What *is* Big Data?

Andy Twigg
Acunu and Oxford

HotStorage11 Panel

Acunu

UNIVERSITY OF OXFORD

Claim: "Big Data" is *not:*
scale-out, web-scale, nosql,

etc...

Some of the biggest big data problems involve mysql.
Some of the best-known NoSQL DBs don't scale out.

...it's about surfing.

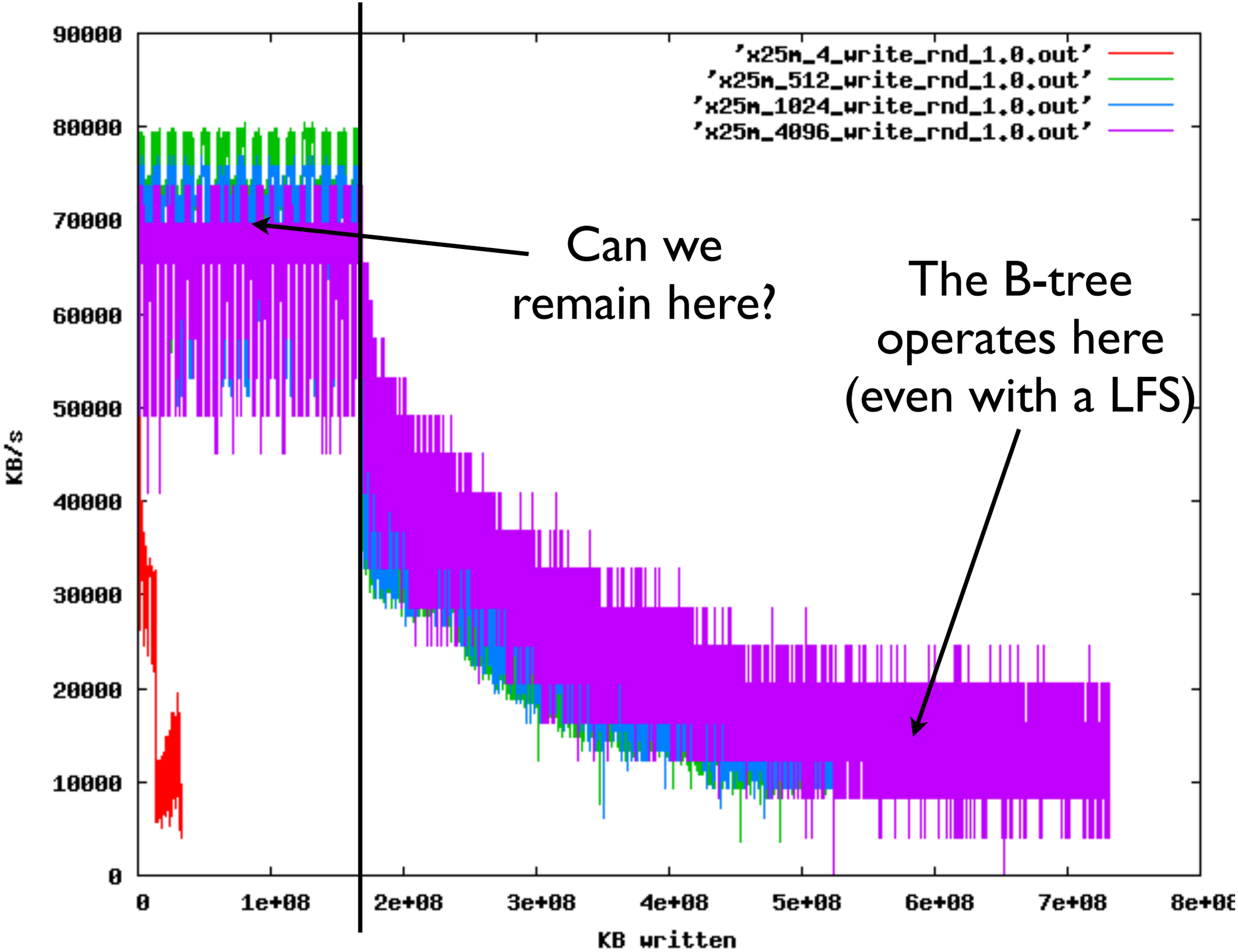
Storage is changing:

