Real-time Web Marketing with
Apache Storm
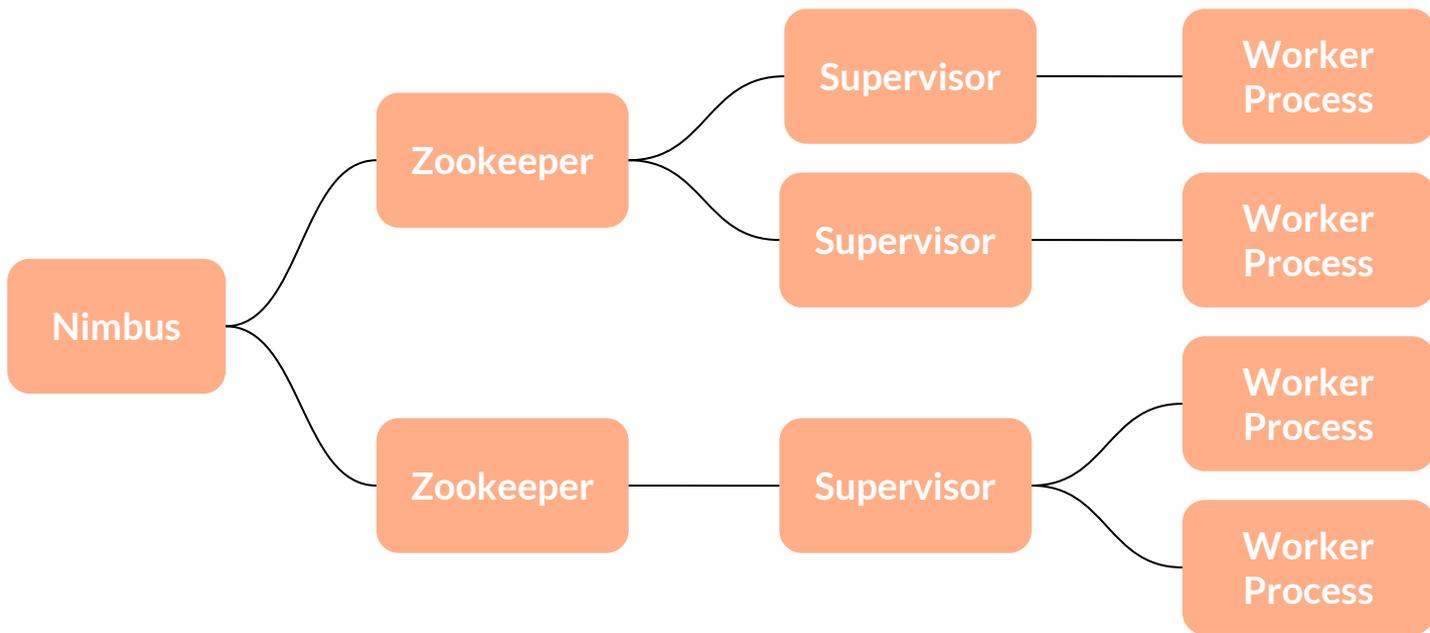
Christopher Little

