NIST Big Data WG meeting

Big Data Definitions and taxonomy subgroup

September 9th, 2013 Agenda

Participants:

Nancy Grady, Wo, William Vohries, Bruno Kelpasas, Bob Marcus, Mark Underwood, Sanjay Misra , Deborah Blackstock, Yui Demchenko (Uva), Natasha Balac, William Miller (Mact USA), Orit, Pw Carey, Pavithhra Kenjige, Gary Mazzfeerro

Agenda

Review of the current document- M0226. - conceptual architecture. The Def&Tax group has synched with Infrastructure Group. Conceptual model of the Big data reference architecture Figure 1 in the Doc # 226. Each box presents a taxonomy role. Data Provider and data consumer

Activities with a data role provider identified: Data consumer, added data steward and data governor



Data Scientist activities are transform activities.

Data Architect pushing out requirements for the system.

Data Modeler – translate logical representation of data into physical representation presented inside capabilities

Data architect and modeler the only ones who have responsibilities that have changed due to big data ( for example data is split across resources)

Different actors can perform different roles (Just like in the movies)

Roles are first level – system manger: could have many different actors – performing that and other roles. Start with a role there is a role for each of the boxes For each box there are many activities – each activities can be perform by one or many actors performing different activities;

How far do we go with this hierarchy?

Many activities in the role of a system manager for example – it is within the role that a bunch of activities happen – should specify the role for a person who is generating security requirements – maybe not an exhausting list but types of activities on a high level

Data provider and consumer could be external or internal to the organization

Data provider – doe they include data source(bits) ?

Activities are the responsibilities of a data provider is to establish API to allow access to data; formal or informal contract to access the data. Could be a passive (ftp) site to use API or could also play a more active role – can negotiate to push data to downstream actors; either push or pull activities are done by data provider.

Some data provider might have a rich data set, provides API. etc – or other might provide http, ftp or other simpler ways

Transformation role

Take requirements, specs from architect – do the development and testing move into operational environment and production;

Activities to establish SW for data curation, analytics, viz, etc.

Security protocols are specified by security

Data Governor – is a better name than system manager

The key for Big Data is in the relationship “--🡪” between the Transformation and capabilities. This is where Big data is truly changed what and how we do. For example if data storage is spread out and we need to carry out the analytics services and need to reach and reduce the data in order to analyze a large amount of data.

Additionally System manager role is important and has changed – needs to improve the communication between the role and actors. It has a new role in orchestration of the big data (for example # of notes needed to use for analytics)

Transformation arrow “🡪” to data provide is it really needed? System Manager would take that role.

Discussion on terminology Gary has send in the email recently; Adding the clarification to some of the terminology. Not focusing on deployment a model and business engagement. This approach is more inclined to responsibility of big data and how it fits in big data – we need to make a consensus as a group. Technical vs. business discussion. Responsibilities of implementing the capabilities should be added; Referring back to the Mindmeister – brainstorming discussion –

Maybe we can add suggested capabilities and attributes/ activities for performing the role of data scientist activities. We should serve to as a guide to the community. Provide terminology and advice how to go down the big data road. Including conceptual technical guidance is important.

We might want to create a few templates in addition to the use cases which are helpful but not granular or specific enough. Prescriptive vs. descriptive guidance for the readers and users.

Big Data Roadmap discussion update - Readiness of big data – Bruno joint the conversation

3 major groups: Architecture, Transformation, Business

These roadmap features were aligned with the roadmap architecture, use cases and actors, taxonomy and other groups; Document M87 is the latest version of the working document –

Nancy brought up the fact that we still have open questions – doe we want to change the ordering of taxonomy levels;

On-line chat notes:

(11:07 AM) Orit joined.

(11:08 AM) William Vorhies (Predictive Modeling, LLC.): Nancy:  No on screen image - blank screen.

(11:08 AM) Pw Carey-\*Compliance Partners, LLC joined.

(11:09 AM) Pw Carey-\*Compliance Partners, LLC: Sorry we're late...&....Good morning....Respectfully yours, Pw

(11:09 AM) Deborah Blackstock (MITRE) joined.

(11:10 AM) Yuri Demchenko (UvA) joined.

(11:11 AM) natasha ucsd joined.

