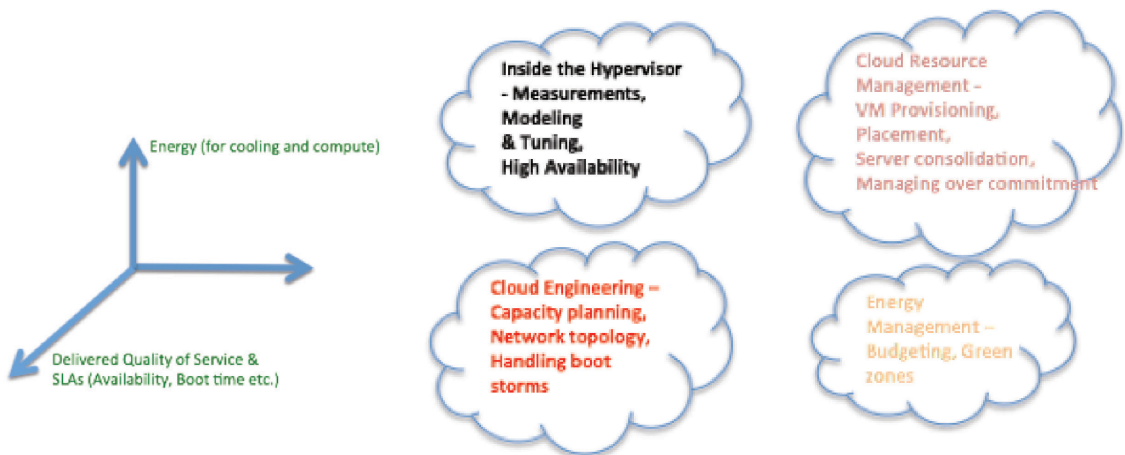


# Behind the cloud-technology to power today's infrastructure clouds



Cloud computing is the concept of computing as a utility bringing the illusion of infinite computing power with a pay-as-you-use model of billing to the consumer. It is analogous to the concept of other utilities such as water and electricity, but applied to computing. Cloud computing services are segregated into three layers based on the abstractions they provide:

- **Infrastructure as a Service (IaaS):** delivers computing infrastructure comprising CPUs, Memory, Storage and Network as a utility to consumers. Amazon is a popular compute public cloud provider that provides computational platforms via their EC2 elastic cloud and storage via the S3 storage cloud.
- **Platform as a Service (PaaS):** delivers software development platforms as a utility for consumers to build applications on. Google's App Engine and salesforce.com are popular platforms to build enterprise applications on.
- **Software as a Service (SaaS):** delivers the highest level of abstraction in terms of consumable software services. Email and eCommerce storefronts are popular examples of SaaS instances that we use everyday.

Our research focuses primarily on the IaaS cloud provider. IaaS clouds are powered by virtualising existing systems and offering them as packages of pre-determined size to consumers as virtual machines. A virtual machine (VM) appears like a computer for all practical purposes to the user but executes as a software layer atop a virtual machine monitor which provides the middleware for multiple virtual machines to multiplex onto a single set of hardware resources.

Some of the interesting problems that cloud providers face and we are working to solve are:

- **VM placement:** When a cloud provider gets a request for a set of VMs, an issue is to try and map these VMs onto the physical machines (PMs) in the data centre. This mapping can be guided by several different objective functions - minimising the set of PMs they occupy, minimising energy consumption of the data centre, maximising the quality of service delivered to the VMs and so on. The placement problem is important to the provider from many different angles - energy consumption, guaranteeing SLAs to their consumers and guarding against data centre resource fragmentation - and so is an important problem.
- **Efficient VM monitors:** The middleware layer that enables multiplexing of VMs onto the set of hardware resources implements many processes to manage and allocate resources. It also enables the provider to overcommit resources and manage the degree of over commitment so as to maximise profits while not sacrificing the quality of service delivered to individual customers. We have focused on making these processes more efficient and on providing finer granularity of control to cloud resource managers.

The cloud is here to stay and the scale at which IaaS providers operate makes it imperative that they have tools and processes to maximise the use of their resources and yet deliver acceptable qualities of services while minimising their operational costs. Have solved many of these and continue to work on other problems that have significantly impacted today's cloud providers.