NEW NBD-PWG Use Case Studies Template May 21 2015

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Instructions** | Please Read! | | | | | Drill Down 1 | | | | | | | | | | | |
| **Owner** | Name, Institution, Editing Rights | | | | | | | | | | | | | | | | |
| **Use Case Title** |  | | | | | | | | | | | | | | | | |
| **Vertical (area)** |  | | | | | | | | | | | | | | | | |
| **Author(s)/Company/Email** |  | | | | | | | | | | | | | | | | |
| **Actors/ Stakeholders and their roles and responsibilities** |  | | | | | | | | | | | | | | | | |
| **Goals** |  | | | | | | | | | | | | | | | | |
| **Use Case Description** |  | | | | | | | | | | | | | | | | |
| **Abstract** | Any Comment | | | | | Drill Down 2 | | | | | | | | | | | |
| **Defining tags** | Specify Tags in Drill Down List | | | | | Drill Down 3 | | | | | | | | | | | |
| **Current**  **Solutions** | **Compute(System)** |  | | | | | | | | | | | | | | | |
| **Storage** |  | | | | | | | | | | | | | | | |
| **Networking** |  | | | | | | | | | | | | | | | |
| **Software** |  | | | | | | | | | | | For ease of compilation, use drill down list 4 | | | | |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** |  | | | | | | | | | | | | | | | |
| **Data Destination** |  | | | | | | | | | | | | | | | |
| **Volume (size)** | Be quantitative below (proviso = any restriction to get #’s) | | | | | | | | | | | | | | | |
| Volume |  | Units | | | | |  | Time | | | |  | | proviso |  |
| **Velocity (e.g. real time)** | Be quantitative below (proviso = any restriction to get #’s) | | | | | | | | | | | | | | | |
| Volume Change per unit time | | |  | | Units per time | | | |  | Time | | |  | proviso |  |
| **Variety (multiple datasets, mashup)** |  | | | | | | | | | | | | | | | |
| **Variability (rate of change)** |  | | | | | | | | | | | | | | | |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity & Data Quality (semantics, syntax)** |  | | | | | | | | | | | | | | | |
| **Visualization** |  | | | | | | | | | | | | | | | |
| **Data Types** |  | | | | | | | | | | | | | | | |
| **Metadata** | Comment on quality and richness of metadata | | | | | | | | | | | | | | | |
| **Curation & Governance** |  | | | | | | | | | | | | | | | |
| **Data Analytics** |  | | | | | | | | | | | | | | | |
| **Workflow** | Comment. Please specify Stages in Drill Down | | | | | | | Drill Down 5 | | | | | | | | | |
| **Security and Privacy Issues** | Comment. Please specify details in Drill Down | | | | | | | Drill Down 6 | | | | | | | | | |
| **Big Data Issues** |  | | | | | | | | | | | | | | | | |
| **User Interface and Mobile access Issues** |  | | | | | | | | | | | | | | | | |
| **List key features and related use cases** |  | | | | | | | | | | | | | | | | |
| **Futures** |  | | | | | | | | | | | | | | | | |
| **More Information (URLs)** |  | | | | | | | | | | | | | | | | |

Notes: No proprietary or confidential information should be included

OLD NBD-PWG Use Case Studies Template Summer 2013

Green edited. Red removed. Additions not indicated

|  |  |  |
| --- | --- | --- |
| **Use Case Title** |  | |
| **Vertical (area)** |  | |
| **Author/Company/Email** |  | |
| **Actors/ Stakeholders and their roles and responsibilities** |  | |
| **Goals** |  | |
| **Use Case Description** |  | |
| **Current**  **Solutions** | **Compute(System)** |  |
| **Storage** |  |
| **Networking** |  |
| **Software** |  |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** |  |
| **Volume (size)** |  |
| **Velocity**  **(e.g. real time)** |  |
| **Variety**  **(multiple datasets, mashup)** |  |
| **Variability (rate of change)** |  |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues, semantics)** |  |
| **Visualization** |  |
| **Data Quality (syntax)** |  |
| **Data Types** |  |
| **Data Analytics** |  |
| **Big Data Specific Challenges (Gaps)** |  | |
| **Big Data Specific Challenges in Mobility** |  | |
| **Security and Privacy**  **Requirements** |  | |
| **Highlight issues for generalizing this use case (e.g. for ref. architecture)** |  | |
| **More Information (URLs)** |  | |
| **Note: <additional comments>** | | |

Notes: No proprietary or confidential information should be included

ADD picture of operation or data architecture of application below table.

Drill Down 1: Instructions for completing fields

This is for original template – needs to be updated and examples given

The following descriptions of fields in the template are provided to help with the understanding of both document intention and meaning of the 26 fields and also indicate ways that they can be improved.

