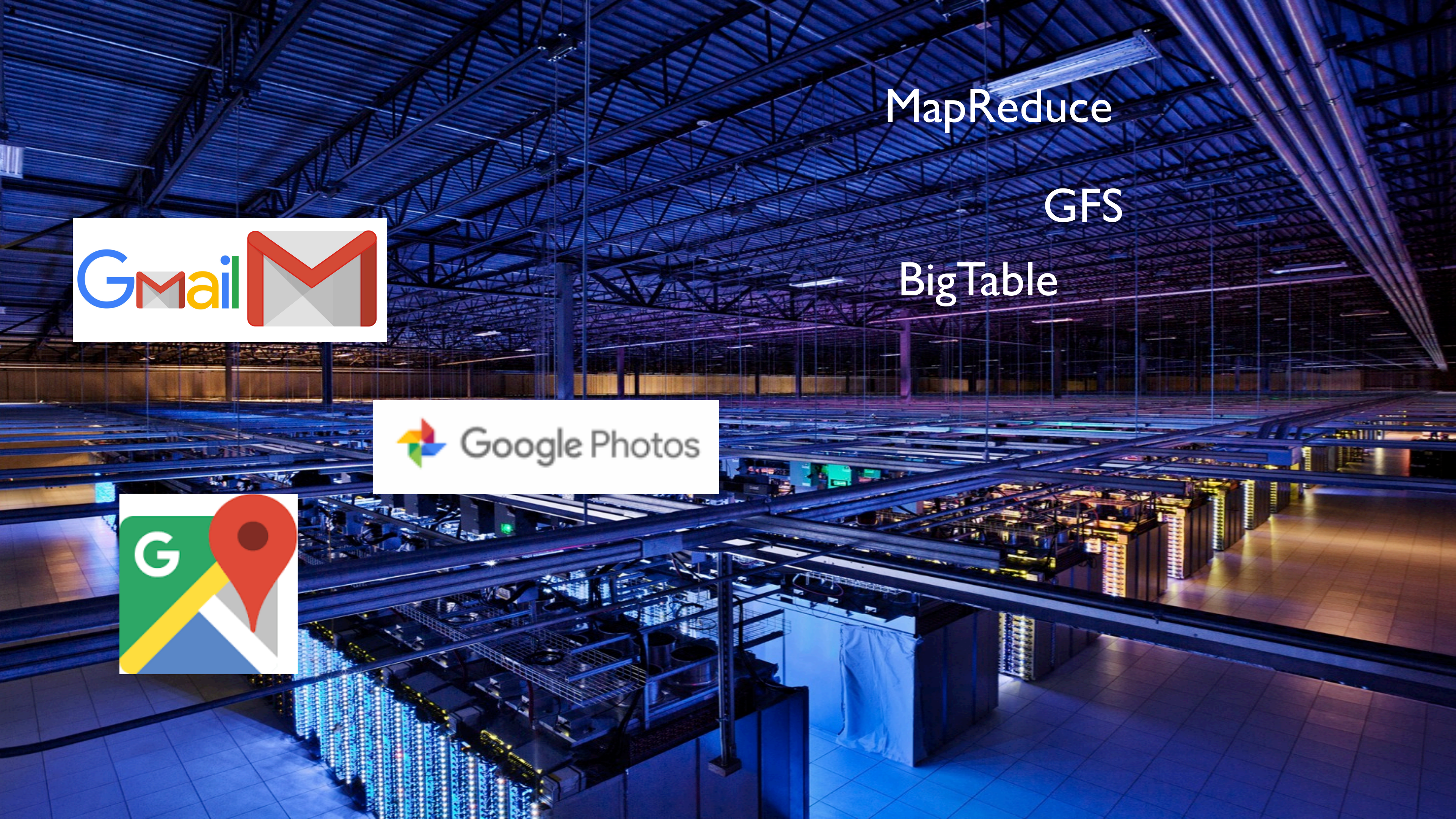
CS 744: BIG DATA SYSTEMS

Shivaram Venkataraman

Fall 2018

ADMINISTRIVIA

- Assignment 1: Due Oct 1
- Sign up for Project meetings
- Group updates



MapReduce

GFS

BigTable

BORG: WORKLOAD

Long-running services (should “never” go down)

