

UF | UNIVERSITY *of*
FLORIDA
The Foundation for The Gator Nation

The Elf On the Yellow Elephant

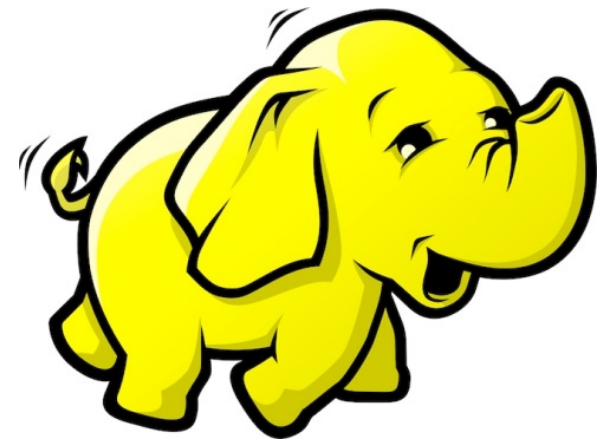
Internet2 Global Summit

Ze Yu@UF



The Yellow Elephant

- Big Data
- Hadoop – The Yellow Elephant
 - Standard tool to process Big Data
- Become the “OS” of the datacenter of Big Data

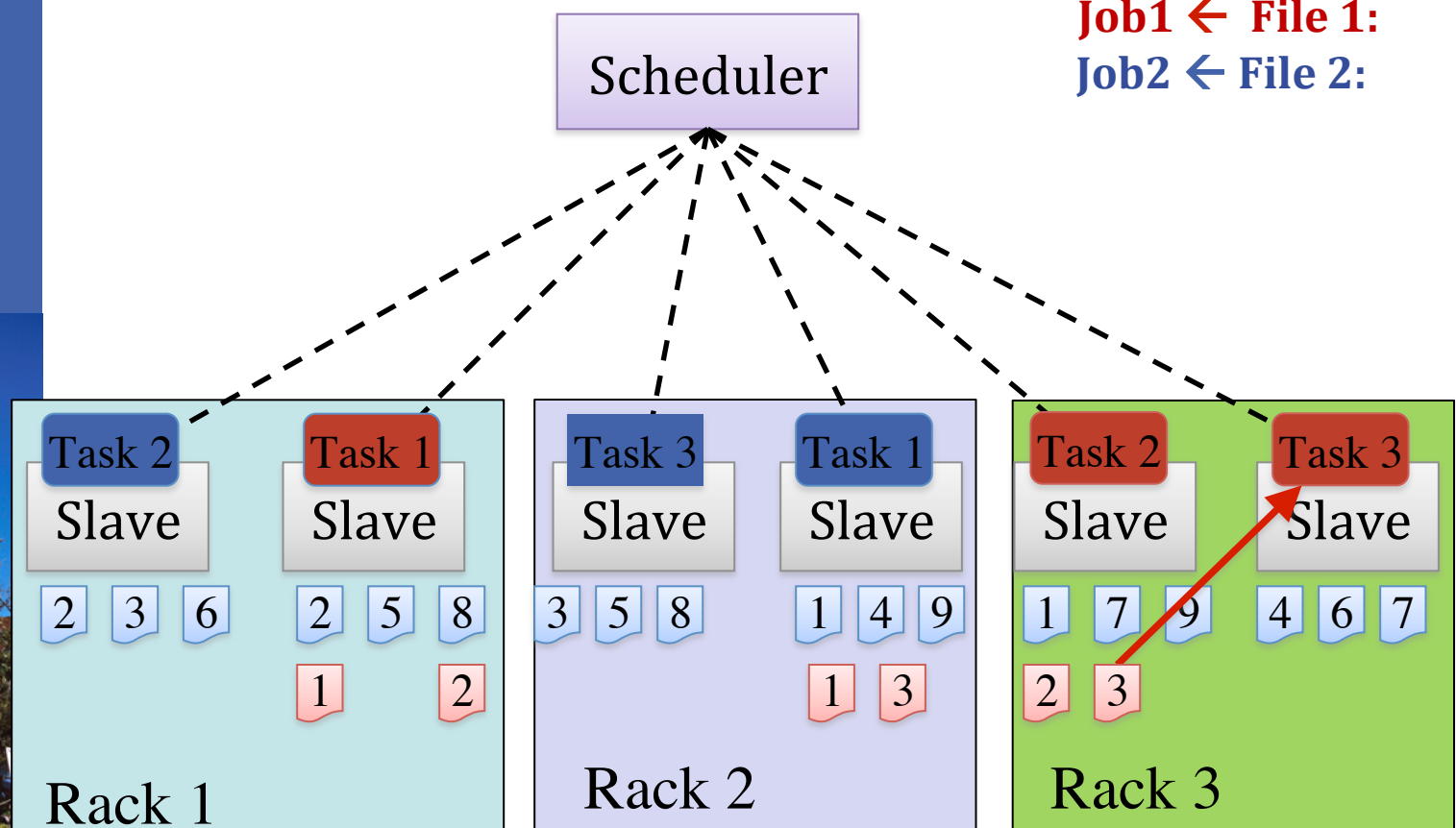


The First Principle of Hadoop – Locality

- Do not move data around
 - Since they are big...
