



Esri International Developer Summit
Palm Springs, CA

BigData Visualization

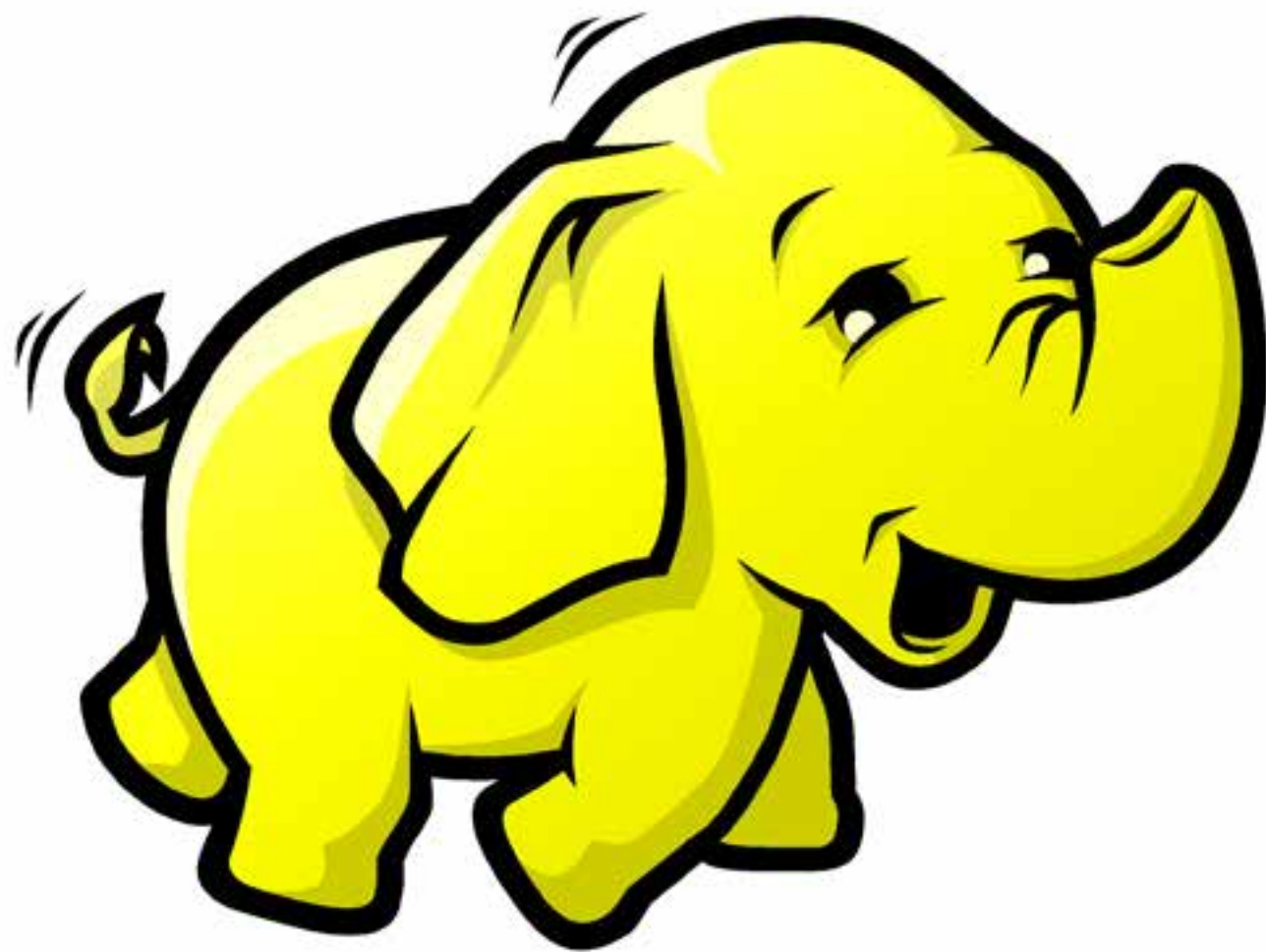
@SpatialAgent
@mraad

BigData ?

”

Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...

