NIST Big Data WG - Definition and taxonomy

Meeting notes Chat

August 26, 2013

* Presentation on Data Scientist (Document #169)
* Roles and Actors discussion – can we breakout key responsibilities and roles
* Taxonomy development – Nancy’s diagram data lifecycle process to be added
* Provenance, Policy, security, etc need to be added
* Actor and transformers for the Data Scientist – components – data source, data collector, and producers ant transformers and data infrastructure and capability management– need a group consensus on the diagram
* For each of the roles what are the major functions and responsibilities (Doc M0072 and M0142)
* ACTION ITEMS:
* Need to dive into roles and responsibilities next week – everyone need to review the existing comments and provide the info on roles and responsibilities
* Additional docs – Teradata and the NIST Big Data Presentations
* Meeting is still on next week

Participants and the On-line Chat:

Karen

Natasha

Sanjay

8:01 AM) Sanjay Mishra(Verizon) joined.

(8:02 AM) \_USER\_NAME\_(\_COMPANY\_) joined.

(8:04 AM) Pw Carey, Compliance Partners, LLC joined.

(8:05 AM) Orit Levin (Microsoft) joined.

(8:07 AM) Karen G: yes Nancy, we can hear you.

(8:08 AM) Karen G: Though, the presentation area is currently blank.

(8:09 AM) Pw Carey, Compliance Partners, LLC: We're experiencing technical difficulities....so we'll log out and log back in....(wish us luck....)

(8:09 AM) \_USER\_NAME\_(\_COMPANY\_)82 joined.

(8:09 AM) Pw Carey, Compliance Partners, LLC disconnected.

(8:09 AM) \_USER\_NAME\_(\_COMPANY\_)82 disconnected.

(8:09 AM) Nancy Grady (SAIC) joined.

(8:10 AM) Karen G: Nancy also appears to have disconnnected. I've not had any connectivity issues during these calls - via FiOS. :)

(8:10 AM) Karen G: Nancy also appears to have disconnnected. I've not had any connectivity issues during these calls - via FiOS. :)

(8:10 AM) Karen G: Hi, Nancy!

(8:10 AM) \_USER\_NAME\_(\_COMPANY\_)94 joined.

(8:11 AM) \_USER\_NAME\_(\_COMPANY\_)94 disconnected.

(8:11 AM) Wo Chang (NIST) joined.

(8:12 AM) Karen G: Correct - cannot see the preso

(8:12 AM) Pw Carey, Compliance Partners, LLC joined.

(8:12 AM) Bill Mandrick joined.

(8:12 AM) Bill Mandrick disconnected.

(8:13 AM) Karen G: Can see desktop view now.

(8:13 AM) Pw Carey, Compliance Partners, LLC: Hurray......&....not a problem....

(8:14 AM) Wo Chang (NIST): Yes, I could see your screen.

(8:14 AM) William Vorhies (Predictive Modeling, LLC.) joined.

(8:14 AM) Bill Mandrick (Data Tactics) joined.

(8:17 AM) Pw Carey, Compliance Partners, LLC: If "\_USER\_NAME\_(COMPANY\_) wants to include their name and company, just log out and log back in...then fill in the text block widows...and Select NEXT... that should do it.....Pw

(8:19 AM) Karen G: For anyone who's just joined, Natasha is reviewing this document: http://bigdatawg.nist.gov/\_uploadfiles/M0169\_v1\_9072641833.pdf

(8:19 AM) Orit Levin (Microsoft) disconnected.

(8:21 AM) Pw Carey, Compliance Partners, LLC: We're still viewing the first page...Agenda....is this correct...?

(8:23 AM) Orit joined.

(8:24 AM) Orit: I don't see any presentation... just the agenda...

(8:24 AM) Karen G: Hi Orit, the presentation is at http://bigdatawg.nist.gov/\_uploadfiles/M0169\_v1\_907264183

(8:25 AM) Karen G: doc 169

(8:25 AM) Karen G: Natasha is speaking, I believe

(8:25 AM) Orit: Thanks!

(8:25 AM) Karen G: yw!

(8:26 AM) Nancy Grady (SAIC) disconnected.

(8:27 AM) Karen G: O’Reilly Strata Survey is very interesting!

(8:29 AM) Bob Marcus (ET-Strategies) joined.

(8:30 AM) Nancy Grady (SAIC) joined.

(8:32 AM) Karen G: I'm interested in to what extent the skills required are industry-specific. I would think that a data scientist in [for example] healthcare, would need to have an excellent understanding of the healthcare industry.

