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

**NBD-PWD-2015/M0464**

**Source: NBD-PWG**

**Status: Draft**

**Title: Web chat from Meeting of Nov. 03, 2015**

**Chat Log D:\\_wo\1DMG\2015\\_BigDataWG\Docs\ChatLog 2015\_11\_03 15\_07.rtf**

**Cavan Capps (to Everyone)**: 1:17 PM: When is the Big Data Conference?

**Tim Zimmerlin (to Everyone)**: 1:19 PM: Northeast Big Data Innovation Hub link: http://northeastbdhub.dsi.columbia.edu

**Tim Zimmerlin (to Everyone)**: 1:20 PM: South BD Hub link: http://southbdhub.org

**Tim Zimmerlin (to Everyone)**: 1:21 PM: Midwest Big Data Hub link: http://midwestbigdatahub.org

**Tim Zimmerlin (to Everyone)**: 1:22 PM: West BD Hub has a broken link.

**Nancy Grady (to Everyone)**: 1:27 PM: FYI, the south hub contains VA, MD, DC

**Tim Zimmerlin (to Everyone)**: 1:28 PM: Nancy, note the listed member orgs for the South Hub.

**shazri (to Everyone)**: 1:28 PM: correct, ETL plays heavy role...especially when making 'technical' decisions....'technical' decisions are usually time sensitive

**shazri (to Everyone)**: 1:29 PM: 'tactical'/'strategic' decision more forgiving at unsync time scales

**shazri (to Everyone)**: 1:30 PM: ETL!

**Tim Zimmerlin (to Everyone)**: 1:30 PM: ETL last or never!

**shazri (to Everyone)**: 1:31 PM: :)

**Tim Zimmerlin (to Everyone)**: 1:31 PM: AWS S3, OpenStack Swift & Ceph, and more are object stores, optimized for raw ingested event messages.

**shazri (to Everyone)**: 1:31 PM: noted ... will check on that

**Mark Underwood (to Everyone)**: 1:31 PM: All - RE ETL, Theme #1 is of course to identify how ETL is different from non BigData approaches. I have written about this vis a vis Syncsort DMX-h (shameless plug) which is Hadoop ETL

**Tim Zimmerlin (to Everyone)**: 1:32 PM: Shazri, please study W3C offerings, esp XML and Schema application languages.

**shazri (to Everyone)**: 1:33 PM: Thanks TIm for the heads up....

**Tim Zimmerlin (to Everyone)**: 1:33 PM: Shazri, each XML document is a self defining object.

**Tim Zimmerlin (to Everyone)**: 1:34 PM: Each XML document is optimized for metadata incl provenance metadata.

**shazri (to Everyone)**: 1:36 PM: Just googled it, i believe it will be helpful

**shazri (to Everyone)**: 1:36 PM: Thanks!

**Tim Zimmerlin (to Everyone)**: 1:36 PM: Mark, I anticipate that the role of ETL is changing because the data flows have changed. E.g., aggregating a set of basic event msgs/records into a larger document.

**William Miller (to Everyone)**: 1:37 PM: Wo may I make a comment!

**Tim Zimmerlin (to Everyone)**: 1:41 PM: Cavan, the Eclipse Foundation already has a large IoT effort with about 20 projects.

**Mark Underwood (to Everyone)**: 1:44 PM: Cavan - that use case (real time feeds) would be worth fleshing out

**William Miller (to Everyone)**: 1:44 PM: Wo I would like to make a session that Cognitive intelligence should be built into the Big Data systems. This is not Data Anlaytics since domain =knowledge must be built-in so the system to learn to differentiate without necessarily having to have human intervention.

**Cavan Capps (to Everyone)**: 1:44 PM: Mark, thanks

**Bob Marcus (to Everyone)**: 1:44 PM: NSF Big Data Hub Webinar on Thurs Nov 5 at 11:00 EST See https://bdhub.info/

**William Miller (to Everyone)**: 1:45 PM: I would also like to advise the groyp on a new capability from IEEE RA called vTEDS (Virutal Transducer Electronic Data Sheets) whcih defines the characteristics of devices for the Internet of Things soon to be avaible from IEEE.

