Bandwidth-Aware Scheduling With SDN in Hadoop: A New Trend for Big Data

Zheng Luo

December 14, 2016

Quick look

This article:

- proposed a scheduling algorithm for Hadoop based on SDN
  - which takes *link bandwidth* into consideration
- on the assumption that
  - nodes are homogenous
  - Finish time of task is known
  - Bandwidth is known
- in order to minimize the completion time of the whole job

Previous works

Most of them(BAR, etc.) assume that:

- A job consists of m equivalent tasks
- There are n nodes, of which some are local and some are remote. Each has different current load L_s^{init}
- On a local node, a task requires C_{loc} to finish
- r^α remote tasks require $C_{rem}(r^\alpha)$ time to finish, where $C_{rem}(\cdot)$ is a monotone increasing function(network load) ($C_{rem}(r^\alpha) = r^\alpha(C_{loc} + q)$ in its experiment)
- Target:

$$\begin{aligned} &= \min \max_{s \in active\_servers} finish\_time_s \\ &= \min \max_{s \in active\_servers} L_s^{init} + \sum_{a(t)=s} C(t) \end{aligned} \tag{1}$$

Architecture

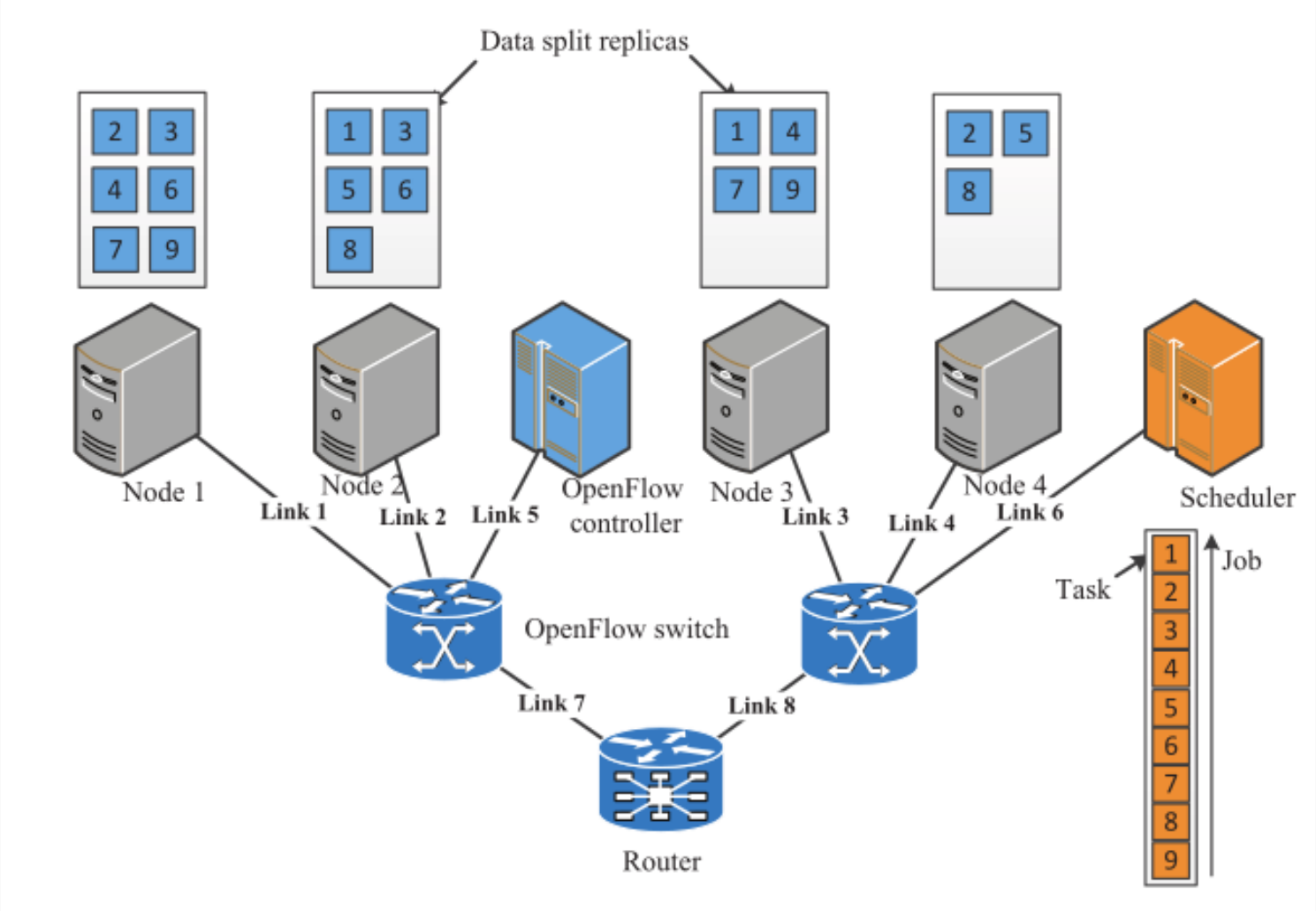

Figure 1: Topology of a Hadoop cluster centrally controlled by SDN

This paper's model

- Time to finish a task j in remote node: $C_{loc} + Size_j / Bandwidth_{src,j}$
- Use time slot method to share link

