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

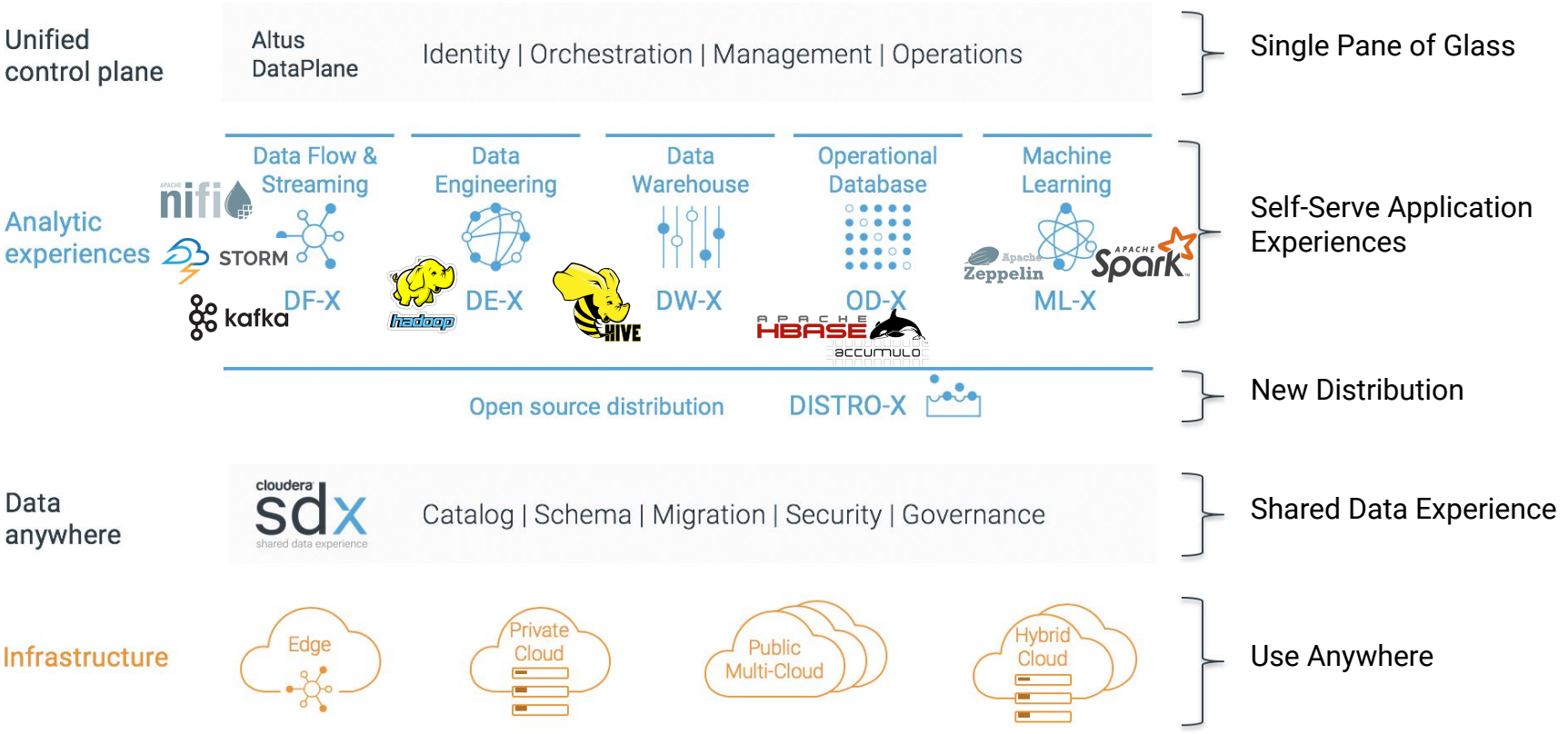
"Open-source Architecture of a real-time Global Situational Awareness System with Cloudera and GeoMesa"

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

Cloudera

Concept: Edge To AI

CLOUDERA DATA PLATFORM



GeoMesa

- GeoMesa Overview
- GeoMesa new features

CLOUDERA

What is GeoMesa?