(11:11 AM) Pw Carey-\*Compliance Partners, LLC: Roles & Responsibilities of the Data Consumer...has different responsibilities depending upon the Cloud Service being provided....(aka: multiple duties and responsibilities)...no?

(11:12 AM) William Miller (MaCT USA) joined.

(11:16 AM) Pw Carey-\*Compliance Partners, LLC: Do we include Audit/Auditors needs in this architecture....?

(11:16 AM) Orit: Thank you, Nancy! Great recap.

(11:16 AM) Pw Carey-\*Compliance Partners, LLC: Just asking.....Pw

(11:20 AM) Orit: Works for me too.

(11:21 AM) Pw Carey-\*Compliance Partners, LLC: So, we're not leaving any requisite party out of this architecture....yes...?

(11:21 AM) Pw Carey-\*Compliance Partners, LLC: High level....via Subgroups...via Low Level....ok....?

(11:22 AM) PavithraKenjige (PK Technologies) joined.

(11:22 AM) Orit: Yes, I agree.

(11:22 AM) Pw Carey-\*Compliance Partners, LLC: Do we need to differentiate between the Cloud Services offered...IaaS, PaaS, & SaaS....?

(11:23 AM) Pw Carey-\*Compliance Partners, LLC: & their impact/influence upon the duties and roles of all concerned parties....?

(11:24 AM) Orit: Push or Pull are implementation details of the APIs.

(11:25 AM) Pw Carey-\*Compliance Partners, LLC: Dear Mr. Orit...can you expand a bit on this....please...?

(11:26 AM) Pw Carey-\*Compliance Partners, LLC: From our perspective...APIs are a security & privacy risk....all the time....yes?

(11:31 AM) Pw Carey-\*Compliance Partners, LLC: Nope we don't see Keith in the Public Chat Room....

(11:33 AM) Orit: To add to the answer to Keith, both Data and Usage services include "publishing" APIs / means to publish what data and what means to access it exist as a part of their service.

(11:36 AM) Orit: Governer has enterprise connotation. I would prefer "orchestrator".

(11:37 AM) Orit: This depends on the system: enterprise vs. loosly coupled

(11:38 AM) Mark Underwood (Krypton Brothers): Agree - orchestrator has that Biztalk / Apache ODE flavor

(11:40 AM) Pw Carey-\*Compliance Partners, LLC: Is the data at risk....?

(11:41 AM) Orit: They may be very different even just "virtual".

(11:41 AM) gary mazzaferro joined.

(11:42 AM) Orit: Still, it is an interface that shoud be shown and described (even if it is not implemented by SW)

(11:42 AM) Pw Carey-\*Compliance Partners, LLC: We agree with Mr. Orit....

(11:43 AM) gary mazzaferro: Nancy has it correct

(11:44 AM) Orit: Ms. Orit ;-)

(11:44 AM) gary mazzaferro: The Transform from the applications perspective

(11:45 AM) Yuri Demchenko (UvA): I have a difficulty to map current tranformation provider (services inside the box) and capability provider to real life existing companies. Do we need to have this?

(11:45 AM) Pw Carey-\*Compliance Partners, LLC: Oops.....Well madame....we still agree with you, regardless of how difficult that is for us to say.....Respectfully yours, Pw Carey, Sir...

(11:46 AM) Orit: Nancy is correct.

(11:46 AM) gary mazzaferro: I have recommended to remove "provider" from the diagram

(11:47 AM) Orit: That's waht the collection block is about.

(11:47 AM) Pw Carey-\*Compliance Partners, LLC: Are they 'Providing' a Service....?

(11:50 AM) Orit: I tend to agree with Nancy!

(11:51 AM) Yuri Demchenko (UvA): To GM: I would agree. In this case we probably need have another diagram to show relations services-capabilities and real life providers/actors.

(11:51 AM) Yuri Demchenko (UvA): Best way is not to mix different type of views.

(11:51 AM) Pw Carey-\*Compliance Partners, LLC: Would 'Data Owner' screw everything up...?

(11:51 AM) Pw Carey-\*Compliance Partners, LLC: Plusss....we like your 'KEY'.....Respectfully yours, Pw

(11:53 AM) William Vorhies (Predictive Modeling, LLC.) disconnected.

