1.1 Benefits of Cloud Computing

- Not an all-or-nothing service
  - You can gradually move to the cloud, called also ***lift and shift***
- You’re able to spend more time on what matters and less time managing underlying details

▼ Cost Effective

- Provides ***pay-as-you-go*** or ***consumption-based*** pricing model
  - No upfront infrastructure cost
  - No need to purchase and manage costly infrastructure/hardware that you may not use to its fullest
  - The ability to pay for additional resources only when they are needed
  - The ability to stop paying for resources that are no longer needed
- Enables better cost predictions using pricing of individual resources/services
  - You can analyze future growth using historical data

▼ Scalability

- Increase or decrease the resources and services used based on the demand or workload at any given time
- Cloud computing supports both:
  - Horizontal scaling
    - Scaling “out”
    - Adding more servers that function together as one unit
  - Vertical scaling

        - Scaling “up”
        - Adding resources to increase the power of an existing server
            - ex: Add more CPUs, or more memory
- Scaling can be done manually or automatically based on specific triggers such as CPU utilization

---

▼ Elasticity

- Cloud computing systems can automatically add & remove resources to meet the current demand
- Example:
    - Add resources for the peak operating hours during which most people access the application
        - Only pay for increased resources during those hours
    - Remove resources when the traffic normalizes
        - Do not pay anymore

▼ Agility

- With On-prem you have to plan ahead and use for a long period of time; not easy to repurpose
- With Cloud, as my demands change I can provision on-demand at any time
- Provides more flexibility

▼ High Availability

- Redundancy is often built into cloud services architecture
    - If one component fails, a backup component takes its place
    - Referred to as ***fault tolerance*** and it ensures that the customer isn’t impacted when a disaster occurs

- Fully redundant data centers located in various regions all over the globe
- Enables local presence close by to give best response time
- Replicate services into multiple regions for redundancy and locality