

1. The IT industry is on the cusp of a major paradigm shift, equivalent to Mainframe/Minicomputer disruption and the PC Server disruption.
2. Central to this and previous shifts has been an order of magnitude cost shift
3. While the shifts have been enabled by technology innovations, success has been much more closely tied to mastery of the new business models/process changes enabled by the technology advances.

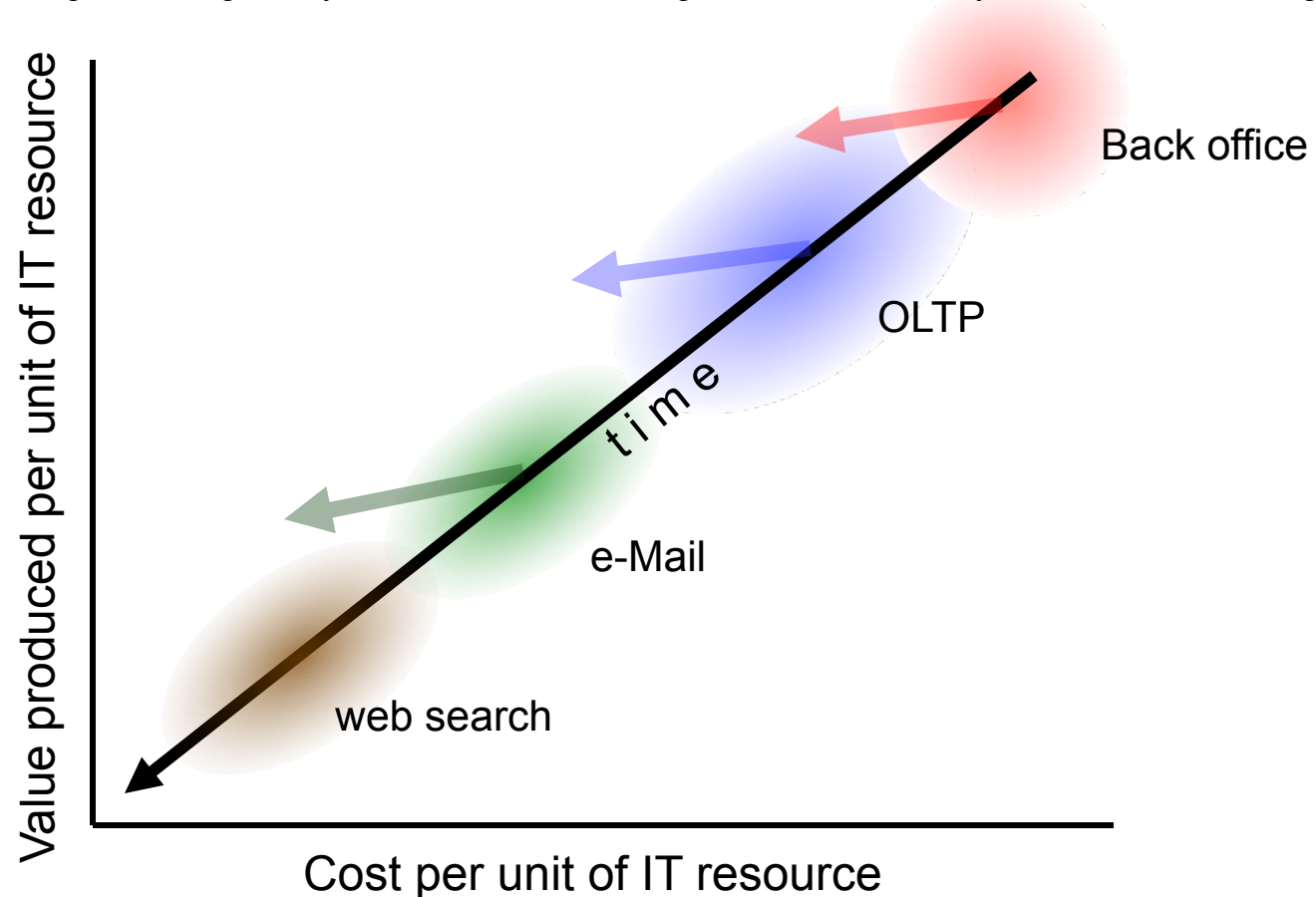
It's maybe best to think of this in the context of a radical redo of the IT delivery supply chain.

4. Time scale of the shift will be comparable to previous shifts – that is at least a 10 to 15 years

e.g., PC Servers were first introduced in 1991 with 0% market share, it was 2005 before reaching 50% market share (revenue).