(8:33 AM) Karen G: or, work very closely with a knowledgeable clinician / business data expert.

(8:32 AM) Karen G: I'm interested in to what extent the skills required are industry-specific. I would think that a data scientist in [for example] healthcare, would need to have an excellent understanding of the healthcare industry.

(8:33 AM) Karen G: or, work very closely with a knowledgeable clinician / business data expert.

(8:38 AM) Pw Carey, Compliance Partners, LLC: Are you going to include 'data provenance'.....?

(8:39 AM) Karen G: Comment: I found the definitions of volume, velocity, etc. very clear.

(8:41 AM) Karen G: Volume, Velocity, Variety, Veracity, that is. I didn't catch 'Viscosity', either. :)

(8:41 AM) Karen G: Yes, there do seem to be a growing number of "V's".

(8:44 AM) Pw Carey, Compliance Partners, LLC: Is there any chance you will include data forensics and audits for risk and fraud....?....Respectfully yours, Pw

(8:44 AM) Karen G: Nancy, Latency does make sense.

(8:45 AM) Pw Carey, Compliance Partners, LLC: And we believe this points to the gaps between BD/Hadoop and OLAP/SOAP big iron client server architectures....yes....?

(8:46 AM) Pw Carey, Compliance Partners, LLC: Gaps that are being worked on....somewhere....

(8:47 AM) William Vorhies (Predictive Modeling, LLC.): Still can only see the Learning and Training Opptys doc on screen

(8:47 AM) William Vorhies (Predictive Modeling, LLC.) disconnected.

(8:47 AM) William Vorhies (Predictive Modeling, LLC.) joined.

(8:49 AM) Bill Mandrick (Data Tactics) disconnected.

(8:50 AM) William Vorhies (Predictive Modeling, LLC.): What's the doc number again - can't see it but will open it separately.

(8:50 AM) Wo Chang (NIST): M0142

(8:51 AM) Bob Marcus (ET-Strategies): Check out http://en.wikipedia.org/wiki/Data\_science

(8:56 AM) Pw Carey, Compliance Partners, LLC: For Mr. Vorhies....M0169...

(9:07 AM) Wo Chang (NIST) disconnected.

(9:07 AM) Pw Carey, Compliance Partners, LLC: We will check it out...Re: Data Science....thanks...Pw

(9:07 AM) Wo Chang (NIST) joined.

(9:13 AM) Karen G: Would it help to revisit the current charter for this workgroup - in order to close out the definition & taxonomy?

(9:14 AM) William Milelr (MaCT USA) joined.

(9:14 AM) William Vorhies (Predictive Modeling, LLC.): can't see it

(9:14 AM) Pw Carey, Compliance Partners, LLC: What's your email address, and we'll send the corrected USE CASE copy along to you......and thank you...

(9:14 AM) Karen G: We still see doc 169 - Coursera?

(9:14 AM) William Vorhies (Predictive Modeling, LLC.): No, can't see it

(9:15 AM) Wo Chang (NIST) disconnected.

(9:15 AM) Pw Carey, Compliance Partners, LLC: We just see pg. 23....

(9:15 AM) Pw Carey, Compliance Partners, LLC: Of the 35 slides....

(9:16 AM) Orit disconnected.

(9:16 AM) Orit joined.

(9:17 AM) Orit disconnected.

9:17 AM) aa joined.

(9:19 AM) natasha (ucsd): Info on Veracity and Viscosity:

(9:19 AM) natasha (ucsd): http://nosql.mypopescu.com/post/19256376891/viscosity-and-virality-beyond-the-three-vs-of-big

(9:20 AM) Mark Underwood (Krypton Brothers) joined.

(9:20 AM) natasha (ucsd): http://www.ami-partners.com/blog/2012/12/06/the-4-dimensions-of-big-data/

(9:21 AM) natasha (ucsd): http://www.forbes.com/sites/raywang/2012/02/27/mondays-musings-beyond-the-three-vs-of-big-data-viscosity-and-virality/

(9:25 AM) natasha (ucsd): 10 V's of Big Data - http://translate.google.com/translate?hl=en&sl=de&u=http://www.bi-news.de/%3Fp%3D255&prev=/search%3Fq%3Dbig%2Bdata%2Bviscosity%26safe%3Dactive%26client%3Dfirefox-a%26hs%3DU0z%26rls%3Dorg.mozilla:en-US:official

(9:25 AM) William Milelr (MaCT USA) disconnected.

(9:26 AM) Karen G: Nancy yes, and the data producer might be an individual, process, or system.

