Schedulers In Practice

(LEC 4)

HW1 Is Out

Announcements

- Final Project Proposals Are Due In 1 Week
  - Only heard from a couple of people thus far?
  - Can get a 1-week extension by e-mailing me with
    - Current sketch/thoughts on possible directions
    - Team composition
    - What is blocking progress

- Proposal format
  - Up to 5 pages (shorter is better)
  - Sec 1: Problem statement + motivation: WHAT & WHY
    - Sec 2: Proposed work: HOW
    - Sec 3: Eval plan: How will you test the result
    - Sec 4: Timeline: WHEN

- Group dynamics

Jobs/Tasks...
Q: What do we run on compute clusters?

→ Q: What is a compute cluster?

Q: What do we run on compute clusters?

Interactive Jobs Differ Significantly

*Assumption about lifetime
- **Resource Requirements**

- **Latency Requirements**

- **Programming Model & Assumptions About # of Machines**

---

Q. **What Do We Run On Compute Clusters?**

```
DATABASE/STORAGE

REVENUE

VIDEO UPLOAD

PAGE RANK

WEB CRAWLER

ENCODE

'Batch Jobs'
```
Batch is willing to wait.

Batch jobs also differ significantly.

Compare interactive vs. batch to Borg levels:

- Monitoring
- Production
- Batch
- Best Effort/Testing

Scheduling: Not just for jobs.
Dataflow Graphs: A Common Paradigm for Writing Data-Intensive Jobs

1. Read(filename).
2. Map(1 word1 (word, 1)). [key]
3. Reduce by key(1 k, v1, v21 (k, v1+v2)).
4. Write (cout)

Trade-offs for when each step starts & stops

Makespan
- TensorFlow
- Stream Processing (e.g., Materialize)
- Database Query Processing

Scheduling Tasks In A Dataflow Job Requires Semantic Information

- Dependencies
- When Can It Start Executing
- Expected Duration
- ...

- Trade-offs Around Semantic Awareness
End Result

- 2-Level Scheduling / "Extensible Scheduling"

TRADE-OFFS?
Putting It All Together / Terminology

Alice  

Beth  

Worker  

Worker  

Worker  

Worker  

Worker  

Worker  

Worker  

Worker  

Scheduler  

Resource Allocator  

PAS  

PBS  

Fairness / Policies In This Model

Q: What Is Fair In This Model?

Q: Interaction B/w Two Levels?
Borg: Priorities Instead of DRF?

Miscellaneous:
- Task Lengths in Mesos
- Why Care About Task Length
Incentivizing Short Tasks?

Today's Reality

Resource Revocation
Isolation Overheads