Batch jobs: few seconds to a few days

BORG CONCEPTS

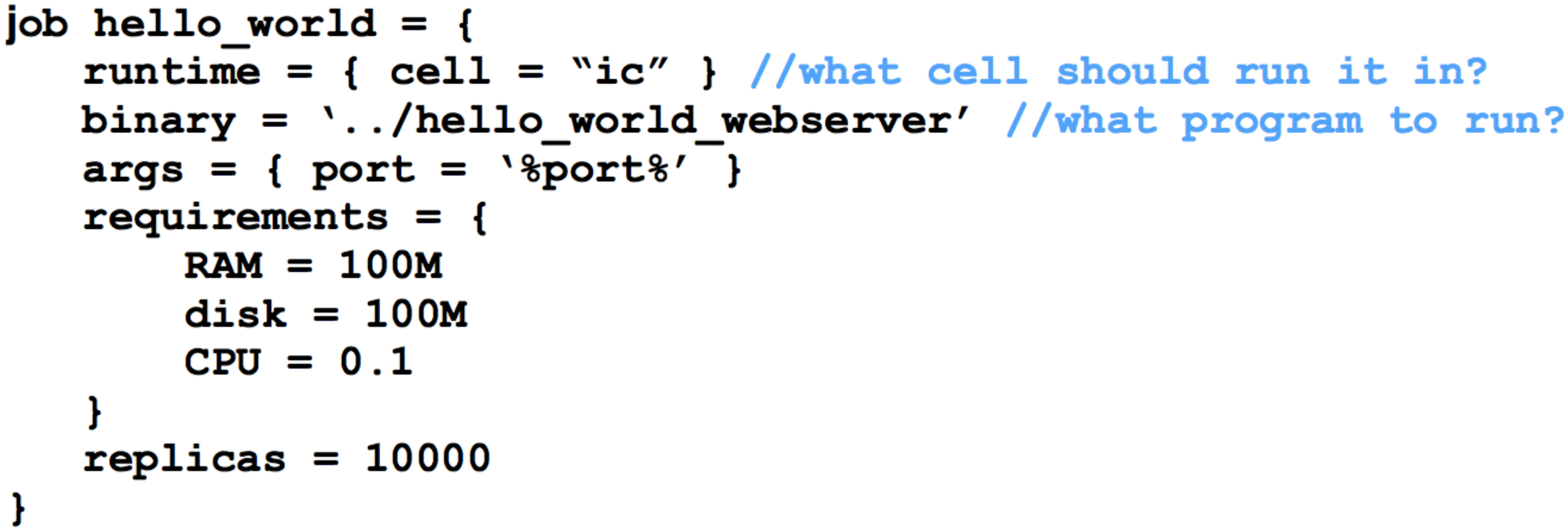
Users submit jobs

Each job is one or more tasks

All tasks that run the same program (binary)

Each job runs in one Borg cell

JOB DESCRIPTION



```
job hello_world = {
   runtime = { cell = “ic” } //what cell should run it in?
   binary = ‘../hello_world_webserver’ //what program to run?
   args = { port = ‘%port%’ }
   requirements = {
       RAM = 100M
       disk = 100M
       CPU = 0.1
   }
   replicas = 10000
}
```