Algorithm

```
tasks.each |task|
    nd_loc = find_earlist_idle_time_point(local_rack.nodes)
    nd_all = find_earlist_idel_time_point(all.nodes)
    if nd_loc == nd_all
        assign(task, nd_loc)
    else
        delta = nd_loc.idle_point - nd_all.idle_point
        if possible_to_allocate_slots(source, nd_all,
            nd_all.idle_point,
            task.size, delta)
            allocate_time_slots(
                source, nd_all, nd_all.idle_point
            )
            assign(task, nd_all)
        else
            assign(task, nd_loc)
        end
    end
end
```

Experiment

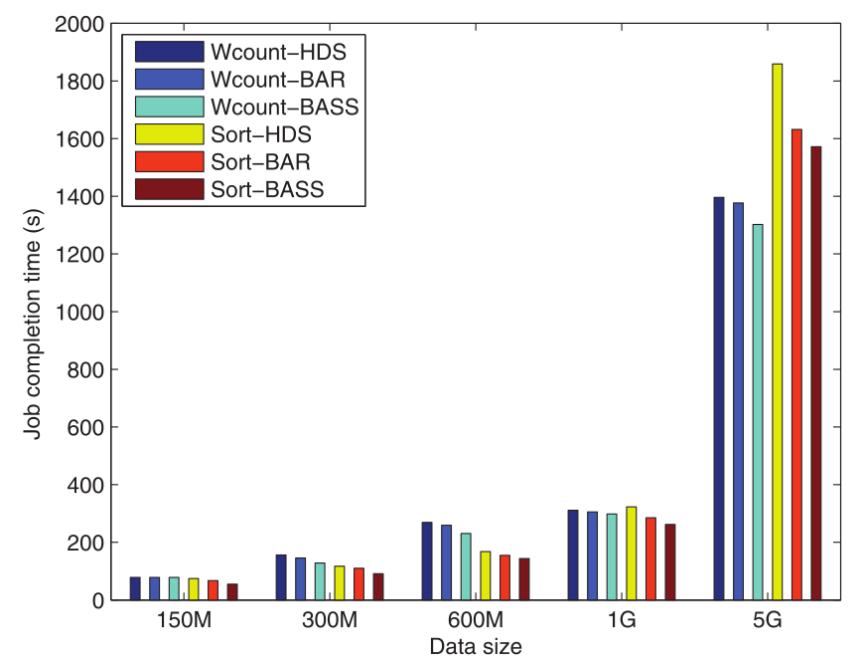

Figure 2: Job completion time for both wordcount and sort jobs

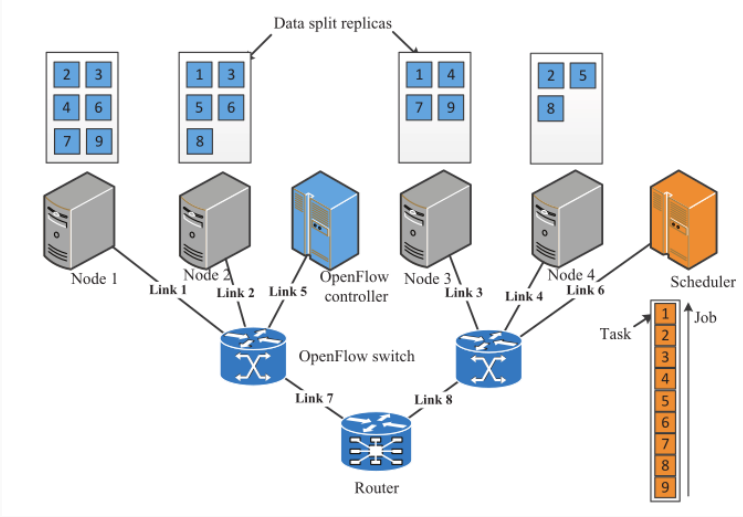

Figure 3: Topology of a Hadoop cluster centrally controlled by SDN

Drawbacks

- Always utilizing full residue bandwidth may not be optimal
  - Use SDN Queue to limit rate? (described in analysis)
- Local optimal != global optimal
  - Add critical-path penalty for node selection?
  - Randomized adjustment? (BAR algorithm)

Thanks for listening

Bandwidth-Aware Scheduling With SDN in Hadoop: A New Trend for Big Data; Peng Qin, Bin Dai, Benxiong Huang, and Guan Xu; IEEE Systems Journal, March 2014