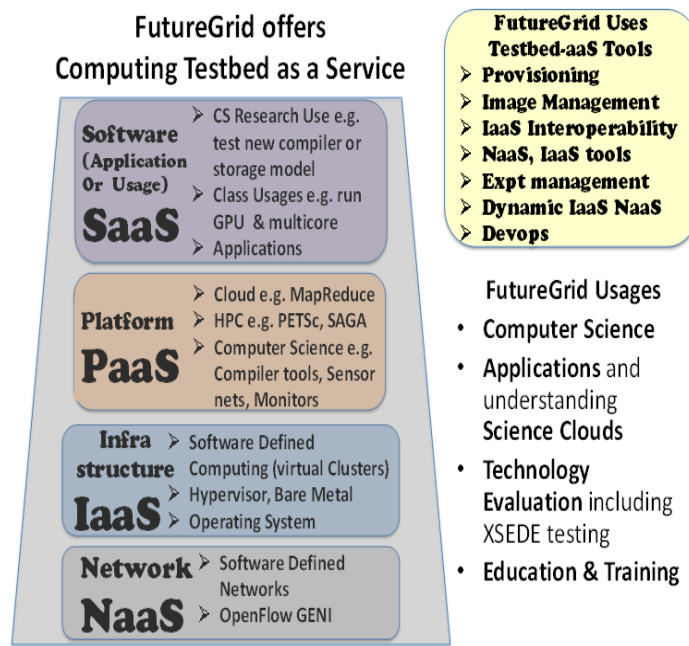# FutureGrid Summary January 2014

Geoffrey Fox (Principal Investigator, Indiana University)

FutureGrid is a national-scale Grid, Cloud and HPC computing test-bed service that includes computational resources at five distributed locations. FutureGrid systems total 4704 cores divided into distributed general purpose clusters at Chicago, Florida, IU and TACC; a Cray XT5m at IU and four small specialized clusters supporting SSD (at SDSC), Large Disk Large memory with and without ScaleMP (at IU) and NVIDIA GPU's (IU). FutureGrid's system model has grown in sophistication and now supports software-defined systems – encompassing virtualized and bare-metal infrastructure, networks, application, systems and platform software – with a unifying goal of providing a Computing Testbed as a Service (CTaaS). The core of FutureGrid's provisioning and management middleware, called Cloudmesh, aggregates resources not only from FutureGrid, but also from OpenCirrus, Amazon, Microsoft Azure, and HP Cloud and GENI resources. Cloudmesh was originally developed to simplify the execution of multiple concurrent experiments on a federated cloud infrastructure, offering both virtual resources and bare-metal provisioning to users.

Users can apply for projects either at FutureGrid or XSEDE portals. A single account provides a user access to all FutureGrid machines - evidence of the type of work that can be done is shown by FutureGrid's experience over 3 and a half years of operation with 355 projects and 2178 users from 55 countries (76% from the USA). 51.8% of these projects related to computer science research (including middleware and cyberinfrastructures), 9.6% to technology evaluation, 13.2% to education, 9.6% to life sciences and 11.5% to other application domains.

We looked in detail at the last 200 projects starting October 25 2011 to understand in more detail usage paradigms. 98 of these projects needed only virtual machine (VM) access and 54 requested both virtualized and non-virtualized nodes. Of the 48 projects not requesting VM's, 8 studied cloud technology like Hadoop and so 160 projects (80%) were cloud related. 16 projects involved GPU access and 30% of all projects used MapReduce in some way. The use of FutureGrid for education has been increasing - 21% of projects in last 2 years have been for education dividing into 29 semester length classes and the 13 remaining split between REU training, Summer Schools, Tutorials, and workshops. Of 42 education requests, 36 were computer science, 3 application oriented and 3 mixed. These education classes covered areas like cloud computing, distributed systems, parallel computing, big data, data-intensive 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.

Coming to the 136 research projects, 109 had a major CS component and 44 an application component with 17 of these jointly classified. Application 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, GIS and accelerator control. 6 social science projects include conflict resolution, disaster management using Twitter, optimization, political science and economics.

Turning to CS-related projects, 23 were in basic virtualization (IaaS) areas. They included provisioning, deployment, new hypervisors including increased performance, elasticity and scheduling, resource management, benchmarking, emulation, and IaaS scaling to support MOOC's. 4 projects studied distributed clouds and storage with federation. 3 projects centered on networking (routing, optimized devices and emulation) and 2 studied software engineering for clouds. 5 projects were aimed at cyberphysical systems with health, power and mobile applications and FutureGrid used for control and support analytics. 2 projects had a P2P focus covering security and fault tolerance. Security was popular with 10 projects covering trusted P2P storage, mobile and other clients, intrusion detection, file sharing, confidentiality and integrity of data, vulnerability, watermarks and hybrid clouds with different security models. There were 4 HPC programming language projects and 8 covering cloud programming with scheduling, fault tolerance and runtime. 5 fault tolerance projects covered P2P, MapReduce, workflow and HPC. 4 projects covered distributed software transactional memory and concurrency control. Data systems were very popular with 13 projects covering cloud storage, data transfer, NoSQL, streaming big data, Apache software stack, analytics, image retrieval, testing, provenance and the semantic web. 2 projects studied enterprise software issues. 9 artificial intelligence projects covered learning networks for images and social media, large scale image classification by clustering, machine learning, text mining, and agents. Network science with 3 projects saw use of NoSQL datastores to study Twitter, a major infrastructure to support graph and other tools and community detection. Finally 13 projects focused on cyberinfrastructure middleware with XSEDE and EMI (European Middleware initiative) systems, software as a service for HPC simulations, HPC clouds, MPI porting, fault tolerance, workflow, scheduling and resource management, tools, logging and performance.

There were 19 Evaluation projects which covered XSEDE testing (an important actual and intended use of FutureGrid), Open Science Grid testing, particle physics, Apache Big Data stack, Solid State disks, comparison of different VM frameworks, familiarization with cloud technology, GPU's and studies aimed at planning institutional initiatives in cloud computing. There were 3 Interoperability projects involving long term support of standard end-points.

FutureGrid interacts with XSEDE on integrating accounting approaches, EOT and software testing.