- Instead, move computation to the data



Locality-Aware Scheduling



Locality Is Hard To Get

- It is a network related property
- The datacenter network is a black box to the big data
 - Inter-datacenter locality-aware deployment is a mission impossible
- Currently, using manual configuration



Summary

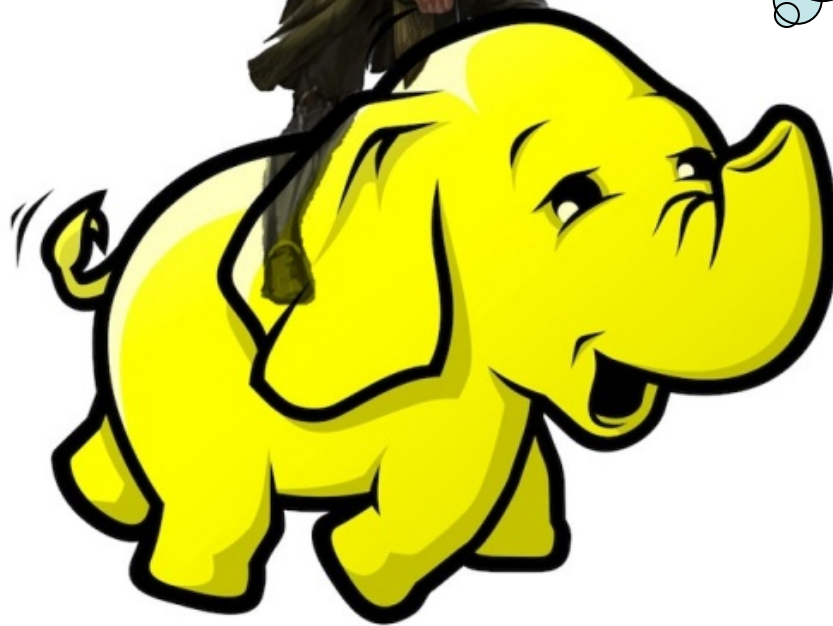
- If we consider Hadoop as the “OS” of datacenter for Big Data, the network is the *missing piece* of this OS





