Real-time Web Marketing with Apache Storm

Christopher Little
Apache Storm
# Alternatives

<table>
<thead>
<tr>
<th>Processing Model</th>
<th>Storm</th>
<th>Hadoop</th>
<th>Spark Streaming</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAG</td>
<td>DAG</td>
<td>MapReduce</td>
<td>DAG</td>
</tr>
<tr>
<td>Processing Unit</td>
<td>Record-at-a-time</td>
<td>Batch</td>
<td>Mini Batch</td>
</tr>
<tr>
<td>Latency</td>
<td>Sub-second</td>
<td>High</td>
<td>Few seconds</td>
</tr>
<tr>
<td>Fault Tolerance</td>
<td>At least once</td>
<td>Exactly once</td>
<td>Exactly once</td>
</tr>
</tbody>
</table>
Architecture (cluster)
Architecture (cluster)

- Nimbus
- Zookeeper
- Supervisor
- Worker Process
Architecture (work)
Architecture (work)

Spout → Bolt

Spout → Bolt → Bolt

(x, y, z)
Architecture (work)

Spout \(x_3\) \(\to\) Bolt \(x_3\) \(\to\) Bolt \(x_1\)

Spout \(x_{10}\) \(\to\) Bolt \(x_5\) \(\to\) Bolt \(x_8\)
Example Topology - Wordcount

Spout x5 → Split x8 → Count x12

(“Mary had a little Mary”) → (“Mary”) → (“Mary”, 1)
(“had”) → (“had”, 1)
(“a”) → (“a”, 1)
(“little”) → (“little”, 1)
(“Mary”) → (“Mary”, 2)
Example Topology - Wordcount

- **Spout** \(x_5\): "Mary had a little Mary"
- **Split** \(x_8\): 
  - ("Mary")
  - ("had")
  - ("a")
  - ("little")
  - ("Mary")
- **Count** \(x_{12}\): 
  - ("Mary", 1)
  - ("had", 1)
  - ("a", 1)
  - ("little", 1)
  - ("Mary", 2)
Example Topology - Wordcount

- **Spout** x5
  - (“Mary had a little Mary”) x5

- **Split** x8
  - (“Mary”) x8
  - (“had”) x8
  - (“a”) x8
  - (“little”) x8
  - (“Mary”) x8

- **Count** x12
  - (“Mary”, 1) x12
  - (“had”, 1) x12
  - (“a”, 1) x12
  - (“little”, 1) x12
  - (“Mary”, 2) x12
Application
20/20 vision puts the important things in sight

Get Started
Schedule your Free LASIK Consultation

$500 OFF LASIK
Schedule Today
Must have treatment by 11/17/20

Save up to $1,200 on LASIK using your Flex-Spending Account

SCHEDULE NOW
“there is a strong relationship between gaze position and cursor position”

- Chen et al. (2001)
Web Analytics

JavaScript Client → Spout → Heat Map

Event Stream
- Clicks
- Mouse movements
- Scroll event
- Form interaction
- Links followed

Realtime Processing
- Filtering
- Windowing
- Aggregation
Session Windowing Example

(k1, (v1, 13:02))
(k2, (v2, 13:14))
(k1, (v3, 13:57))
(k1, (v4, 13:20))

ParDo

AssignWindows

GroupByKey

MergeWindows

(k1, (v1, [13:02, 13:32]))
(k2, (v2, [13:14, 13:44]))
(k1, (v3, [13:57, 14:27]))
(k1, (v4, [13:20, 13:50]))

(k1, ([(v1, [13:02, 13:32])
,(v3, [13:57, 14:27])
,(v4, [13:20, 13:50]])
)
(k2, ([(v2, [13:14, 13:44]])
)

ParDo

MergeWindows

MergeWindows

(k1, ([(v1, v4], [13:02, 13:50]))
(k1, (v3, [13:57, 14:27]))
(k2, (v2, [13:14, 13:44]))
Real-time Web Marketing with Apache Storm

Christopher Little