On a summery day in Pleasanton, a large group filled our room to hear four great speakers. BMC Software hosted our event, providing a great breakfast, lunch, technical speaker and a product presentation.

We began the day with Joel Smith, SAS, who has had a long career in IT beginning as a systems programmer, migrating to the vendor world to do benchmarking, disk and supercomputer work at Amdahl. He has been at SAS for 12 years. His talk, “What is the Big Deal about Big Data,” was particularly timely given that “Big Data” is the big buzz, but many do not understand the implications. Smith began by talking about big numbers, such as 200PB of data held by Yahoo and the SmartGrid where data is collected on every appliance in your house resulting in the equivalent of 10MM servers. “Big” means the point at which the volume, velocity and variety of data exceeds an organization’s storage or compute capacity for accurate and timely decision-making. With that much data, how do you mine it? This information can provide such value as telling you what customers are saying about your product/service, help project sales and define such metrics as stock to be held available or optimal routing schemas.

Smith then moved into defining how this data is stored, accessed and analyzed differently, given software such as Hadoop. Now that memory is cheaper, we can afford to store more data in memory, reducing access time. He defined various options for getting value from your data, giving specific examples from major internet companies. As an example, Yahoo has 42,000 Hadoop nodes over 20 clusters holding 200 PB of data.

Next is figuring out how to manage this data from a performance basis. He spoke about Ganglia, an open-source, scalable, distributed monitoring system. It monitors clusters and grids providing live and historical stats. Built on Apache, it will auto-discover nodes. Listeners pick up metrics from other nodes. This is information we need to know because as big data grows, IT will have to manage it.

Our next speakers were Larry and Coburn Watson, a father-son team representing An-de-le Computing. Larry provides consulting services and his son, Coburn, is a key player in the Cloud group at Netflix. Both have extensive resumes and education as scientists, so their topic, “Improve Performance Testing by Employing the Scientific Method,” was grounded in their personal experience. They showed us how a typical performance test might be conducted and the results that were discovered. Then, after showing us how the scientific process works (formulate a hypothesis, design the tests, then execute them to prove/disprove the hypothesis), they explained the differences in the results they obtained. When you do it this way, you discover the flaws inherent in the original testing design, as well as discovering ways to make the testing more accurate, effective and efficient.
At each step of the way, they showed the fallacies we too often buy into, such as not understanding that if you want to have a certain transaction mix you have to factor in the time it takes to execute the transaction. There is a great deal of important detail to mine – see their paper for the complete story.

As always in November, our business meeting was focused on Board elections. A slate was presented and each candidate was elected unanimously.

Vice-chairman – Don Murray
Treasurer – Keith McAndrew
Directors: Mel Boksenbaum, Bill Jouris
Rob Creighton stepped up to fill Don’s director position

After grabbing lunch, we came back to hear Randy Knaub, BMC Software, talk about Proactive Operations – why you need it and how BMC can solve this challenge for you with integrated solutions.

Jie Lu, BMC Software, began our afternoon session with “CPU Measurement Inside the Virtual Machine.” Lu has been working on the challenges of performance management and capacity planning for BMC since ‘97 and has noted that over the years measuring CPU has become much more difficult, though it is still a critical metric to measure. The OS knows the number as it is the layer closest to the hardware. But virtualization inserts an abstraction between the hardware and the OS so the number is not as accurate. How do you make it better? Many vendors claim to have fixed this problem, but Lu believes the number is still disputable.

He devoted some time to helping us understand how CPUs measure time, the options they have and why the numbers are not all that useful. As we can’t modify the guest OS, the challenges are high. For some of us, this brought back memories of when MVS began to run under a hypervisor – the first time we realized we might not know as much as we thought about CPU measurement.

We were happy to welcome back Michelson Award winner, Adrian Cockcroft, Director of Cloud Systems Architecture for Netflix. He spoke about “Cloud Capacity Planning and Monitoring at Netflix.” One of the great aspects of hearing a real customer story is that we can all relate to the underlying business problem. Cockcroft began by giving us background on how Netflix is structured – what work runs where. At Netflix the goal is extremely fast product development, which means investing in this effort as well as accepting a certain amount of risk. This drove them to the Cloud. Once you are there, certain challenges like classic capacity planning are eliminated. They chose to run corporate IT on SaaS, built their own PaaS to make development more products and now use the cloud to run their movie streaming applications. Amazon is their cloud provider, and they use Amazon Hadoop to manage their large data stores. To manage their systems, they rely on AppDynamics to get real-time measurement and corrective action.

Everything is backed up in threes. Testing is constant to keep availability high. A chaos “monkey” takes things down, and the latency “monkey” slows things and introduces errors. A key feature is “hinting” where if one part goes down, the component holding the data marks it and sends it when possible.
He continued with the challenges of cloud management, especially given that with cloud computing, IP addresses change all the time, but most tools rely on the consistency of the IP address. A gee-whiz number is that they have been able to log 1MM writes/sec on Cassandra, a key system.

Another innovation is that they use open source code which does two things. One, the developers keep their code very clean because others will see it, and the Apache OSS community contributes ingenuity to further speed innovation.

Heads up on futures – soon, there will be a Netflix button on your remote.

After National CMG, our next meeting will be February 5, 2013; location TBD.

Denise P. Kalm
NCCMG Secretary
Notes scribed on my iPad – what a great device!