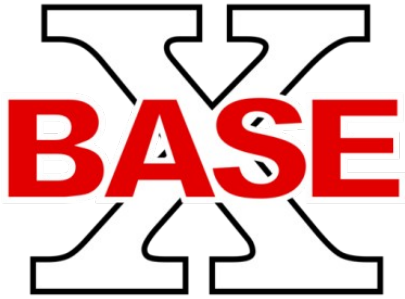



# Distribution and Replication in BaseX

What to expect in the future

BaseX User Meeting

Pre-Conference Day, XML Prague 2013

Dirk Kirsten

dk@basex.org

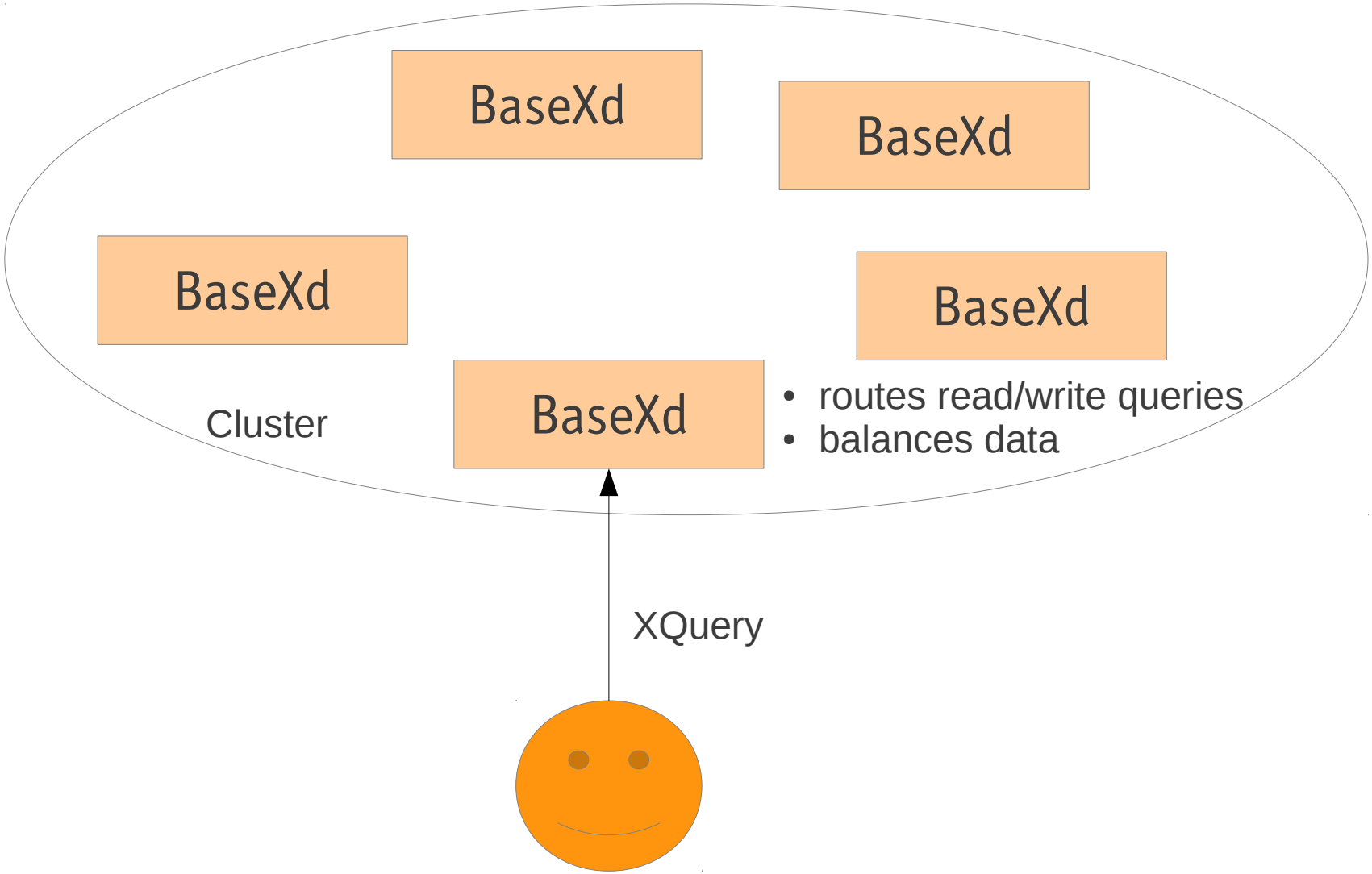
## Why do you need that?

