

This enab  
that use b

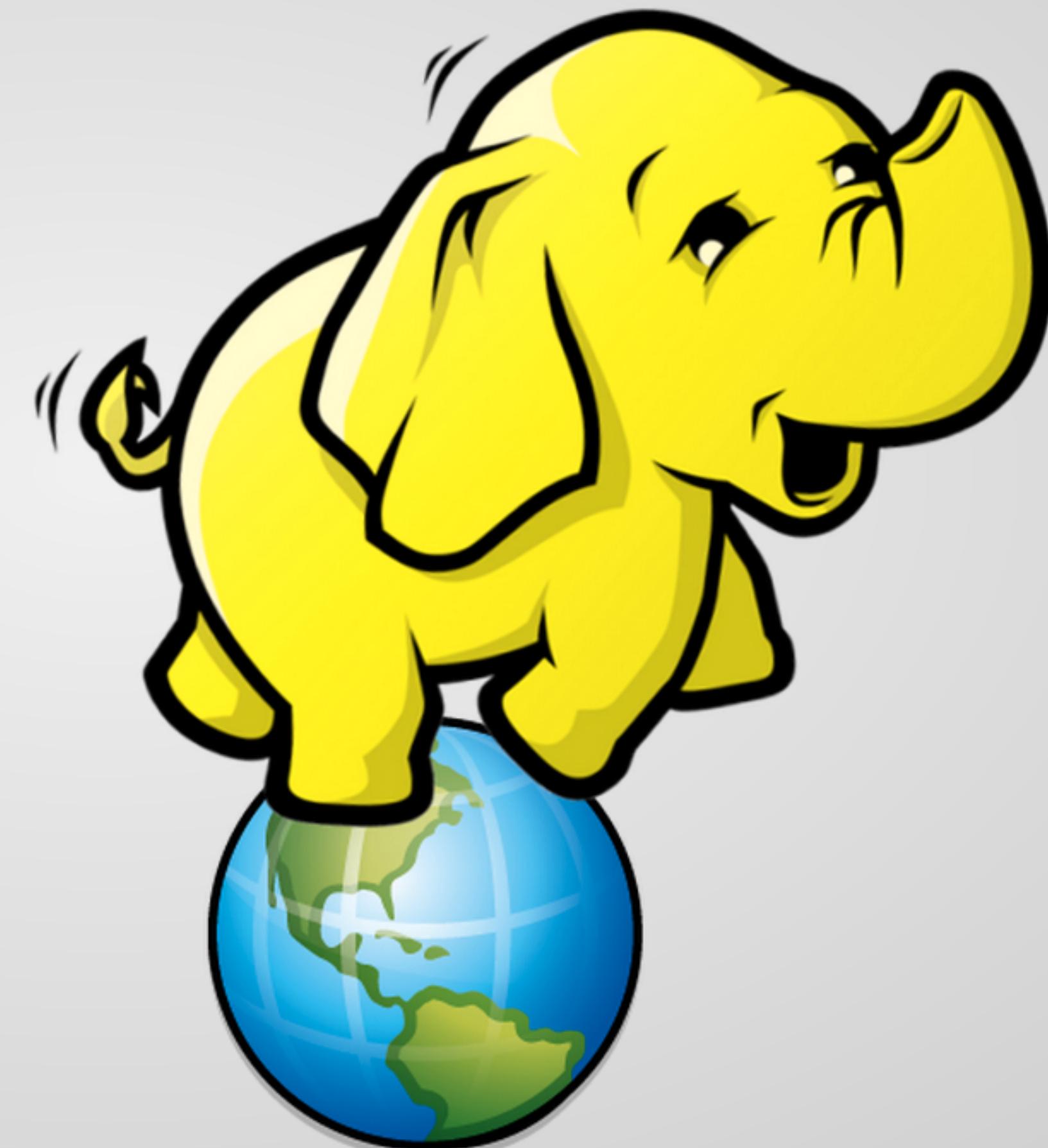
Why

You  
pro

...and your  
from within

Who i

# Big Data: Using ArcGIS with Apache Hadoop



David Kaiser @ddkaiser  
Michael Park

Esri DevSummit 2013

Session Offering ID: [301](#)