(11:54 AM) Mark Underwood (Krypton Brothers): Slightly different subject: are we going to reconcile the mindmap with this Stack/Value Chain overview. The mindmap has more detail which I find helpful.

(11:57 AM) Orit: These are technical roles, not business roles.

(11:59 AM) Orit: Yes.

(11:59 AM) Bob Marcus (ET-Strategies): I believe that it will be difficult for readers to understand the Taxonomy and RA without some simple Use Case examples

(12:00 PM) Pw Carey-\*Compliance Partners, LLC: Dear Bob....we agree with your opinion....

(12:00 PM) Mark Underwood (Krypton Brothers): Bob - the mindmap has examples on its peripheries

(12:00 PM) Pw Carey-\*Compliance Partners, LLC: The less open to interpretation the better....no?

(12:02 PM) Orit: These are not "deployment models".

(12:03 PM) Orit: These are "technical roles".

(12:03 PM) Pw Carey-\*Compliance Partners, LLC: If we know that....does our audience know that....?

(12:03 PM) Orit: If we replace roles with the same "capabilities, the architecture wouldn't cahnge a bit.

(12:06 PM) Orit: The documents will explain that this is a Tech view, not Business view.

(12:06 PM) Orit: Actors (that are not on the figure) are more related to Business view.

(12:08 PM) Pw Carey-\*Compliance Partners, LLC: Wouldn't that be addressed via a Glossary...?

(12:10 PM) Bob Marcus (ET-Strategies): Some simple Use Cases showing the interactions among blocks and abstractions are essential for reader understanding!!

(12:12 PM) Orit: Bob, we hear you. I am adding it to the "Open Issues" table. My thought is to suggest to have another White paper to collect them.

(12:13 PM) Pw Carey-\*Compliance Partners, LLC: That sounds reasonable to us.....Pw

(12:14 PM) Pw Carey-\*Compliance Partners, LLC: But we are acting as a point of knowledge/trust....yes...?

(12:15 PM) Mark Underwood (Krypton Brothers): Orit - that was the approach taken in the Sec and Priv group - it's a separate section. Tho mapping it cleanly to the canonical diagrams is nontrivial.

(12:16 PM) Orit: Mark, good point! Thanks!

(12:19 PM) Pw Carey-\*Compliance Partners, LLC: Good points...Madame....

(12:25 PM) Pw Carey-\*Compliance Partners, LLC: How do I get started...where should I get started....what do I need to understand to make a good decision for my organization.....yes...?

(12:25 PM) Orit: Let's try to agree on the Tax. Let's have the other discussions (roadmap, refarch, etc.) calls. We need the Taxonomy be nailed down...

(12:26 PM) Pw Carey-\*Compliance Partners, LLC: You all may not realize it, but this organization is being used as a pseudo certification authority....by all sorts of folks...who have no connect with this organization....true story...Respectfully yours, Pw

(12:27 PM) Pw Carey-\*Compliance Partners, LLC: We agree with Ms. Orit....too

(12:27 PM) Mark Underwood (Krypton Brothers): yes, given short timeframe remaining as well

(12:27 PM) Orit: Zooming into the Transformation Block is a discussion to have in the RefArch call...probably...

(12:28 PM) Bob Marcus (ET-Strategies): How are readers expected to use the Taxonomy and Reference Architecture?

(12:29 PM) Orit: Bob, let's fisrt make it clear to us, then we write the manual/guide to the readers.

(12:29 PM) Orit: as you suggested earlier.

(12:31 PM) Pw Carey-\*Compliance Partners, LLC: Correct....several examples will help the reader/user/audience....

(12:32 PM) Bob Marcus (ET-Strategies): I think that looking at these deliverables from a readers perspective would help clarify the discussion by making it more pragmatic and less philosophical

(12:32 PM) Pw Carey-\*Compliance Partners, LLC: All good points....Pw

(12:33 PM) Pw Carey-\*Compliance Partners, LLC: Are being suggested by the current Speaker....

(12:33 PM) Mark Underwood (Krypton Brothers): Yeah, Bruno - My Cx thought they were ready for BigData but had no master data baseline in place

(12:36 PM) Pw Carey-\*Compliance Partners, LLC: Would that translate into a Template for each Use Case....?

