Cloud Infrastructure Guide

What is Cloud Computing?

Cloud computing is the delivery of computing services including servers, storage, databases, networking, software, analytics, and intelligence over the Internet (the cloud) to offer faster innovation, flexible resources, and economies of scale. Organizations can rent access to computing resources rather than owning and maintaining physical data centers and servers.

Cloud Service Models

Infrastructure as a Service (IaaS) provides virtualized computing resources over the internet. With IaaS, organizations rent IT infrastructure including servers, virtual machines, storage, networks, and operating systems from a cloud provider on a pay-as-you-go basis. Major providers include Amazon Web Services EC2, Google Compute Engine, and Microsoft Azure Virtual Machines. IaaS offers maximum flexibility and control over computing resources.

Platform as a Service (PaaS) provides a complete development and deployment environment in the cloud. PaaS includes infrastructure (servers, storage, networking) plus middleware, development tools, database management systems, and business intelligence services. Developers can build, test, and deploy applications without worrying about underlying infrastructure. Examples include Heroku, Google App Engine, and AWS Elastic Beanstalk.

Software as a Service (SaaS) delivers software applications over the internet on a subscription basis. The cloud provider hosts and manages the software application and underlying infrastructure, handling maintenance, updates, and security patches. Users access the application through web browsers. Common examples include Google Workspace, Microsoft 365, Salesforce, and Dropbox.

Benefits of Cloud Computing

Cloud computing offers numerous advantages including cost savings through reduced capital expenditure, scalability to handle varying workloads, reliability through redundancy and backup systems, performance through global networks of data centers, security through enterprise-grade protection, and speed in deploying new resources. Organizations can focus on their core business rather than managing IT infrastructure.