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

**NBD-WD-2013/M0220v2**

|  |  |
| --- | --- |
| **Source:** | Reference Architecture Subgroup |
| **Status:** | Meeting Summary with updates |
| **Title:** | Reference Architecture Subgroup Meeting Summary Sep 5, 2013 |
| **Date:** | September 8, 2013 |
| **Author(s):** | Orit Levin (Microsoft) |

Meeting Summary from September 5, 2013

Discussion and comments regarding document M0100v7:

**Slide 3: Deliverable #1: The White Paper (outline posted as M0151)**

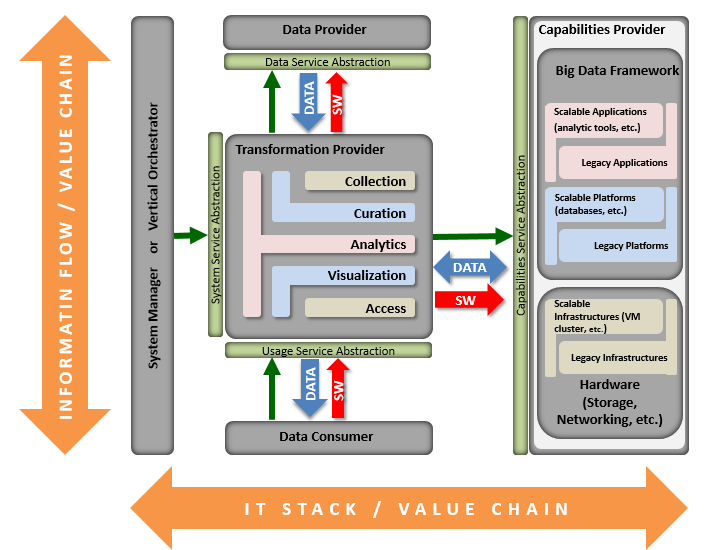
* **Section 2:** The original contributors of the listed RA proposals are responsible for providing the text for the corresponding subsections
* **Section 3**: The task of explaining the listed architectures in the subsections has been assigned as following:
  + IBM: SAFIA-D(CLOUD-AGE IT)
  + Oracle: Harry J Foxwell
  + Booz-Allen: David Boyd (Data Tactics)
  + Others: Wo will reach out to known experts and assign the subsections accordingly.

The text in each subsection should include the following:

* Description of the approach with its goals and applicability
* Figure representing the conceptual architectural model
* List of the key components with their brief description

**Slides 4-8: Deliverable #2: The Common RA**

* The main open issue regarding the role of the Data Sources has been resolved: Data Provider will be used as the abstraction for the new info feeds into the BD system and expose its interfaces through the “Data Service Abstraction”.
* The new baseline diagram to be used for follow up discussions across all subgroups is shown below:



From the chat conversation: the Key to the diagram need to be added ASAP. (See M0226 for an updated version)

**Slide 9: Open Issues updated with reached agreement and resolutions:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Abstract** | **Alternative Approaches** | **Resolution** | **Notes** |
| Data Sources role | 1. Data Sources are included in “Data Manager”, which uses the “services” provided by the “transformation” to drive the system. 2. Data Sources are a stand-alone Level 1 component, which provides “data services” to other RA components. | RESOLVED  Alternative #2 | Blocking the progress |
| Terminology and Taxonomy used for the Level 1 and Level 2 RA blocks | 1. Collector, Manager, Producer, Consumer, etc. 2. Collection, Management, Sources, Use, etc. | RESOLVED  Alternative #1 | Relates to the “actors” vs. “roles” vs. “activities” open issue in Def&Tax subgroup |
| Granularity of interfaces shown in terms of data flow and tools flows | 1. Shown together by a single arrow 2. Shown separately | RESOLVED  Alternative #2 |  |
| Placement of Management and “Security & Privacy” blocks | 1. Inside the “Capabilities Provider” 2. Inside the “System Manager” 3. Each block (Capabilities, Security&Privacy, Management) is stand alone | TBD |  |
| “Management” block resolution | 1. Management only, no additional details on the figure 2. Management block contains System Management and Lifecycle Management sub-blocks 3. System Management and Lifecycle Management blocks are stand alone | TBD |  |
| “Capabilities’” sub-block’ vocabulary | TBD: Needs targeted input and explicit discussion from the Roadmap subgroup | TBD | Pending sync with the Roadmap subgroup |

**Action Items:**

* Wo: assign additional contributors to White Paper Section 3
* Orit: compile Chat comments and add to the Open Issues List (Done See Below)
* Orit: Update Deliverable #2 BD Ref Arch with the new baseline diagram and align its Outline accordingly (Done as M0226).

**Main points from the chat discussion:**

1. “Def&Tax” subgroup will provide the definitions for each RA components. “RefArch” subgroup will explain the diagram and the data flows (including description of arrows).
2. To help reader understanding, there is a need to illustrate the RA by mapping several simple Use Cases to the RA.
3. Start with the high level RA. Then more information can be provided using different techniques such as:
   1. Adding details by zooming into specific blocks/components
   2. Overlaying supplementary/orthogonal views, such as in the “Security & Privacy” case that will form a backplane to the RA since it touches all its interfaces and components.
4. Clarification is needed regarding the boundaries of the RA diagram. What is external and what is internal. For example, whether Data Sources are internal or external to the RA.
5. The five top blocks represent different roles. An actor may perform several roles. Multiple / different actors can perform the same role.
6. “Data Service Abstraction” and “Usage Service Abstraction” may be similar or same because when the BD systems are “chained” or stacked on the top of each other, then a Data Consumer of one (top) system would become a Data Provider to the other (bottom) system.
7. The actual data transfer between the Data producer and the Transformation Provider can be implemented using various manual and/or SW techniques (e.g., push or pull). It will be based on the agreement between the two parties; it is the System’s Manager (a.k.a., Orchestrator) responsibility to identify the Data Providers with their interfaces and then instruct the Transformation (e.g. its “collection” sub-function) to trigger the “Data Service Abstraction” invocation. [In the case of “push” (e.g., YouTube), end user plays the Orchestrator, Data Provider, and Data Consumer” roles. YouTube provides the Transformation/Collection and Visualization functions. Orchestrator logins and fires the YouTube UI. Then the Data Provider makes its file available using one of the agreed formats. Then the Data Consumers can access and watch the video.]
8. Handling of data between a data producer and the transformation provider needs to include validation and error checking and potential additional functionality.
9. The vertical system manager can/should include security, life cycle management, etc. It an enterprise case (e.g., in the federal enterprise architecture), these would be a part of “governance”.
10. It is not clear / or not specified how the role of “Broker” relates to the RA.