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

**NBD-PWD-2017/M0601**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of Feb. 14, 2017**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2017\_02\_14 15\_34.rtf**

**William Miller (Private)**: 1:14 PM: Have you attended the new IEEE Big Data WG?

**William Miller (Private)**: 1:14 PM: I am planning to join

**David W. Boyd (to Everyone)**: 1:22 PM: Nancy: There is a taxonmy of data organization in the Ref Arch that would be a good start for a taxonomy that covers NOSQL and SQL and File Systems

**Tim Zimmerlin (to Everyone)**: 1:23 PM: Nancy, the current top level application architecture dichotomy is server based orchestration versus serverless orchestration.

**Tim Zimmerlin (to Everyone)**: 1:24 PM: ...near future cluster, including cloud, includes special purpose processors like GPU, low power ARM, FPGA.

**Tim Zimmerlin (to Everyone)**: 1:25 PM: ...in memory SPARK is another key feature.

**Tim Zimmerlin (to Everyone)**: 1:26 PM: ...near real time stream processing versus batch processing is another dichotomy.

**Russ Reinsch (to Organizer(s) Only)**: 1:28 PM: Note to self: compare our taxon w others, relationships.

**Tim Zimmerlin (to Everyone)**: 1:30 PM: Recommendation 1: Hew very close to NIST Cloud Arch & NIST Security Arch.

**Tim Zimmerlin (to Everyone)**: 1:31 PM: Recommendation 2: Start taxonomy with orchestration, including orchestration objects & properties.

**Tim Zimmerlin (to Everyone)**: 1:32 PM: Recommendation 3: Point to full system of systems architecture...with edge clouds, hyperclouds, IoT, Smart vehicles, Smart cities, CPS. Data buffering at layer 3/4.

**Tim Zimmerlin (to Everyone)**: 1:35 PM: Sensors energy budget is primary design limit; however, 100mW is already assigned for radio. This means crypto energy is acceptable to IoT device budget.

**Tim Zimmerlin (to Everyone)**: 1:36 PM: ...sensors will look similar to Arduino, but much smaller.

**Russ Reinsch (to Organizer(s) Only)**: 1:36 PM: Ck researchgate for something pubbed by mark

**Mark Underwood (to Everyone)**: 1:38 PM: https://www.researchgate.net/publication/307122744\_Semantic\_Interoperability\_for\_the\_Web\_of\_Things

**Mark Underwood (to Everyone)**: 1:41 PM: Nancy: Reference is here for the current draft of the cybersec framework... Of course it's in Vol 4 https://www.nist.gov/cyberframework/draft-version-11

**Mark Underwood (to Everyone)**: 1:46 PM: FYI Links to current CSCC documents http://www.cloud-council.org/resource-hub.htm

**Russ Reinsch (to Organizer(s) Only)**: 1:47 PM: Goodier

**Russ Reinsch (to Organizer(s) Only)**: 1:48 PM: CIO.gov taxonomy survey

**Mark Underwood (to Everyone)**: 1:48 PM: Maybe this? https://policy.cio.gov/taxonomies/

**Russ Reinsch (to Organizer(s) Only)**: 1:50 PM: ck pg. 47

**Russ Reinsch (to Organizer(s) Only)**: 1:54 PM: Vol. 2 section 2 somet about search and retrieve

**Cavan Paul Capps (to Everyone)**: 1:55 PM: I would think that HPC and Big Data are different, in that Hadoop is designed more for fault tolerance than performance

**Tim Zimmerlin (to Everyone)**: 2:00 PM: Cavan, not today. Comet & Wrangler are hybrid HPC / HTC.

**Tim Zimmerlin (to Everyone)**: 2:01 PM: ...HPC folks think of "big data" as HTC in their playbook.

**Geoffrey Fox (to Everyone)**: 2:03 PM: i am away for 10 minutes or so , best Geoffrey

**Tim Zimmerlin (to Everyone)**: 2:07 PM: Nancy, brilliant chart. Coupling, loose vs tight. Shared data, nothing to much.

**Tim Zimmerlin (to Everyone)**: 2:08 PM: ...I often start my analysis with ALGORITHMS. Algorithms allow me to see the whole playing field.

**Tim Zimmerlin (to Everyone)**: 2:09 PM: ...identifying algorithms drives subsequent software reuse.

**Russ Reinsch (to Organizer(s) Only)**: 2:09 PM: Nancy: HPDA is high perf data a.

**James And Kay Goodier (to Everyone)**: 2:10 PM: the reference was https://cio.gov/wp-content/uploads/2017/01/CIO-Council-State-of-Federal-IT-Report-January-2017.pdf. there are taxonomies and big data references in this report.