- scale-out horizontally
  - easily (and cheaply) add more servers (commodity hardware)
  - manage large volumes of data
  - distribute the load → faster execution
- replication is de facto a must-have for any reliable application
  - by scaling out and an increased number of servers failures become more likely and replication even more important

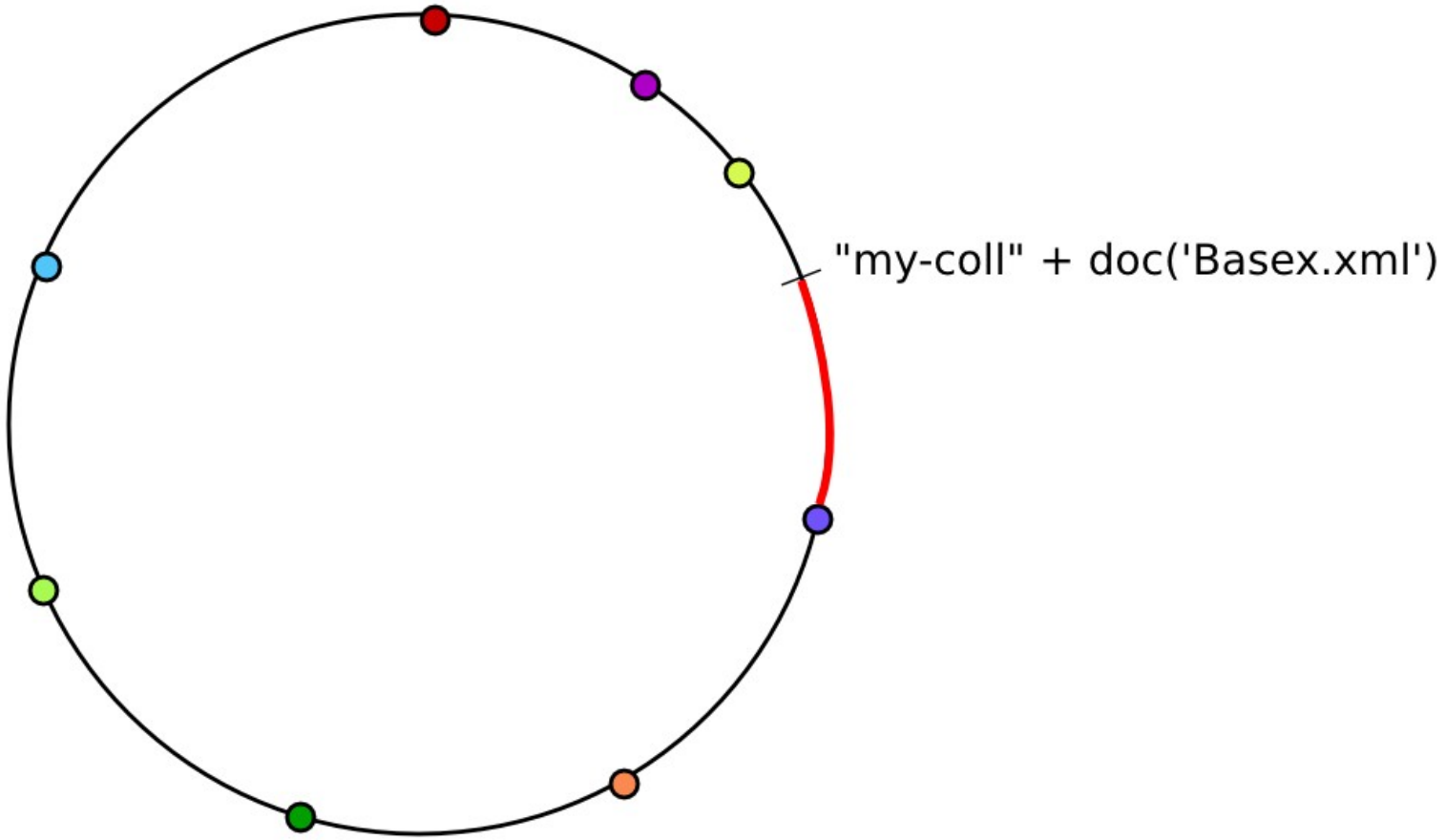
## What can you do nowadays?

