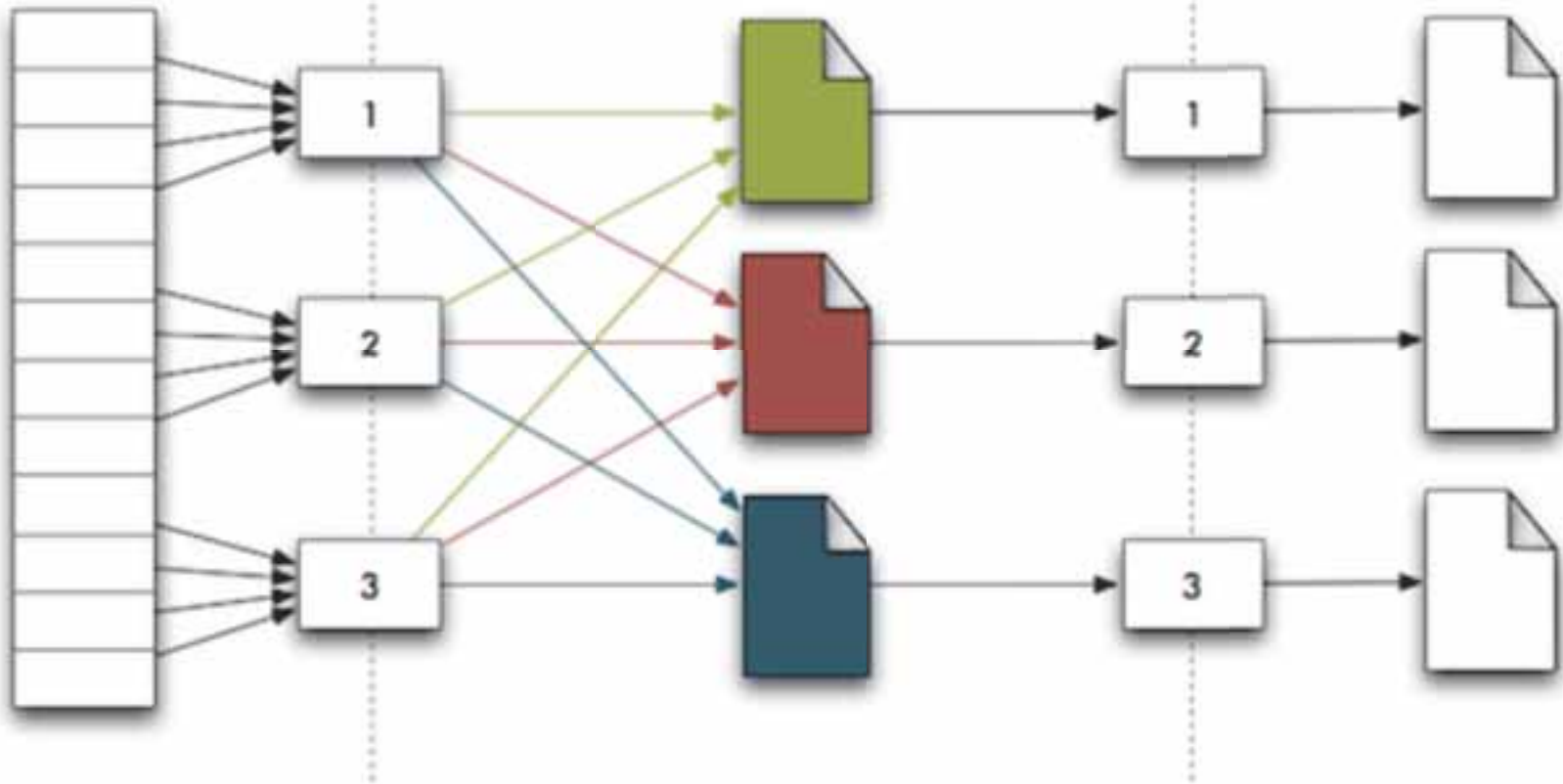
Map

Reduce

Input Files

Intermediary Files

Output Files



Apache Hive



“SQL”



