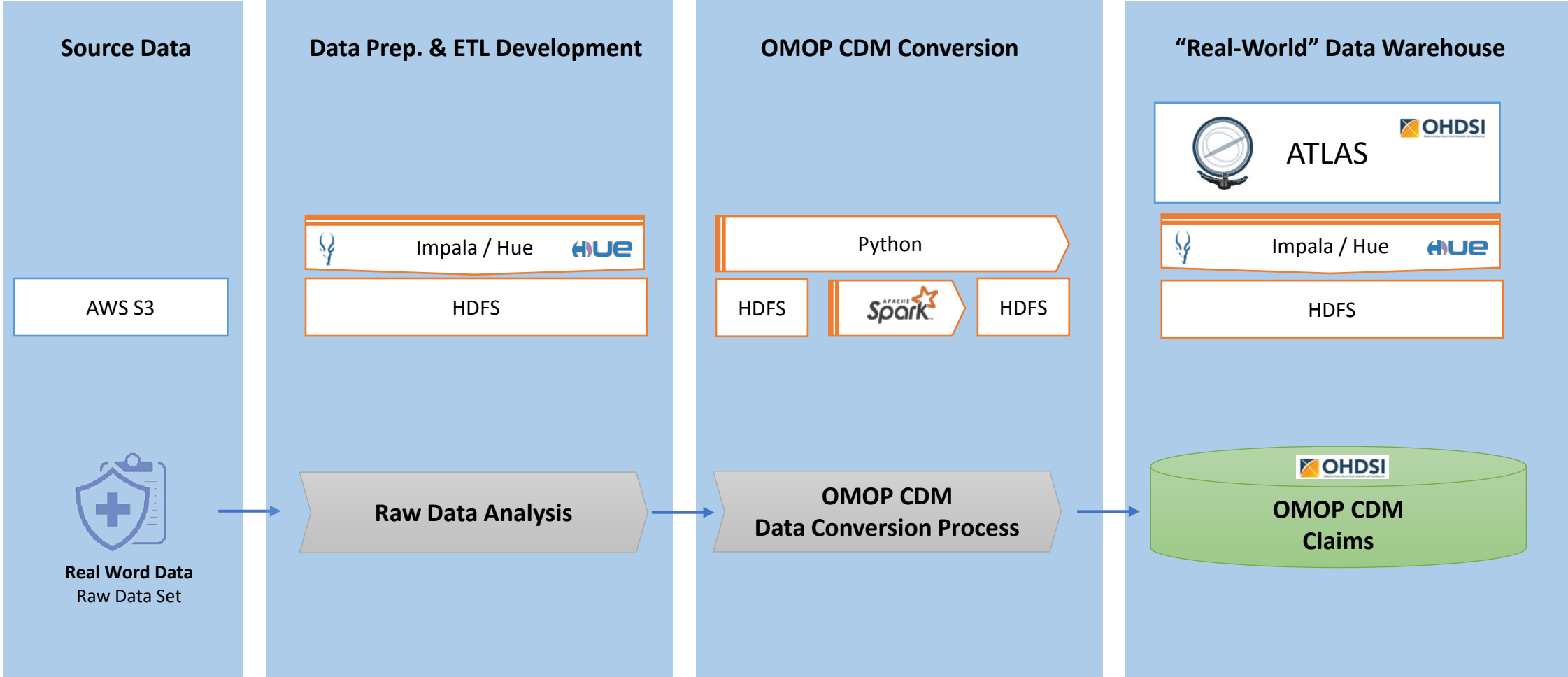


OMOP CDM Conversion with *hadoop*

Gregory Klebanov
CTO - Odysseus Data Services, Inc.



CDM Conversion with Hadoop - Architecture



Why Apache Hadoop? Benefits and Improvements

General

- Portable between various Hadoop distributions – no vendor lock-in (AWS, Cloudera, Apache Foundation)
- Great scalability for Big Data processing
- Less software / hardware components to support
- Overall lower infrastructure cost

Data Prep and ETL Development Improvements

- No need to load data into RDBMS – immediate access via Impala / Hue
- Data Analysts and ETL Developers can continue to use SQL
- Scalable for complex large data queries

OMOP CDM Conversions

- Faster OMOP CDM conversions - allows for parallel processing
- Great scalability for Big Data processing

Data Analytics

- No need to load data into RDBMS – immediate access via Impala / Hue
- Shorter deployment cycle

Some initial challenges

Building a skill set:

- Training staff
- Finding experienced Hadoop engineers and architects

Some learning curve:

- Infrastructure set up and sizing
- Thinking differently about an approach
- Understanding ways of working with Big Data

Technology is still evolving