- Distribution:
  - Nothing, really...
- Replication:
  - Built-in Backup Tool
  - *rsync* to copy files to other server

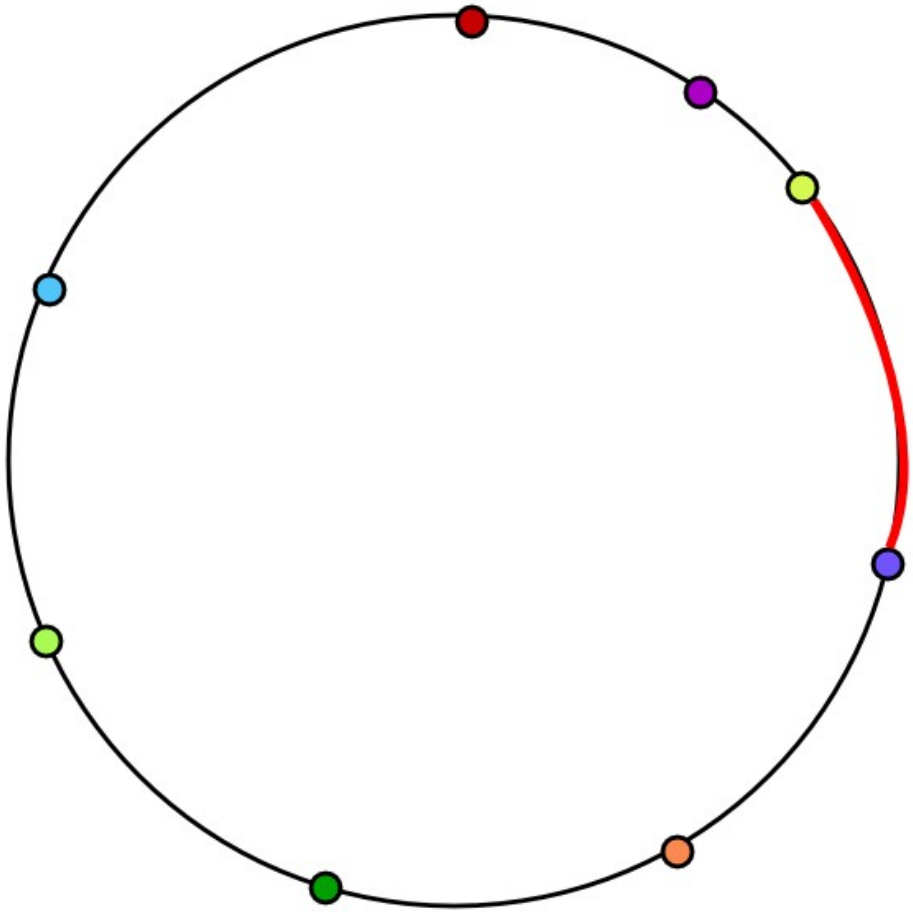
