

SSDs EXCEL IN BIG DATA ARCHITECTURE

Boosting cost/performance with SSDs is all about **LOCATION, LOCATION, LOCATION**

Here are 8 critical spots in big data architecture...

1

SCALE UP

One larger multicore, all-SSD node in a Hadoop cluster can take on bigger jobs or chaining, freeing up smaller nodes for easier jobs¹

2

VIRTUALIZATION

SSDs mitigate ill effects of smaller I/O blocks and larger number of requests in Xen²

5

MAP REDUCE

Targeting intermediate shuffle results produced by MapReduce algorithms at an SSD reduces severe loads on overwhelmed HDDs³

6

TIERED STORAGE

Aiming more HDFS requests at PCIe SSDs in a hybrid cluster increases average I/O rate⁴

7

SMALL "HOT" FILES

Caching Facebook messages on an SSD triples HBase performance through reduced latency⁵

8

TIGHTLY COUPLED

MIT tests FPGA fabric and local ARM cores on SSDs to pre-process data and speed searches⁶

3

HYBRID DATABASE

Keeping smaller files and indices in RAM, with larger files stored on SSDs, speeds up NoSQL⁸

4

SPILLING RDDS

Spark running on Amazon EC2 nodes backed by Amazon S3 all-SSD storage sets records⁷

BIG DATA CLUSTERS WITH SSDS IN CRITICAL LOCATIONS COMPLETE MORE REQUESTS,

SSDs with V-NAND unlock big data performance with:

↑
HIGHER IOPS

↓
LOWER LATENCY

+MORE DWPD

For more details on findings from these research studies samsung.com/enterprisessd