Cloud Computing Services and Architectures

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. Cloud computing services are typically categorized into three main types: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). Infrastructure as a Service (IaaS) provides virtualized computing resources over the internet. IaaS providers, such as AWS, Microsoft Azure, and Google Cloud Platform, offer virtual machines, storage, networks, and operating systems on a pay-as-you-go basis. This model allows businesses to avoid the upfront cost and complexity of buying and managing physical servers and datacenter infrastructure. Platform as a Service (PaaS) provides a platform allowing customers to develop, run, and manage applications without the complexity of building and maintaining the infrastructure typically associated with developing and launching an app. Examples include AWS Elastic Beanstalk, Microsoft Azure App Services, and Google App Engine. Software as a Service (SaaS) delivers software applications over the internet, on a subscription basis. With SaaS, cloud providers host and manage the software application and underlying infrastructure, and handle any maintenance, like software upgrades and security patching. Examples include Google Workspace, Microsoft Office 365, and Salesforce. Cloud architectures can be public, private, or hybrid. Public clouds are owned and operated by third-party cloud service providers. Private clouds are used exclusively by a single business or organization. Hybrid clouds combine public and private clouds, allowing data and applications to be shared between them.