- Dan Ariely

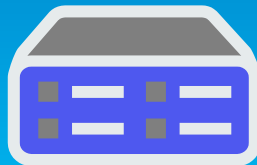


Hadoop Basic Stack

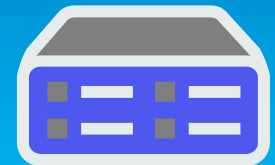
MapReduce

Yet Another Resource Negotiator (YARN)

Hadoop Distributed File System (HDFS)



Commodity Servers



The Zoo

- Hive - Ad Hoc Query - “SQL” to MapReduce
- Pig - High Level Data Analysis Language
- Impala - MPP SQL Engine
- Mahout - Machine Learning Toolbox
- HBase - Columnar Key/Value Database
- Cascading - Flow Data Analysis
- Avro - Data Serializer
- Zookeeper - Centralized State Management

All Services - Cloudera Manager

hadoop1:7180/cm/services/status

All Services - Cloudera Manager

cloudera manager Home Services Hosts Activities Diagnose Audits Charts Administration Search by Service, Support admin

February 10 2014, 2:54:47 PM EST

01:25 01:30 01:35 01:40 01:45 01:50 01:55 02:00 02:05 02:10 02:15 02:20 02:25 02:30 02:35 02:40 02:45

Now

All Services

Try Cloudera Enterprise for 60 Days Add Cluster

Cluster 1 - CDH4

Actions

Name	Status	Role Counts	Actions
hbase1	Good Health	4 RegionServers, 1 Master	Actions
hdfs1	Good Health	1 SecondaryNameNode, 1 NameNode, 1 Balancer, 4 DataNodes	Actions
hive1	Good Health	1 Hive Metastore Server, 4 Gateways	Actions
hue1	Stopped	1 Beeswax Server, 1 Hue Server	Actions
impala1	Good Health	1 Impala Catalog Server Daemon, 4 Impala Daemons, 1 Impala StateStore Daemon	Actions
mapreduce1	Good Health	1 JobTracker, 4 TaskTrackers	Actions
oozie1	Stopped	1 Oozie Server	Actions
sqoop1	Stopped	1 Sqoop Server	Actions
zookeeper1	Good Health	3 Servers	Actions

Cloudera Management Services

View Maintenance Mode Status

Name	Status	Role Counts	Actions
mgmt1	Good Health	1 Event Server, 1 Host Monitor, 1 Activity Monitor, 1 Alert Publisher, 1 Service Monitor	Actions



GIS Tools for Hadoop

Big Data Spatial Analytics for the Hadoop Framework



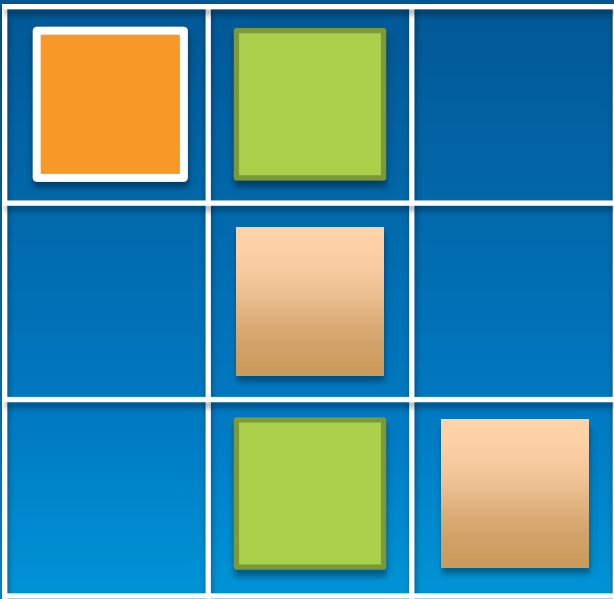
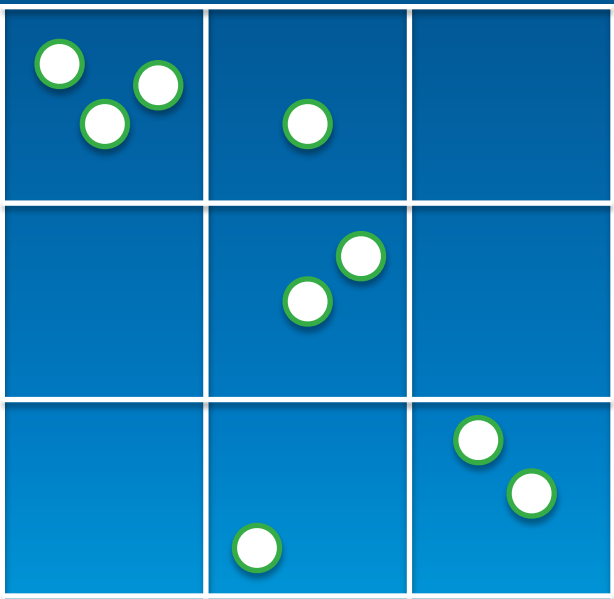
View project on
[GitHub](#)

