Guidance on how to use NBD-WG deliverables for Big Data Projects

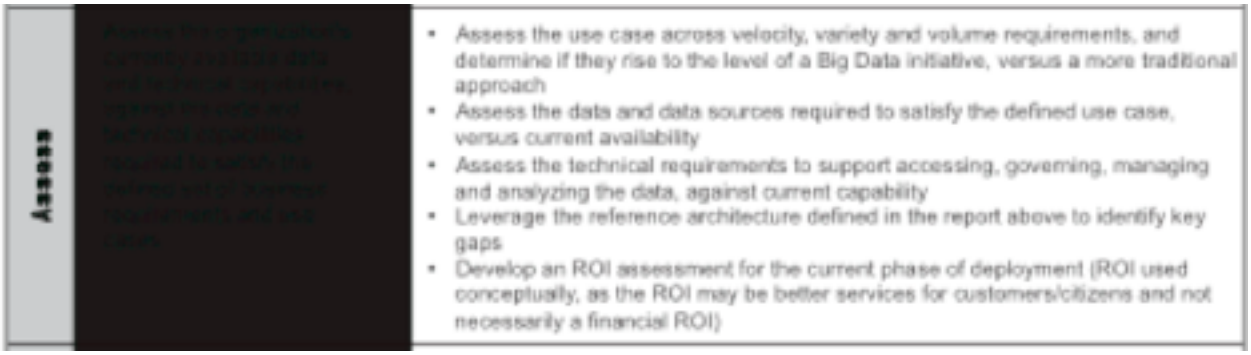
TechAmerica has published a report on Demystifying Big Data that has been well received by government Big Data stakeholders (e.g. NITRD). The report includes a reasonable 4 step Process for Big Data projects including Define, Assess, Plan, Execute. As an exercise, I mapped NBD-WG deliverables to steps in their Process. The numbers i n the mapping are the specific bullet point in the step. This mapping was inspired by a discussion today in the Roadmap Subgroup on the value of providing guidance to readers on how to use the NBD-WG deliverables. This is only a personal first cut to stimulate discussion.

Define - Relevant Deliverables

Definition/Taxonomy - Helps decide if Big Data is better than conventional solutions (3)
Requirements/Use Cases - Documented use cases suggest possible opportunity (1,2)
Reference Architecture - Helps compare Big Data and traditional solutions (4)
Roadmap - Can help in creation of overall vision (5)

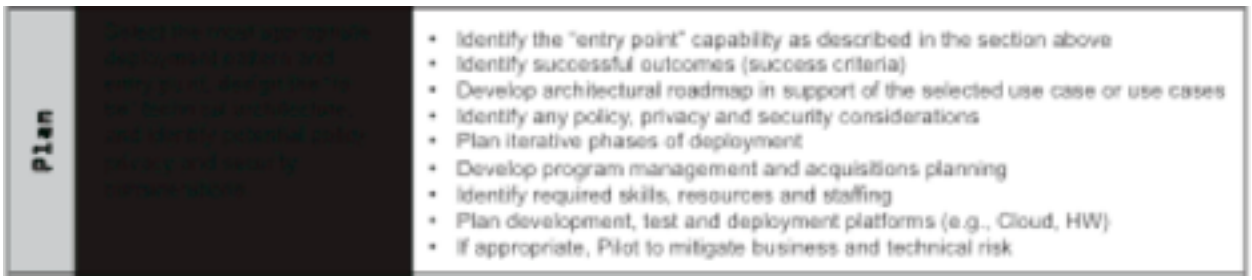
Assess - Relevant Deliverables

Requirements/Use Cases - Helps assess the use case for Big Data implementation (1)
Requirements/Use Cases - Requirements helps assess technical requirements (3)
Reference Architecture - Combined with requirements to identify technical gaps (4)
Roadmap - Can supply analysis of technical gaps and ways to handle them (4)



Plan - Relevant Deliverables

Reference Architecture, Roadmap, Use Cases - Can assist architectural roadmap (3)
Privacy and Security - Help identify policy security, privacy considerations (4)
Roadmap - Might help planning iterative phases of deployment (5)
Roadmap - Help identify skills, resources, and staffing (7)



Execute - Relevant Deliverables

Out of scope for current NBD-WG deliverables. However project experiences can feed back into future NBD-WG deliverables.

Complete Table from TechAmerica's Demystifying Big Data Page 30 for reference

Define		• Identify key business challenges, and potential use cases to address • Identify areas of opportunity where access to Big Data can be used to better serve the citizenry, the mission, or reduce costs • Ask – does Big Data hold a unique promise to satisfy the use case(s) • Identify the value of a Big Data investment against more traditional analytic investments, or doing nothing • Create your overall vision, but chunk the work into tactical phases (time to value within 12-18 month timeframe) • Don't attempt to solve all Big Data problems in the initial project – seek to act tactically, but in the strategic context of your key business imperatives
Assess		• Assess the use case across velocity, variety and volume requirements, and determine if they rise to the level of a Big Data initiative, versus a more traditional approach • Assess the data and data sources required to satisfy the defined use case, versus current availability • Assess the technical requirements to support accessing, governing, managing and analyzing the data, against current capability • Leverage the reference architecture defined in the report above to identify key gaps • Develop an ROI assessment for the current phase of deployment (ROI used conceptually, as the ROI may be better services for customers/citizens and not necessarily a financial ROI)
Plan		• Identify the "entry point" capability as described in the section above • Identify successful outcomes (success criteria) • Develop architectural roadmap in support of the selected use case or use cases • Identify any policy, privacy and security considerations • Plan iterative phases of deployment • Develop program management and acquisitions planning • Identify required skills, resources and staffing • Plan development, test and deployment platforms (e.g., Cloud, HW) • If appropriate, Pilot to mitigate business and technical risk
Execute		• Deploy the current phase project plan • Build out the Big Data platform as the plan requires, with an eye toward flexibility and expansion • Deploy technologies with both the flexibility and performance to scale to support subsequent use cases and corresponding data volume, velocity and variety
Review		