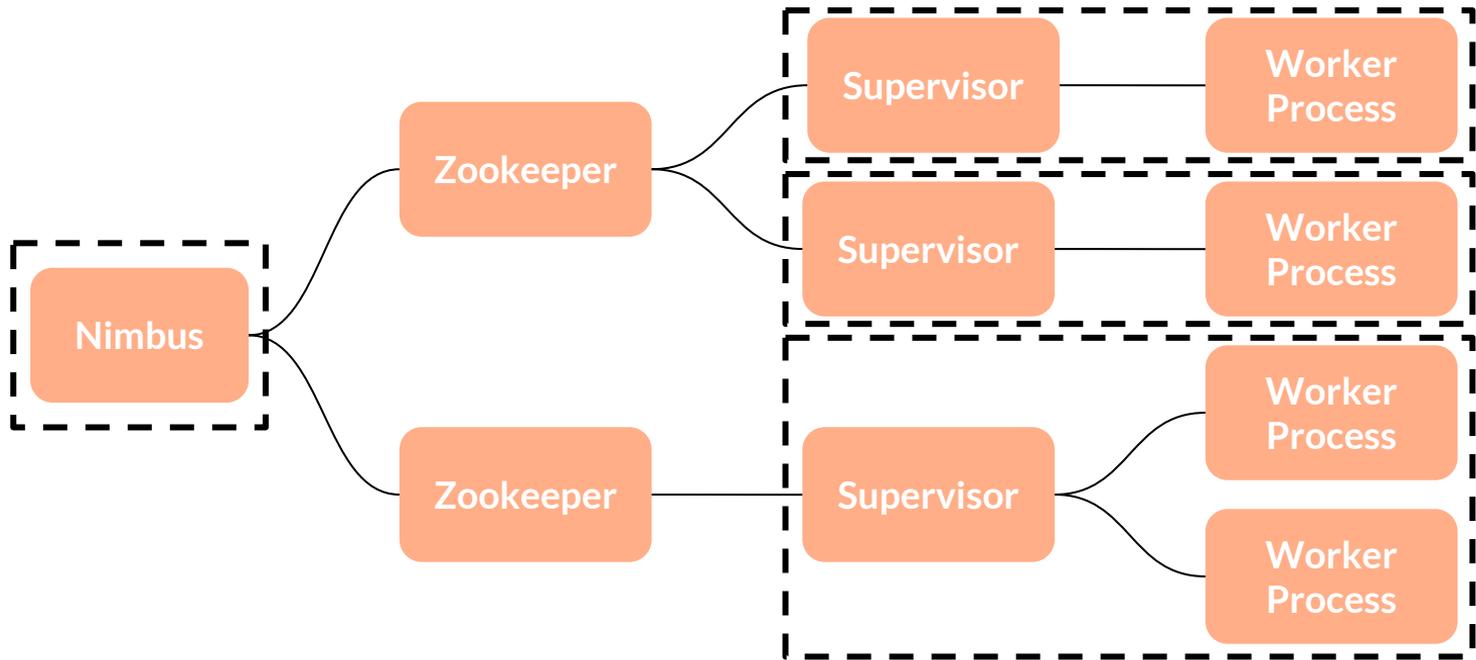# Apache Storm

Alternatives

	Storm	Hadoop	Spark Streaming
Processing Model	DAG	MapReduce	DAG
Processing Unit	Record-at-a-time	Batch	Mini Batch
Latency	Sub-second	High	Few seconds
Fault Tolerance	At least once	Exactly once	Exactly once

Architecture (cluster)