GIS To

Hadoop us

with limite

Esri has re  
spatial-dat

This enab  
that use bo

Why

# Follow along with this presentation

<http://esri.github.com/gis-tools-for-hadoop/devsummit2013>

Use WiFi network: **Esri2013**

Big Data: Using ArcGIS with Apache Hadoop



ntation

## GIS Tools for Hadoop

Hadoop users often have data with spatial value, but with limited options for spatially analyzing this data

mit2013

Esri has released an open-source framework to enable spatial-data processing in your Hadoop applications

This enables you, as a developer, to build analytical tools that use both Hadoop and ArcGIS.

e Hadoop

# Why is this important?



This enables you, as a developer, to build analytical tools  
that use both Hadoop and ArcGIS.

the Hadoop



## Why is this important?

Your Hadoop applications can  
provide spatial analysis

...and your users can leverage your Hadoop applications  
from within the ArcGIS Geoprocessing environment

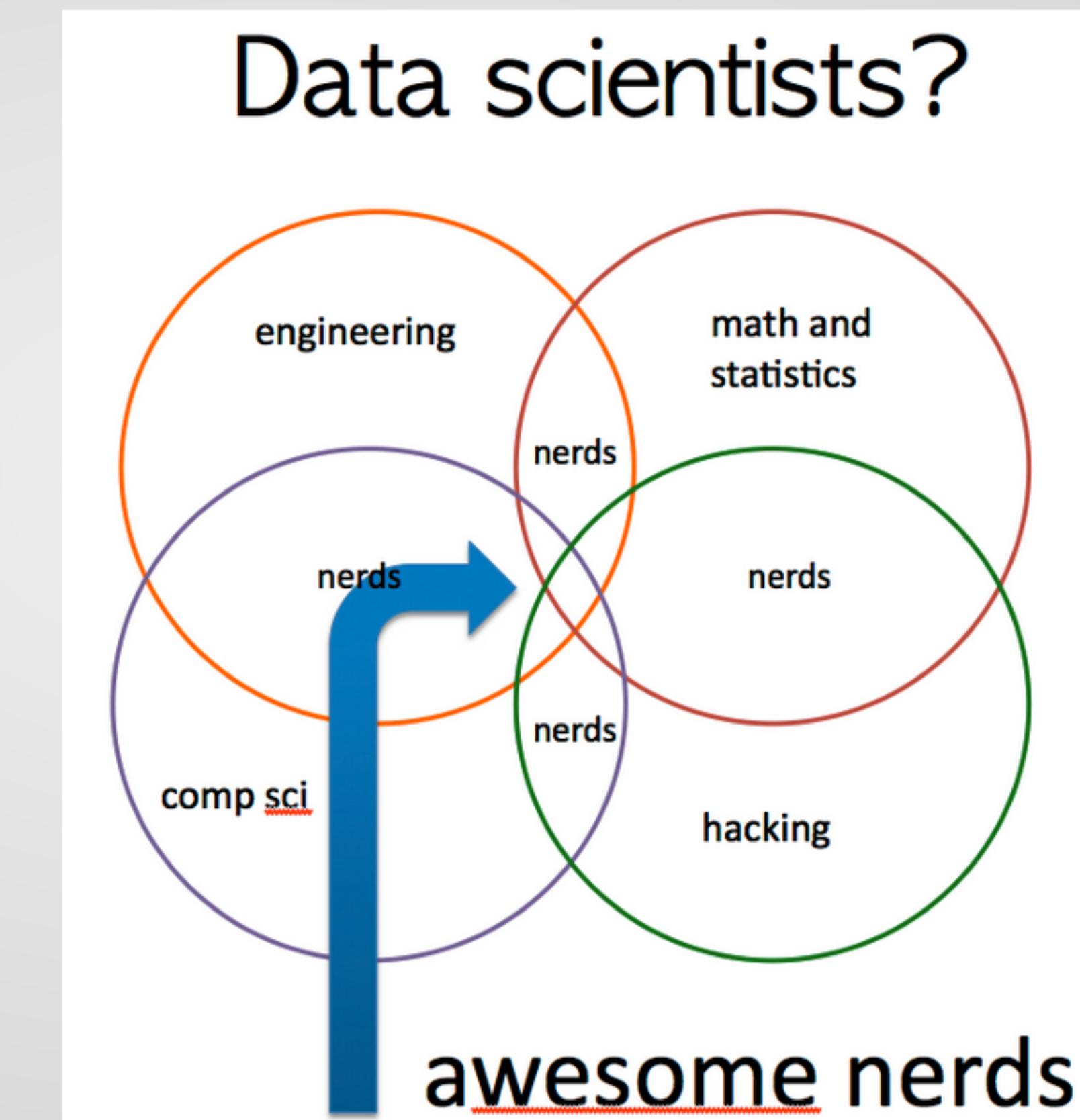
## Who is a Data Scientist?