1. New workloads are dominated by web-scale applications **enabled by** and consuming vast quantities of very cheap IT resources (*CPU, memory, storage, ...*)

- ✓ Over time, as compute costs have fallen, successive generations of new applications have been enabled.
- Each generation has required more compute resource to produce a given (significant) absolute value.
- Successive significant new applications (as measured by revenue) consume vastly more resources than preceding significant applications.
- With resource scaling now being mostly based on cores & storage – this means vastly more cores and storage



Next Generation IT Architecture and IT Supply Chain

1. Architecture and design of the “data center is the system”

- Very large scale 400K to 1M “cores” (50K to 100K servers)
- Managed as single integrated system including datacenter networking and storage
- 10X to 100X improvement in IT cost of delivery
- Design to handle diverse workload/jobs
- System is general purpose (i.e., not strongly workload optimized), but may include workload optimized components
 - By analogy, S/390 was general purpose, but included workload optimized components such as vector unit (for scientific workloads), and encryption unit (for banking)
 - Need a balance between workload specific components and general purpose based on utilization of components and effective cost.

2. Clusters of data centers

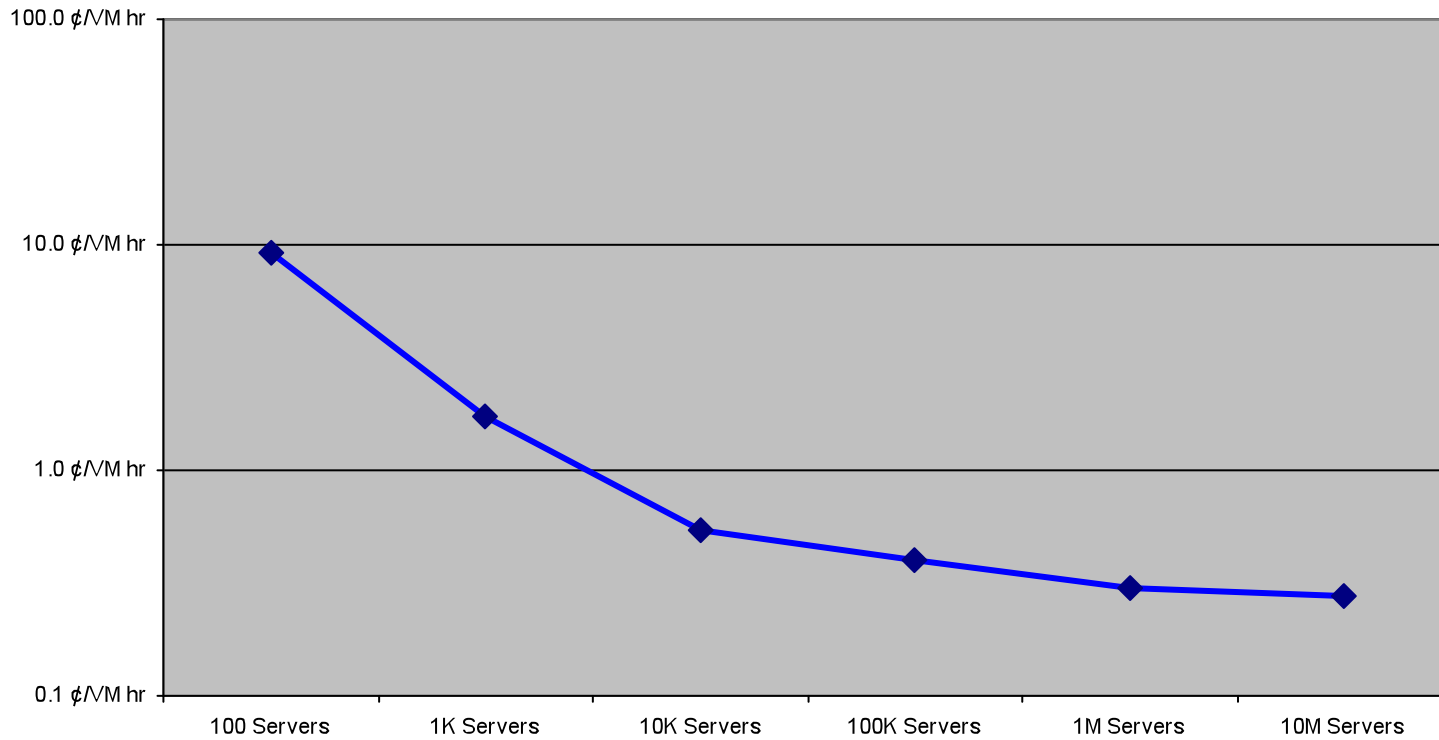
- Geographically distributed for lot of reasons