GIS Tools For Hadoop

- Geometry API
 - Point / Line / Polygon
 - Operations - Contains, Intersect, Buffer
 - I/O - WKT, GeoJSON, Shape
- Hive Spatial UDF
 - ST_POINT, ST_CONTAINS
- GeoProcessing Extensions

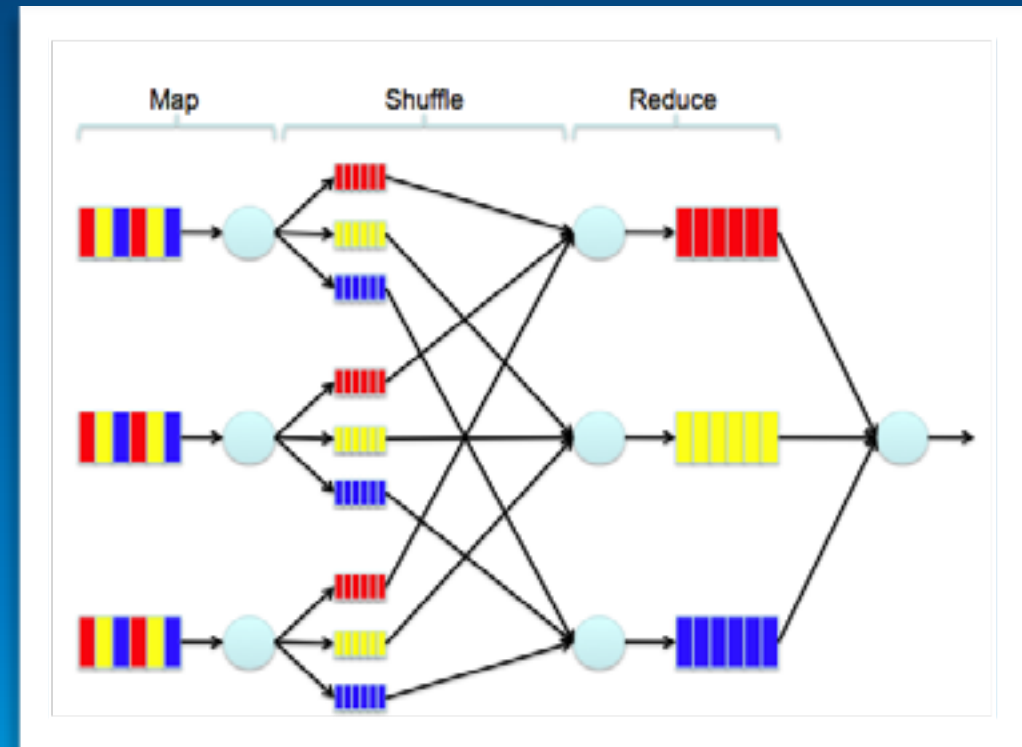
Hello, MapReduce !

