Hadoop Working Group

July 7, 2017

Attendees:

* Shawn Dolley, John Hope
* John Saunders – Abbvie
* Anthony Reckard – QuintilesIMS
* Malcolm McRoberts – NantHealth
* Mui Van Zandt – QuintilesIMS
* Siva Muthukrishnan – QuintilesIMS
* Lee Evans – LTS
* Gregory Klebanov – Odysseus
* Kirill - Odysseus
* Dave Ficenec – Takeda

<http://www.ohdsi.org/web/wiki/doku.php?id=projects:workgroups:hadoop-wg> shows the following info on future sessions

* July 7th, 2017 - Mui van Zandt & Greg Klebanov lead post mortem of the Hackathon
* July 21th, 2017 - regular meeting
* August 4th, 2017 - no meeting, summer break
* August 18, 2017 - break over, regular meeting, seeking presentations
* September 1, 2017 - planning for Symposium, regular meeting, seeking presentations

\*One note – Mui Van Zandt is presenting to the total Community on Tuesday at noon eastern, connection/dial in info on the OHDSI home page and in the Forum, on the hackathon event.

Reminder: October 18-20 Symposium

Topics requested by the working group for presentations:

* Cloudera present on Oozie, Autosys and Scheduling needs to be presented on one of the 3 meetings before mid Sept. (request from Mui)
* Kudu (request from Malcolm)
* Machine Learning (Greg) (patient-level predict; brainstorm about potential speakers)
* NLP (Malcolm – Charity GA Tech, CHOA could speak on, Linguamatics)
* Data profiling, Data Quality tools, MDM (request from Malcolm)

New business: none

* Greg and Mui gave update on what happened in the hackathon
* Lee Evans gave an update on details and that now Impala is running in the OHDSI cloud and he and Tom are working on standing up ATLAS with the 1% synpuf data set on Hadoop Apache HDFS and OHDSI Atlas and Apache Impala
* Various people talked about the hackathon
* Dialogue on how we will handle the lack of delete support in Apache Impala, either by using Postgres or by using Apache Kudu in the future: seems like Lee, Tom, and Greg have a sort of handle on the current thinking but nothing set in stone, Malcolm has interest and I think we need to have another iteration of that discussion in our WG meeting