(9:26 AM) William Miller (MaCT USA) joined.

(9:27 AM) Orit joined.

(9:27 AM) Orit disconnected.

(9:27 AM) William Vorhies (Predictive Modeling, LLC.) disconnected.

(9:28 AM) Karen G: Also really like where you're 'going' with synching up with the ref architecture definitions. Wish I could see the diagram you're discussing. :)

(9:28 AM) William Miller (MaCT USA): Version 29.0.1547.57 m

(9:28 AM) William Miller (MaCT USA): there is a new version of Google Chrome that fixs a problem with Megagmeeting access

(9:28 AM) Karen G: Also really like where you're 'going' with synching up with the ref architecture definitions. Wish I could see the diagram you're discussing. :)

(9:28 AM) William Miller (MaCT USA): Version 29.0.1547.57 m

(9:28 AM) William Miller (MaCT USA): there is a new version of Google Chrome that fixs a problem with Megagmeeting access

(9:31 AM) William Miller (MaCT USA): http://www.asterdata.com/bridging-the-gap-video/

(9:31 AM) William Miller (MaCT USA): Thi si the link to the Hybrid model from Teradata

(9:34 AM) William Miller (MaCT USA): http://csrc.nist.gov/groups/SMA/forum/documents/june2012presentations/fcsm\_june2012\_cooper\_mell.pdf

(9:34 AM) William Miller (MaCT USA): Big Data Presenation from NIST

9:25 AM) William Milelr (MaCT USA) disconnected.

(9:26 AM) Karen G: Nancy yes, and the data producer might be an individual, process, or system.

(9:26 AM) William Miller (MaCT USA) joined.

(9:27 AM) Orit joined.

(9:27 AM) Orit disconnected.

(9:27 AM) William Vorhies (Predictive Modeling, LLC.) disconnected.

(9:28 AM) Karen G: Also really like where you're 'going' with synching up with the ref architecture definitions. Wish I could see the diagram you're discussing. :)

(9:28 AM) William Miller (MaCT USA): Version 29.0.1547.57 m

(9:28 AM) William Miller (MaCT USA): there is a new version of Google Chrome that fixs a problem with Megagmeeting access

(9:31 AM) William Miller (MaCT USA): http://www.asterdata.com/bridging-the-gap-video/

(9:31 AM) William Miller (MaCT USA): Thi si the link to the Hybrid model from Teradata

(9:34 AM) William Miller (MaCT USA): http://csrc.nist.gov/groups/SMA/forum/documents/june2012presentations/fcsm\_june2012\_cooper\_mell.pdf

(9:34 AM) William Miller (MaCT USA): Big Data Presenation from NIST

(9:41 AM) Bob Marcus (ET-Strategies): The NIST Presentation is at M0065

(9:45 AM) Mark Underwood (Krypton Brothers): That language works for me - I was thinking message broker or agent terminology for that

(9:45 AM) Pw Carey, Compliance Partners, LLC: Dear Mr. Miller...thank you....

(9:46 AM) Bob Marcus (ET-Strategies): Roles are at M0072

(9:46 AM) Mark Underwood (Krypton Brothers): ack'd

(9:46 AM) Pw Carey, Compliance Partners, LLC: Thanks...we'll review M0072...

9:45 AM) Pw Carey, Compliance Partners, LLC: Dear Mr. Miller...thank you....

(9:46 AM) Bob Marcus (ET-Strategies): Roles are at M0072

(9:46 AM) Mark Underwood (Krypton Brothers): ack'd

(9:46 AM) Pw Carey, Compliance Partners, LLC: Thanks...we'll review M0072...

(9:48 AM) Pw Carey, Compliance Partners, LLC: We should review M0142...yes...&...we'll have a Monday Meeting on the 2nd...yes..?

(9:49 AM) Pw Carey, Compliance Partners, LLC: Pizza for everyone.....

(9:50 AM) Karen G: Hah. Virtual izza.

(9:50 AM) Karen G: pizza

(9:50 AM) Pw Carey, Compliance Partners, LLC: low cal....

(9:51 AM) Karen G: veggie plz

(9:51 AM) Pw Carey, Compliance Partners, LLC: Thank you...

(9:51 AM) Pw Carey, Compliance Partners, LLC: This is a 'veggie free zone'

(9:52 AM) aa: Thanks Nancy!

(9:52 AM) Karen G: Thanks Nancy!

(9:52 AM) Bob Marcus (ET-Strategies) disconnected.

(9:52 AM) William Miller (MaCT USA) disconnected.