# Who is a Data Scientist?



## Finding Your Data Scientist

# Finding Your Data Scientist



[www.hilarymason.com](http://www.hilarymason.com)

Geo

have data with spatial value, but  
ons for spatially analyzing this data

an open-source framework to enable  
essing in your Hadoop applications

as a developer, to build analytical tools  
oop and ArcGIS.

his important?

Hadoop applications can  
spatial analysis

can leverage your Hadoop applications  
cGIS Geoprocessing environment

Data Scientist?

HELLO  
my name is

Data Scientist

our Data Scientist

Data scientists?



www.hilarymason.com

## Geometry API for Java

Simple API Functions for Java

com.esri.core.geometry.\*

Relationship Analysis

SESSION OFFERING ID: 501



E-mail:

David Kaiser <dkaiser@esri.com> @ddkaiser  
Michael Park <mpark@esri.com>

- Send a Github 'pull request' so we can pull the tool back into our project

Let us know how you are doing

- Add your content to the Wikis on Github
- Troubles? Open new Issues in Github

Geoprocessing Tools for Hadoop

<http://github.com/Esri/geoprocessing-tools-for-hadoop>

To build the source:

ant

Download the GIS Tools Project

Clone or Fork the project from Github

<http://github.com/Esri/gis-tools-for-hadoop>

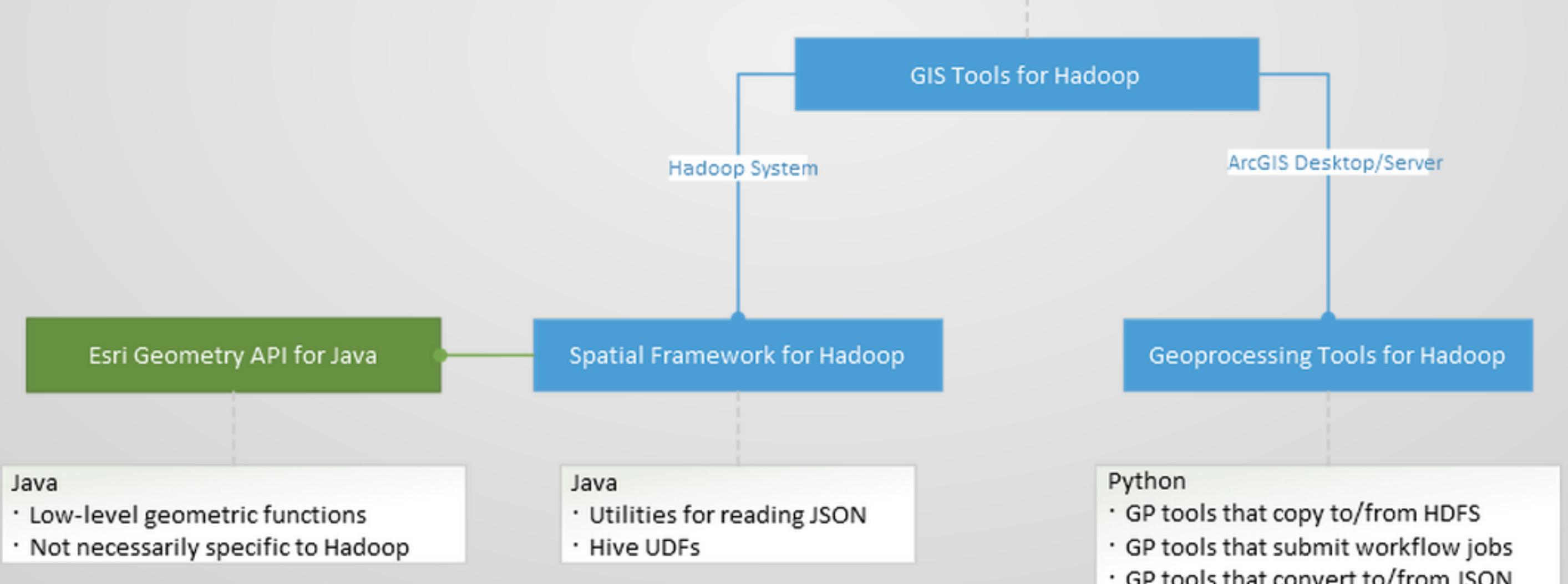
Pre-built samples in the 'samples' director

Place your completed tools in the 'tools' directory if you want to share them

# GIS Tools for Hadoop

The Hadoop 'Tools' are a combination of  
custom Hadoop applications and ArcGIS GP Tools.

- Sample tools that demonstrate the GIS capabilities provided for Hadoop
- Templates that can be used to build tools that solve specific big data problems