Density Analysis - Cell Count



MapReduce Recap

- Map
 - Extract
 - Filter
 - Transform
- Reduce
 - Group By
 - Aggregate

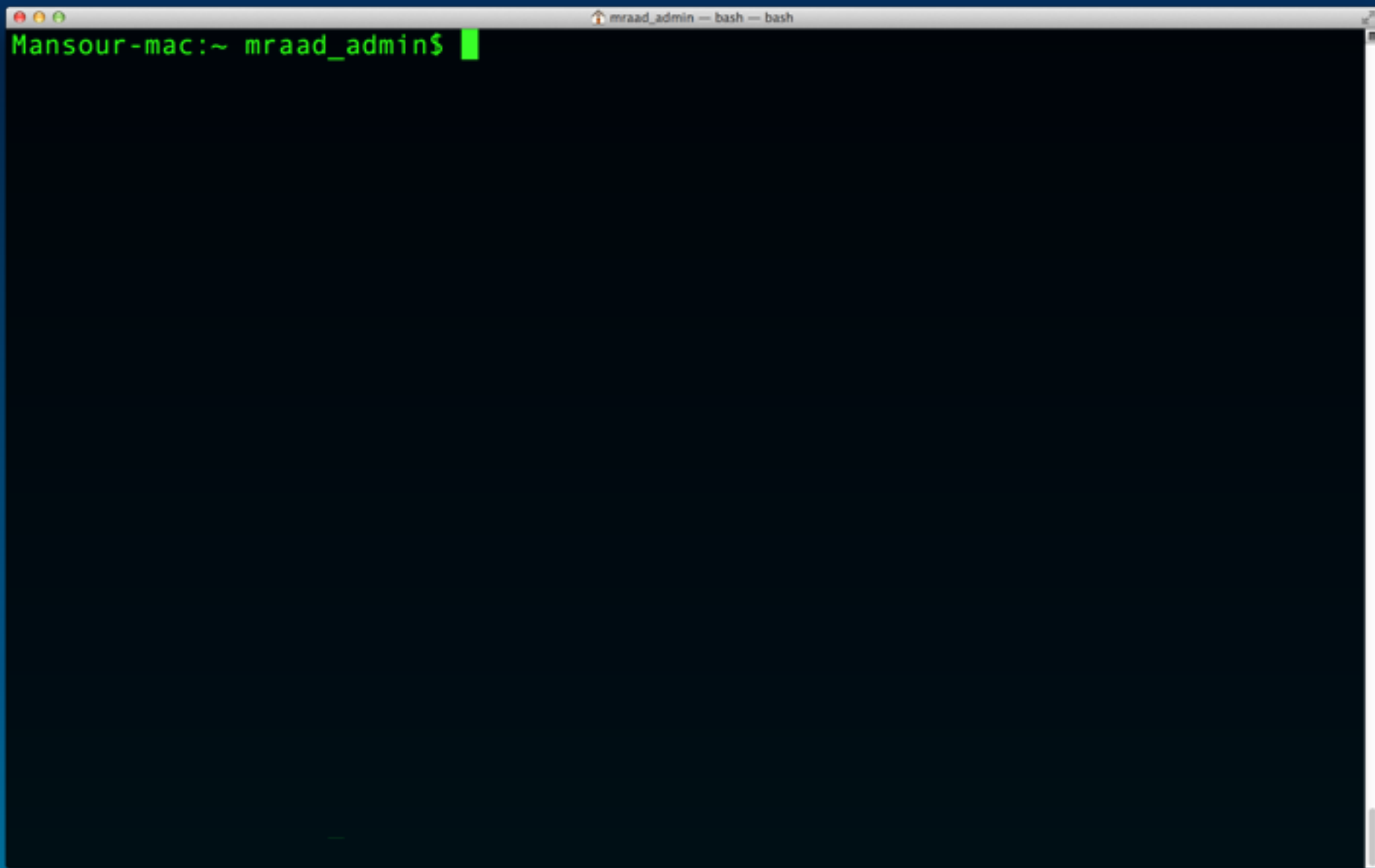


Cell Count

```
function map(lineno,text) {  
  (x,y) = tokenize(text)  
  if(inGrid(x,y)){  
    (cellX,cellY) = toCell(x,y)  
    emit((cellX,cellY),1)  
  }  
}
```

```
function reduce((cellX,cellY),iterator){  
  sum = 0  
  for( one in iterator){  
    sum = sum + one  
  }  
  emit((cellX,cellY), sum)  
}
```

Cooler UX...