JOB PROPERTIES

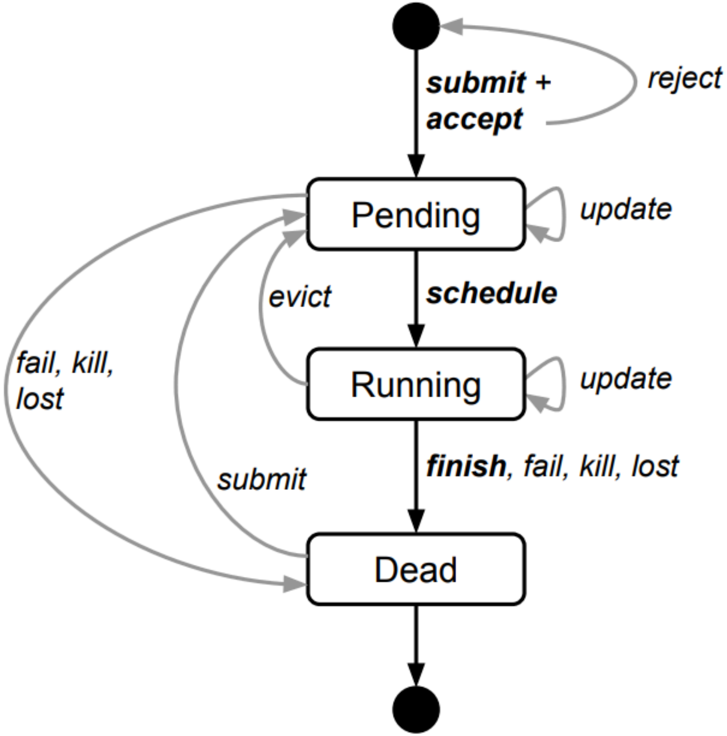
Name

Constraints

Properties

- Resource requirements
- No slots!
- Static Binaries

JOB LIFECYLE



QUOTAS, PRIORITIES, BNS

Priority

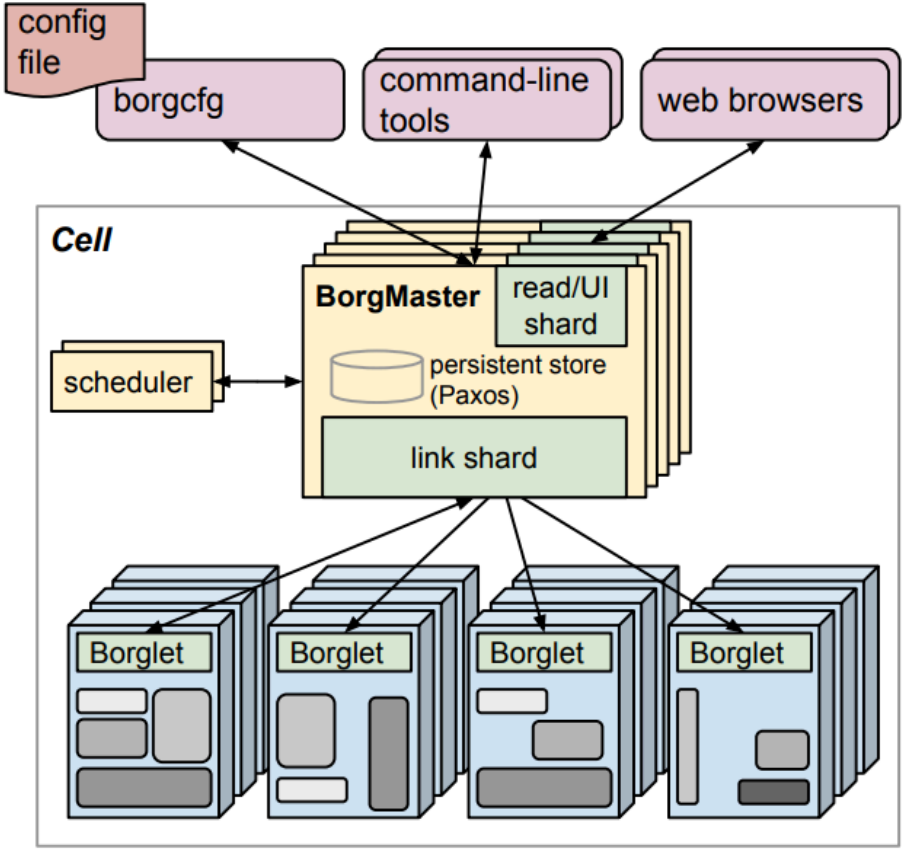
- High priority can preempt lower priority