Elf

Using Software
Defined Network to
manage network for
Big Data



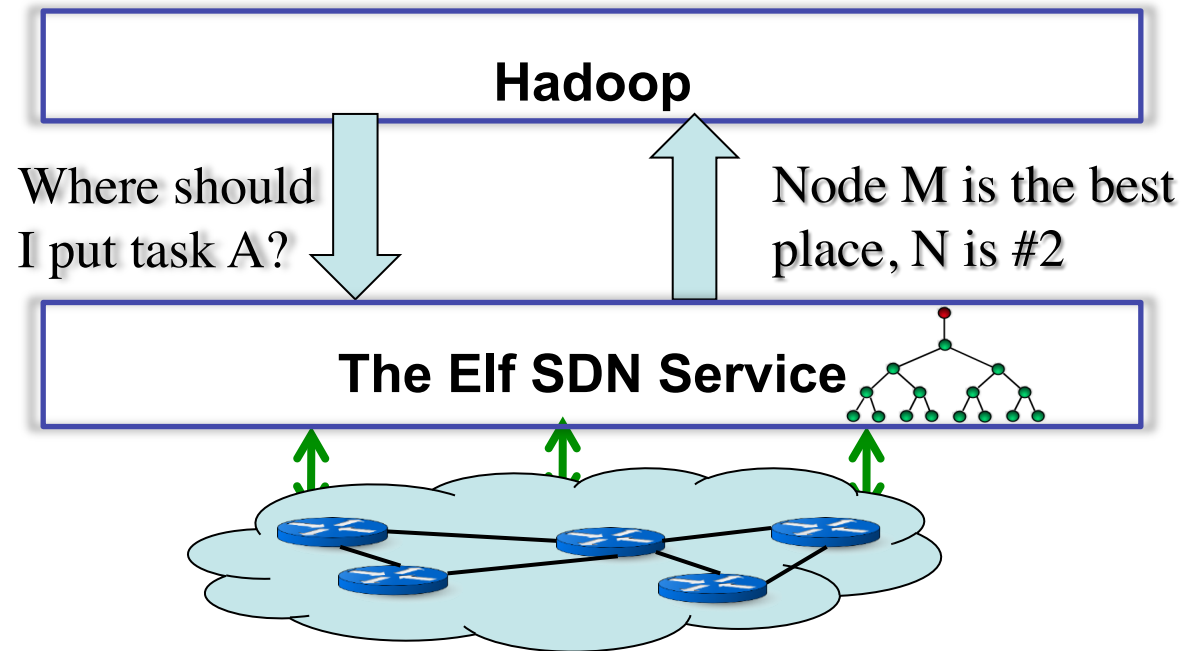
Why SDN

- Programmable control
 - No more manual work
- Global view
 - Knows everything about the network
- Easy for integration
 - The rendezvous of compute & network
- Flexible
 - Handle the dynamics of the network



