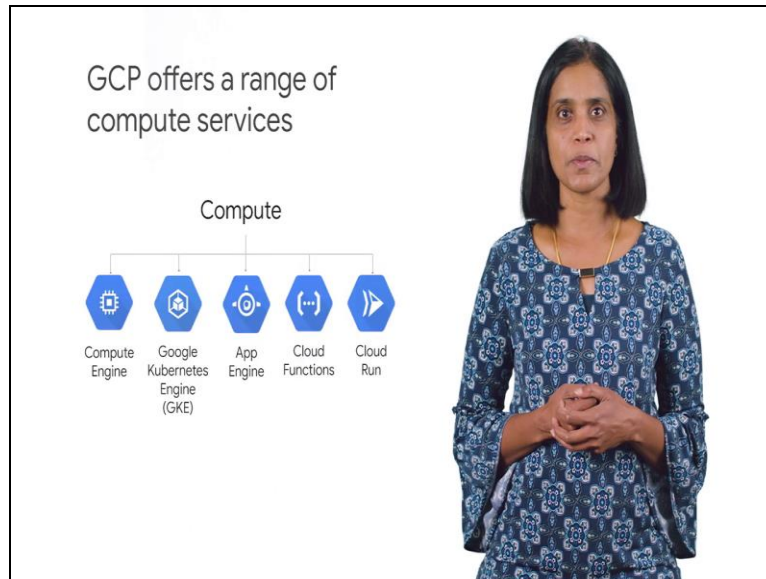


Google Cloud Computing Foundation Course
Sowmya Kannan
Department of Computer Science
Indian Institute Technology Kharagpur

Lecture-5
Google Cloud Architecture

(Refer Slide Time: 00:07)



Next you will focus on Google specific offerings in the cloud. Google cloud platforms products and services can be broadly categorized as compute, storage, Big Data, machine learning networking and operations or tools. Leveraging compute can include virtual machines via compute engine, running docker containers in a managed platform using google kubernetes engine, deploying applications in a managed platform like App Engine or running event-based server less code using cloud functions.

A variety of managed storage options are available as well. For unstructured storage there is cloud storage for managed relational databases there is cloud sequel or cloud spanner and for no sequel there are options like cloud datastore or cloud BigTable.

(Refer Slide Time: 01:00)

GCP offers services for getting value from data

Big data

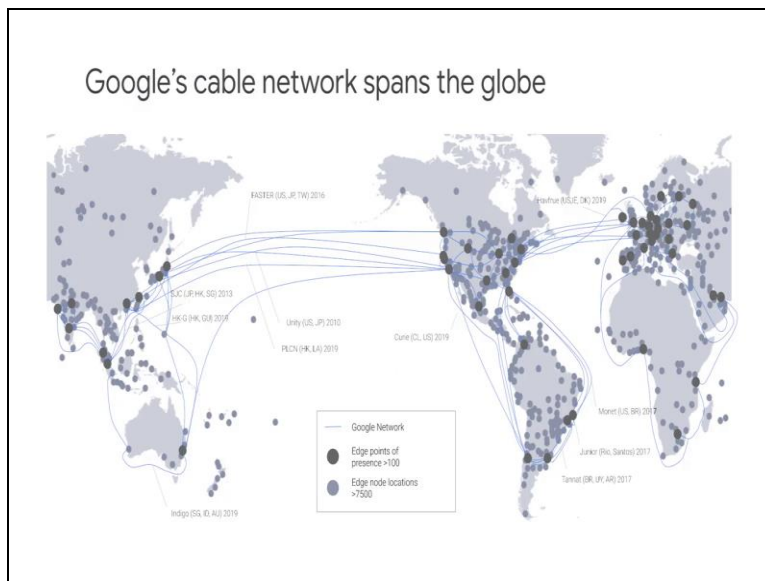
- BigQuery
- Cloud Pub/Sub
- Cloud Dataflow
- Cloud Dataproc
- Cloud Datalab

Machine learning

- Cloud ML Engine
- Cloud Vision API
- Cloud Speech API
- Cloud Translation API
- Cloud AutoML

Managed services dealing with big data and machine learning are available as well.

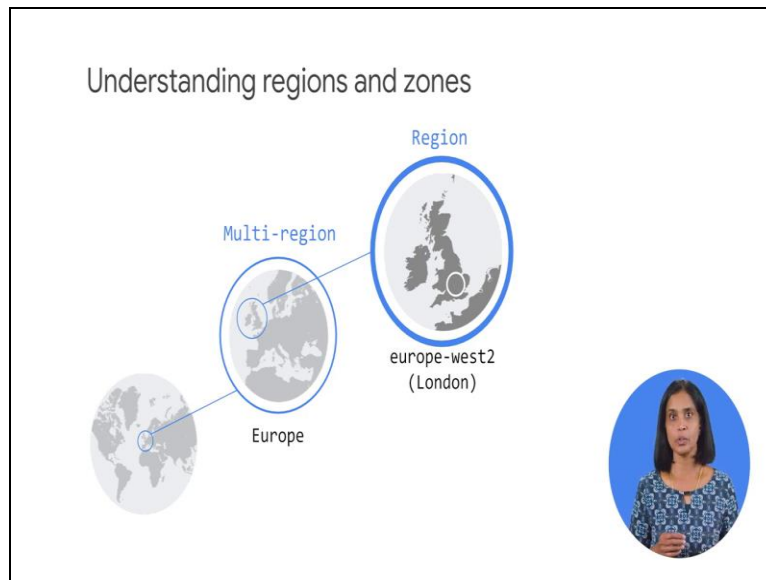
(Refer Slide Time: 01:07)



Our data centers around the world are interconnected by the Google Network which by some publicly available estimates carries as much as 40% of the world's Internet traffic today. This is the largest network of its kind on earth and it continues to grow it is designed to provide the highest possible throughput and the lowest possible latencies for applications. The network interconnects with the public Internet at more than 90 internet exchanges and more than 100 points of presence worldwide.

When an internet user sends traffic to a Google resource we respond to the users request from an edge network location that will provide the lowest delay or latency. Our edge caching network places content close to end-users to minimize latency. Applications in GCP can take advantage of this edge network too.

(Refer Slide Time: 02:14)



Google cloud divides the world into 3 multi regional areas. The Americas, Europe and Asia Pacific, next the three multi regional areas are divided into regions which are independent geographic areas on the same continent. Within a region this fast network connectivity generally round-trip network latency of under one millisecond that is at the 95th percentile. As you can see one of the regions in Europe is Europe West to London.

Finally regions are divided into zones which are deployment areas for GCP resources within a focused geographic area. You can think of a zone as a data center within a region although strictly speaking a zone is not necessarily a single data center. Compute engine virtual machine instances reside within a specific zone. If that zone became unavailable so would your virtual machine and the workload running on it. