- What are the main differences and commonalities between the HPC and BDA requirements/technologies/working-assumptions in this area?

What are HPC and BDA?

HPC is more or less to solve some PDEs?

Differences:

Have some commonalities in their targeting objects (e.g. graphs) but their characteristics is sometimes quite different.

BDA sometimes lacks locality and thus sometimes need fully global communications, fully dense matrices to represent a problem.

Commonalities/Similarities:

Common mission is to design efficient algorithms/programs to solve problems in both fields.

We might not find commonalities in problems/data themselves, but we could find them in methodologies. In this sense HPC could help BDA.

In some iterative learning process, we can find (e.g.) sparse decomposition steps very similar to done in HPC.

Large sparse linear algebra should be a good target for both communities.

Uncertainty of data.

- Are there common needs/problems/interfaces could serve as the basis (or as stepping stones) along a path to (some reasonable level of) infrastructure and application convergence?

Big data libraries are still poor(?) in terms of parallelization.

Large sparse matrices.

HPC will need BDA-like approach (e.g. graph-oriented?).

Studies on integrated library, esp. API.

Common(?) interface of visualization, graphs, ...

- Are there interdomain testbeds that combine BDA and HPC workflows in ways that could help uncover pathways toward convergence?

Classic/ordinary HPC systems can be good testbeds, at least for analytics, ...

NVM and other emerging technologies are beneficial for both field (maybe different impacts).

A large system with appropriate I/O (e.g. HDFS).

- What is/are the technology or new research that may be a game changer?

  New algebra (for tensor).
  We could find new interesting field midst between algebra for HPC and BDA.
  Large dimension problems.
  Stochastic programming (e.g. having uncertainty as the first-level object?).
  Scalable library of ...

- What action would be your number one priority to be taken rapidly to ensure success of the converge of Extreme computing and Big Data infrastructures?

  Parallel algorithm (and its implementation, IMO) research.
  On-the-fly machine learning algorithms, on-the-fly quantification of acquired data.
  Scalable parallelized libraries of dense/sparse algebra.
  Statistics tools should come to HPC world.
  Good and killer application for HPC/BDA collaboration, interaction, ...

- What action would be your number one priority to be taken rapidly to ensure the emergence of efficient Extreme computing and Big Data applications?

  Training and skill development.

- How would you measure the success of the BDEC initiative?

  Funding ($xx M).