Architecture (cluster)

```
Nimbus ─── Zookeeper ─── Supervisor ─── Worker Process
       │              └── Supervisor ─── Worker Process
       └── Zookeeper ─── Supervisor ─── Worker Process
                                    └── Worker Process
```
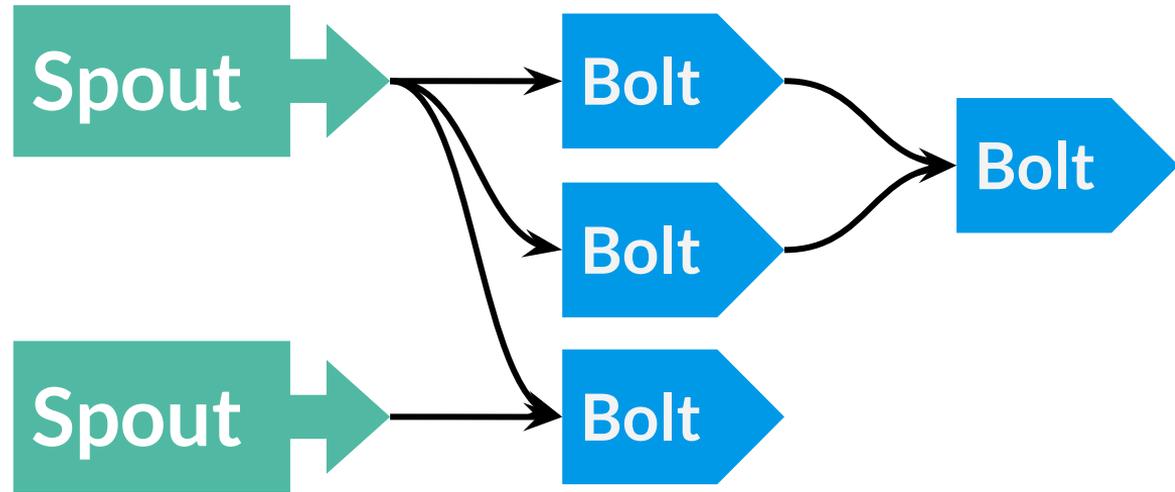
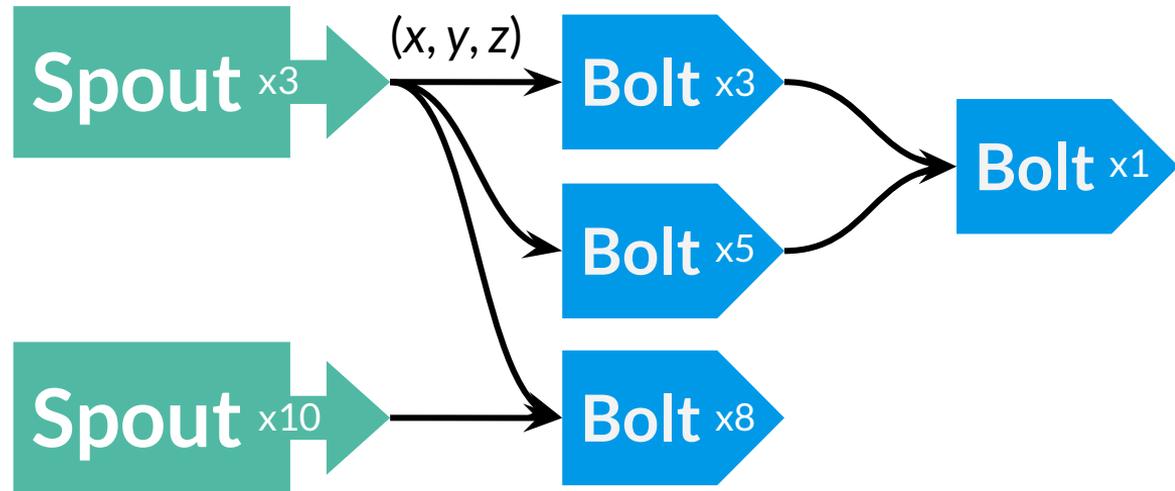# Architecture (work)
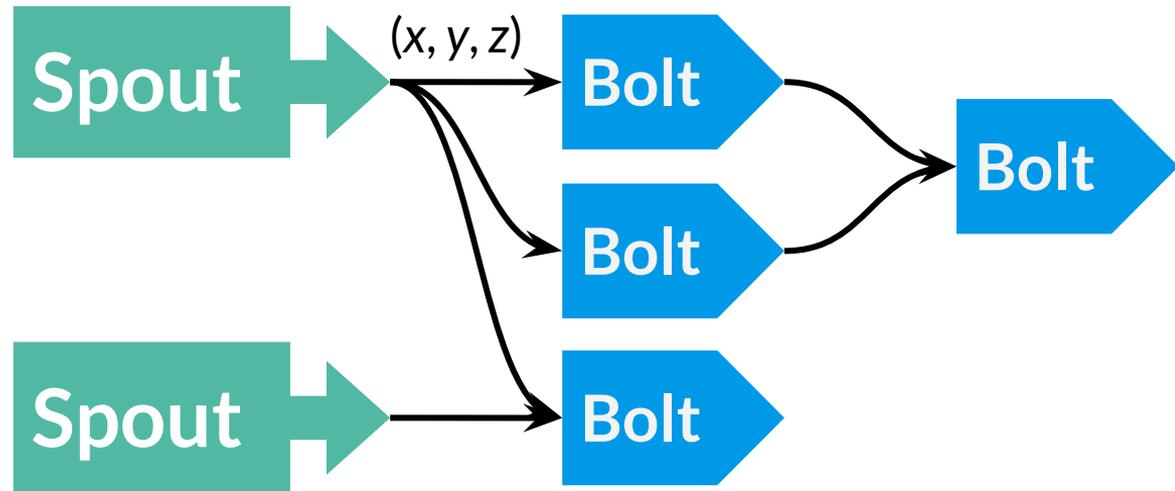
Architecture (work)

