**Cloudonomics Law #1: Utility services cost less even though they cost more.**  
An on-demand service provider typically charges a utility premium — a higher cost per unit time for a resource than if it were owned, financed or leased. However, although utilities cost more when they are used, they cost nothing when they are not. Consequently, customers save money by replacing fixed infrastructure with clouds when workloads are spiky, specifically when the peak-to-average ratio is greater than the utility premium.

**Cloudonomics Law #2: On-demand trumps forecasting.**  
The ability to rapidly provision capacity means that any unexpected demand can be serviced, and the revenue associated with it captured. The ability to rapidly de-provision capacity means that companies don’t need to pay good money for non-productive assets. Forecasting is often wrong, especially for black swans, so the ability to react instantaneously means higher revenues, and lower costs.

**Cloudonomics Law #3: The peak of the sum is never greater than the sum of the peaks.**  
Enterprises deploy capacity to handle their peak demands – a tax firm worries about April 15th, a retailer about Black Friday, an online sports broadcaster about Super Sunday. Under this strategy, the total capacity deployed is the sum of these individual peaks. However, since clouds can reallocate resources across many enterprises with different peak periods, a cloud needs to deploy less capacity.

**Cloudonomics Law #4: Aggregate demand is smoother than individual.**  
Aggregating demand from multiple customers tends to smooth out variation. Specifically, the “coefficient of variation” of a sum of random variables is always less than or equal to that of any of the individual variables. Therefore, clouds get higher utilization, enabling better economics.

**Cloudonomics Law #5: Average unit costs are reduced by distributing fixed costs over more units of output.**  
While large enterprises benefit from economies of scale, larger cloud service providers can benefit from even greater economies of scale, such as volume purchasing, network bandwidth, operations, administration and maintenance tooling.

**Cloudonomics Law #6: Superiority in numbers is the most important factor in the result of a combat (Clausewitz).**  
The classic military strategist Carl von Clausewitz argued that, above all, numerical superiority was key to winning battles. In the cloud theater, battles are waged between botnets and DDoS defenses. A botnet of 100,000 servers, each with a megabit per second of uplink bandwidth, can launch 100 gigabits per second of attack bandwidth. An enterprise IT shop would be overwhelmed by such an attack, whereas a large cloud service provider — especially one that is also an integrated network service provider — has the scale to repel it.

**Cloudonomics Law #7: Space-time is a continuum (Einstein/Minkowski)**  
A real-time enterprise derives competitive advantage from responding to changing business conditions and opportunities faster than the competition. Often, decision-making depends on computing, e.g., business intelligence, risk analysis, portfolio optimization and so forth. Assuming that the compute job is amenable to parallel processing, such computing tasks can often trade off space and time, for example a batch job may run on one server for a thousand hours, or a thousand servers for one hour, and a query on Google is fast because its processing is divided among numerous CPUs. Since an ideal cloud provides effectively unbounded on-demand scalability, for the same cost, a business can accelerate its decision-making.

**Cloudonomics Law #8: Dispersion is the inverse square of latency.**  
Reduced latency — the delay between making a request and getting a response — is increasingly essential to delivering a range of services, among them rich Internet applications, online gaming, remote virtualized desktops, and interactive collaboration such as video conferencing. However, to cut latency in half requires not twice as many nodes, but*four*times. For example, growing from one service node to dozens can cut global latency (e.g., New York to Hong Kong) from 150 milliseconds to below 20. However, shaving the next 15 milliseconds requires a thousand more nodes. There is thus a natural sweet spot for dispersion aimed at latency reduction, that of a few dozen nodes — more than an enterprise would want to deploy, especially given the lower utilization described above.

**Cloudonomics Law #9: Don’t put all your eggs in one basket.**  
The reliability of a system with *n* redundant components, each with reliability *r,*is 1-(1-*r*)*n*. So if the reliability of a single data center is 99 percent, two data centers provide four nines (99.99 percent) and three data centers provide six nines (99.9999 percent). While no finite quantity of data centers will ever provide 100 percent reliability, we can come very close to an extremely high reliability architecture with only a few data centers. If a cloud provider wants to provide high availability services globally for latency-sensitive applications, there must be a few data centers in each region.

**Cloudonomics Law #10: An object at rest tends to stay at rest (Newton).**  
A data center is a very, very large object. While theoretically, any company can site data centers in globally optimal locations that are located on a core network backbone with cheap access to power, cooling and acreage, few do. Instead, they remain in locations for reasons such as where the company or an acquired unit was founded, or where they got a good deal on distressed but conditioned space. A cloud service provider can locate greenfield sites optimally.

Video Exapanation: https://www.youtube.com/watch?v=he4QnJB\_ycM&t=40s