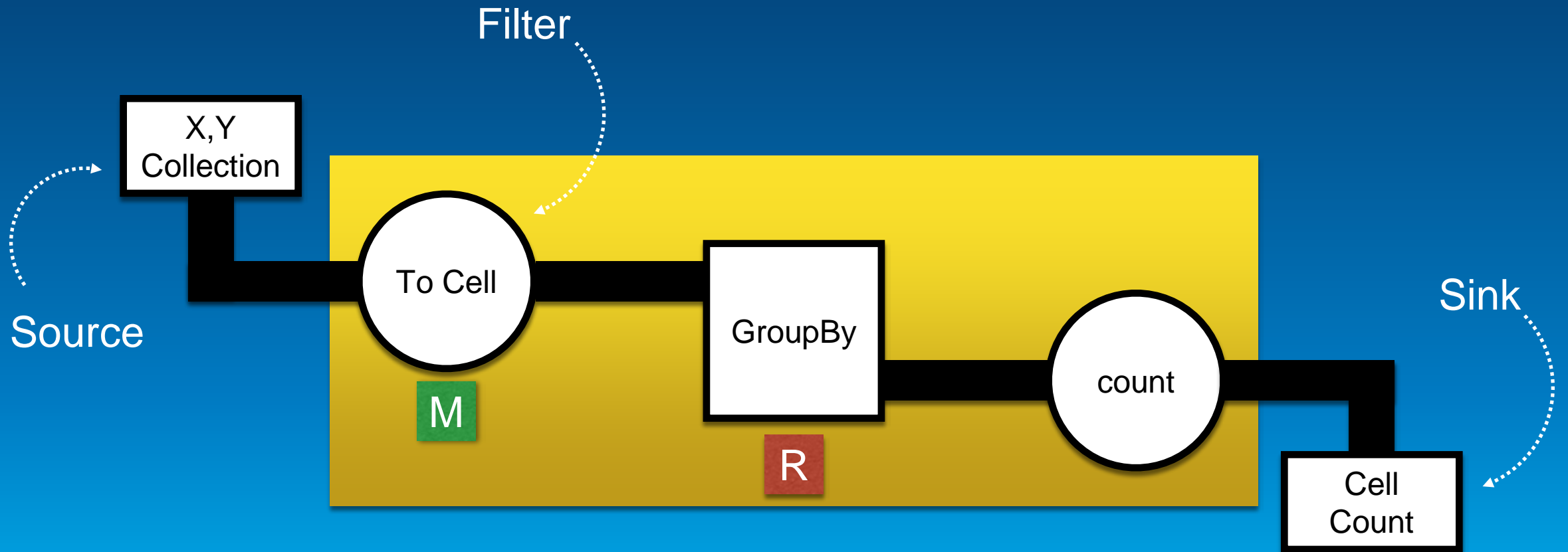
In Action Demo

MapReduce Is Hard...

Thinking Of Data As Water



Cascading Pipeline



```
33
34     final Fields inFields = new Fields("ID", "X", "Y").
35         applyTypes(long.class, double.class, double.class);
36
37     final Tap inTap = new Hfs(new TextDelimited(inFields, false, "\t"), args[0]);
38
39     final Tap outTap = new Hfs(new TextDelimited(true, ","), args[2], SinkMode.REPLACE);
40
41     final SpatialDensity spatialDensity = new SpatialDensity();
42
43     Pipe pipe = new Each("start", new Fields("X", "Y"), spatialDensity);
44
45     pipe = new GroupBy(pipe, spatialDensity.getFieldDeclaration());
46
47     pipe = new Every(pipe, Fields.GROUP, new Count(new Fields("POPULATION")));
48
49     final Properties properties = AppProps.appProps().
50         setJarClass(App.class).
51         buildProperties();
52
53     properties.put(SpatialDensity.KEY_SHP, args[1]);
54
55     final FlowConnector connector = new HadoopFlowConnector(properties);
56
57     final Flow flow = connector.connect(inTap, outTap, pipe);
58
59     flow.complete();
60
```

Spatial Join !

Cascading In Action

**How About No Programming ?
What About SQL ?**



Hive and Impala



```
drop table if exists zipcodes;
```

```
create external table if not exists zipcodes(  
id int,  
lon double,  
lat double  
) row format delimited  
fields terminated by '\t'  
lines terminated by '\n'  
stored as textfile  
location '/user/cloudera/zipcodes';
```

```
SELECT  
T.X-180+0.5 AS LON,  
T.Y-90+0.5 AS LAT,  
COUNT(*) AS POPULATION  
FROM (  
SELECT  
FLOOR(LON+180) AS X,  
FLOOR(LAT+90) AS Y  
FROM ZIPCODES  
) T GROUP BY T.X,T.Y;
```

