**Use Cases from NBD(NIST Big Data) Requirements WG**

<http://bigdatawg.nist.gov/home.php>

**NBD(NIST Big Data) Requirements WG Use Case Template Aug 11 2013**

|  |  |  |  |
| --- | --- | --- | --- |
| **Use Case Title** | | Belle II Experiment | |
| **Vertical (area)** | | Scientific Research: High Energy Physics | |
| **Author/Company/Email** | | David Asner & Malachi Schram, PNNL, [david.asner@pnnl.gov](mailto:david.asner@pnnl.gov) & [malachi.schram@pnnl.gov](mailto:malachi.schram@pnnl.gov) | |
| **Actors/Stakeholders and their roles and responsibilities** | | David Asner is the Chief Scientist for the US Belle II Project  Malachi Schram is Belle II network and data transfer coordinator and the PNNL Belle II computing center manager | |
| **Goals** | | Perform precision measurements to search for new phenomena beyond the Standard Model of Particle Physics | |
| **Use Case Description** | | Study numerous decay modes at the Upsilon(4S) resonance to search for new phenomena beyond the Standard Model of Particle Physics | |
| **Current**  **Solutions** | **Compute(System)** | | Distributed (Grid computing using DIRAC) |
| **Storage** | | Distributed (various technologies) |
| **Networking** | | Continuous RAW data transfer of ~20Gbps at designed luminosity between Japan and US  Additional transfer rates are currently being investigated |
| **Software** | | Open Science Grid, Geant4, DIRAC, FTS, Belle II framework |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** | | Distributed data centers  Primary data centers are in Japan (KEK) and US (PNNL) |
| **Volume (size)** | | Total integrated RAW data ~120PB and physics data ~15PB and ~100PB MC samples |
| **Velocity**  **(e.g. real time)** | | Data will be re-calibrated and analyzed incrementally  Data rates will increase based on the accelerator luminosity |
| **Variety**  **(multiple datasets, mashup)** | | Data will be re-calibrated and distributed incrementally. |
| **Variability (rate of change)** | | Collisions will progressively increase until the designed luminosity is reached (3000 BB pairs per sec).  Expected event size is ~300kB per events. |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues)** | | Validation will be performed using known reference physics processes |
| **Visualization** | | N/A |
| **Data Quality** | | Output data will be re-calibrated and validated incrementally |
| **Data Types** | | Tuple based output |
| **Data Analytics** | | Data clustering and classification is an integral part of the computing model. Individual scientists define even level analytics. |
| **Big Data Specific Challenges (Gaps)** | | Data movement and bookkeeping (file and event level meta-data). | |
| **Big Data Specific Challenges in Mobility** | | Network infrastructure required for continuous data transfer between Japan (KEK) and US (PNNL). | |
| **Security & Privacy**  **Requirements** | | No special challenges. Data is accessed using grid authentication. | |
| **Highlight issues for generalizing this use case (e.g. for ref. architecture)** | |  | |
| **More Information (URLs)** | | http://belle2.kek.jp | |
| **Note:** <additional comments> | | | |

**Note: No proprietary or confidential information should be included**