3. Its an integrated HW, software, and process “System Design”

- HW optimized for scale that is much bigger than traditional rack or blade designs (Google container datacenter design probably has breakeven order point in vicinity of 500 containers, ~500K servers, to justify DE \$'s to develop)
- Management software is an even bigger problem than optimized HW design
- Need platform services on top of all of that
- Radical redo of other aspects of IT supply chain/business models

Achievable Cost vs Scale

Total Cost per "Small" VM hr
("Small VM/hr as arbitrary unit of throughput)

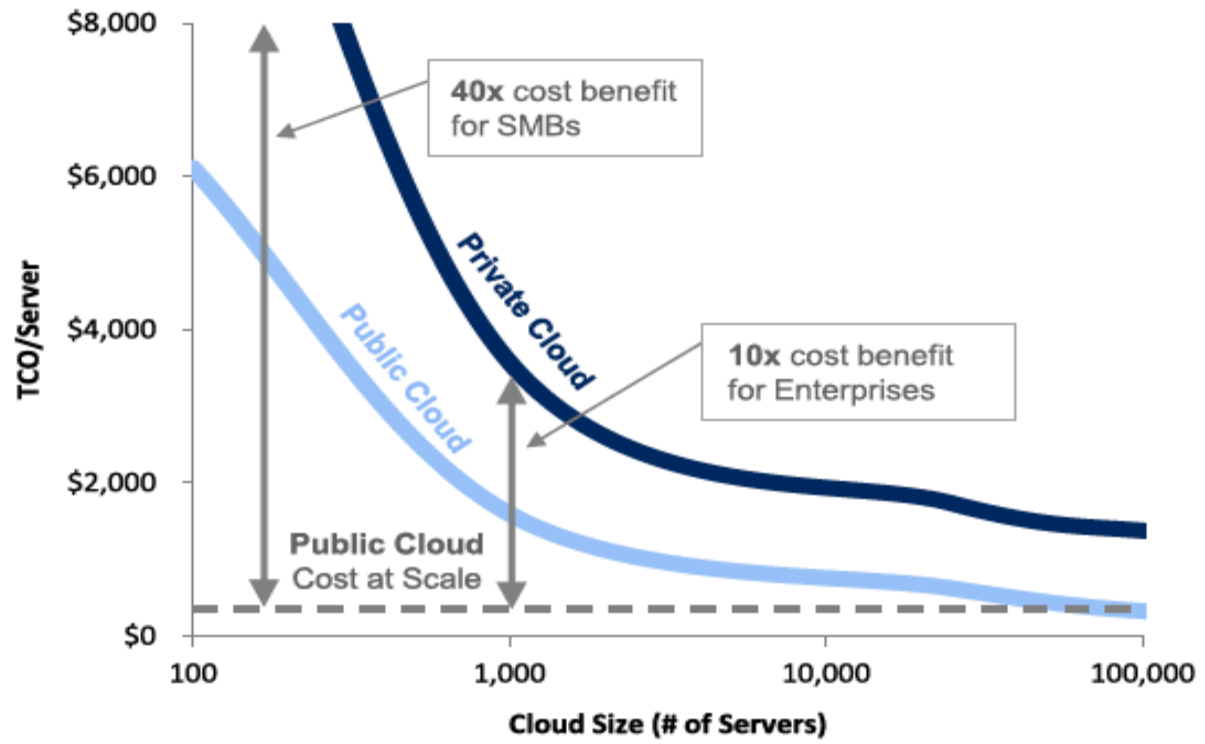


Legacy IT delivery does not have a cost problem because it is done badly,
but because legacy IT delivery is operating at wrong scale