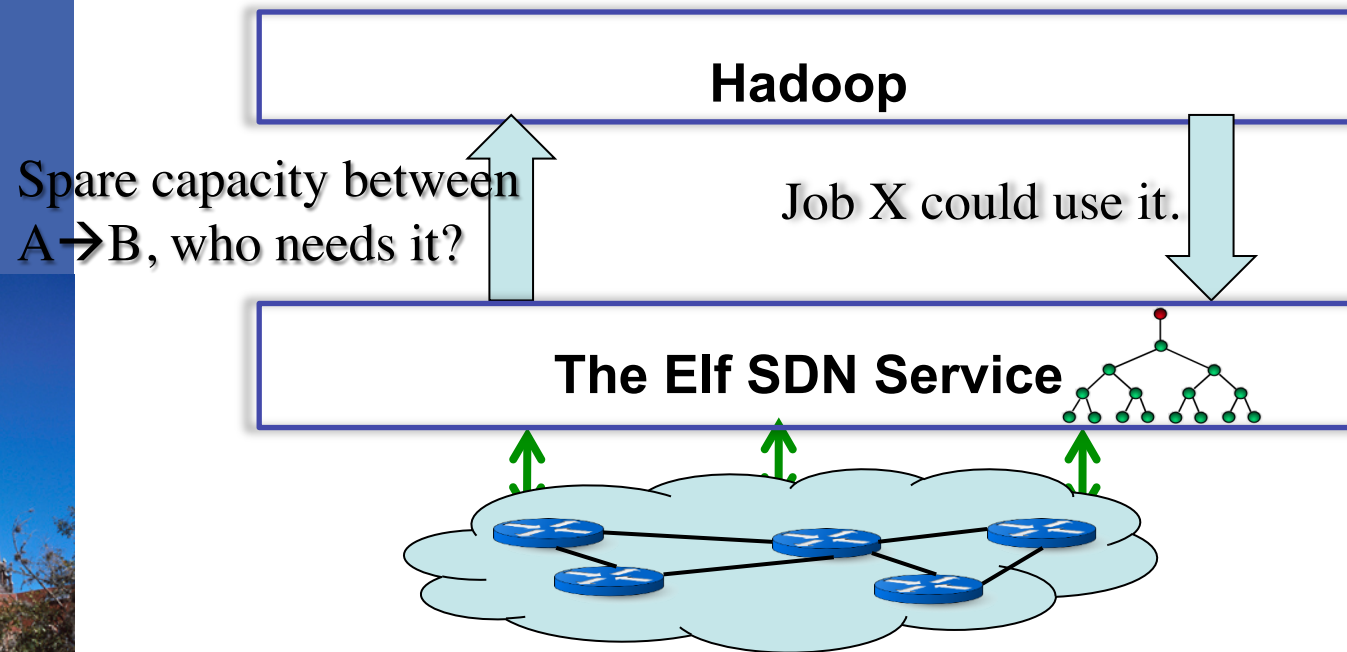
Elf - Example

Locality-Aware Computation



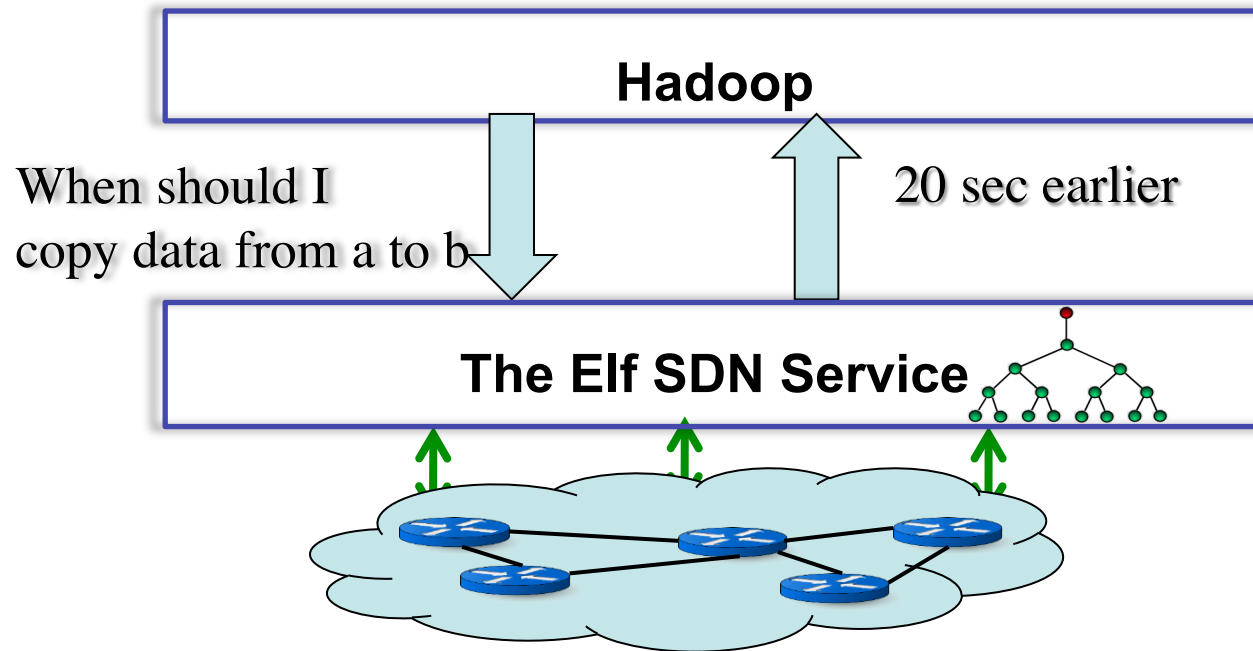
Elf - Example

Network-Driven Computation



Elf - Example

Timely Data Prefetching



Takeaways

- Big data ecosystem relies more and more on being network-aware
- SDN can help Big Data as its network controlling knob



Thanks



My Wish List of SDN

- More performance-related control primitives added to SDN
 - Bandwidth control e.t.c
 - Both northbound and southbound API