# BaseX Distributed



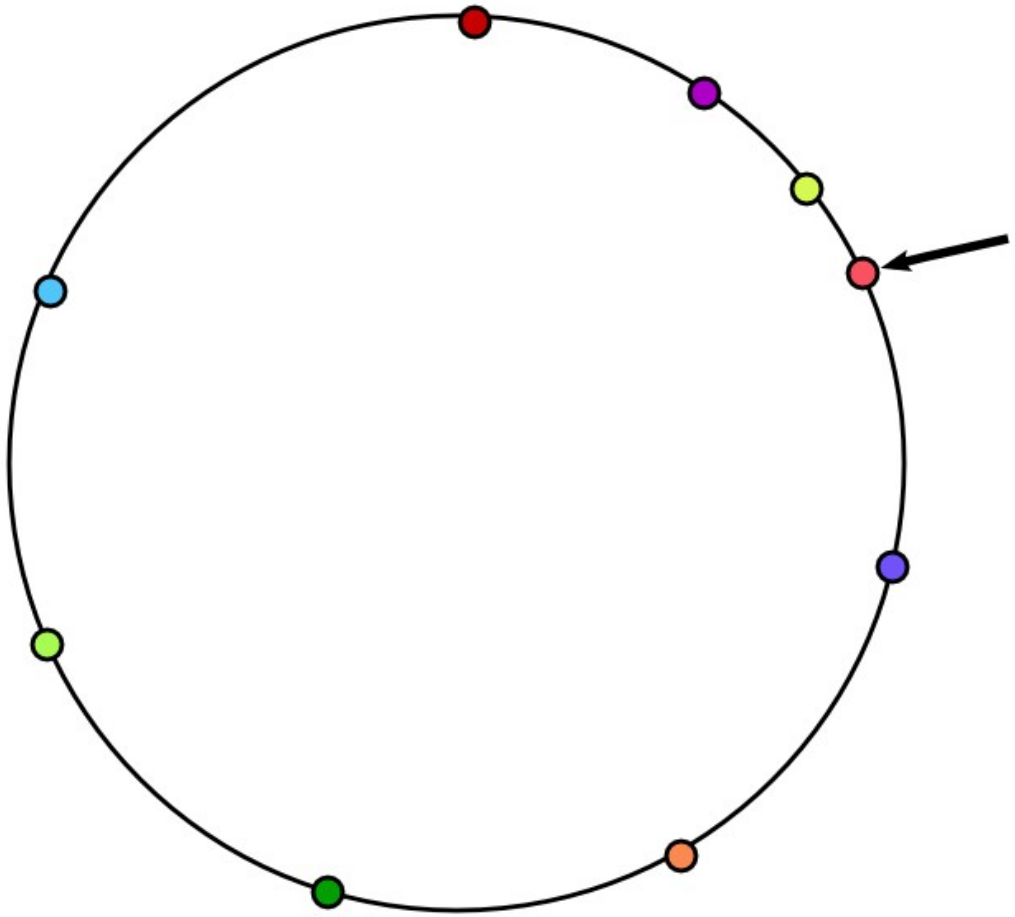
# BaseX Distributed



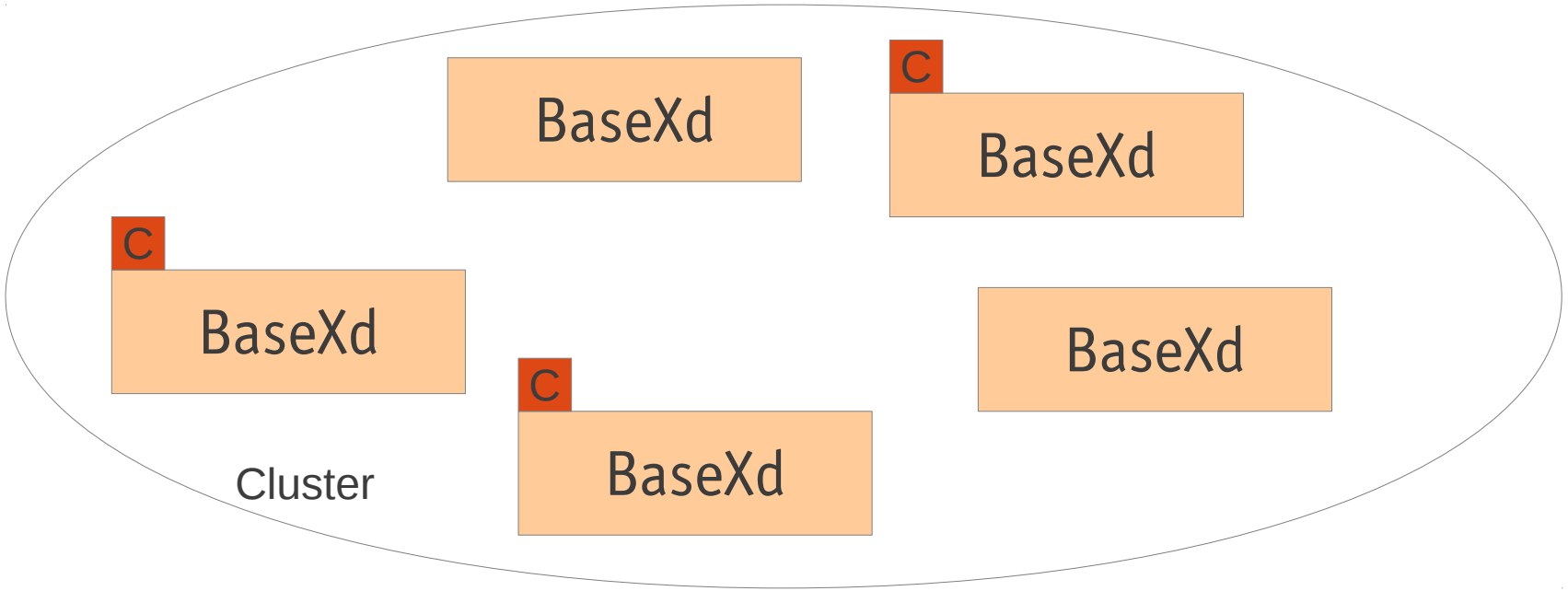
# BaseX Distributed