Geoprocessing Tools for Hadoop

Features To JSON, JSON To Features

- Provide serialization to and from JSON formats

Copy To HDFS, Copy From HDFS

- Moves files between ArcGIS and Hadoop

Execute Workflow

- Starts a workflow using the Hadoop Oozie workflow engine

Demo

## Developing Custom MapReduce Apps

### Spatial Framework for Hadoop

Enables developers to:

- spatially enable MapReduce applications

Enables Hadoop users to:

# Geometry API for Java

## Simple API Functions for Java

com.esri.core.geometry.\*

### Relationship Analysis

- equals
- disjoint
- touches
- crosses
- within
- contains
- overlaps

### Operations

- buffer
- cut
- clip
- convexHull
- intersect
- union
- difference

Spatial

Enables

- spatial

Enables

- run spa

Provides

- JSON U

- Hive U

Uses the

# Spatial Framework for Hadoop

Enables developers to:

- spatially enable MapReduce applications

Deve

Enables Hadoop users to:

- run spatial Hive queries with ST\_Geometry functions

Provides Java API's for:

- JSON Utility classes
- Hive UDF's

Uses the **Geometry API for Java**

## Spatial Data in Hadoop

JSON files store collections of 'features'

- **Unenclosed JSON** is the dominant style; simple and appendable
- **Enclosed JSON** can optionally be used as a 'feature class'  
(A collection that should be analyzed as a complete set)

Accessing geometries from Hadoop Data Sources

- **com.esri.hadoop.json** - access JSON data as arrays of 'features'
- **com.esri.core.geometry** - construct geometry from arguments

<https://github.com/Esri/spatial-framework-for-hadoop/wiki/JSON-Formats>

```
WHERE ST_Contains (co,
ST_Point (earthquakes.lo
```