MapReduce Job

```
hive> select * from cities  
where country= 'lebanon' ;
```



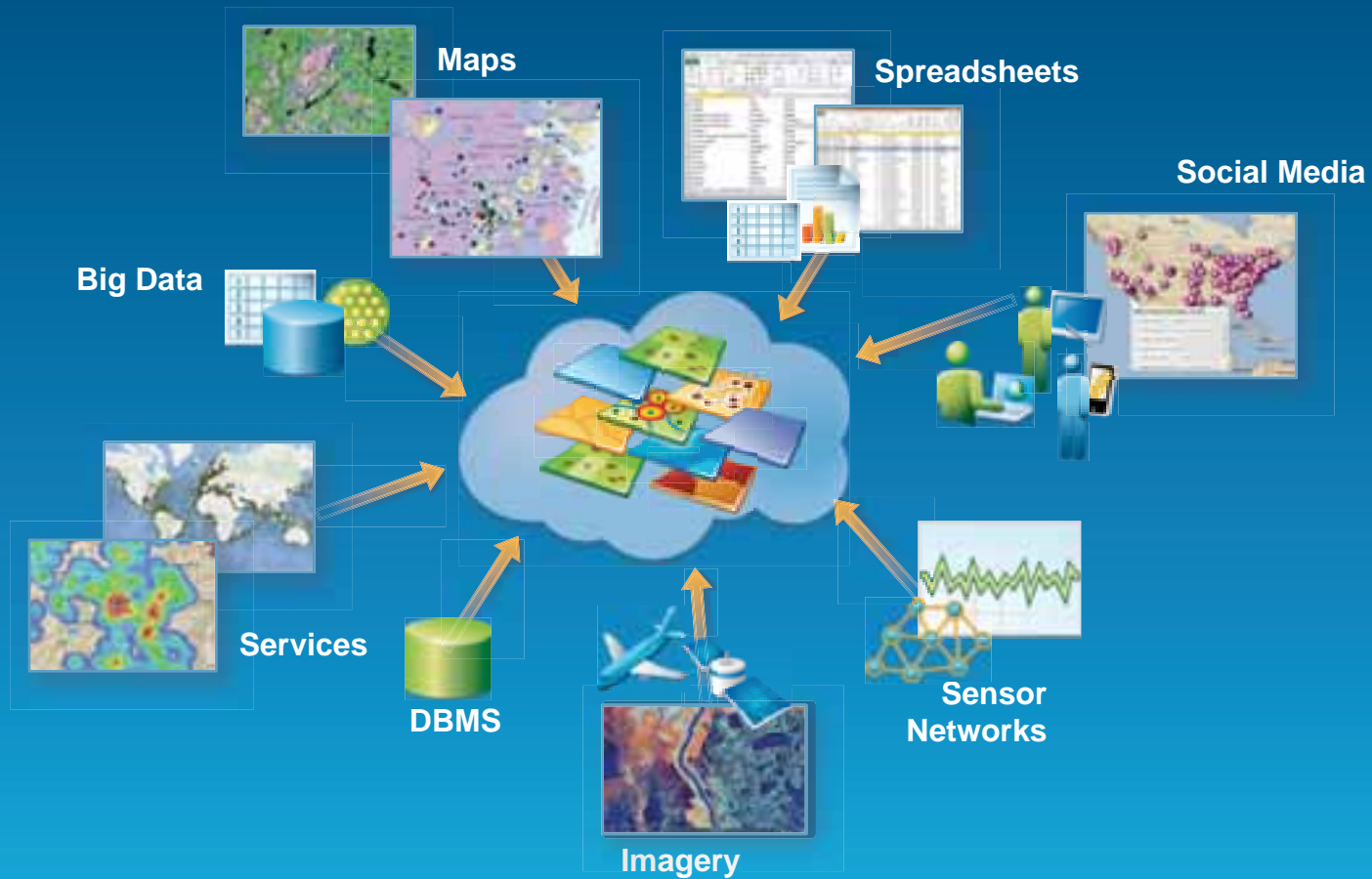
MapReduce



HDFS
CSV
TSV
JSON
BINARY

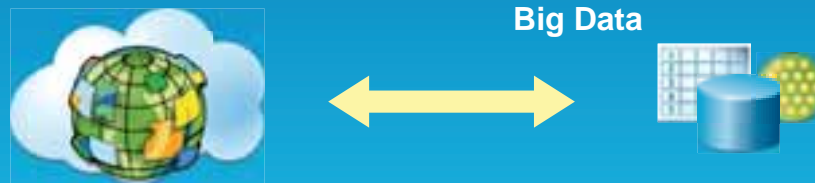
What About Spatial ?

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

- <http://esri.github.com/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

The screenshot shows the GitHub repository page for 'GIS Tools for Hadoop'. The header features a yellow elephant icon and a globe, with the text 'GIS Tools for Hadoop' and 'Big Data Spatial Analytics For the Hadoop Framework'. A GitHub logo is also present. The main content area includes a description of the project, a 'README to Go' section, and a list of key features and users.

GIS Tools for Hadoop
Big Data Spatial Analytics For the Hadoop Framework

Looking at data without location, most of the time seems like looking at just part of a story. Including location and geography in analysis reveals patterns and associations that otherwise go missed. As Big Data emerges as a new frontier for analysts, including location in Big Data is becoming significantly important.

Data that includes location, and that is enhanced with geographic information in a structured form, is often referred to as Spatial Data. Doing Analytics on Spatial data requires an understanding of geometry and operations that can be performed on it. Creating interfaces to include spatial data and spatial analysis is the goal of this Open Source effort.

GIS Tools for Hadoop is an open source toolkit intended for Big Spatial Data Analytics. The toolkit provides different options:

- **Earl Geometry API for Java:** A general geometry library can be used to extend Hadoop core with vector geometry types and operations, and enable developers to build MapReduce applications for spatial data.
- **Spatial Framework for Hadoop:** Extends Hive and is based on the Earl Geometry API, to enable Hive Query Language users to leverage a set of analytical functions and geometry types. In addition to some utilities for JSON used in Avro/GIS.
- **Geospatial Tools for Hadoop:** Contains a set of tools to use Avro/GIS Geospatial tools, based on the Earl Geometry API and Spatial Framework for Hadoop. Developers can download the source code of the tools and customize it, they can also create new tools and contribute it to the open source project. Through these tools Avro/GIS users can move their spatial data and execute a pre-defined workflow inside Hadoop.

Powered by Esri
The code base provided by Esri. Right-click to download source to desktop.

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)