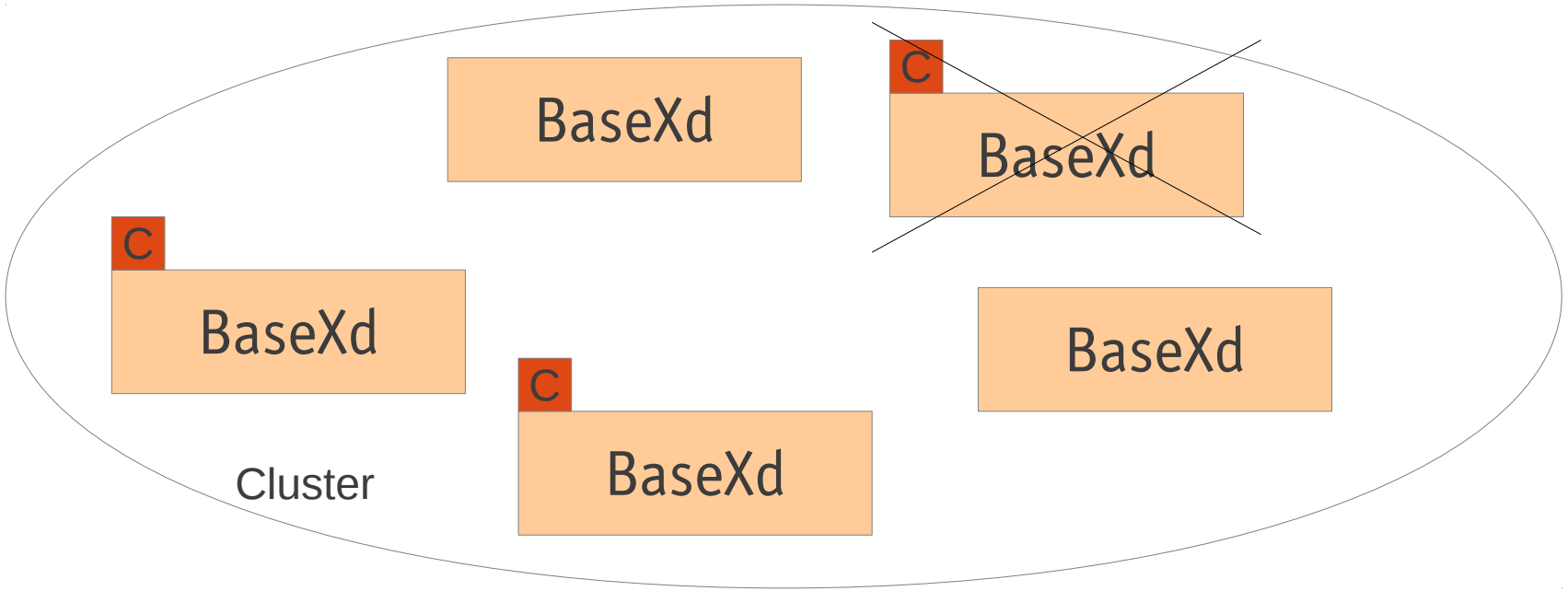
# BaseX Distributed



# BaseX Distributed - config server



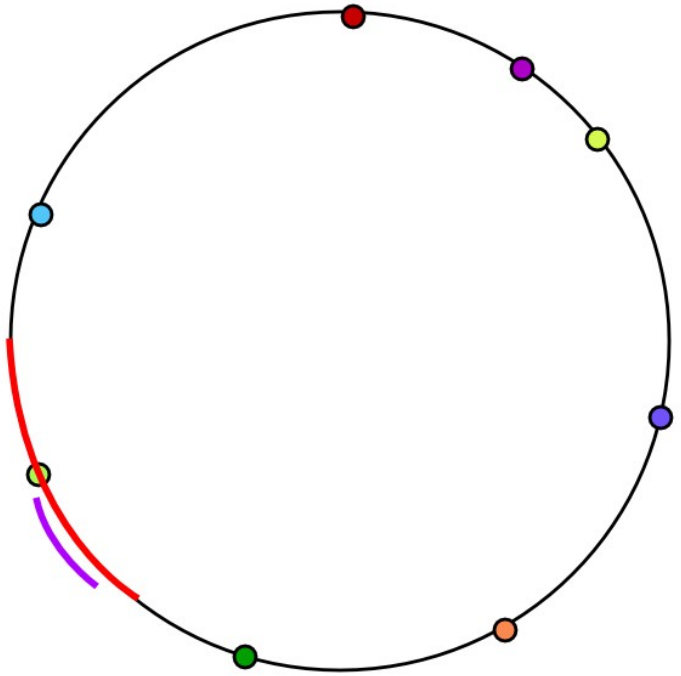
# BaseX Distributed - config server




- x Nodes elect a new leader
- 3 nodes have to server as config server → otherwise just read-only
  - no rebalancing
  - documents can still be read or write

