**NIST Big Data Working Group (NBD-WD)**

**NBD-WD-2013/M01xx**

|  |  |
| --- | --- |
| **Source:** | Reference Architecture Subgroup |
| **Status:** | Draft |
| **Title:** | Reference Architecture Subgroup Meeting Summary August 15, 2013 |
| **Date:** | August 21, 2013 |
| **Author(s):** | Orit Levin (Microsoft), James Ketner (AT&T), Don Krapohl (Augmented Intelligence) |

**Meeting Summary from July 15, 2013**

Discussion and comments regarding document M0100v4:

Slide 3:

* Add clarification that the Ref Arch subgroup gets the requirements from the Requirements subgroup and then defines the RA applicable to address these requirements.
* To make 'Big Data' platform effective, unique and usable, as well as different from what we have in pre-Big Data era (large filesystems, data warehouses, relational, etc.), 'Big Data' architecture should be capable of different kinds of data representations or storage of the same data set.

Slide 6:

* The innovations enabling “horizontal scaling” in support of Big Data are applicable beyond DBs. They form (and should be shown) as a horizontal layer beneath DB and other capabilities if needed.
* The two boxes on the right seem to indicate that you can either use Relational database or you can use NoSQL. This representation may be misleading because in some scenarios (e.g., data analysis) you may want to combine data from Relational database and a NoSQL database. Or, you may want to extract from NoSQL and load into a Relational database for querying. As a result, these boxes should not be separate.

Slide 7 (same as in M0124):

* Cloud Computing is a “deployment model”. Many BD frameworks using “horizontal scaling technology” are being built today in-house. Therefore, the deployment aspect (model) is less important to the RA than the technological aspect (model).
* We could consider in the future to show a separate picture of the deployment models using the same RA layout, but it is confusing to show these different views (models) on the same picture.
* To emphasize the “mix and match” capabilities/infrastructure/support functions spirit, it may be beneficial to show a single layered block instead of the “(CC or BD) Framework” and “Traditional Framework” blocks.
* The conceptual RA needs to be less busy, e.g., fewer main and nested boxes.
* We need to start writing the text to clarify the intended meaning of data Sources and Usage and their relation to “Collection” and “Access”.
* Additional ideas for alternative naming for “Transformation”, “Capability Management”, “Service Abstraction” are welcome. To better reflect the meaning of each block, the names may also evolve in the course of the document writing.

M0100v4

* The outline was presented. It is a very important document, since this is the foundation for one of the subgroup deliverables.
* It is an initial draft only and comments are welcome with suggestions for modifications and refinement.
* The Outline will evolve to reflect the evolving text as well.
* Please, provide your comments on the Ref Arch list ASAP and preferably before Wed, Aug 21 (that date of the joint conf call).

Action Items:

* Orit: Initiate discussion threads on the Ref Arch list in order to refine (I) Common RA diagram and (II) Common RA outline
* Wo: Update the RA diagram to reflect today’s discussion
* All: For the “White Paper” deliverable, we are looking for contributions for the Outline (and the original content/text) describing different BD RA approaches.