Zookeeper

What it actually is.

Tobias Schlottke
November 2011

Zookeeper

What it actually is.

Tobias Schlottke
November 2011

What it is

„ZooKeeper is a centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services.
All of these kinds of services are used in some form or another by distributed applications. Each time they are implemented there is a lot of work that goes into fixing the bugs and race conditions that are inevitable.
Because of the difficulty of implementing these kinds of services, applications initially usually skimp on them, which make them brittle in the presence of change and difficult to manage. Even when done correctly, different implementations of these services lead to management complexity when the applications are deployed."

WTF?

In simple words

„A service which helps to coordinate distributed systems.“
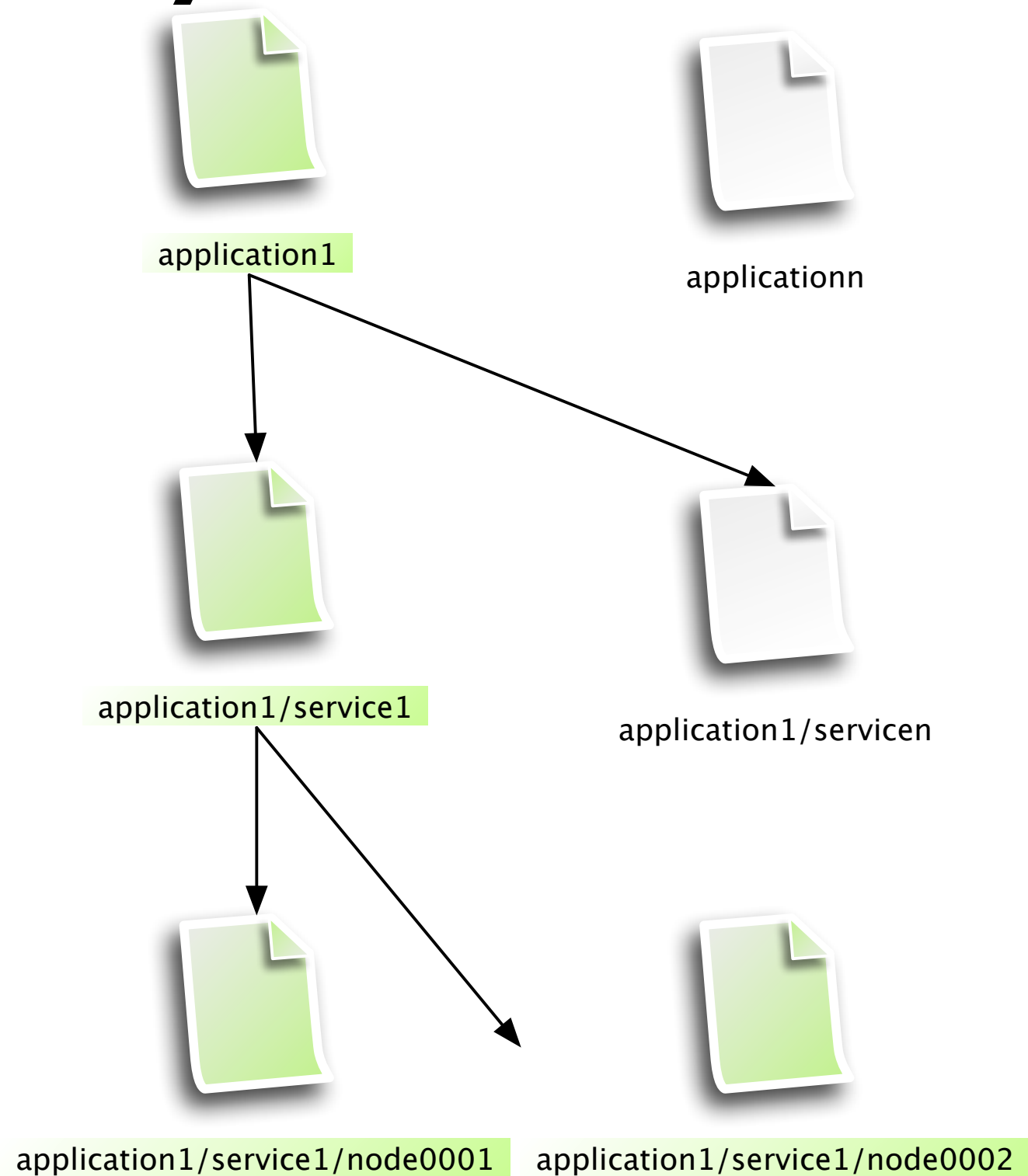
Invented by

Invented by

YAHOO!®

More details please!

