NIST Big Data Public Working Group Overview (Co-Chairs)

Robert Marcus  
ET-Strategies

Wo Chang  
NIST

Chaitanya Baru  
NSF

NIST Campus  
Gaithersburg, Maryland  
June 1, 2017
Agenda

• Why Big Data? Why NIST?
• NBD-PWG Goal and Approach
• Subgroups Tasks, Interactions, and Deliverables
  – Definitions and Taxonomies Subgroup
  – Use Case and Requirements Subgroup
  – Reference Architecture Subgroup
  – Security and Privacy Subgroup
  – Standards Roadmap Subgroup
• Next Steps (Workplan)
Why Big data? Why NIST?

- **Why Big Data?** There is a broad agreement among commercial, academic, and government leaders about the remarkable potential of “Big Data” to spark innovation, fuel commerce, and drive progress.

- **Why NIST?** Recommendation from January 15 – 17, 2013 Cloud/Big Data Forum and (b) A lack of consensus on some important, fundamental questions is confusing potential users and holding back progress. Questions such as:
  - *What are the attributes that define Big Data solutions?*
  - *How is Big Data different from the traditional data environments and related applications that we have encountered thus far?*
  - *What are the essential characteristics of Big Data environments?*
  - *How do these environments integrate with currently deployed architectures?*
  - *What are the standardization challenges that need to be addressed to accelerate the deployment of robust Big Data solutions?*

NBD-PWG is being launched to address these questions and is charged to develop consensus definitions, taxonomies, secured reference architecture, and standards roadmap for Big Data that can be embraced by all sectors.

June 1, 2017
NBD-PWG Goal and Approach

Goal

*Develop a secured reference architecture that is vendor-neutral, technology- and infrastructure-agnostic to enable any stakeholders (data scientists, researchers, etc.) to perform analytics processing for their given data sources without worrying about the underlying computing environment.*

Approach – Three Stages (refers to three versions)

1. Identify the high-level NIST Big Data Reference Architecture (NBDRA) key components, which are technology-, infrastructure-, and vendor-agnostic [Done]

2. Define general interfaces between the NBDRA key components [Ongoing]

3. Validate the NBDRA by building Big Data general applications through the general interfaces.
NBD-PWG Goal and Approach

NIST Big Data Standards Roadmap

Data Sources
- Sensors
- Simulations
- Modeling
- Etc.

Data Consumers
- End users
- Repositories
- Systems
- Etc.

BDRA Interface

BDRA Ecosystem Components

Computing Resources

Analytics Resources

Support Infrastructure
- Data Types Services
- Database Services
- Distributed File System Services
- Infrastructure Services

Analytics Application
- Analytics Services
- Visualization & BI Services
- Value-added Content Services
- Security and Privacy Services
Subgroups Focus and Tasks (see Subgroup Co-Chairs presentation)
Subgroups Interaction:

Activities were carried out in parallel and the numbers represent how content flow.
Subgroups Tasks, Interactions, and Deliverables

Deliverable: Stage-1 – High-level Reference Architecture

https://bigdatawg.nist.gov/V1_output_docs.php (Sep. 16, 2015)
Subgroups Tasks, Interactions, and Deliverables

Deliverable: Stage-2 – Reference Architecture Interface

# Next Steps (Workplan)

## Stage-2 NBDIF Workplan

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 1 &amp; 2</td>
<td>Workshop to discuss and collect Version 2 comments</td>
</tr>
<tr>
<td>June 3 – July 15</td>
<td>NBD-PWG to resolve all comments and edits</td>
</tr>
<tr>
<td>July 15</td>
<td>Publish draft documents for official public comment period (45 days)</td>
</tr>
<tr>
<td>Sept. 1 – Sept. 30</td>
<td>NBD-PWG reviews and addresses public comments</td>
</tr>
<tr>
<td>Oct. 1 – Oct. 30</td>
<td>NIST internal review</td>
</tr>
<tr>
<td>Nov. 15</td>
<td>Publish final Version 2 of all volumes of NBDIF</td>
</tr>
</tbody>
</table>

June 1, 2017
Subgroup Co-Chairs Presentation on NBDIF

- Volume 1, Definitions (Nancy Grady, SAIC)
- Volume 2, Taxonomies (Nancy Grady, SAIC)
- Volume 3, Use Cases and General Requirements (Geoffrey Fox, U. Indiana, Piyush Mehrotra, NASA)
- Volume 6, Reference Architecture (David Boyd, InCadence Corp.)
- Volume 4, Security and Privacy (Arnab Roy, Fujitsu; Mark Underwood, Krypton Brothers)
- Volume 8, Reference Architecture Interface (Gregor von Laszewski, U. of Indiana)
- Reference Architecture Software Implementation Environment and Demonstration (Gregor von Laszewski, U. of Indiana)
- Volume 7, Standards Roadmap (Russell Reinsch, Center for Government Interoperability)
- Volume 9, Adoption and Modernization (Russell Reinsch, Center for Government Interoperability)
Volume Presentation Outline

• For each volume (defers Q/As to breakout group)
  – Scope of the volume
  – Brief recap of version 1
  – Highlights of version 2 accomplishments
  – Summary of version 2 areas needing contributions
  – Topics that could be considered for version 3

• Breakout groups for Q/As between 3:00PM – 5:00PM
  – Lecture Room B – Use Case & Requirements and Security & Privacy
  – Lecture Room C – Definitions and Taxonomies
  – Lecture Room D – Standards Roadmap and Adoption & Modernization