Quotas

- Used for admission control
- Infinite quota at priority zero

Service Discovery using BNS

ARCHITECTURE



MASTER, BORGLET

BorgMaster

- Single Leader, five-ways replicated
- Paxos group – using Chubby locks

Borglet

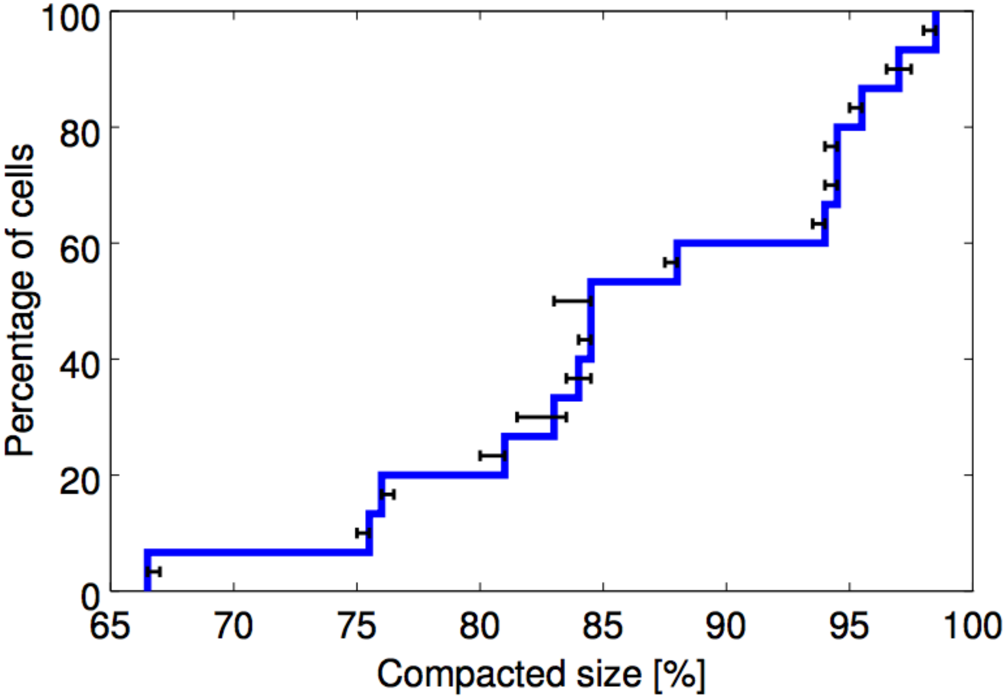
- Daemon on each machine
- Borgmaster pulls updates from Borglets
- Health checks used to detect failures