* **Use Case Title:** Title provided by the use case author
* **Vertical (area):** Intended to categorize the use cases. However, an ontology was not created prior to the use case submissions so this field was not used in the use case compilation.
* **Author/Company/Email:** Name, company, and email (if provided) of the person(s) submitting the use case.
* **Actors/ Stakeholders and their roles and responsibilities:** Describes the players and their roles in the use case.
* **Goals:** Objectives of the use case.
* **Use Case Description:** Brief description of the use case.
* **Current Solutions:** Describes current approach to processing Big Data at the hardware and software infrastructure level.
* **Compute(System):** Computing component of the data analysis system.
* **Storage:** Storage component of the data analysis system.
* **Networking:** Networking component of the data analysis system.
* **Software:** Software component of the data analysis system.
* **Big Data Characteristics:** Describes the properties of the (raw) data including the four major ‘V’s’ of Big Data described in *NIST Big Data Interoperability Framework: Volume 1, Big Data Definition* of this report series.
* **Data Source:** The origin of data, which could be from instruments, Internet of Things, Web, Surveys, Commercial activity, or from simulations. The source(s) can be distributed, centralized, local, or remote.
* **Volume:** The characteristic of data at rest that is most associated with Big Data. The size of data varied drastically between use cases from terabytes to petabytes for science research (100 petabytes was the largest science use case for LHC data analysis), or up to exabytes in a commercial use case.
* **Velocity:** Refers to the rate of flow at which the data is created, stored, analyzed, and visualized. For example, big velocity means that a large quantity of data is being processed in a short amount of time.
* **Variety:** Refers to data from multiple repositories, domains, or types.
* **Variability:** Refers to changes in rate and nature of data gathered by use case.
* **Big Data Science:** Describes the high level aspects of the data analysis process
* **Veracity:** Refers to the completeness and accuracy of the data with respect to semantic content. *NIST Big Data Interoperability Framework: Volume 1, Big Data Definition* discusses veracity in more detail.
* **Visualization:** Refers to the way data is viewed by an analyst making decisions based on the data. Typically visualization is the final stage of a technical data analysis pipeline and follows the data analytics stage.
* **Data Quality:** This refers to syntactical quality of data. In retrospect, this template field could have been included in the Veracity field.
* **Data Types:** Refers to the style of data such as structured, unstructured, images (e.g., pixels), text (e.g., characters), gene sequences, and numerical.
* **Data Analytics:** Defined in *NIST Big Data Interoperability Framework: Volume 1, Big Data Definition* as “the synthesis of knowledge from information”. In the context of these use cases, analytics refers broadly to tools and algorithms used in processing the data at any stage including the data to information or knowledge to wisdom stages, as well as the information to knowledge stage.
* **Big Data Specific Challenges (Gaps):** Allows for explanation of special difficulties for processing Big Data in the use case and gaps where new approaches/technologies are used.
* **Big Data Specific Challenges in Mobility:** Refers to issues in accessing or generating Big Data from Smart Phones and tablets.
* **Security and Privacy Requirements:** Allows for explanation of security and privacy issues or needs related to this use case.
* **Highlight issues for generalizing this use case:** Allows for documentation of issues that could be common across multiple use-cases and could lead to reference architecture constraints.
* **More Information (URLs):** Resources that provide more information on the use case.
* **Note: <additional comments>:** Includes pictures of use-case in action but was not otherwise used.

**Drill Down 2: Abstract**

Please complete a standalone abstract of length of about 1/3 page excluding pictures with

1. Title
2. Contributors with institutions
3. URL
4. Application
5. Current Data Analysis Approach
6. Future of Application and Approach
7. Any pictures illustrating use case or its analysis with captions

Note items 1 and 2 can be generated from Use Case Title and Author(s)/Company/Email fields of main form

**Drill Down 3: List of tags**

*Here* ***YN*** *= Yes, No, Don’t Know or N/A*

***Ch*** *= Choose 1 or more from list, Don’t Know or N/A. Note these choices could be replaced by multiple Yes-No items*

***Free Form*** *– user text*

**Data Tags – this references input data**

Uses Geographical Information Systems? YN

Related to Internet of Things? YN

Data Real time Streaming, Batched Streaming, Repository? Ch

Data comes from HPC simulations? YN

Data Fusion important? YN

Data transient/permanent? Ch

Data shared/dedicated between different applications/users? Ch

User data system based on Objects or Files? Ch

Uses HDFS or wide area file system or parallel file system? Ch

Uses SQL or NOSQL or NewSQL or Graph? Ch

Specify other data access/storage issues? Free Form answer

**Analytics Tags**

Basic statistics used? YN

Search/Query/Index Important? YN

Classification Important? YN

Recommender Engine Used? YN

Graph algorithms used? YN

Clustering algorithms used? YN

Alignment algorithms used? YN

(Deep) Learning algorithms used? YN

Algorithm O(N^2) ? YN

Data regular? YN

Data dynamic? YN

Specify other data analytics used? Free Form answer

**Programming Model Tags**

Event-based? YN

Dataflow? YN

Pleasingly parallel or “global linked processing”? Ch

Classic MapReduce? YN

Spark or Iterative MapReduce? YN

Graph processing? YN

MPI and/or BSP? YN

Workflow software used? YN

Python or Scripting front ends? YN

Agent-based? YN

Shared memory architectures important? YN

Use case I/O dominated? YN

Specify bytes/flops or other I/O versus computing issues? Free Form answer

Specify other compute/programming issues? Free Form answer

**Drill Down 4: Software**

Please provide an easy to parse list. See <http://hpc-abds.org/kaleidoscope/> for some example software systems

**Drill Down 5: Workflow**

Please specify workflow or stages of the analysis. For each stage give items below. If this feature, there is redundancy with features in main form. Need to resolve; one possibility is that upper level form has major stages or summary of all.

1. Name of stage
2. Data Source(s)
3. Nature of Data
4. Software used
5. Analytics
6. Infrastructure used
7. Percentage of use case effort
8. Other comments

**Drill Down 6: Security & Privacy**

Add detailed list of S&P working group