## BaseX Distributed

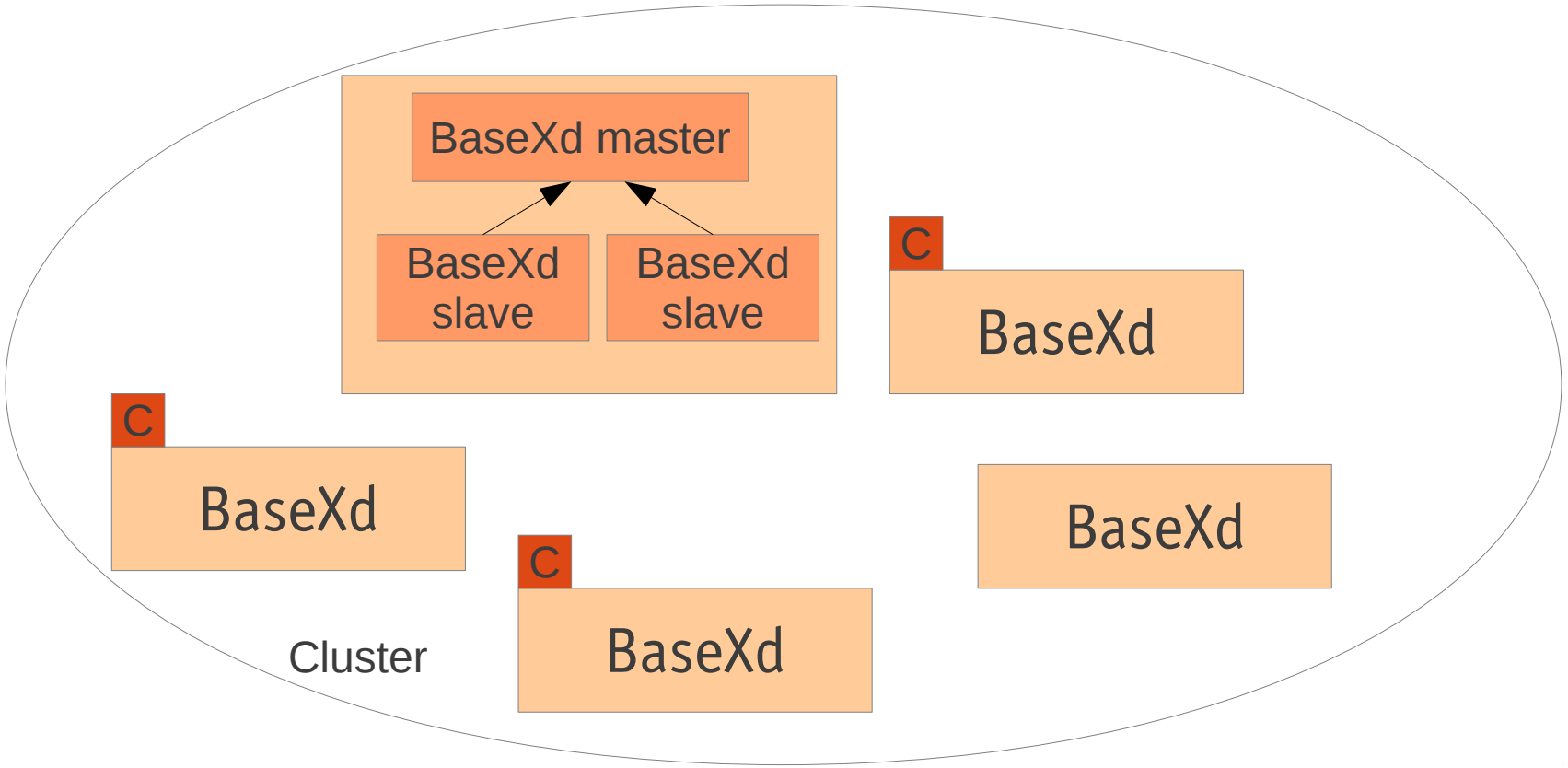
```
for $doc in collection('webpage')
where starts-with(base-uri($doc), 'my-doc')
return $doc//p[@class = 'whatever']
```



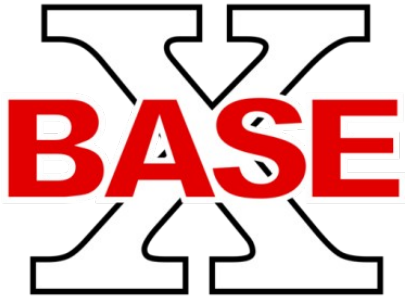
## BaseX Distributed

```
for $p in dist:query( ('webpage', 'webpage2'),
   for $i in $coll
   return $i/my/node/now
)
return $p/content/text()
```

# Replication




- traditional master/slave system
- all writes are happening on the master
- slaves automatically receive updates
- failover management in case of failure of the master




Any questions, suggestions or use cases?
Now or write to dk@basex.org