SCHEDULER

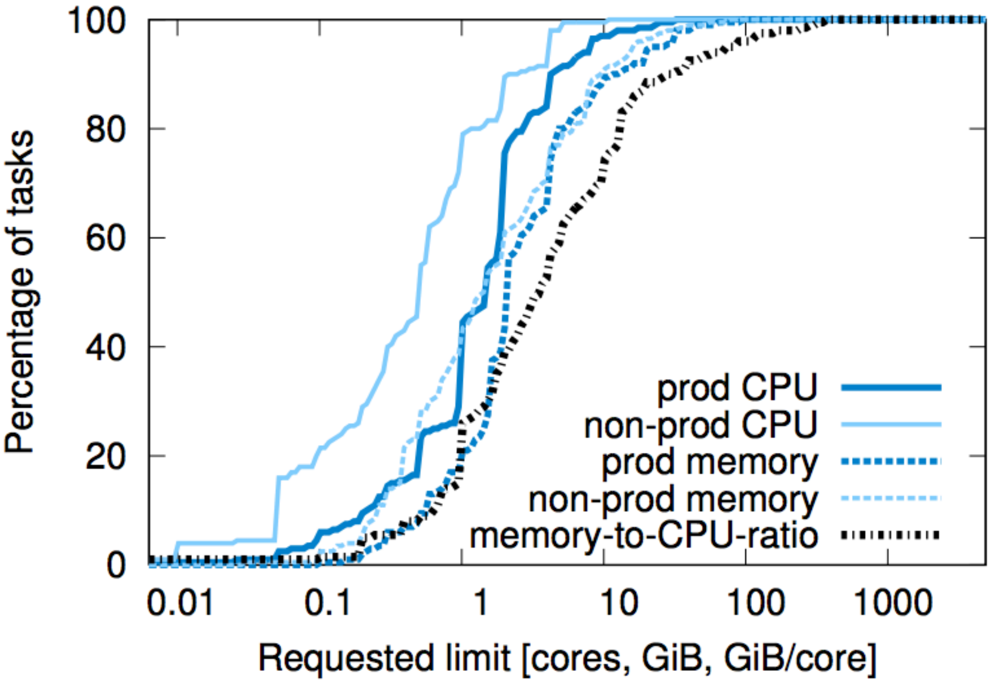
- Feasibility checking pass, Scoring pass
- Task cache (static binaries)
- Scalability
    - Split master into multiple processes
    - Use replicas for communication
    - Randomize machines used for scoring

    …

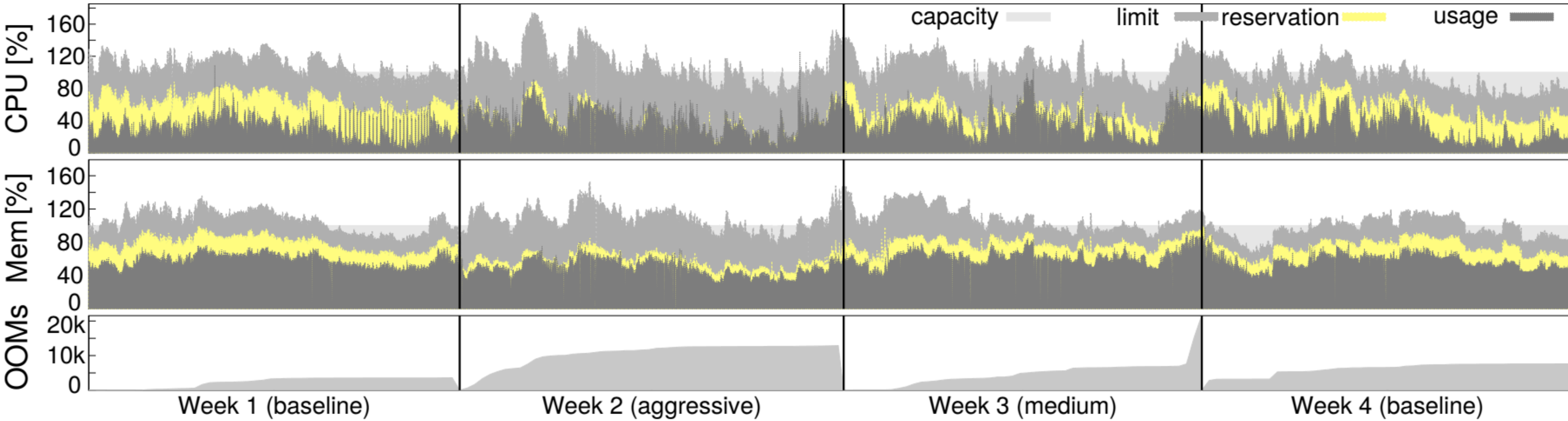
UTILIZATION: CELL COMPACTION



REQUEST SIZE: NO SWEET SPOT



RECLAMATION



LESSONS, DISCUSSION

- Jobs are restrictive, Allocs are useful

- IP address per container

- Kernel of distributed operating system

QUESTIONS / DISCUSSION ?