Big Data + Analytics: Making Analytics as Services

Wo Chang
Digital Data Advisor
NIST Big Data Public Working Group, Co-Chair
ISO/IEC JTC 1/SC42/WG2 Big Data, Convenor
IEEE BDGMM, Chair
wchang@nist.gov

May 26, 2020
Standards Roadmap Subgroup

NIST Big Data Interoperability Framework (NBDIF)

[Link: https://bigdatawg.nist.gov/V3_output_docs.php]

- NIST SP1500-1: Definitions
- NIST SP1500-2: Taxonomies
- NIST SP1500-3: Use Cases & Requirements
- NIST SP1500-4: Security & Privacy
- NIST SP1500-5: Architecture Survey – White Paper
- NIST SP1500-6: Reference Architecture
- NIST SP1500-7: Standards Roadmap
- NIST SP1500-8: Reference Architecture Interface
- NIST SP1500-9: Adoption & Modernization
ISO/IEC JTC 1/SC 42(AI)/WG2 Big Data Standards Activities

- Members: 220+ from 24 NBs: Australia, Austria, Belgium, Canada, China, Denmark, Finland, France, Germany, India, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Netherlands, Norway, Russian Federation, Singapore, Sweden, United Arab Emirates, UK, US

- ISO/IEC Projects

<table>
<thead>
<tr>
<th>NBDIF Volumes</th>
<th>ISO/IEC Documents (publication date)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vol. 1: Big Data Definitions</td>
<td>20546 Big data Overview and Vocabulary (Feb. 2019)</td>
</tr>
<tr>
<td>Vol. 3: Big Data Use Cases and Requirements</td>
<td>20547-2 Big Data Use Cases and Derived Requirements (April 2018)</td>
</tr>
<tr>
<td>Vol. 5: Big Data Architecture Survey White Paper Skip</td>
<td></td>
</tr>
<tr>
<td>Vol. 6: Big Data Reference Architecture</td>
<td>20547-3 Big Data Reference Architecture (March 2020)</td>
</tr>
<tr>
<td>Vol. 7: Big Data Standards Roadmap</td>
<td>20547-5 Big Data Standards Roadmap (April 2018)</td>
</tr>
<tr>
<td>Vol. 8: Big Data Reference Architecture Interfaces (new) Explore</td>
<td></td>
</tr>
<tr>
<td>Vol. 9: Big Data Adoption and Modernization (new) Explore</td>
<td></td>
</tr>
</tbody>
</table>

Next Step: Explore Big Data + Analytics – Making Analytics as Services

Plus other "data" related projects such as
- Process Management Framework for big data analytics
- A series of data quality for analytics and machine learning
- Data exploration
Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 28, 2020
Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 28, 2020
Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 28, 2020
Links

[1] FAIR Data (https://www.rd-alliance.org/groups/go-fair-ig)


**NBD-PWG Next Step**

*Motivation and Industry Trends*

With Big Data’s compound annual growth rate at 61 percent and its ever-increasing deluge of information in the mainstream, the collective sum of world data will grow from 33 zettabytes (ZB, $10^{21}$) in 2018 to 175 ZB **by 2025**. The presence of such a rich source of information requires a massive analysis that can effectively bring about much insight and knowledge discovery.

**Programming Libs**

- Python
- scikit-learn
- Shiny
- dplyr
- ggplot2
- knitr
- lubridate
- Others...

**ML Frameworks**

- TensorFlow
- PyTorch
- Caffe2
- Keras
- Cognitive Toolkit
- Others...

**Analytics Services**

- IBM Watson Analytics
- Google Analytics
- Microsoft Azure Cognitive Services
- Knowledge
- Speech
- Others...
NBD-PWG Next Step

Explore Big Data + Analytics – Making Analytics as Services

NBD-PWG is exploring how to extend NBDIF for packaging scalable analytics as services to meet the challenges of so much information. These services would be reusable, deployable, and operational for Big Data, High Performance Computing, and AI machine learning (ML) and deep learning (DL) applications, regardless of the underlying computing environment.

Topics for discussion include:

- **Exploration**: Determine and document the level of interest from industry, government, and academia in extending the NBDIF to develop scalable analytics as services that are reusable, deployable, and operational, regardless of the underlying computing environment.

- **Key Focus Areas**
  1. Compile and organize use cases, analytic services from traditional statistical, AI/ML/DL, and emerging analytics application domains; identify and document technical requirements.
  2. Package analytic algorithms with well-defined input and output parameters as service payloads that can be reusable, deployable, and operational across multi-cores, CPUs, and GPU computing platforms.
  3. Encapsulate service payload with well-defined format, interface, and end-to-end access control for open and secure computing environment.
  4. Establish federated registries to locate and consume analytics services with persistent identifiers across organizations.
  5. Provide resource management for application orchestration and workflow between processes.

*Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 26, 2020*
NBD-PWG Next Step: Making Analytics as Services

Focus 1 - Compile Analytic Service Use Cases

Selection of use cases: (a) available datasets and (b) available analytics codes

Source [1]: http://1.bp.blogspot.com/-PKiTQa0mrn4/T_mGb6A13yI/AAAAAAAA8Q/TtH7xyjQ3FA/s640/analytics+tools+landscape.bmp

Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 26, 2020
NBD-PWG Next Step: Making Analytics as Services

Focus 2 - Package Analytic Algorithms with well-defined I/O Parameters

Input Parameters
- Inputs
  - Data Source-1
  - Data Source-n
  - Var-1
  - ...
  - Var-n
- Results
  - Device
  - protocol
  - Format
  - ...

Output Parameters
- Output
  - JobID
  - Date/Time
  - Out-1
  - ...
  - Out-n

Analytic Algorithm

Explore the following:
- Portable Format for Analytics (PFS): http://dmg.org/pfa

Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 26, 2020
NBD-PWG Next Step: Making Analytics as Services

Focus 3 - Encapsulate Analytic Service as Payload

Enable Big Data analytics services/tools for **interoperability, portability, reusability, and extensibility**.

Practical Aspect: Analytics services/tools can be **reusable, deployable, and operational** (max. use of resources) on any of Big Data, HPC, machine/deep learning, etc. computing environment.
NBD-PWG Next Step: Making Analytics as Services

Focus 4 – Federated Analytic Services Registry

Metadata is generated per analytic model within the Repository and pushed, via a Registration Service, into a Federated Metadata Registry, creating a digital object for analytic as a service.

Federated Metadata Registry provides Analytics Management and Discovery Services for users.
Reference Architecture Subgroup

NBDIF Vol. 6: NIST Big Data Reference Architecture (NBD-RA)
NBD-PWG Next Step: Making Analytics as Services

Focus 5: Resource Management – Orchestration and Workflow

KEY:
- DATA: Big Data Information Flow
- SW: Software Tools and Algorithms Transfer
- Interface: Programmable Interface

Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 26, 2020
Beyond: Enable Convergence of Data + Analytics + Compute

Repositories
Datasets

Collection of Metadata, Federated Registries

Collection of Analytic Services

Exascale Big Data Analytics & Systems

Containers/Microservices/etc.

Many CPUs/Cores/Threads/GPUs/FPGAs/ASIC/etc.

Big Data + Analytics: Making Analytics as Services, Wo Chang, NIST/ITL, May 28, 2020
Questions?

Please contact: wchang@nist.gov