Achievable Cost vs Scale

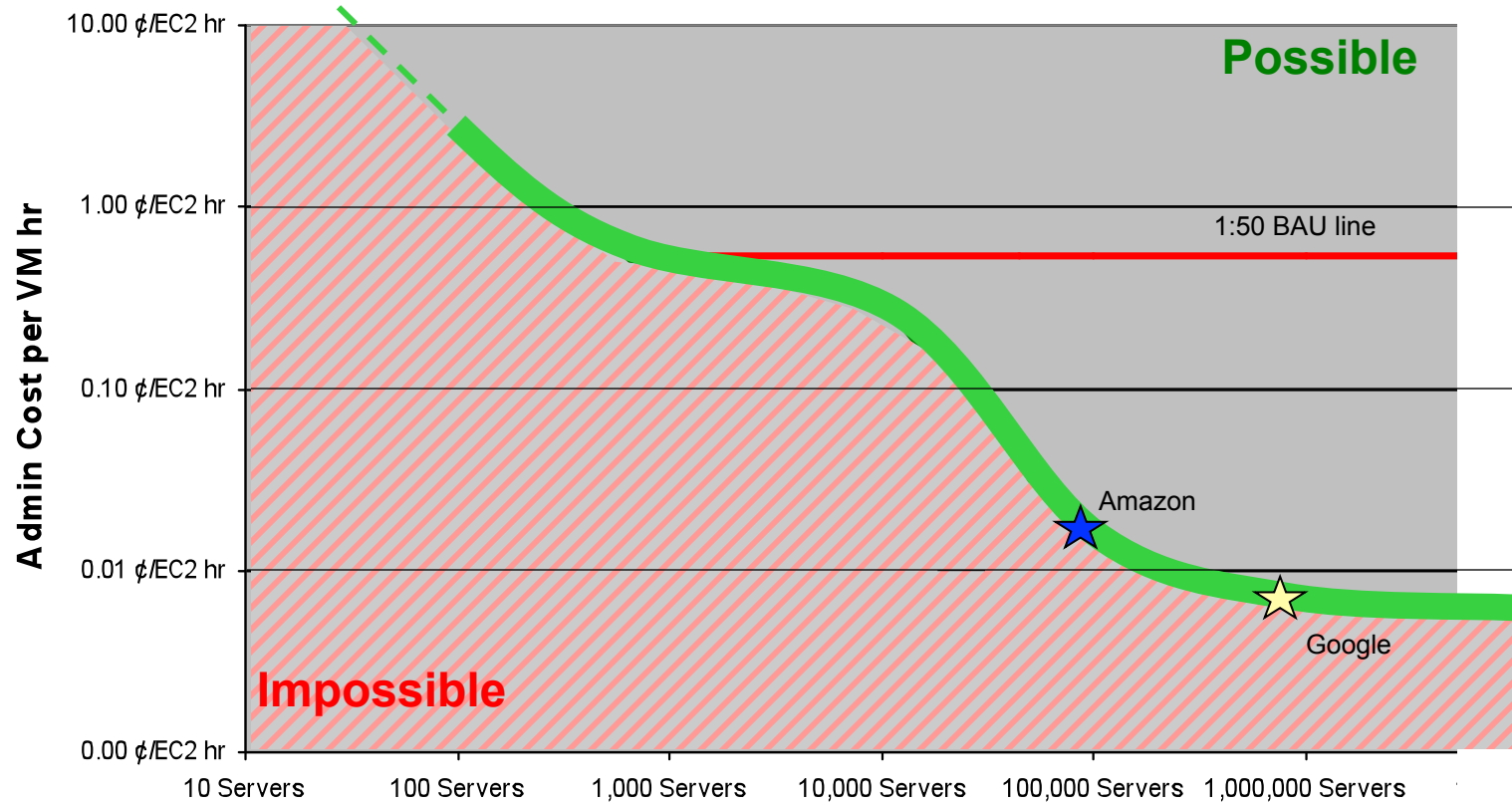
Microsoft

FIG. 22: COST BENEFIT OF PUBLIC CLOUD



Source: Microsoft.

Administration Cost Model with Scale



Some thoughts, conclusions

1. One cannot ignore cost & scale implications, even in the legacy/private cloud context – the cost delta is too large (currently ~50X)
 - Ultimately, even high cost private cloud customers will have majority of their IT consumption (by MIPS, GB's, ...) sourced from a more cost competitive delivery model
 - New growth apps/workloads simply are not viable with current BAU delivery costs
2. Economies of scale is driving re-integration of the industry
 - Relatively few very large cloud providers, or cloud provider brands
 - Vastly larger number of solutions providers (inherent to solution domain skills).
 - Interface between solutions and cloud providers likely above IaaS more like PaaS
3. Each cloud brand/platform is going to be highly proprietary.
 - Technology is immature, its not difficult to think of high value innovations
 - Competitors can get to market with high value differentiation much faster than standardization processes can homogenize the different brands.
 - Solutions will find each brand highly sticky, and it will be difficult to move solutions from one platform to another.
 - Few solution providers will take a 6 month time to market hit and a 40% performance hit (cost hit) to avoid using proprietary platform service.