

# Replication Factor

↓  
how many times  
the data is duplicated  
across all nodes

## Cassandra

- Highly scalable database
- No single point of failure
- Fault tolerant
- Tunable consistency
- Flexible data model
- Elastic and incremental expansion
- Multi-data centre support
- Distributed real-time counter support
- Lightweight transaction

Rebalancing simplified by doubling nodes

New  
installations  
should use  
vNodes

Virtual Nodes  
Nikolas Favre-Felix

↓  
does  
not  
scale  
well

No need to  
manually  
rebalance

### Virtual nodes in Cassandra 1.2+

- More than one token per node
- Random token assignment
- Incremental cluster resize, one node at a time
- Streaming to/from all nodes, not just neighbours

Partition sizes  
may differ between nodes  
Balanced cluster - keys are  
distributed randomly

60 nodes x 256 tokens ← issues