Goal: To get a sense of priority for the group in aggregate both on Hadoop topics to discuss and work toward common sense of solution/best approach, and sense of priority on what coding should and could happen next

Please rank order the following with a number (e.g. 3). Where 1 is your most important, 2 is your second most important. Please do not include, if you can bear it, words like “High” after the number or comments like “I don’t care about this”. Just use 1 through ‘n’.

**What Use Cases are most important to you?**

1. ETL platform – some parts of Hadoop could be used for ETL or to stage ETL for data as it is moving around, even if base data not reposed in Hadoop File System
2. Storage/warehouse – use Hadoop project(s) such as Hadoop File System as a place to store base data, in lieu of Oracle or Postgres
3. Analytics engine – a method for running analytics. One benefit mentioned was speed, i.e. using compute power in the commodity hardware or other underlying hardware in Hadoop system
4. Natural Language Processing - Existing member, in the OHDSI NLP Working Group, in using Hadoop for NLP. Presumably to dump clinical notes in their native format and parse text using open source NLP.
5. OMOP CDM but not for real world evidence - use of CDM solely as a data model, untethered to any metrics that OHDSI/OMOP has today, for anyone with claims or E.H.R. type data
6. Search - SOLR is part of Hadoop ecosystem (based on Lucene which “everyone uses”).
7. Sensor data – Hadoop might be a good fit for the kind of data that comes off sensors or wearables, note, it may not be mature in OMOP CDM today
8. Method for non-SQL approach to tasks – won’t ever be rid of SQL, however Hadoop enables easier use of Java/R/Python others than data
9. Creating pre-aggregated data sets (such as frequently used cohorts or cohorts of interest), there is a need there and could be a benefit. Could tie to Achilles.
10. Other (please describe it in a general way, so I can try to match it with the same write in ‘others’ from other people)

**What should folks with time and coding skills work on to build out next?**

1. Changes to csv file to see if metrics can/system can create Hadoop-centric or Hadoop-OK SQL
2. Windowing functions on approaches and different Hadoop product fit
3. A reference architecture
4. Temp table documentation on approaches and different Hadoop product fit
5. Testing current code of OMOP CDM in Impala
6. Benchmarking current OMOP CDM code for Impala
7. Tuning current OMOP CDM code for Impala
8. Writing OMOP CDM in Hbase
9. Writing OMOP CDM in some other Hadoop centric data store
10. Writing ETL ingest for some common claims-type data source such as HGMD
11. Changing SQL Render in some way to make it more conducive to Hadoop
12. Assessing support for S3 or a cloud reference architecture showing S3
13. Looking at other Hadoop approaches (e.g. Hortonworks, Splice Machines)
14. Other (please write it in a general way to enable aggregation with similar votes)