## Framework for Hadoop

MapReduce applications

ers to:

queries with ST\_Geometry functions

for:

es

PI for Java

Hadoop

ns of 'features'

e dominant style; simple and appendable

ctionally be used as a 'feature class'

ould be analyzed as a complete set)

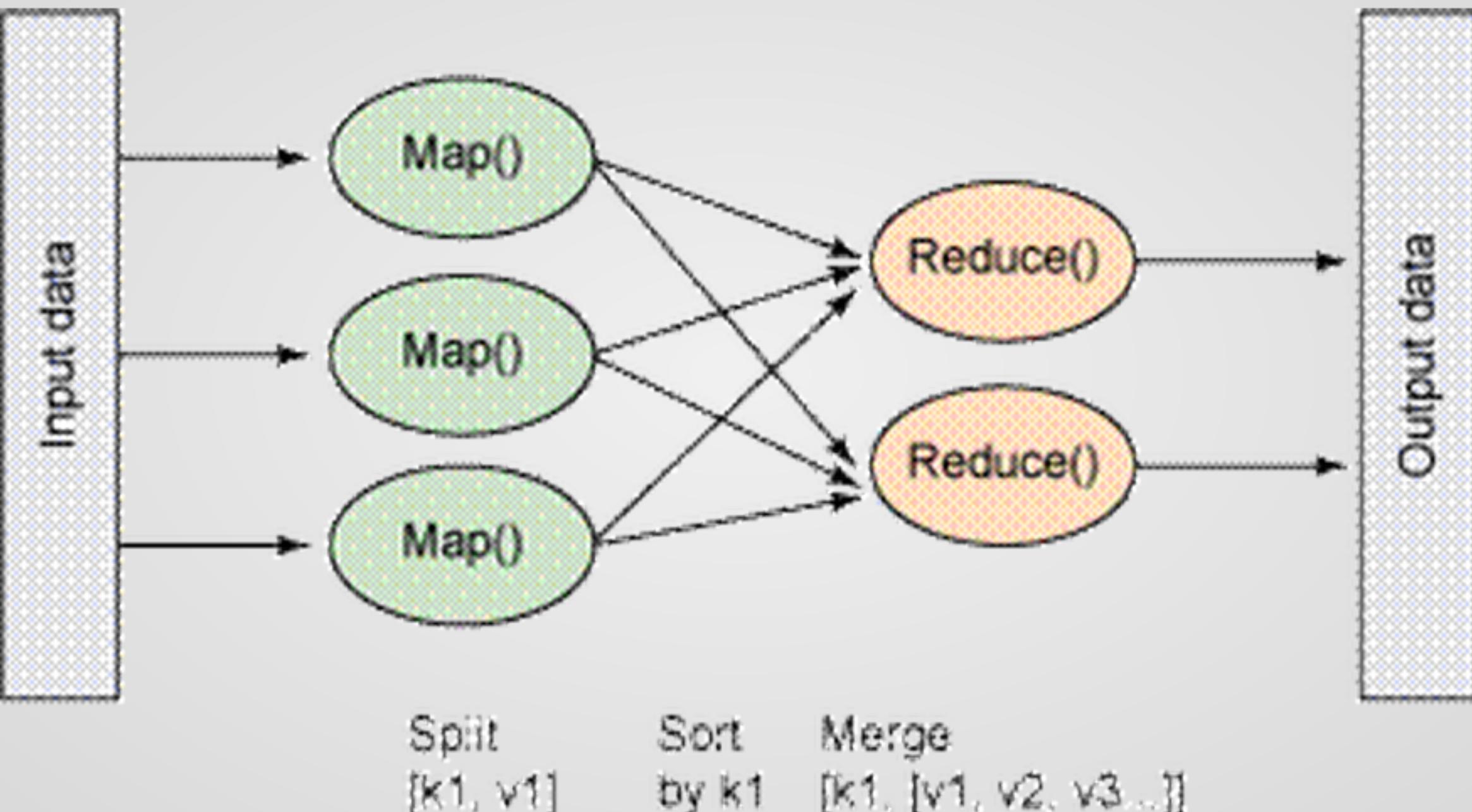
m Hadoop Data Sources

- access JSON data as arrays of 'features'

try - construct geometry from arguments

[al-framework-for-hadoop/wiki/JSON-Formats](https://github.com/Esri/gis-tools-for-hadoop/wiki/JSON-Formats)

