Attendees:

* Shawn Dolley
* Tony Reckard
* Greg Klebanov
* John Hope
* Lee Evans
* Tom White
* Corina
* Mike Warfe

Action items:

* Action Item: Set up Git repository for this working group to keep all work – Mui
* Action Item: Inventory of what tools are being used/will be used by each group - Mui

**Technical Goals**

* OHDSI Stack –
	+ OHDSI stack working on Hadoop (CDM, Achilles, ATLAS – Feature Extraction)
		- Feature Extraction – name for another project on OHDSI, a project that Frank mentioned that we may want to consider when we are looking at Hadoop
		- CDM includes ETL - Utilization of technical to ETL
			* When to use Spark versus SQL versus Impala
		- Atlas Plus has multiple features: Vocabulary, Concept Set, Cohorts, Population Estimate, Patient Level Estimate, Incident Rate
		- Survey to talk ask who would be interested in working on these sections and who has actual business cases for this section. To start after the symposium
			* Talk to Frank regarding the architecture group – Maybe use the time during the
	+ Provide docker containers for running the OHDSI stack on Hadoop
	+ Set up a cloud instance of the OHDSI stack on Hadoop
* Split out the group into technical and implementation/education and use this call to come back to see how each group is doing – put in survey. Any volunteers to lead each sub-group?
* Running R and Python on Hadoop
	+ Feature Extraction to be moved here.
* SOLR with Hadoop
	+ What are the use case for SOLR that needs to be supported
* Kudu support
	+ Tom - Trying it out and documenting how to do it. Explain why it is needed.
		- Schedule and book some time to work devote full attention/working session

**Educational/Implementation Goals**

* Hadoop Education
	+ OHDSI stack in Hadoop and how to use them
		- Hadoop with CDM
		- Hadoop with Atlas
	+ Hadoop Tutorial – 2019
		- Talk to Frank this year for integrating Hadoop into the Architect Tutorial 2018.
	+ Q&A session – “I have this struggle/problem. Has anyone else had the same problem? If so, let’s talk?”
* Hadoop Documentation
	+ Hadoop configuration
		- Documenting known configurations
	+ Known version of components that work together – chase Greg
* Hadoop Implementation
	+ What is Hadoop really good at? Documentation
		- Bath cohorts
	+ Lessons learned from tools installation or ETL conversions
	+ How to build business case to move over to Hadoop from other environment
* Data Governance and what other kind of data do we want to bring in with Hadoop. Data Lake and demonstrating what that would look like with Observation data and OHDSI