**Current Draft:**

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Use Case Title** | | Large Scale Geospatial Analysis and Visualization | |
| **Vertical (area)** | | Defense – but applicable to many others | |
| **Author/Company/Email** | | David Boyd/Data Tactics/  [dboyd@data-tactics.com](mailto:%20dboyd@data-tactics.com) | |
| **Actors/Stakeholders and their roles and responsibilities** | | Geospatial Analysts  Decision Makers  Policy Makers | |
| **Goals** | | Support large scale geospatial data analysis and visualization. | |
| **Use Case Description** | | As the number of geospatially aware sensors increase and the number of geospatially tagged data sources increases the volume geospatial data requiring complex analysis and visualization is growing exponentially. Traditional GIS systems are generally capable of analyzing a millions of objects and easily visualizing thousands. Today’s intelligence systems often contain trillions of geospatial objects and need to be able to visualize and interact with millions of objects. | |
| **Current**  **Solutions** | **Compute(System)** | | Compute and Storage systems - Laptops to Large servers (see notes about clusters)  Visualization systems - handhelds to laptops |
| **Storage** | | Compute and Storage - local disk or SAN  Visualization - local disk, flash ram |
| **Networking** | | Compute and Storage - Gigabit or better LAN connection  Visualization - Gigabit wired connections, Wireless including WiFi (802.11), Cellular (3g/4g), or Radio Relay |
| **Software** | | Compute and Storage – generally Linux or Win Server with Geospatially enabled RDBMS, Geospatial server/analysis software – ESRI ArcServer, Geoserver  Visualization - Windows, Android, IOS – browser based visualization. Some laptops may have local ArcMap. |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** | | Very distributed. |
| **Volume (size)** | | Imagery – 100s of Terabytes  Vector Data – 10s of Gigabytes but billions of points |
| **Velocity**  **(e.g. real time)** | | Some sensors delivery vector data in NRT. Visualization of changes should be NRT. |
| **Variety**  **(multiple datasets, mashup)** | | Imagery (various formats NITF, GeoTiff, CADRG)  Vector (various formats shape files, kml, text streams: Object types include points, lines, areas, polylines, circles, ellipses. |
| **Variability (rate of change)** | | Moderate to high |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues)** | | Data accuracy is critical and is controlled generally by three factors:   1. Sensor accuracy is a big issue. 2. datum/spheroid. 3. Image registration accuracy |
| **Visualization** | | Displaying in a meaningful way large data sets (millions of points) on small devices (handhelds) at the end of low bandwidth networks. |
| **Data Quality** | | The typical problem is visualization implying quality/accuracy not available in the original data. All data should include metadata for accuracy or circular error probability. |
| **Data Types** | | Imagery (various formats NITF, GeoTiff, CADRG)  Vector (various formats shape files, kml, text streams: Object types include points, lines, areas, polylines, circles, ellipses. |
| **Data Analytics** | | Closest point of approach, deviation from route, point density over time, PCA and ICA |
| **Big Data Specific Challenges (Gaps)** | | Indexing, retrevial and distributed analysis  Visualization generation and transmission | |
| **Big Data Specific Challenges in Mobility** | | Visualization of data at the end of low bandwidth wireless conections. | |
| **Security & Privacy**  **Requirements** | | Data is sensitive and must be completely secure in transit and at rest (particularly on handhelds) | |
| **Highlight issues for generalizing this use case (e.g. for ref. architecture)** | | Geospatial data requires unique approaches to indexing and distributed analysis. | |
| **More Information (URLs)** | | Applicable Standards: <http://www.opengeospatial.org/standards>  <http://geojson.org/>  <http://earth-info.nga.mil/publications/specs/printed/CADRG/cadrg.html>  Geospatial Indexing: Quad Trees, Space Filling Curves (Hilbert Curves) – You can google these for lots of references. | |
| **Note:** The has been some work with in DoD related to this problem set. Specifically, the DCGS-A standard cloud (DSC) stores, indexes, and analyzes some big data sources. However, many issues still remain with visualization. | | | |