# Developing Custom MapReduce Apps



## Simple MapReduce

```
void setup() {
    IStream = hdfs.open(new Path("..."));
    featureClass = EsriFeatureClass.read(IStream);
}

void Map(Long key, Text value) {
    float longitude = Float.parseFloat(value);
    float latitude = Float.parseFloat(value);
    Geometry point = new Point(longitude, latitude);

    for (EsriFeature feature : featureClass) {
        if (GeometryEngine.contains(feature.getGeometry(), point)) {
            String name = feature.getName();
            context.write(new Text(name), value);
            found = true;
        }
    }
}
```

<https://github.com/Esri/gis-tools-for-hadoop>

Demo

# Simple MapReduce Code Sample

```
void setup() {
    iStream = hdfs.open(new Path(config.get("input")));
    featureClass = EsriFeatureClass.fromJson(iStream);
}

void Map(Long key, Text value) {
    float longitude = Float.parseFloat(values[COL_LONG]);
    float latitude = Float.parseFloat(values[COL_LAT]);
    Geometry point = new Point(longitude, latitude);

    for (EsriFeature feature : featureClass.features) {
        if (GeometryEngine.contains(feature.geometry, point)) {
            String name = feature.attributes.get(LABEL_ATTR);
            context.write(new Text(name), data);
            found = true;
            break;
        }
    }
}
```

<https://github.com/Esri/gis-tools-for-hadoop/tree/master/samples>

Demo

# ST\_Geometry in Hive

```
SELECT counties.name, count(*) cnt FROM counties
```

```
JOIN earthquakes
```

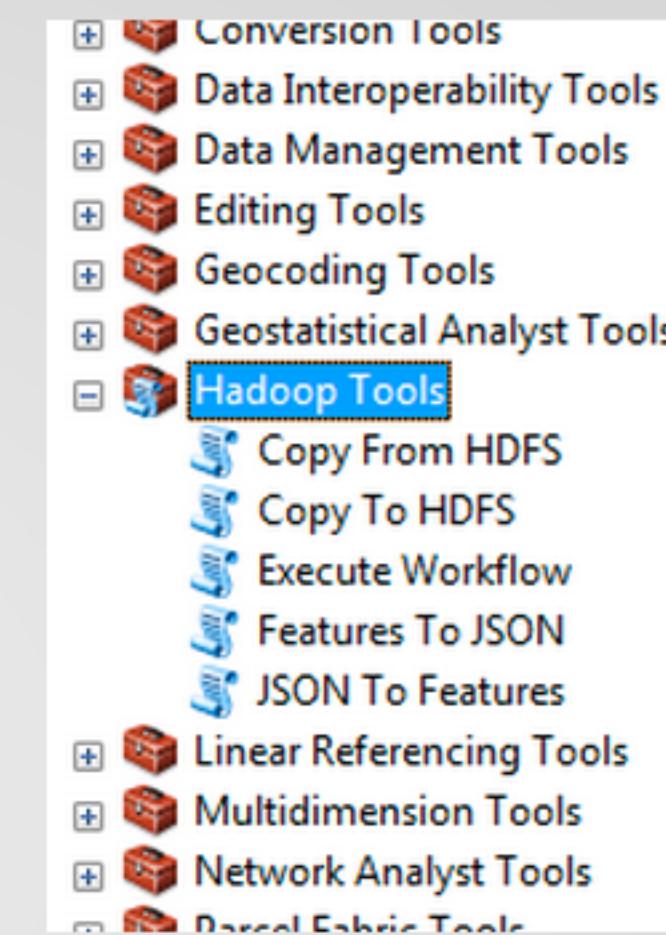
```
WHERE ST_Contains(counties.boundaryshape,  
ST_Point(earthquakes.longitude, earthquakes.latitude))
```

```
GROUP BY counties.name
```

```
ORDER BY cnt desc;
```

