with commercial

on Mone-hous Accel / demo under

data-intensive

distributed + parallel

for programing

Problem:

Show

Sanjay Goemans, Tom Leighton

The Goemans

into GFS (Next Class - continue)

Help Reduce Examination Model

Agenda

CSE 533, Fall 2015
Features of Warehouse Datacenter

The Datacenter as a Scale Computer

- An intro to the Design of Barredo & Holzle

- Big is the Network

- Expensive, Efficient, Simple

- Commodities

- Cost Effective

- Every thing is becoming

- Rack

- Chp

- Infra

- H/w

- T/w
- S/w (system) plays a very important role.
- fault tolerance
- distribution of computation
- synchronization (atomicity)
- energy/power usage
- load balancing
- programmer productivity

Map Reduce (Dean & Ghemawat).
- Page Ranking $m(0.2)$
- $n_1(0.2) \rightarrow n_3(0.2) \rightarrow n_5(0.2) \rightarrow n_0(0.3)$
- aggregation (Reduce)
each m.r. phase/molecule

(See later)

(See later)

- Treat at high weight
- Reduce - Chernod
Execution
Minimum on file and more

- Find another replica of the
  relief of each charge

- Write on GPS - multiple

- Happier feel (familial detection)

E.T. - Anything can fail! 

- Optimizations to reduce communication

- Ensure correctness (check pedal)

- Ensure health - Rehabs / 
  Located

- Heaviness / Load bearing

- Fault tolerance
- S test
- S test
- R test produced by a different o/p for a different R, may correlate

A test. executed. However the "o/p of a part" when Fe
- weakens guarantee
- non-deterministic m.r.
- deterministic m.r.
- same o/p as

Same thing in presence of failure:
Big ideas behind M-R. (Ref: Data-Intensive Text Processing with Map Reduce J. Lin & C. Dyer)

- Scale "out", not "up"
- Assume failures are common
- Move process to data
- Process data sequentially and avoid random access
- Hide system level details from application developer
- Seamless scalability
  - data: if given twice amount of data, processing time should double
  - resource: give twice the amount of resource, the computation time should half