**Cavan Paul Capps (to Everyone)**: 2:11 PM: Or all nodes have GPUs on them

**James And Kay Goodier (to Everyone)**: 2:11 PM: Nancy, I like your model, but I would add storage and utility nodes

**Mark Underwood (to Everyone)**: 2:12 PM: Thanks Kay

**Mark Underwood (to Everyone)**: 2:12 PM: (I love that it's a WordPress site)

**James And Kay Goodier (to Everyone)**: 2:12 PM: Me too!

**Mark Underwood (to Everyone)**: 2:13 PM: I'm not seeing that CIO page from my browser :(

**Cavan Paul Capps (to Everyone)**: 2:13 PM: We are working out how to put up Tensor Flow (using GPUs) on Spark... not sure how well they will integrate

**Tim Zimmerlin (to Everyone)**: 2:14 PM: HPC today is focued / limited to sparse matrix, no fault tolerance with cores or communications.

**James And Kay Goodier (to Everyone)**: 2:14 PM: can you see Cio.gov?

**Tim Zimmerlin (to Everyone)**: 2:15 PM: Yes, been on the Web site. Will read afterwards.

**James And Kay Goodier (to Everyone)**: 2:16 PM: go to Fed IT Updates at top

**Tim Zimmerlin (to Everyone)**: 2:17 PM: HPC today runs at around 3 to 5 percent rated system speed...with real world workloads.

**Tim Zimmerlin (to Everyone)**: 2:18 PM: ...as Gregor mentioned virtual clusters are trending orchestration objects.

**Mark Underwood (to Everyone)**: 2:19 PM: Yes, I cio.gov home is ok

**Gregor (laszewski@gmail.com) (to Everyone)**: 2:19 PM: how do i access the portal?

**Gregor (laszewski@gmail.com) (to Everyone)**: 2:20 PM: is ther a url for that?

**James And Kay Goodier (to Everyone)**: 2:20 PM: https://cio.gov/wp-content/uploads/2017/01/CIO-Council-State-of-Federal-IT-Report-January-2017.pdf

**Mark Underwood (to Everyone)**: 2:21 PM: OK, got it now, thx

**Quyen Nguyen (to Everyone)**: 2:21 PM: my email is: quyen.nguyen@census.gov. Nancy and David, please send me email about Vols 2 and 6. I am glad to participate in contributing to the Vols, as Wo suggested.

**Gregor (laszewski@gmail.com) (to Everyone)**: 2:22 PM: Also anyone has an answer for wher I find the portal?

**Gregor (laszewski@gmail.com) (to Everyone)**: 2:22 PM: e.g. it seems that there is some portal where documents are developed. Where do i find it and how do i get access to it?

**Wo Chang (to Cavan Paul Capps)**: 2:23 PM: Can I talk to you after this telecon?

**James And Kay Goodier (to Everyone)**: 2:24 PM: Mark, I found postings of recent audits of privacy - very enlightening. How can I share these with you?

**Mark Underwood (to Everyone)**: 2:25 PM: James and Kay - Email, or we can schedule a phone call if you think that is better mark.underwood@kryptonbrothers.com

**James And Kay Goodier (to Everyone)**: 2:35 PM: thanks

**Wo Chang (to Cavan Paul Capps)**: 2:35 PM: bigdataarch@nist.gov

**Wo Chang (to Everyone)**: 2:36 PM: bigdataarch@nist.gov

**Tim Zimmerlin (to Everyone)**: 2:41 PM: Gregor, the old / new NIST Big Data Web portal is https://bigdatawg.nist.gov

**Tim Zimmerlin (to Everyone)**: 2:42 PM: ...this Web portal only has old version 1 documents so far, but new documents will be uploaded soon.

**Russ Reinsch (to Everyone)**: 2:55 PM: Time for me to check out. Cheers -

**Cavan Paul Capps (to Everyone)**: 2:58 PM: Kay, can you share the privacy docs with me as well?

**Ann Racuya-Robbins (to Everyone)**: 3:00 PM: Kay I would be interested in the doc as well.

**Robert Reyling (to Everyone)**: 3:00 PM: As would I!

**Cavan Paul Capps (to Everyone)**: 3:01 PM: I am working with some folks at the privacy tools center on new developments and working to get a comprehensive lit review

**Tim Zimmerlin (to Everyone)**: 3:03 PM: Very good meeting! Be back next week...

**Mark Underwood (to Everyone)**: 3:05 PM: Example startup.... @aporeto application segmentation for containers

**Ann Racuya-Robbins (to Everyone)**: 3:06 PM: Thank you all.

**Gregor (laszewski@gmail.com) (to Everyone)**: 3:07 PM: ann please send me your microsoft onedrive account and i add you to the document

**Wo Chang (to Everyone)**: 3:08 PM: wchang@nist.gov

**Anil Srivastava (Private)**: 3:23 PM: Will a recording of the webconference and/or presentations be archived?