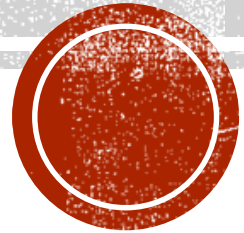
CLOUDONOMICS
(THE ECONOMICS OF CLOUD COMPUTING)

Dr. Manjunath Hegde
Dept. of Computer Applications
Manipal Institute of Technology, Manipal



INTRODUCTION (1/3)

- Economics is the study of what constitutes balanced human behavior in the attempt to fulfil needs and wants.
- Economics has always been a powerful force in driving industry transformations and as more and more customers evaluate cloud computing investments that will significantly affect Return on Investment.
- It can be win – win
  - You can make money in the Cloud by letting somebody save money in the Cloud and that's not a contradiction
- You need (and want) to consider the big picture
  - The Cloud is not just about the CPU price per hour, but it's about preparing your IT to be as flexible as possible, whether you're a user or a provider
- The basic mechanics are easy to understand
  - You don't need a degree in finance to take the first two points home

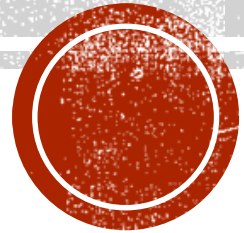
INTRODUCTION (2/3): SOME TERMS MUST BE KNOWN

- Marginal cost and value
  - Think: peanuts for an additional passenger on a plane vs. the price he pays for the ticket
- Opportunity cost
  - Think: pleasure gained when going to the movies vs. studying (and getting good grades and for these being admitted to college)
- Time value of money
  - Think: 100 Rs is the same as 105 Rs in a year if you get 5% interest, but what's 100 Rs in 2 hours if you're hungry now?

INTRODUCTION (3/3)

- There are many reasons for organizations to move from traditional IT infrastructure to Cloud Computing.
- One of the most cited benefits is the economics of the Cloud.
- Many people point out the cost savings that Cloud Computing brings to an organization, we believe attention should be drawn to four distinct mechanisms through which these cost savings are generated;
  - By lowering the opportunity cost of running technology
  - By allowing for a shift from capital expenditure to operating expenditure
  - By lowering the total cost of ownership (TCO) of technology
  - By giving organizations the ability to add business value by renewed focus on core activities
- In this paper we detail these four mechanisms and introduce several case studies and examples to show the increased economic value that Cloud Computing brings to an organization.

THE 10 LAWS OF CLOUDONOMICS



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.
- Under this strategy, the total capacity deployed is the sum of these individual peaks.
- Since clouds can reallocate resources across many enterprises with different peak periods, a cloud needs to deploy less capacity.
- For example: Consider the daily internet usage quota is 2 gB

  So sum of the peak for quota is 14 GB/week

  But peak of the sum may different (Monday 1.5GB, Tuesday 1GB etc.)

CLOUDONOMICS LAW #4 & #5

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.

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).

- 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,
- 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.

CLOUDONOMICS LAW #9

Don't put all your eggs in one basket.

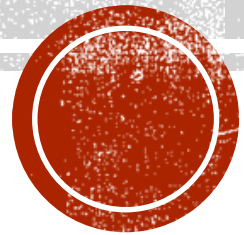
- The reliability of a system with n redundant components, each with reliability r, is 1-(1-r)n.
- So if the reliability of a single data center is 99 percent, two data centers provide four nines (99.99 percent) and three data centers provide six nines (99.9999 percent).
- While no finite quantity of 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.

THE 10 LAWS OF BEHAVIORAL CLOUDONOMICS



BEHAVIOURAL LAW #1

Risk and Loss Aversion

- There are emotional and perceptual asymmetries between losses and gains.
- A loss is more painful than a commensurate gains
- Certainly chief information officers must exercise due diligence regarding proposed cloud initiatives. Also aware about asymmetries causes that overweighted relative to benefits such as total cost reduction and enhanced agility.

BEHAVIOURAL LAW #2

Flat-Rate Bias

- One effect of loss aversion is that consumers often prefer flat-rate plans when pay-per-use would cost less.
- There are several reasons for the "flat-rate bias.
- The "insurance effect": Customers want to smooth monthly bills and appease their aversion to loss.
- The "taxi meter effect": Watching the taxi meter running up a bill reduces the pleasure associated with an activity. Even without an actual meter, "mental accounting" can have the same impact.
- The "convenience effect" is related to the ease of selecting the default option, which is often flat rate.
- The "overestimation effect" occurs when customers believe that their usage levels will be higher than they actually are. Consumers are likely to overestimate their ability to forecast future usage, leading them to overpay by selecting the wrong plan

BEHAVIOURAL LAW #3

Need for Control and Autonomy

- Author David Rock reports that people have a deep-seated need for control over their environment, or they may exhibit "learned helplessness" and shortened life spans.
- Owning a data center may provide a perception of greater control over assets, making dashboards, portals, transparent policies and fine-grained management essential for cloud providers.
- Autonomy —"I can do it myself!"— is an important driver for cloud computing. Developers can autonomously procure infrastructure resources, and platform services enable the democratization of IT.

BEHAVIOURAL LAW #4

Fear of Change

- It is also observes that people are often uncomfortable with uncertainty and therefore fear change.
- The cloud offers not just new technologies but new business and organizational models.
- Consequently, overcoming the inertia of the traditional owned asset model may require free trials, unalterable and explicit privacy policies, and/or multiyear price guarantees.

BEHAVIOURAL LAW #5

The Endowment Effect

- People value goods that they already own more than they would pay to acquire them.
- It is showed that for the same hard-to-acquire IPL tickets, people were willing to pay up to about 1700 Rs, but weren't willing to sell them for less than 2,500 Rs.
- Add in the choice-supportive bias, which rationalizes selected options and discounts unselected ones, and a stubborn fondness for existing IT technology and organization assets can be understood.

BEHAVIOURAL LAW #6 & #7

The Status Quo Bias and Escalation of Commitment

- Moreover, we tend to prefer things the way they've always been, and invest additional amounts in past strategies which we have pursued.
- Again, this can lead to inertia slowing the adoption of new approaches.

Hyperbolic Discounts and Instant Gratification

- People tend to discount future risks and benefits hyperbolically, that is, more steeply than accounting texts teach
- A chocolate chip cookie is much more valuable now than in an hour. This is good for the cloud, which promises instant gratification via on-demand services.
- Moreover, the "pain" of payment is deferred, thus discounted.

BEHAVIOURAL LAW #8 & #9

The Zero-Price Effect

- Ariely argues that zero is special.
- People would rather receive a free 10 Rs gift certificate (a 10 Rs gain) than pay 7 Rs for a 20 Rs one (13 Rs gain).
- This also benefits the cloud, since up-front costs are typically eliminated.

Need for Status

- Rock points out that humans and other social primates have exquisite, fine-grained status detectors.
- For cloud adoption, the status associated with managing a large IT organization with a substantial asset base needs to be replaced by the status accruing to being perceived as an innovator through the use of cloud services.

BEHAVIOURAL LAW #10

Paradox of Choice

- Web commerce has enabled a shift from a few big hits to a long tail of boundless choice.
- However, too many choices may cause paralysis by analysis, reducing sales.
- Fewer bundles, rather than unlimited configurability, may best boost service provider revenues.