**NIST Big Data Public Working Group (NBD-PWG)**

**NBD-PWD-2016/M0525**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of May 3, 2016**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2016\_05\_03 14\_33.rtf**

**Tim Zimmerlin (to Everyone)**: 1:28 PM: I want to know how those students spent their time, what frustrated them most of all.

**Ann Racuya-Robbins (to Everyone)**: 1:29 PM: I would like to participate in that discussion.

**Tim Zimmerlin (to Everyone)**: 1:29 PM: I would welcome their recommendations for improvements, their grading of the SW tools.

**Russell Reinsch (to Organizer(s) Only)**: 1:30 PM: Meeting notes not entered into the appropriate google doc: Tim recommend a gap analysis on requirements.

**David Boyd (to Everyone)**: 1:31 PM: Sorry I am unmuted have to troubleshoot.

**Tim Zimmerlin (to Everyone)**: 1:31 PM: Geoffrey, how about those students writing a 250 to 500 assessment of the state of big data analytics?

**David Boyd (to Everyone)**: 1:31 PM: Yes I think we should look at their applications and see if there are other activities or functional components we missed.

**David Boyd (to Everyone)**: 1:32 PM: Most of their BD Framework elements have already been mapped into the RA. It is the applications where things get interesting.

**Tim Zimmerlin (to Everyone)**: 1:32 PM: Statistically, I would like to know how many students plan to continue in big data...how many never want to do big data ever.

**Mark Underwood (to Everyone)**: 1:34 PM: Apps + orchestration (not to start that debate up again)

**David Boyd (to Everyone)**: 1:34 PM: Microphone just not working,

**David Boyd (to Everyone)**: 1:35 PM: Mark, yeah some of the orchestration is the other component.

**Tim Zimmerlin (to Everyone)**: 1:35 PM: Note that these students (per Geoffrey) are advanced users...far ahead of most users. Stonebraker calls naive users "civilians" in contrast to SW developers.

**Mark Underwood (to Everyone)**: 1:38 PM: URL for Chamleon (sp?) ?

**Tim Zimmerlin (to Everyone)**: 1:40 PM: Mark, https://www.chameleoncloud.org

**Russell Reinsch (to Everyone)**: 1:40 PM: Have to download the app onto your desktop

**Russell Reinsch (to Everyone)**: 1:41 PM: << Ann

**Tim Zimmerlin (to Everyone)**: 1:41 PM: Mark, both Chameleon and CloudLab are NSF follows to FutureGrid (now renamed FutureSystems).

**Tim Zimmerlin (to Everyone)**: 1:42 PM: https://www.cloudlab.us/

**Ann Racuya-Robbins (to Everyone)**: 1:43 PM: File has been sent Wo.

**Tim Zimmerlin (to Everyone)**: 1:46 PM: Wrangler and Comet are two more US grids on Internet2. Both are hybrids of HPC and HTC, very fast, highly scalable.

**Tim Zimmerlin (to Everyone)**: 1:48 PM: Internet2 connects many 100s of US and EU and other foreign campuses. Most all of CERN depends exclusively or extensively on Internet2.

**Tim Zimmerlin (to Everyone)**: 1:49 PM: Regional networks like Chicago's StarLight are connected by Internet2 with IU NOC coordination and SW tool based orchestration.

**Tim Zimmerlin (to Everyone)**: 1:51 PM: These are in place for the four NSF regional hubs.

**Tim Zimmerlin (to Everyone)**: 1:55 PM: Feedback is good!

**Mark Underwood (to Everyone)**: 1:57 PM: Thx Tim

**Tim Zimmerlin (to Everyone)**: 2:11 PM: Yep, VoltDB webinar...

**Mark Underwood (to Everyone)**: 2:13 PM: I also audited that talk

**Tim Zimmerlin (to Everyone)**: 2:14 PM: "analytics" is overloaded. Stonebraker might have been disrespecting traditional statistics ala ANOVA. With machine learning supposedly improving on that.

**Tim Zimmerlin (to Everyone)**: 2:16 PM: Russell, I suggest you google "analysis of variance" or ANOVA. Hopefully, Wikipedia has some good pages.

**Tim Zimmerlin (to Everyone)**: 2:17 PM: Mean, Variance, Standard Deviation, Maximum Likelihood, Neumann / Peason decision theory.

**Tim Zimmerlin (to Everyone)**: 2:18 PM: ...err...Pearson

**Tim Zimmerlin (to Everyone)**: 2:20 PM: ...err...Neyman

**REYLING, ROBERT A CTR USAF AFMC AFLCMC/HN (to Everyone)**: 2:22 PM: Wo begins wrap-up at 2:22 PM

**Tim Zimmerlin (to Everyone)**: 2:22 PM: Criminal data trespass, Criminal data sharing, Civilian liability for either normal or abnormal (e.g., mistakes) data sharing.

**Russell Reinsch (to Organizer(s) Only)**: 2:22 PM: Note to self: provide link to Volt presentation and time of comment about analytics; + context for discussion.

**Tim Zimmerlin (to Everyone)**: 2:23 PM: I little effort...I am not a lawyer.

**Ann Racuya-Robbins (to Everyone)**: 2:25 PM: do you have a link Tim?

**Russell Reinsch (to Everyone)**: 2:27 PM: Where are we seeking feedback on preliminary white paper template? Item 6.b.i

**Mark Underwood (to Everyone)**: 2:28 PM: Tim, recall the RFI proposal from last week - maybe you can contribute some language

**Tim Zimmerlin (to Everyone)**: 2:32 PM: Mark, I do not understand your comment. Let's continue with emails.

**Mark Underwood (to Everyone)**: 2:32 PM: Tim, ack'd

**Tim Zimmerlin (to Everyone)**: 2:33 PM: Ann, I have extensive RSS / Atom