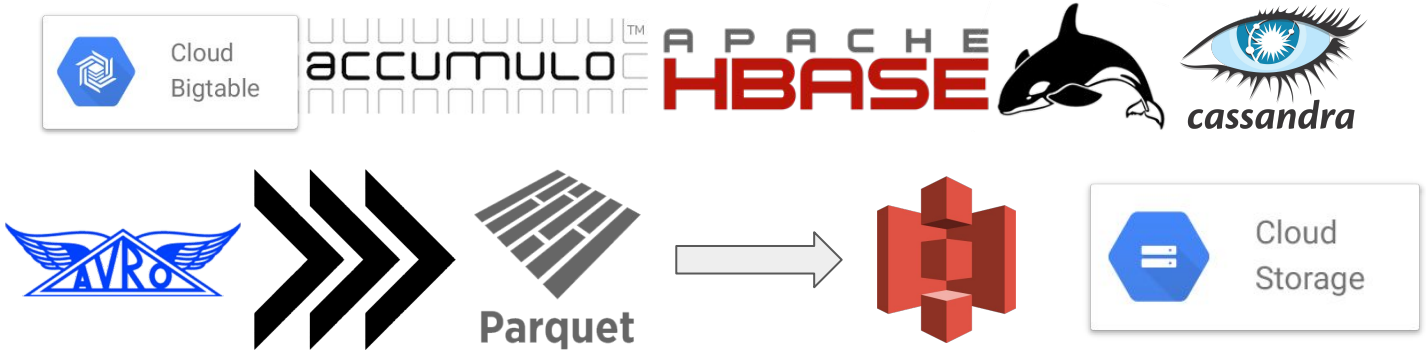
A suite of tools for streaming, persisting, managing, and analyzing spatio-temporal data at scale

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

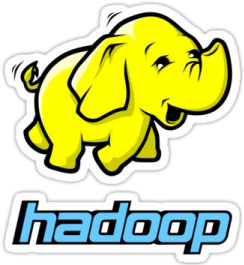


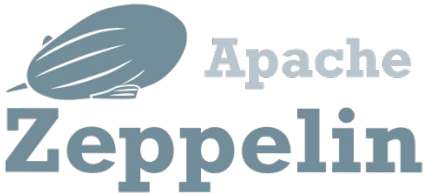
What is GeoMesa?

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

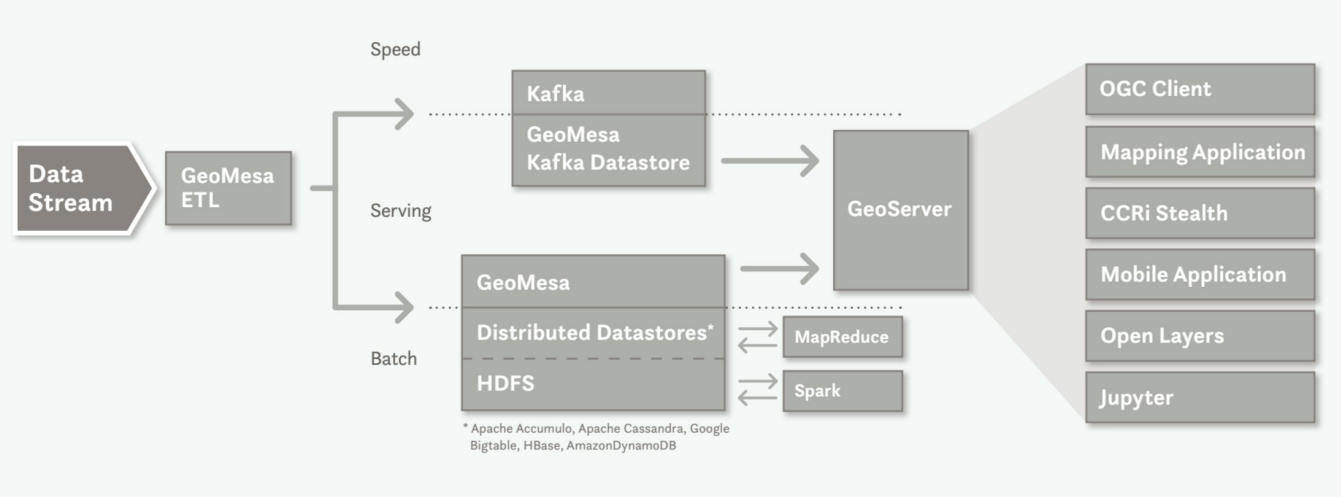
What is GeoMesa?

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

CLOUDERA

Backup Slides

GeoMesa release history

Version	Release date	Release Notes
2.0	April 20, 2018	GeoMesa 2.0.0
2.1	November 12, 2018	GeoMesa 2.1.0
2.2	December 28, 2018	GeoMesa 2.2.0
2.3	March 29, 2019	GeoMesa 2.3.0

[Internal Scratch work] List of features to discuss

Feature	Version	
GeoMesa Redis DataStore	2.3.0	
GeoMesa FileSystem DataStore	2.3.0	Improvements
JTS PySpark Integration	2.3.0	No need for a GeoMesa datastore
GeoTools 20 / GeoServer 2.14 support	2.2.0	
Spark 2.3 support	2.1.0	
Apache Kudu DataStore	2.1.0	

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

Component / Capability (e.g. Streaming)

Feature Name (e.g. Confluent Integration

Project and Version (e.g. GeoMesa 2.0)