Architecture (work)

Example Topology - Wordcount

Spout x5 → **Split** x8 → **Count** x12

("Mary had a little Mary")

("Mary")
("had")
("a")
("little")
("Mary")

("Mary", 1)
("had", 1)
("a", 1)
("little", 1)
("Mary", 2)

Example Topology - Wordcount

Spout x5 → **Split** x8 → **Count** x12

("*Mary had a little Mary*")

("*Mary*")
("*had*")
("*a*")
("*little*")
("*Mary*")

("*Mary*", 1)
("*had*", 1)
("*a*", 1)
("*little*", 1)
("*Mary*", 2)

Example Topology - Wordcount

Shuffle()　　　　　　　**Group**(*0*)

Spout x5 　→　 **Split** x8 　→　 **Count** x12

(*"Mary had a little Mary"*)

(*"Mary"*)
(*"had"*)
(*"a"*)
(*"little"*)
(*"Mary"*)

(*"Mary", 1*)
(*"had", 1*)
(*"a", 1*)
(*"little", 1*)
(*"Mary", 2*)
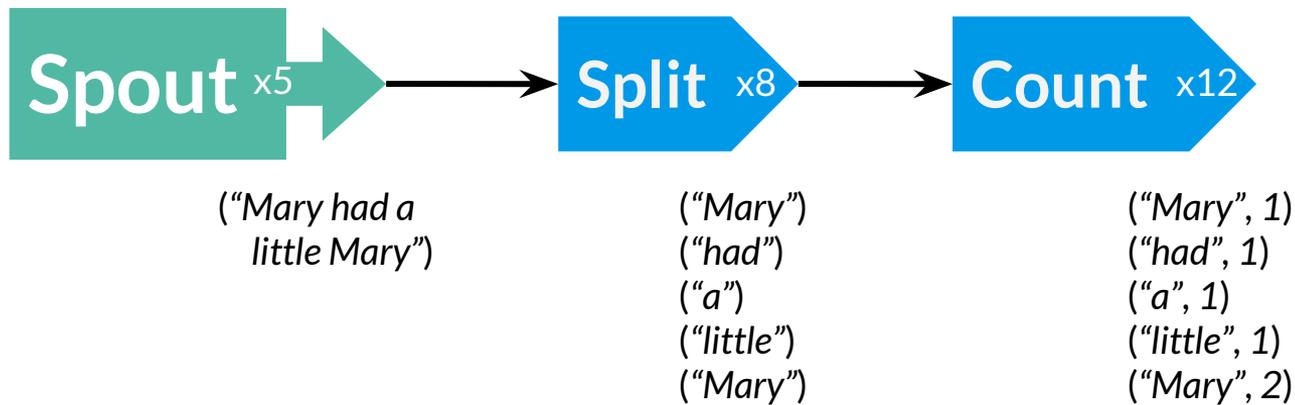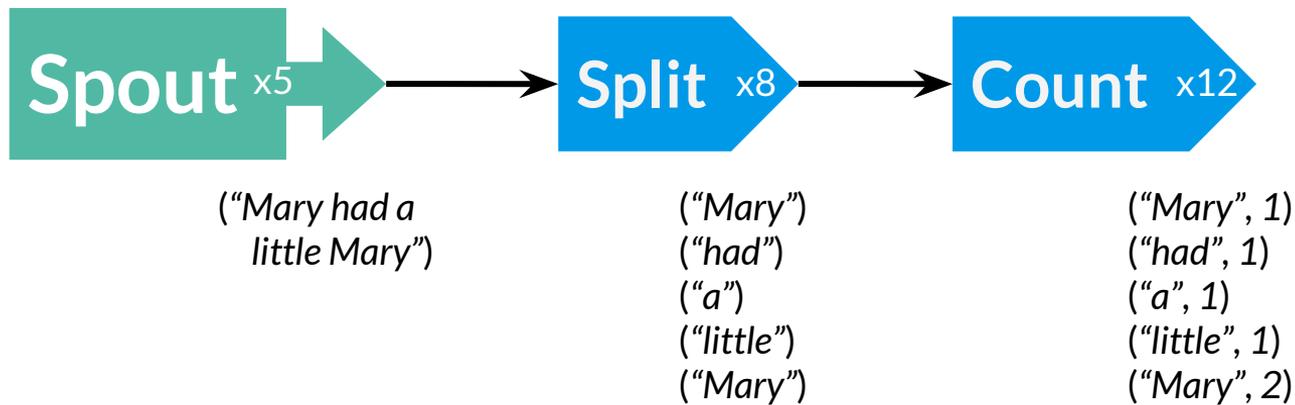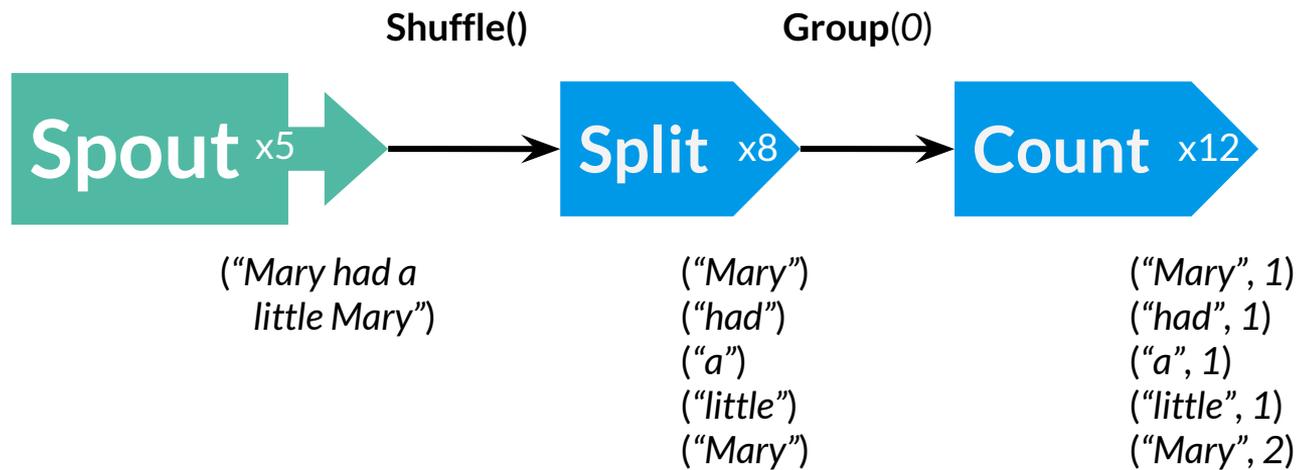
Application

LasikPlus
Imagine your life in focus

0.5%   # 8

1.1%   # 6

PATIENT STORIES
1.1%   # 6

TREATMENT
0.5%   # 8

6.7%   # 1

AFFORDABILITY
4.0%   # 2

0.5%   # 8

RESOURCES
0.5%   # 8

1.3%   # 5

ABOUT LASIKPLUS
0.8%   # 7

20/20 vision puts the important things in sight

Get Started

Schedule your Free
LASIK Consultation

SCHEDULE NOW ▶
2.1%   # 3

$500 Off LASIK
Schedule Today

Must have treatment by 11/17/12

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

LEARN MORE ▶
1.6%   # 4

8

"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))                              (k1, (v1, [13:02, 13:32]))
(k2, (v2, 13:14))        ParDo                  (k2, (v2, [13:14, 13:44]))
(k1, (v3, 13:57))   ──────────────▶            (k1, (v3, [13:57, 14:27]))
(k1, (v4, 13:20))      AssignWindows            (k1, (v4, [13:20, 13:50]))
```

GroupByKey

MergeWindows

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

Real-time Web Marketing with
Apache Storm

Christopher Little