**Tim Zimmerlin (to Everyone)**: 1:45 PM: You need a WebEx password to join: bdhubs

**Wo Chang (to Everyone)**: 1:52 PM: The ATRAC Federal Big Data conf: http://www.fedsummits.com/big-data/december-2015/agenda/

**Mark Underwood (to Everyone)**: 1:54 PM: Off Ann's topic: DataTorrent has open sourced their unified stream and batch platform - to compete w/ Spark and Storm - See Apache Apex

**Mark Underwood (to Everyone)**: 1:55 PM: https://www.datatorrent.com/project-apex/

**Tim Zimmerlin (to Everyone)**: 1:57 PM: pregnant data?!? fecund data?!?

**Tim Zimmerlin (to Everyone)**: 2:00 PM: just data, not "big data" exclusive of small datasets.

**Frank Farance (to Everyone)**: 2:02 PM: datum (pl., data or datums)

designation whose concept is a value

datum (pl., data or datums)

designation whose concept is a value

**Frank Farance (to Everyone)**: 2:04 PM: value 2

value concept

concept with a defined notion of equality to that concept

**Frank Farance (to Everyone)**: 2:06 PM: NOTE Although any signifier can designate a value, typically representation systems are used to afford computability among signifiers because digital computers are signifier processors, not concept processors. For example, a decimal positional numeration system (289 = 2×102 + 8×101 + 9×100) is used by hand calculators and a binary position numeration system (10001 = 1×24 + 0×23 + 0×22 + 0×21 + 1×20) is used by modern digital computers - both position numeration systems afford automated arithmetic computation by employing a pre-determined set of rules for processing signifiers (not processing concepts).

**Frank Farance (to Everyone)**: 2:07 PM: EXAMPLE Given the concepts red (defined as "visible light in the range of wavelengths 650 to 700 nanometers") and yellow (defined as "visible light in the range of wavelengths 525 to 570 nanometers"), it is undefined whether red equals yellow and, hence, these concepts by themselves are not values. These concepts can become values once a notion of equality is defined (equality can be defined within the concepts' definition or defined external to the concepts' definition). If the notion of equality is defined as "has visible electromagnetic radiation?", then one can ask whether red equals yellow (true) and whether red equals infrared (false, because infrared is not visible). If the notion of equality is defined as "has overlapping ranges of wavelengths?", then red does not equal yellow (their wavelengths don't overlap) but cornflower equals blue (cornflower is a slice of the blue portion of the visible spectrum). If the notion of equality is defined as "has the same range of wavelengths?", then none of red

**Bob Marcus (to Everyone)**: 2:12 PM: Comparison of Apex Streaming mentioned by Mark with Storm and Spark at https://www.datatorrent.com/project-apex/

**Tim Zimmerlin (to Everyone)**: 2:13 PM: I agree with Frank in general, but I often ignore the distinctions.

**Tim Zimmerlin (to Everyone)**: 2:14 PM: As far as Ann's question: data resides inside computers AND knowledge resides inside human minds.

**Michael Valivullah (to Everyone)**: 2:16 PM: Different levels of abstraction.Data is ato

**Wo Chang (to Everyone)**: 2:17 PM: I am back.

**Michael Valivullah (to Everyone)**: 2:17 PM: Data is atomic.

**Michael Valivullah (to Everyone)**: 2:19 PM: Data is atomic. Information is inference from data. Knowledge is next level based on expereince and information. Wisdon is next level. An expert has it.

**Tim Zimmerlin (to Everyone)**: 2:33 PM: I agree with Frank because the sentence emphasizes "cluster".

**Russell Reinsch (to Everyone)**: 2:34 PM: Take up a vote

**Russell Reinsch (to Everyone)**: 2:35 PM: Im surprised that NIST does not use some technique or s/w for arriving at consensus. Delphi, etc.

**Tim Zimmerlin (to Everyone)**: 2:35 PM: Frank is correct here because the sentence assumes an upgrade has occurred from traditional to a new architecture.

**Tim Zimmerlin (to Everyone)**: 2:37 PM: Frank is correct here because editorializing is outside the scope of the NBD-PWG.

**Tim Zimmerlin (to Everyone)**: 2:44 PM: Ann, people agree on the values you espouse; however, big data is new technology purposed for good and bad ends.

**William Miller (to Everyone)**: 2:47 PM: http://fuzzieee2015.org/wp-content/uploads/2014/10/A-Special-Session-FUZZIEEE2015-Jie-Lu.pdf

**William Miller (to Everyone)**: 2:47 PM: Please take a look at this link

**Tim Zimmerlin (to Everyone)**: 2:48 PM: William, just did. Yep...

**Mark Underwood (to Everyone)**: 2:59 PM: Ann and Nancy = Heart

**Robert Reyling (to Everyone)**: 3:03 PM: concur with Mark's applause for effort!

**shazri (to Everyone)**: 3:05 PM: thanks!