<https://github.com/Esri/gis-tools-for-hadoop/tree/master/samples/point-in-polygon-aggregation-hive>

# Geoprocessing Tools for Hadoop



## Features To JSON, JSON To Features

- Provide serialization to and from JSON formats

## Copy To HDFS, Copy From HDFS

- Moves files between ArcGIS and Hadoop

## Execute Workflow

- Starts a workflow using the Hadoop Oozie workflow engine

## Demo

ST\_  
SELECT

# Download the GIS Tools Project

Clone or Fork the project from Github

<http://github.com/Esri/gis-tools-for-hadoop>

Pre-built samples in the 'samples' directory

Place your completed tools in the 'tools'  
directory if you want to share them

# Get the Source Code

Geometry API

<http://github.com/Esri/geometry-api-java>

Spatial Framework for Hadoop

<http://github.com/Esri/spatial-framework-for-hadoop>

Geoprocessing Tools for Hadoop

<http://github.com/Esri/geoprocessing-tools-for-hadoop>

To build the source:

**ant**

# Contributing Your Work

Fork the **gis-tools-for-hadoop** project

- Hack on the code
  - > Make new tools
  - > Do **awesome** spatial analysis on big data
- Send a Github 'pull request' so we can pull the tool back into our project

Let us know how you are doing

- Add your content to the Wikis on Github
- Troubles? Open new Issues in Github

# We want your feedback!

## Session Feedback

<http://esriurl.com/survey>

Session Offering ID: **301**



E-mail:

David Kaiser <[dkaiser@esri.com](mailto:dkaiser@esri.com)> @ddkaiser

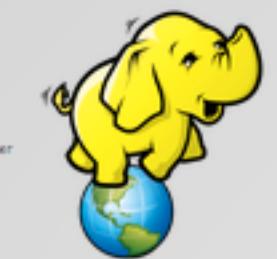
Michael Park <[mpark@esri.com](mailto:mpark@esri.com)>

## Follow along with this presentation

<http://esri.github.com/gis-tools-for-hadoop/devsummit2013>

Use WiFi network: Esri2013

## Big Data: Using ArcGIS with Apache Hadoop



David Kaiser @ddkaiser  
Michael Park  
  
Esri DevSummit 2013  
  
Session Offering ID: 301

## GIS Tools for Hadoop

Hadoop users often have data with spatial value, but with limited options for spatially analyzing this data

Esri has released an open-source framework to enable spatial data processing in your Hadoop applications

This enables you, as a developer, to build analytical tools that use both Hadoop and ArcGIS.

## Why is this important?

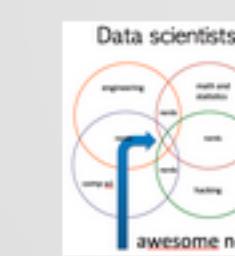
Your Hadoop applications can provide spatial analysis

...and your users can leverage your Hadoop applications from within the ArcGIS Geoprocessing environment

## Who is a Data Scientist?



## Finding Your Data Scientist



## We want your feedback!

### Session Feedback

<http://esriurl.com/survey>

Session Offering ID: 301

## Contributing Your Work

Fork the gis-tools-for-hadoop project

- Hack on the code
- > Make new tools
- > Do awesome spatial analysis on big data
- Send a Github pull request so we can pull the tool back into our project