(12:36 PM) Orit: I agree with the "template" line of thinking.

(12:37 PM) Pw Carey-\*Compliance Partners, LLC: For each Use Case seems a bit much.....Pw

(12:39 PM) Mark Underwood (Krypton Brothers): To Wo: Does the mindmap have versioning - in case you want to roll back what we contribute?

(12:39 PM) Pw Carey-\*Compliance Partners, LLC: Great idea...the use of a few templates...right up there with the use of a KEY....Respectfully yours, Pw

(12:41 PM) Pw Carey-\*Compliance Partners, LLC: Bless you.....

(12:43 PM) gary mazzaferro: Templates are often called "Industry Reference Architectures"

(12:43 PM) Orit: Nancy et al, thinking about external vs. internal, each of the five blocks can be internal or external to the "system". That's one of the "new things" about Big Data. You can delegate any of the activities and can be just loosely coupled sustem overall.

(12:44 PM) Pw Carey-\*Compliance Partners, LLC: "Industry Reference Architectures".....gaud....that's awful....according to a brief survey we just conducted....Pw

(12:48 PM) gary mazzaferro: IRAs is not my term but valuable to solution providers

(12:49 PM) Pw Carey-\*Compliance Partners, LLC: Interesting......

(12:50 PM) Pw Carey-\*Compliance Partners, LLC: Within M0087 is there any mention of Audits/Auditors....?

(12:50 PM) William Miller (MaCT USA) disconnected.

(12:50 PM) William Miller (MaCT USA) joined.

(12:51 PM) Pw Carey-\*Compliance Partners, LLC: Or....GRC (Governance, Risk & Compliance) & CIA (Compliance, Integrity & Availability)....?

(12:58 PM) Mark Underwood (Krypton Brothers): PW -in the mindmap, this is in the Capability Provider "quadrant" - but needs some fleshing out

(12:58 PM) Pw Carey-\*Compliance Partners, LLC: The definition for 'Governance' doesn't appear to address the legislated requirements being expanded by Fed. State, & Local Agencies...."Governance: The readiness of governance policies and processes to be applied to the technologies adopted as part of a Big Data initiative.  Additionally, readiness of governance policies and processes for application to the data managed and operated on as part of a Big Data initiative.

(12:59 PM) Pw Carey-\*Compliance Partners, LLC: Correct....it does need some fleshing out.....thanks Pw

(1:02 PM) Pw Carey-\*Compliance Partners, LLC: Welcome and appreciate the collaboration.....Respectfully yours, Pw

(1:02 PM) Mark Underwood (Krypton Brothers): Narratology problem here is big

(1:03 PM) Pw Carey-\*Compliance Partners, LLC: We believe that eventually well arrive at a 'well polished pearl'....yes?

(1:03 PM) Mark Underwood (Krypton Brothers): In scope or not?

(1:03 PM) Pw Carey-\*Compliance Partners, LLC: What is a 'Narratology' problem....please...it's a new term for us folks way out here....?

(1:04 PM) Mark Underwood (Krypton Brothers): It's the lingustic/psychological/AI discipline of "storytelling"

(1:05 PM) Pw Carey-\*Compliance Partners, LLC: You mean "What we have here is a Failure to Communicate"....?

(1:06 PM) Pw Carey-\*Compliance Partners, LLC: Dear Mr. Chang, would care for a cough drop.....?

(1:07 PM) Pw Carey-\*Compliance Partners, LLC: Rather....would you care for a cough drop....?

(1:07 PM) Deborah Blackstock (MITRE) disconnected.

(1:07 PM) Mark Underwood (Krypton Brothers): ciao

(1:07 PM) gary mazzaferro disconnected.

(1:07 PM) Mark Underwood (Krypton Brothers) disconnected.

(1:08 PM) Pw Carey-\*Compliance Partners, LLC: & we believe in Cross-Pollination...too....

(1:08 PM) PavithraKenjige (PK Technologies) disconnected.

(1:08 PM) Yuri Demchenko (UvA) disconnected.

(1:08 PM) William Miller (MaCT USA) disconnected.

(1:08 PM) Orit disconnected.