Hive and Impala In Action

ArcGIS & Hadoop

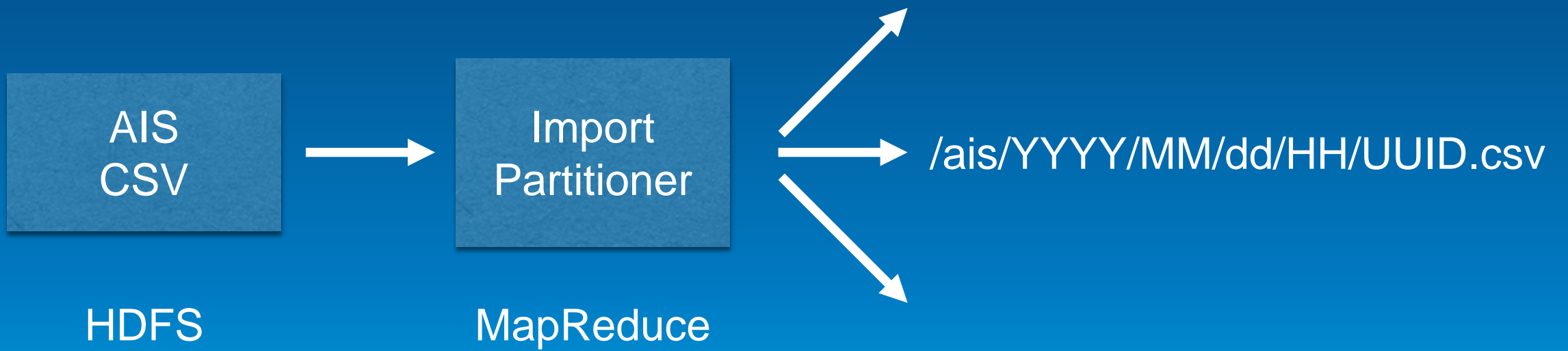
AIS DATA

- **14.8 Million data points**
- **1 Month**
- **MMSI, Zulu Time, Lat, Lon, Vessel ID, Draught**
- **Port of Miami - Free**