- Naming nodes

- Configuring nodes

- Synchronizing nodes

- Organize groups of nodes

- Heartbeat

- Democracy / Leader election

But how to organize this stuff?

Filesystem Schema

application1

applicationn

application1/service1

application1/servicen

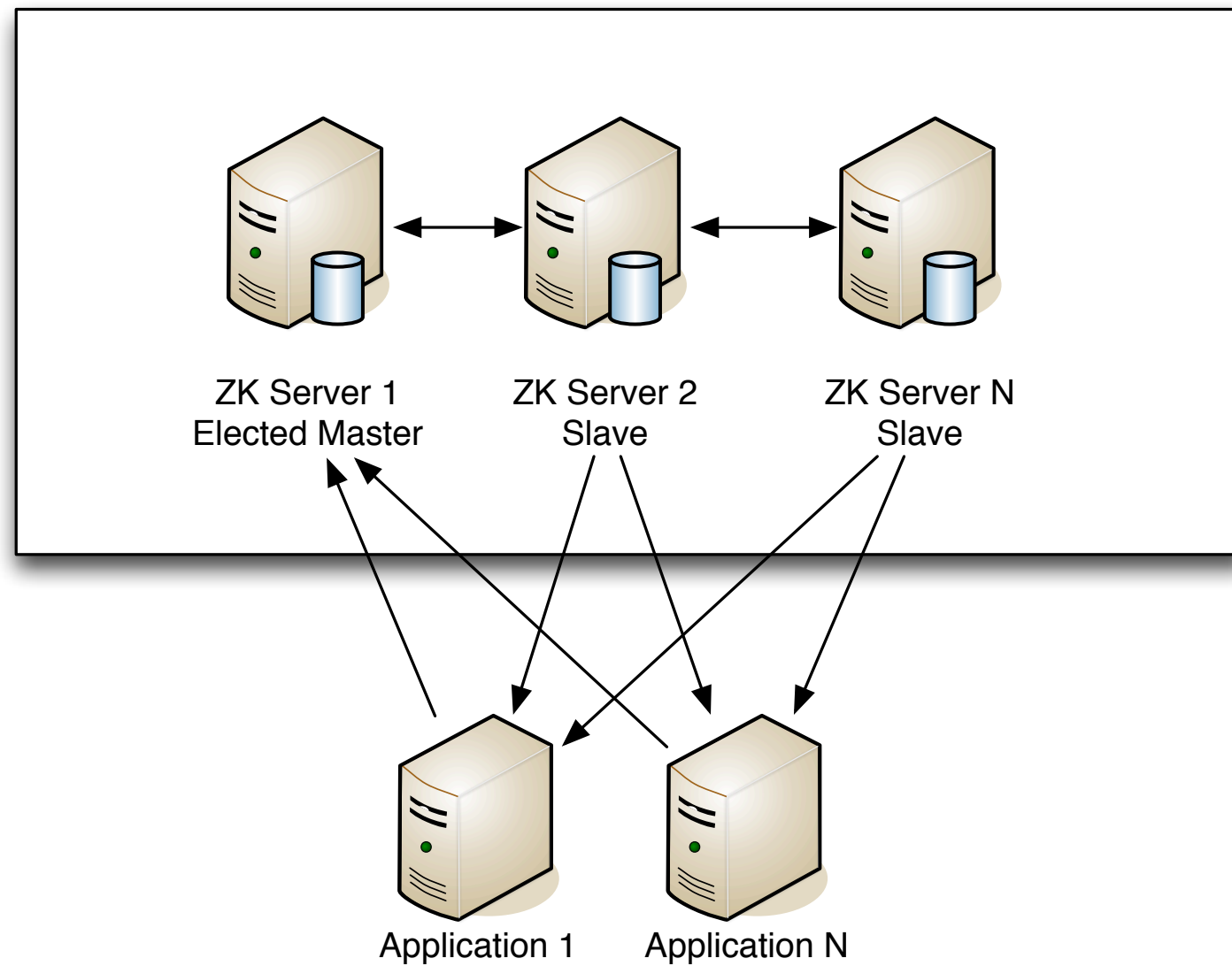application1/service1/node0001   application1/service1/node0002

„Filesystem" Features

- Organized in a node tree `app1/service1`

- „Ephemeral" nodes

- „Sequential" nodes `0123`

- Notification system

- High availability

- Exchange format does not matter

Other things you get for free

- „Atomicity“

- „Consistency“

- „Reliability“

- „Timeliness“

- „Conditional updates“ (Versioning)