Let us know how you are doing

- Add your content to the Wikis on Github
- Troubles? Open new issues in Github

## Get the Source Code

### Geometry API

<http://github.com/Esri/geometry-api-java>

Spatial Framework for Hadoop

<http://github.com/Esri/spatial-framework-for-hadoop>

Geoprocessing Tools for Hadoop

<http://github.com/Esri/geoprocessing-tools-for-hadoop>

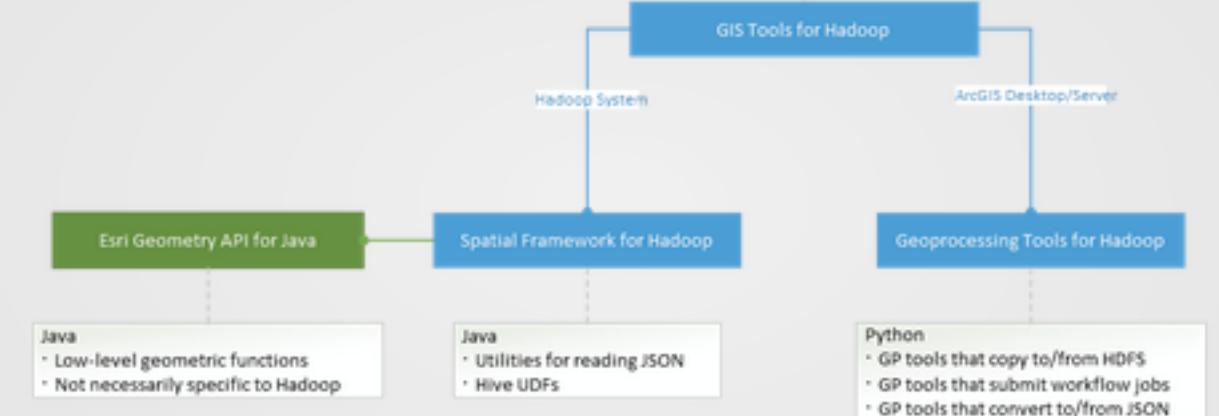
To build the source:

ant

# GIS Tools for Hadoop

The Hadoop 'Tools' are a combination of custom Hadoop applications and ArcGIS GP Tools.

- Sample tools that demonstrate the GIS capabilities provided for Hadoop
- Templates that can be used to build tools that solve specific big data problems



## Download the GIS Tools Project

Clone or Fork the project from Github

<http://github.com/Esri/gis-tools-for-hadoop>

Pre-built samples in the 'samples' directory

Place your completed tools in the 'tools' directory if you want to share them

## Geoprocessing Tools for Hadoop



Features To JSON, JSON To Features

- Provide serialization to and from JSON formats

Copy To HDFS, Copy From HDFS

- Moves files between ArcGIS and Hadoop

Execute Workflow

- Starts a workflow using the Hadoop Oozie workflow engine

Demo

ST\_Geometry in Hive

SELECT counties.name, count(\*) c FROM counties

JOIN earthquakes

WHERE ST\_Contains(counties.boundaryshape,

ST\_Point(earthquakes.longitude, earthquakes.latitude))

GROUP BY counties.name

ORDER BY c DESC;

[https://github.com/Esri/gis-tools-for-hadoop/blob/master/hadoop-project/src/main/resources/st\\_geometry\\_in\\_hive.hql](https://github.com/Esri/gis-tools-for-hadoop/blob/master/hadoop-project/src/main/resources/st_geometry_in_hive.hql)

Input data

Map()

Map()

Map()

Reduce()

Reduce()

Output data

Sort by k1 [v1, v2, v3, ...]

Sort by k1 [v1, v2, v3, ...]

Merge [k1, v1, v2, v3, ...]

Simple MapReduce Code Sample

<https://github.com/Esri/gis-tools-for-hadoop/blob/master/hadoop-project/src/main/java/com/esri/hadoop/mapreduce/MapReduceExample.java>

Input data

Map()

Map()

Map()

Reduce()

Reduce()

Output data

Sort by k1 [v1, v2, v3, ...]

Merge [k1, v1, v2, v3, ...]