**FutureGrid Summary**

FutureGrid was a national-scale Grid, Cloud and HPC computing test-bed service of modest size that includes a number of computational resources at five distributed locations. FutureGrid experience and architecture was built around software defined systems at all levels of the stack shown in figure 1 - encompassing VM and bare-metal infrastructure, networks and application, systems and platform software – with a unifying goal of providing Computing Testbeds as a Service. FutureGrid systems totaled 4704 cores divided into distributed general purpose clusters at Chicago, Florida, IU and Texas; a Cray XT5m at IU and four small specialized clusters supporting SSD (at SDSC), Large Disk Large memory (at IU) and general purpose GPU’s (IU). FutureGrid’s system model grew in sophistication and ultimately supported software-defined systems – encompassing virtualized and bare-metal infrastructure, networks, application, systems and platform software – with a unifying goal of providing Computing Testbeds as a Service (CTaaS). Cloudmesh aggregates resources not only from FutureGrid, but also from OpenCirrus, Amazon, Microsoft Azure, and HP Cloud and GENI resources. FutureGrid's software system Cloudmesh was originally developed in order to simplify the execution of multiple concurrent experiments on a federated cloud infrastructure and in addition to virtual resources, FutureGrid exposed bare-metal provisioning to users. The importance of DevOps tools like Cloudmesh and close integration of Software and Systems administration staff were important lessons from FutureGrid. During its operation, FutureGrid supported 417 projects with 2601 registered users.

The six figures show the Computing Testbed as a Service architecture; the distributed network that was FutureGrid superimposed on a map of the USA; the Cloud, Infrastructure, Grid, HPC and Testbed services offered by FutureGrid; a mosaic of FutureGrid machines; four snapshots of FutureGrid monitoring capabilities; and a word cloud constructed from FutureGrid project titles.

We found it possible to classify the FutureGrid projects into four major areas: Computer Science and Middleware (56%); Domain Science (21%); Training Education and Outreach (14%); and Computer Systems Evaluation (9%). The numbers in parentheses indicate percentages of total projects and illustrate the importance of computer science projects in FutureGrid’s portfolio. Looking at 200 FutureGrid projects in a two year window 10/11-11/13, there were 136 research projects (others were in education, technology evaluation and interoperability) of which 109 had a major CS component and 44 an application component with 17 of these jointly classified.

98 of the 200 projects only needed access to virtual machines (VM’s) and 54 requested both VM’s and physical nodes. Of the 48 projects not requesting VM’s, 8 studied cloud technology like Hadoop. In total, 160 projects (80%) were cloud related. 16 projects involved GPU access and 30% of all projects used MapReduce.

We identified an initial list of 25 broad cloud computing research areas needing FutureGrid-type capabilities, by pooling our experience with topics studied in these FutureGrid projects (whose numbers are shown in parentheses): Core Virtualization (17); Networking (3); Wireless (0); P2P (2); Cyber-Physical CPS and Mobile Systems (5); Real-Time (0); Storage (2); Distributed Clouds & Systems (8); Resource management (9); Security and Privacy (10); Fault-Tolerance (5); Cyberinfrastructure (11) Programming Models (12); Libraries (5); Data Systems (10); Streaming Data (2); Artificial Intelligence (7); Network/Web Science (3); Software Engineering (2); Education (42, 90% of which on computer science); Open Source Software Testbed (0); Interoperability (3); Energy & Sustainability (0); Domain Science (44); and Technology Evaluation (19).

Application (domain science) projects in the sample of 200 projects, included 18 from bioinformatics including genomics, radiology, cardiovascular simulation, surgery control, health sensors, iPlant cyberinfrastructure and text mining. Only 10 application projects had a simulation (major focus of most HPC systems) focus including combustion, CFD, subsurface modeling, climate, weather, ocean, environment, earthquakes and supply chains. Physical science and engineering data intensive applications (8 projects) include astronomy, particle physics, aerospace reliability, ocean observation, hydroinformatics, Geographic Information Systems and accelerator control. 6 social science projects include conflict resolution, disaster management using Twitter, optimization, political science and economics.

The broader impact of FutureGrid included a major successful effort to support educational activities whose importance had not been appreciated in original FutureGrid planning. In the sample of 200 projects from the middle of the project, there were 42 education requests of which 36 were computer science (CS), 3 application oriented and 3 mixed. In the 42 projects, 36 had a CS only flavor, 3 an application only focus, 3 both application and CS. Education classes covered areas like cloud computing, distributed systems, parallel computing, big data, data­intensive computing and datamining, business analytics, autonomic computing, cyberinfrastructure, storage, software carpentry, data centers and large scale infrastructure, MapReduce, high performance computing, networking, science clouds, and particular tools supported on FutureGrid. We focused on Minority Serving Institutions (MSI) with involvement of student summer research experiences and attendance of MSI's at FutureGrid events in all project years.