# CSL862 Major

Answer all 5 questions          24/11/2011          Max Marks: 36

1.  **The Google File System**
    a.  What effect (if any) will the following have on the performance/reliability of GFS. Provide a qualitative assessment (whether the effect will be high/low/unnoticeable/etc.)

        i.  Write-back versus Write-through Linux Filesystem Buffer Cache. Comment on both performance and reliability/durability. [3]

        ii. Improvements in TCP/IP Networking Stack for pipelining data to chunk replicas [1]

iii.       Larger L2 cache in each machine  [1]

b.  In Figure 3:
     i.        How is the "Network Limit" calculated?  [2]

     ii.      Why is aggregate read rate higher than aggregate write rate?  [1]

iii. Explain the behavior of the curve of "aggregate append rate" even with increasing number of clients.   [1]

c. Consider the following command executed on "/gfs" which mounts a Google File System:

    find   /gfs   -name "foo.c"

Explain what happens on GFS when this command is executed on a GFS client. Also, comment on it's performance.   [2.5]

2. **BigTable: A Distributed Storage System for Structured Data**
   a. What happens if a tablet server hangs?   [2]




   b. What happens if the table grows too large to fit in one tablet?  [1]

c.  Why does a tablet server use a commit log ("tablet log")? Why not simply make the change directly in local persistent storage? Why not simply make the change in the GFS file?  [3]

d.  In Figure 6 (right-hand side) graph, why does aggregate rate not scale linearly with the number of tablet servers?  [1]

e. "If Chubby becomes unavailable for an extended period of time, Bigtable becomes unavailable". Give two concrete examples of what could go wrong if Chubby becomes unavailable. [2]

### 3. MapReduce

Recently, the Computer Services Center (CSC) has setup a cloud computing service called Baadal. Here are the specifications of Baadal as mentioned on CSC webpage:

a. 32 blade servers each with 2x6 core Intel(R) Xeon(R) CPU X5670 @ 2.93GHz and 16 GB RAM.
b. 16 blade servers each with 2x4 core Intel(R) Xeon(R) CPU E5540 @ 2.53GHz and 12 GB RAM
c. A 10Gbps ethernet backbone
d. 50 TB of virtualized storage based on a NetApp 3210V NAS and HP EVA6400 SAN with FC disks.
e. Open source virtualization technology based on KVM.
f. Baadal: an indigineously developed cloud orchestration and virtualization management software that can work with multiple virtualization technologies like KVM, Xen, and VMware

Professor X works in the area of databases and is highly interested in Cloud Computing and decides to use Baadal for his research. He sets up 19 GFS VMs each with two VCPUs, 2GB memory, and 80GB virtual disk. He then starts to use GFS/MapReduce to implement his data-mining algorithms. His program executes on all VMs simultaneously. Does this setup make sense? Why or why not? [3]

b.  Give a one-line explanation for each of the following design decisions in MapReduce


    i.    The MapReduce library in the user program first splits the input files into M pieces of typically 16 to 64 MB per piece.  [1.5]


    ii. The outputs of the Map function are written to local disk (not GFS)   [1.5]

iii. On failure, completed reduce tasks do not need to be re-executed  [1]

c.  How will Figure 2 change if MapReduce used NFS instead of GFS?   [2.5]

## 4. Nested Virtualization

    a. What happens when L1 (guest hypervisor) executes vmlaunch?  [2]

    b. Figure 9: Why is Multi-dimensional paging (aka EPT-on-Shadow) performing better than Shadow-on-EPT?  [2]

c.  In Section 4.3, the paper analyzes the flow of handling the cpuid instruction. Briefly explain what is happening in Step 5:

   L1 modifies VMCS1-2 (repeat n-times)

[2]