DEMO Steps

- GP Toolbox
- Track Assembly
- Hex Generation
- Density Analysis

Import Job

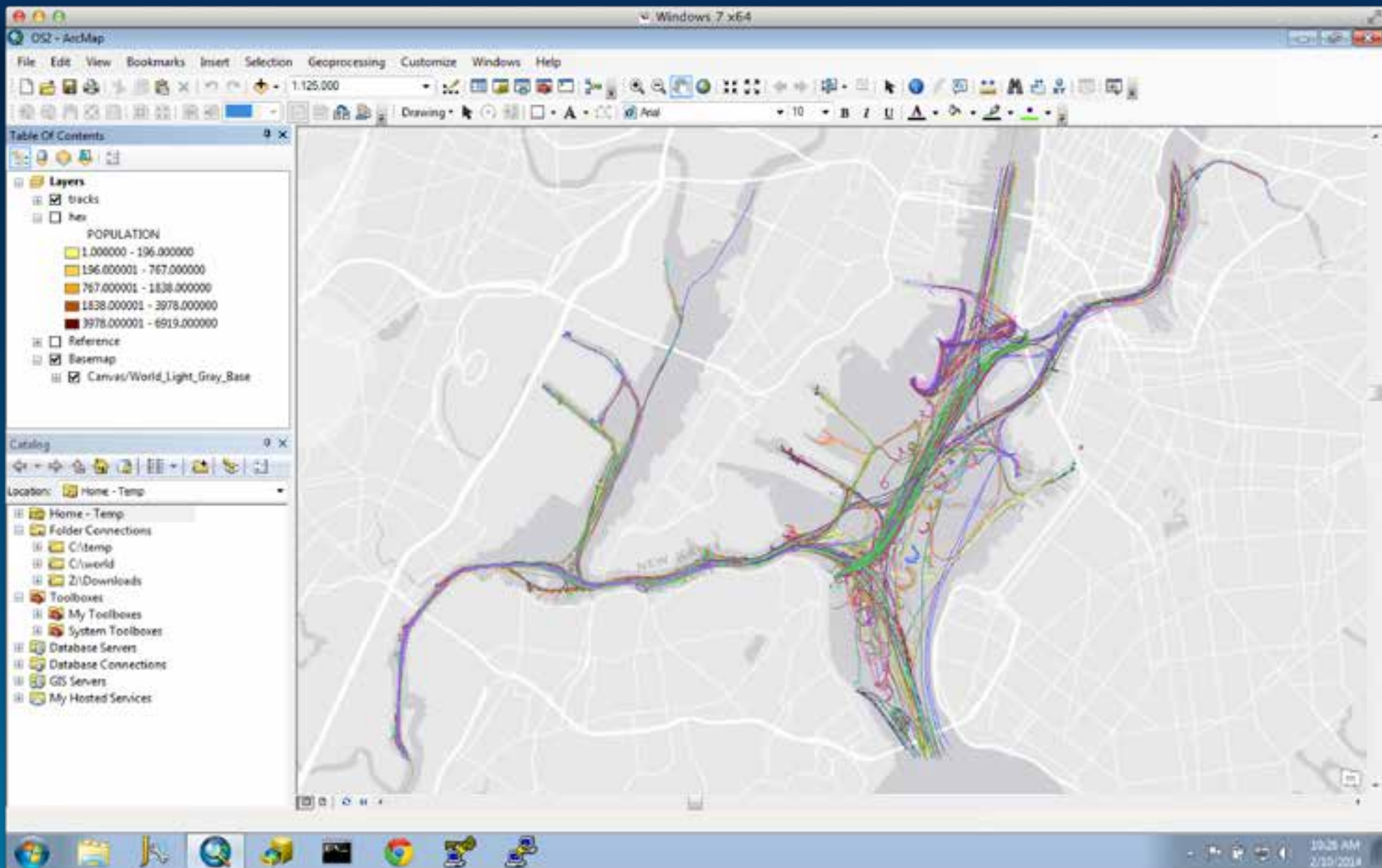



```
12  */
13  final class ImportReduce
14      extends Reducer<DateHour, Text, NullWritable, Text>
15  {
16      private MultipleOutputs<NullWritable, Text> m_mos;
17
18      @Override
19      protected void setup(final Context context) throws IOException, InterruptedException
20      {
21          m_mos = new MultipleOutputs<NullWritable, Text>(context);
22      }
23
24      @Override
25      protected void reduce(
26          final DateHour key,
27          final Iterable<Text> values,
28          final Context context) throws IOException, InterruptedException
29      {
30          final String outputPath = String.format("/por/%4d/%02d/%02d/%02d/%s",
31              key.yy, key.mm, key.dd, key.hh, UUID.randomUUID().toString());
32          for (final Text value : values)
33          {
34              m_mos.write(NullWritable.get(), value, outputPath);
35          }
36      }
37
38      @Override
39      protected void cleanup(final Context context) throws IOException, InterruptedException
40      {
41          m_mos.close();
42      }
43  }
```

```
mraad_admin -- ec2-user@ip-10-198-67-75:~ -- ssh -- bash
[ec2-user@ip-10-198-67-75 ~]$ impala-shell
Starting Impala Shell without Kerberos authentication
Connected to ip-10-198-67-75.us-west-1.compute.internal:21000
Server version: impalad version 1.2.3 RELEASE (build 1cab04cdb88968a963a8ad6121a2e72a3a623eca)
Welcome to the Impala shell. Press TAB twice to see a list of available commands.

Copyright (c) 2012 Cloudera, Inc. All rights reserved.

(Shell build version: Impala Shell v1.2.3 (1cab04c) built on Fri Dec 28 19:39:39 PST 2013)
[ip-10-198-67-75.us-west-1.compute.internal:21000] > select hour,count(hour) from ais where year=2009 and month=1 group by hour order by hour limit 24;
Query: select hour,count(hour) from ais where year=2009 and month=1 group by hour order by hour limit 24
+-----+
| hour | count(hour) |
+-----+
| 0    | 614439      |
| 1    | 609928      |
| 2    | 608169      |
| 3    | 609392      |
| 4    | 607497      |
| 5    | 605260      |
| 6    | 606295      |
| 7    | 611110      |
| 8    | 609829      |
| 9    | 601292      |
| 10   | 610953      |
| 11   | 623449      |
| 12   | 630816      |
| 13   | 627904      |
| 14   | 625762      |
| 15   | 629808      |
| 16   | 627557      |
| 17   | 627228      |
| 18   | 630212      |
| 19   | 631419      |
| 20   | 630062      |
| 21   | 626892      |
| 22   | 624329      |
| 23   | 618161      |
+-----+
Returned 24 row(s) in 1.93s.
[ip-10-198-67-75.us-west-1.compute.internal:21000] > █
```



