Use of Big Data Analytics Jobs is Growing
Impact of communication on job performance

Facebook jobs spend **25%** time in communication![1]

[1] Based on information from full Facebook trace used in Aalo. Aalo Sigcomm’15 slides.
Impact of communication on job performance

Facebook jobs spend **25%** time in communication![1]

[1] Based on information from full Facebook trace used in Aalo. Aalo Sigcomm’15 slides.
Coflow abstraction (HotNets-XI)
Coflow abstraction (HotNets-XI)

**Coflow:**
Collection of all flows that share the same goal
Coflow abstraction (HotNets-XI)

**Coflow:**
Collection of all flows that share the same goal

[A Map Reduce job]

Coflow abstraction (HotNets-XI)

**Coflow:**
Collection of all flows that share the same goal

A Map Reduce job

---

Coflow abstraction (HotNets-XI)

Coflow:
Collection of all flows that share the same goal

Coflow Completion Time (CCT):
Completion time of its last flow

Coflow Scheduling Problem

• Coflow scheduling problem
  • Minimize average Coflow Completion Time (CCT)
State-of-the-art Online Coflow Schedulers

Saath (CoNEXT’17), Graviton (HotCloud’16) and Aalo (Sigcomm’15)

<table>
<thead>
<tr>
<th></th>
<th>Q_2</th>
<th>Q_1</th>
<th>Q_0</th>
</tr>
</thead>
<tbody>
<tr>
<td>t_1</td>
<td></td>
<td></td>
<td>C_i</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t_2</td>
<td></td>
<td>C_i</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t_3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Low Priority (Larger slots)
- High Priority (Smaller slots)
State-of-the-art Online Coflow Schedulers
Saath (CoNEXT’17), Graviton (HotCloud’16) and Aalo (Sigcomm’15)

Low Priority (Larger slots)

Q1
Q0

C_i
C_k

Global Co-ordinator

DC Network

Sender Node-1
Receiver Node-1

Sender Node-2
Receiver Node-2

High Priority (Smaller slots)
Drawback of Learning by Priority Queue

• This approach in essence is “try-and-miss”
• Every coflow starts from highest priority queue and sends a fixed amount of data (try)
• If it is not able to finish (miss) it is demoted to a lower priority queue.
• The above process repeats.
Drawback of Learning by Priority Queue

• This approach in essence is “try-and-miss”
• Every coflow starts from highest priority queue and sends a fixed amount of data (try)
• If it is not able to finish (miss) it is demoted to a lower priority queue.
• The above process repeats.

\[
\begin{align*}
&\text{Low Priority (Larger slots)} \\
&Q_2 \\
&Q_1 \\
&Q_0 \\
&\text{High Priority (Smaller slots)} \\
&\quad C_i \\
&\quad \ldots \\
&\quad \ldots \\
&\quad \ldots \\
&t_1
\end{align*}
\]
Drawback of Learning by Priority Queue

- This approach in essence is “try-and-miss”
- Every coflow starts from highest priority queue and sends a fixed amount of data (try)
- If it is not able to finish (miss) it is demoted to a lower priority queue.
- The above process repeats.
Drawback of Learning by Priority Queue

• This approach in essence is “try-and-miss”
• Every coflow starts from highest priority queue and sends a fixed amount of data (try)
• If it is not able to finish (miss) it is demoted to a lower priority queue.
• The above process repeats.
Philae

• A Coflow has many flows.
• Sampling is a time proven technique to estimate average with high accuracy.
Philae

- A Coflow has many flows.
- Sampling is a time proven technique to estimate average with high accuracy.
Philae

Average CCT Speedup

150 nodes 900 nodes

Average CCT Speedup

Reduce Stage
Shuffle (Communication)
Map Stage

A Map Reduce job
Philae: Your Coflow Has Many Flows
Sampling Them for Fun and Speed

Full talk in track – II at 9:15 am on Friday July 12th.

Akshay Jajoo   Y. Charlie Hu   Xiaojun Lin