Amazon Web Services Response to: The Networking and Information Technology Research and Development (NITRD) Program Request for Input (RFI) – National Big Data R&D Initiative

November 14, 2014

Submitted By:

Amazon Web Services, Inc.
12900 Worldgate Dr. Suite 800
Herndon, VA 20170
Jamie Baker
Sr. Civilian Account Manager, AWS
Public Sector
301-461-0569
bakjames@amazon.com
Cage Code: 66EB1
DUNS Number: 965048981
NAICS: 518210

Submitted To:

Wendy Wigen
703-292-4873
wigen@nitrd.gov

The purpose of this document is to provide the Networking and Information Technology Research and Development (NITRD) Big Data Senior Steering Group (BDSSG) with feedback to inform the development of a framework, set of priorities, and ultimately a strategic plan for the National Big Data R&D Initiative. We will restrict our feedback on the initial framework document, *The National Big Data R&D Initiative: Vision and Priority Actions*, to a few major items.

1. Who are you, contact information, and what is your experience working with big data, and your role in the big data innovation ecosystem? – Amazon Web Services, Inc. (AWS) is a leading commercial cloud provider in the USA, and a leading platform for development and operations of open source and commercial Big Data solutions. More information here: http://aws.amazon.com/big-data/. Contact information can be found on this document's cover page. In addition to a robust set of foundational infrastructure services, we provide robust security capabilities and managed services used widely in the commercial and academic research Big Data space.

2. What are the gaps that are not addressed in the Visions and Priority Actions document? – The initial framework document touches on the building and expansion of access to cyberinfrastructure, but does not explicitly relate this point to a later goal of long-term sustainability of the program, and resources created within such programs. AWS strongly believes that an operational expenditure (OpEx) approach to IT infrastructure increases the scalability, security and flexibility of resources available to end users, while also adding the benefits of multi-site collaborations through cloud technology.

3. What do you think are the most high impact ideas at the frontiers of big data research and development? What new research, education, and/or infrastructure investments do you think will be game-changing for the big data innovation ecosystem? – Big Data systems have the following characteristics: developed at a fast pace; infrastructure is based on commodity hardware and refreshed often; processes are tuned for horizontal scale. Commercial clouds, like AWS, offer an infrastructure platform that is well suited to all of these characteristics by providing rapid access to flexible and low cost IT resources in an on-demand, pay-as-you-go model with no upfront investments in hardware. Instead, you can provision exactly the right type and size of

> *The benefits of cloud computing include:*
> - *Self-service provisioning;*
> - *On-demand control and elasticity;*
> - *No up-front capital expenses or commitments for infrastructure/facilities;*
> - *Pay-as-you-go pricing;*
> - *Improved agility; and,*
> - *Lower cost of ownership than traditional on-premises offerings.*

computing resources needed to solve the problem at hand. With Cloud Computing, you can access as many resources as you need, almost instantly, and only pay for what you use. The following are five advantages and benefits of adopting cloud computing:

- *Trade capital expense for variable expense* – Instead of having to invest heavily in data centers and servers before you know how you're going to use them, you can use cloud computing and only pay when you consume computing resources, and only pay for how much you consume.

- *Benefit from massive economies of scale* – By using cloud computing, you can achieve a lower variable cost than you can get on your own. Because usage from hundreds of thousands of customers are aggregated in the cloud, cloud computing providers such as AWS can achieve higher economies of scale which translates into lower pay as you go prices.

- *Flexibility – Stop guessing capacity* – Eliminate guessing on your infrastructure capacity needs. When you make a capacity decision prior to deploying an application, you often either end up sitting on expensive idle resources or dealing with limited capacity. With Cloud Computing, these problems go away. You can access as much or as little as you need, and scale up and down as required with only a few minutes notice.

- *Increase speed and agility* – In a cloud computing environment, new IT resources are only a click away, which means you reduce the time it takes to make those resources available to your developers from weeks to just minutes. This results in a dramatic increase in agility for the organization, since the cost and time it takes to experiment and develop is significantly lower.

- *Stop spending money on building, running, and maintaining data centers* – Focus on projects that are core to your organization, not the infrastructure. Cloud computing lets you focus on your own customers, rather than on the heavy lifting of racking, stacking and powering servers.
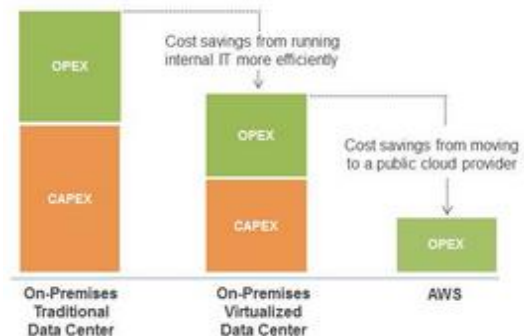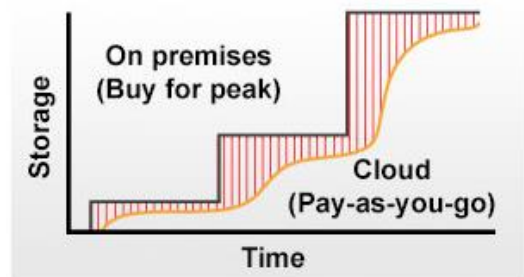
4. How can the federal government most effectively enable new partnerships, particularly those that cross sectors or domains? – Federal agencies can promote the creation of central data sharing hubs among the members of a particular initiative on commercial cloud infrastructures. This model would enable collaborators to bring their own software, processes, and financial resources to the same environment as the data hub. In this model, AWS's cloud infrastructure would host high-value data sets, pre-configured software, and reference architectures that enable researchers within a community to bootstrap their research.

5. A short explanation of why you feel your contribution/ideas should be included in the strategic plan. – As the leading commercial cloud provider in the big data ecosystem, we feel that we have a depth of experience and knowledge to offer to the NITRD computing infrastructure at the scale of Big Data, while controlling for costs of the data analysis. We would welcome the opportunity to provide the NITRD BDSSG detailed information on considering an OpEx approach to IT infrastructure, the scalability, security and flexibility of the cloud, and the benefits of researchers collaborating through cloud technology.