Spatial
Framework

Run Hive Queries
with spatial
operators



Esri Geometry
API

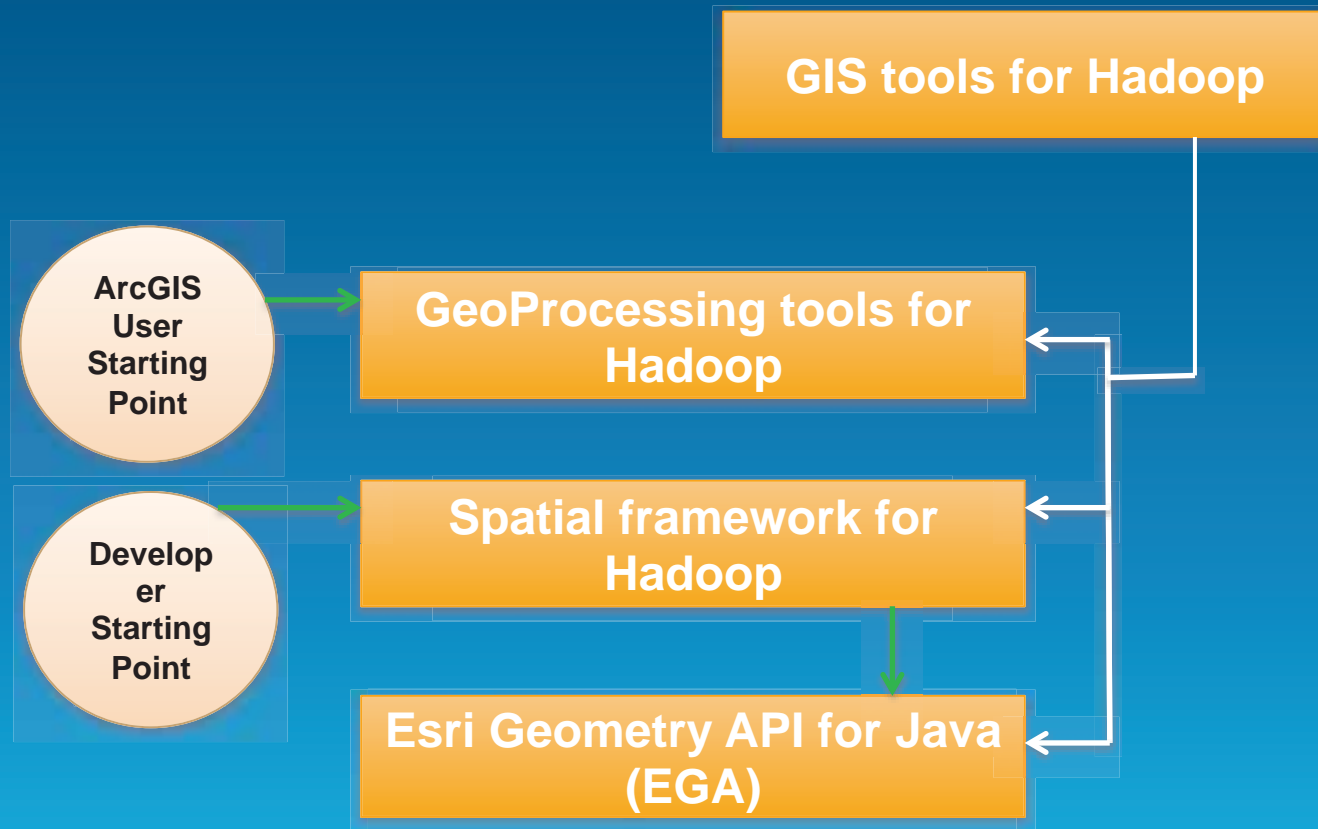
Build Map/
Reduce Spatial
Apps in Java



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?

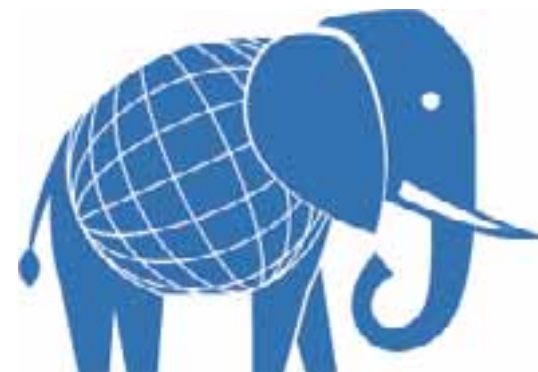
Data and Query



- Use Case, problem definition
- Dataset
 - Formats
 - Conversion
 - Location
 - Loading
- Query
 - Hive Script

Cloud

AWS Elastic Map Reduce GP Tools



On Premise



Cloudera

Map Reduce

cloudera[®]
Ask Bigger Questions



Big Data

The Data warehouse

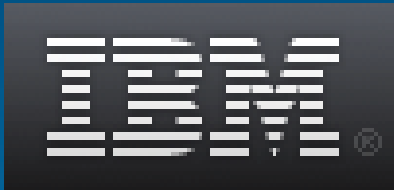
Unlocking Data-Driven White Spaces through Collaboration The Power of Contextual Data



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

Esri + IBM | Partnership

Big Data Partners





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/ucsessionsurveys

Paper – pick up and put in drop box



Understanding our world.