3 fundamental shifts

1: storage technology

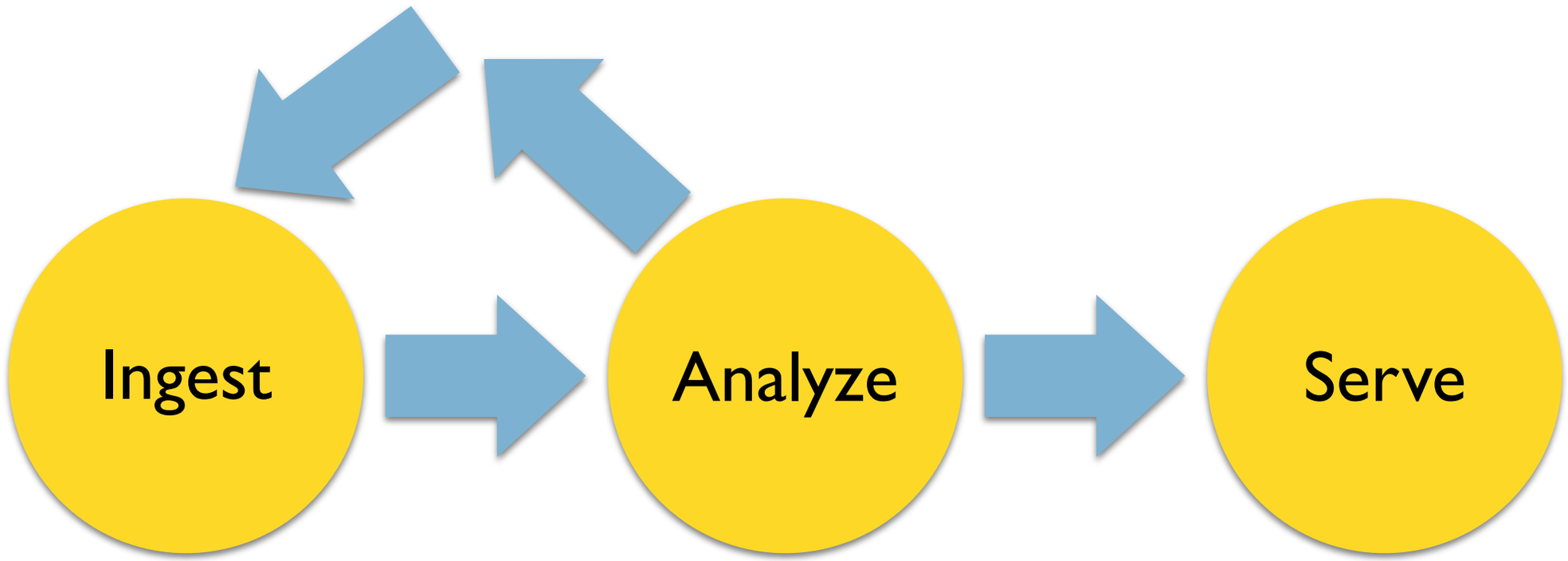
$100
60 GB
40,000 IOPS
150 MB/s

$100
2000 GB
100 IOPS
130 MB/s



~device capacity

2: know your workloads!



High-entropy update Range queries Random reads

- small entries (e.g. cookies, ticks)
- little value (you don't care losing 1%)
- high-entropy, high rate updates (>100,000/s)
- range queries (analytics)

3: commodity h/w

Lots of people rely on EC2. How do you get reliability there?

Enterprises like support and reliability. But what is the easiest way to get 5 9's reliability, assuming s/w works?

Scale-out is unavoidable.

The challenge is getting to grips with the problems scale-out brings - CAP, Vector clocks, etc.
Read the literature! Don't reinvent the wheel!

Acunu

bitbucket.org/acunu

www.acunu.com/download

Apache, Apache Cassandra, Cassandra, Hadoop, and the eye and elephant logos are trademarks of the Apache Software Foundation.