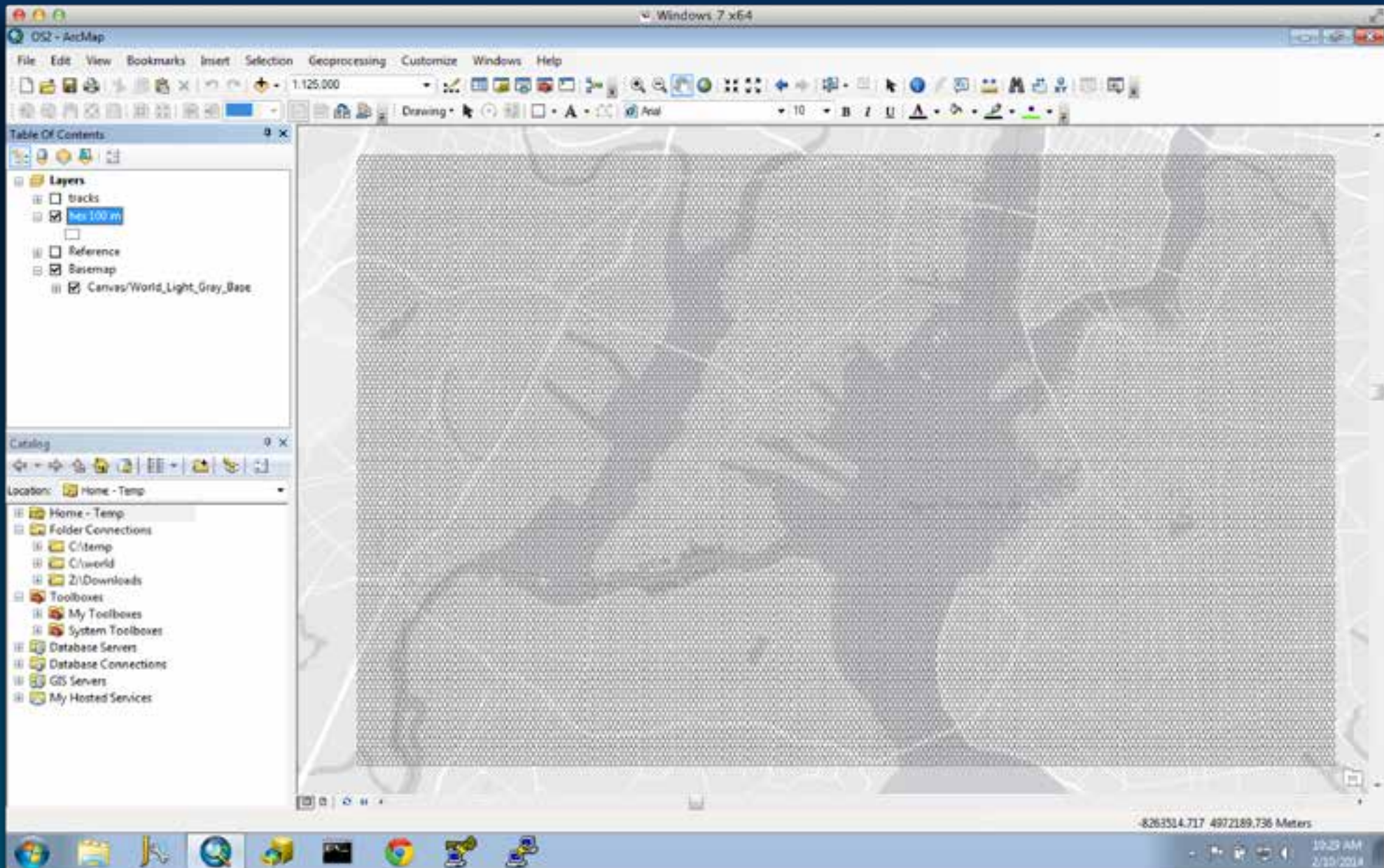


Table Of Contents

- Layers
 - Tracks
 - hex 100 m
 - POPULATION
 - 1.000000 - 196.000000
 - 196.000001 - 767.000000
 - 767.000001 - 1838.000000
 - 1838.000001 - 3978.000000
 - 3978.000001 - 6919.000000
 - Reference
 - Basemap
 - Canvas/World_Light_Gray_Base

Catalog

Location: Home - Temp

- Home - Temp
- Folder Connections
 - C:\temp
 - C:\world
 - Z:\Downloads
- Toolboxes
 - My Toolboxes
 - System Toolboxes
- Database Servers
- Database Connections
- GIS Servers
- My Hosted Services



Cancelled

-8263415.498 4971329.838 Meters

WebMaps

<http://coolmaps.esri.com/BigData>

/ShippingTracks

/ShippingVolume

/ShippingRank

/ShippingGlobe

Return of KillerApps ?

<https://github.com/mraad>

@SpatialAgent

@mraad



Understanding our world.