Open-source Architecture of a real-time Global Situational Awareness System with Cloudera and GeoMesa

Jim Hughes, CCRi
David Kaiser, Cloudera
Talk outline

● Introductions
  ○ David Kaiser, Cloudera
  ○ Jim Hughes, CCRi

● Motivation

● Cloudera
  ○ What is Cloudera?
  ○ What is new with Hortonworks and Cloudera?

● GeoMesa
  ○ What is GeoMesa?
  ○ What is new with GeoMesa in versions 2.1 / 2.2 / 2.3?

● Conclusion: Cloudera and GeoMesa are great together!

● Thank you very much
Introductions
David Kaiser

- 10 Years @ Esri
  - Released *GIS Tools for Hadoop*
- 5.5 Years @ Hortonworks
  - (Now merged with Cloudera)
- Open-Source Involved since 1997
  - Linuxcare, Linux OS & Support startup
  - Linux User Group (LUG) 15+ years
- Extensive Background in Data Platforms
Jim Hughes

- CCRi’s Director of Open Source Programs
- GeoMesa core committer
- SFCurve Project Lead
- JTS committer
- Contributor to GeoTools and GeoServer
Motivation
Motivation: Unpacking the title

“Open-source Architecture of a real-time Global Situational Awareness System with Cloudera and GeoMesa”
Motivation: Unpacking the title

“Open-source Architecture of a real-time Global Situational Awareness System with Cloudera and GeoMesa”

Open-source means that the core software has no cost and its source can be examined.
Motivation: Unpacking the title

“Open-source Architecture of a real-time Global Situational Awareness System with Cloudera and GeoMesa”

Real-time means two things:

a) data can be streamed efficiently

b) queries run in an interactive amount of time.
Motivation: Unpacking the title

“Open-source Architecture of a real-time Global Situational Awareness System with Cloudera and GeoMesa”

Global-scale requires technologies which are scalable and resilient.

Such systems also need to be efficient in storage and compute.
Motivation: Unpacking the title

“Open-source Architecture of a real-time Global Situational Awareness System with Cloudera and GeoMesa”

Situational Awareness means that the data in the system matters to some decision maker.

Multiple relevant datasets are present; their fusion and further analysis provides insight.
Motivation: Unpacking the title

“Open-source Architecture of a real-time Global Situational Awareness System with Cloudera and GeoMesa”

Implied: One or more of the data sets will be spatial, so the technologies need to understand spatial types and functions.
Concept: Edge To AI

CLOUDERA DATA PLATFORM

- Unified control plane
- Analytic experiences
- Data anywhere
- Infrastructure
- Altus DataPlane
- Identity | Orchestration | Management | Operations
- Data Flow & Streaming
- Data Engineering
- Data Warehouse
- Operational Database
- Machine Learning
- Single Pane of Glass
- Self-Serve Application Experiences
- New Distribution
- Shared Data Experience
- Use Anywhere
- Open source distribution
- DISTRO-X
- Catalog | Schema | Migration | Security | Governance
GeoMesa

- GeoMesa Overview
- GeoMesa new features
What is GeoMesa?

A suite of tools for streaming, persisting, managing, and analyzing spatio-temporal data at scale
What is GeoMesa?

A suite of tools for **streaming**, persisting, managing, and analyzing spatio-temporal data at scale
What is GeoMesa?

A suite of tools for streaming, **persisting**, managing, and analyzing spatio-temporal data at scale.
What is GeoMesa?

A suite of tools for streaming, persisting, managing, and analyzing spatio-temporal data at scale.
What is GeoMesa?

A suite of tools for streaming, persisting, managing, and **analyzing** spatio-temporal data at scale.
Proposed Reference Architecture
Demo
Thanks!

Jim Hughes

- jhughes@ccri.com
- Geomesa.org
- http://gitter.im/locationtech/geomesa

David Kaiser

- dkaiser@cloudera.com
More Talks about GeoMesa

4:00pm: Optimizing Big Data Formats for Vector Data

4:30pm: Space-Filling Curves and The Art of Composition
# GeoMesa release history

<table>
<thead>
<tr>
<th>Version</th>
<th>Release date</th>
<th>Release Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>April 20, 2018</td>
<td>GeoMesa 2.0.0</td>
</tr>
<tr>
<td>2.1</td>
<td>November 12, 2018</td>
<td>GeoMesa 2.1.0</td>
</tr>
<tr>
<td>2.2</td>
<td>December 28, 2018</td>
<td>GeoMesa 2.2.0</td>
</tr>
<tr>
<td>2.3</td>
<td>March 29, 2019</td>
<td>GeoMesa 2.3.0</td>
</tr>
</tbody>
</table>
### List of features to discuss

<table>
<thead>
<tr>
<th>Feature</th>
<th>Version</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeoMesa Redis DataStore</td>
<td>2.3.0</td>
<td></td>
</tr>
<tr>
<td>GeoMesa FileSystem DataStore</td>
<td>2.3.0</td>
<td></td>
</tr>
<tr>
<td>JTS PySpark Integration</td>
<td>2.3.0</td>
<td>Improvements</td>
</tr>
<tr>
<td>GeoTools 20 / GeoServer 2.14 support</td>
<td>2.2.0</td>
<td>No need for a GeoMesa datastore</td>
</tr>
<tr>
<td>Spark 2.3 support</td>
<td>2.1.0</td>
<td></td>
</tr>
<tr>
<td>Apache Kudu DataStore</td>
<td>2.1.0</td>
<td></td>
</tr>
</tbody>
</table>
GeoMesa Roadmap Upcoming Features

- Version 3.0
- Support for HBase 2
- Support for Accumulo 2
- Support for Hadoop 3
- Support for Java 11
- Upgrade to Scala 2.12
Feature Name

Feature Overview