The Zookeeper Service

ZK Server 1
Elected Master

ZK Server 2
Slave

ZK Server N
Slave

Application 1   Application N

The Zookeeper Service

- All Servers store a copy of data in Mem

- Master election at startup

- All updates go to the leader

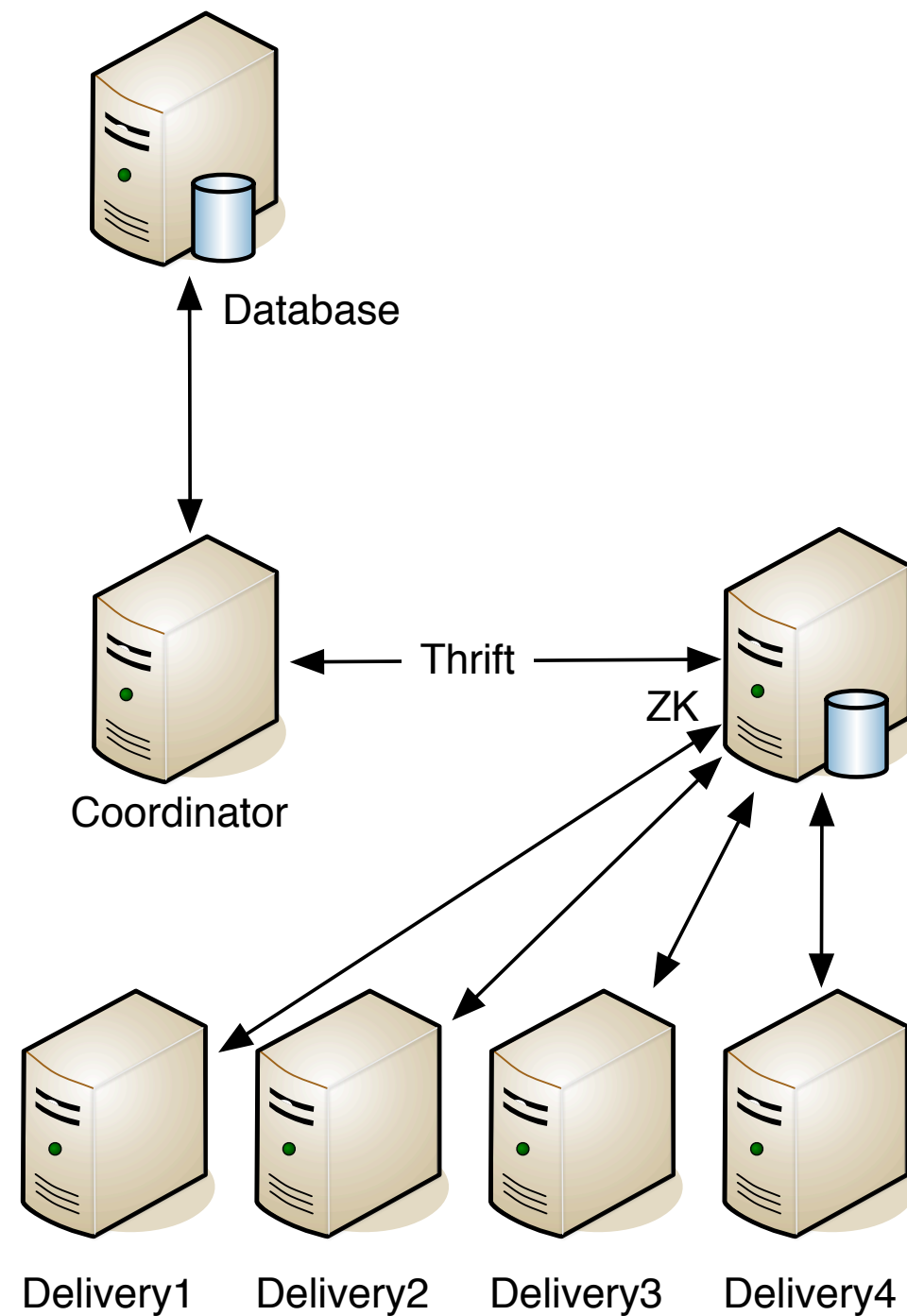- Responses are sent when a majority persited

Snares

- Data must fit in memory

- Data size for each node should be <1MB

- Client libraries not all that good so far

- Ephemeral nodes are not allowed to have children

Usecase

„I am an adserver. I deliver 20.000 ads a second and for this, you will most likely design me as a distributed service"

Adserver Usecase

Thats it!
Pretty simple, huh?