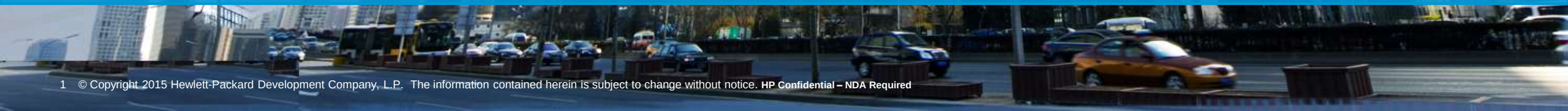
Tech Disruption: Challenging Analytics Architectures

Joseph George
Director, HPC and Big Data Solutions
@jbgeorge

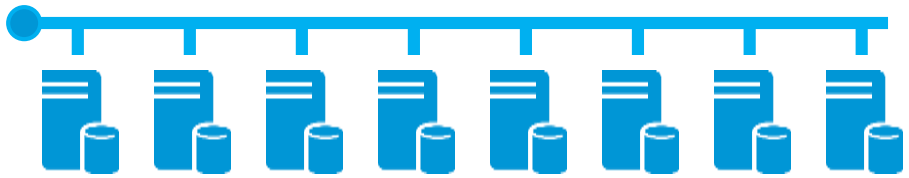
September 10 2015



Disrupting the Market with Scalable Asymmetric Architecture

Current traditional Big Data approach (Symmetric)

- Compute and storage are always collocated
- All servers are identical
- Data is partitioned across servers on direct-attached storage (DAS)



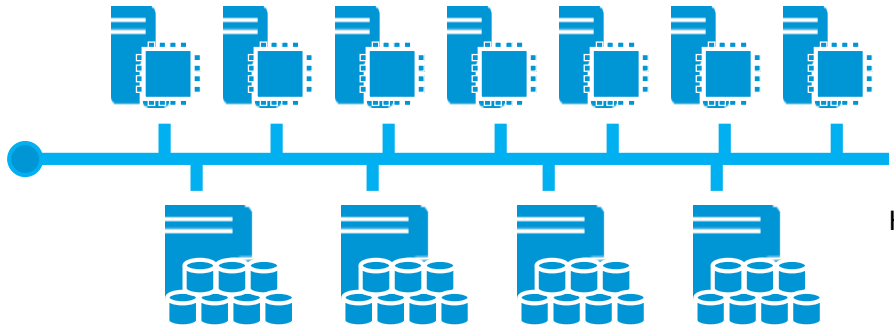
YARN Applications, HDFS, ORC Files, Parquet, Hbase, Cassandra

Two Socket, 2U Servers

New Big Data approach (Asymmetric)

- Separate compute and storage tiers connected by Ethernet networking
- Standard Hadoop installed asymmetrically with storage components on the storage servers and yarn applications on the compute servers

Compute Optimized Servers



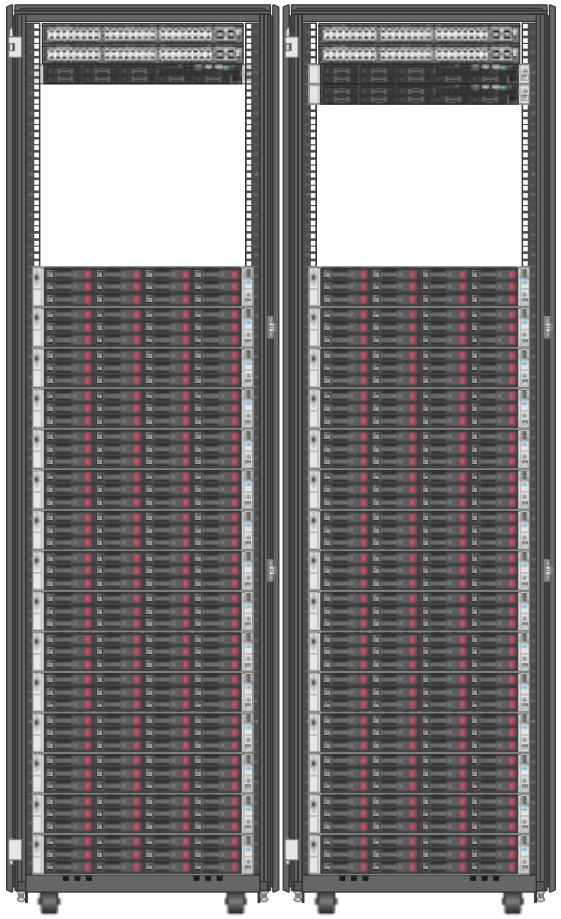
YARN Applications

HDFS, ORC Files, Parquet, Hbase, Cassandra

Storage Optimized Servers

The Disruption of Scalable Asymmetric Architecture

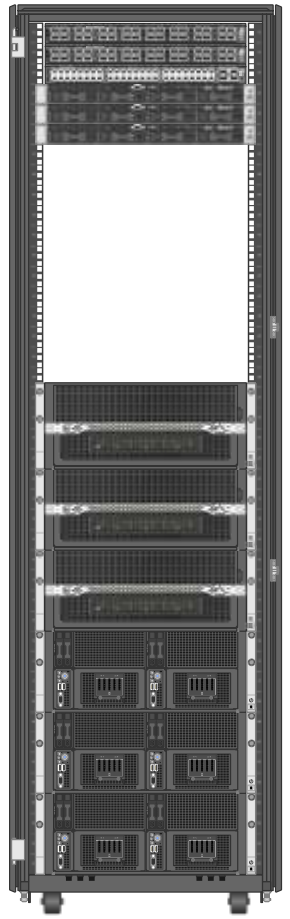
Equivalent price/performance (MapReduce workload)



Traditional Architecture

	Traditional Symmetric	Scalable Asymmetric	Delta
Total memory	3.84 TB	4.6 TB	20% more memory
Network fabric	10GbE	40GbE	
Density (rack U)	69	32	2.16x greater density *
Datacenter floor tiles	2	1	½ the floor tiles *
Power (W)	26,475	14,296	46% reduction *
5-year power / cooling costs (itio@ $.20/kWh)	$232K	$125K	$107K savings *

- *Does not include further benefits from cluster and workload consolidation into a single data lake*



Scalable Asymmetric Architecture

Benefits and Market Feedback

Hadoop community & vendors supportive

IT Management

- Ability to scale compute and storage separately
- Lower chance of obsolescence by software changes
- The ability to consolidate clusters with less types of servers to support
- 2x reduction in floor space
- 2x reduction in power
- Software licensing cost (if applicable)

www.hp.com/go/Hadoop

HPC Users

- Optimized data flow therefore optimized application performance
- Analytics applications used to collect and pre-process non-HPC data sources.
- Compute elements optimized by function
- Interdisciplinary HPC applications have direct access to all data

Hadoop Developers

- Consolidation = less data movement
- Potential to make IT more responsive to requests for capacity
- 100% Hadoop – no patches, not proprietary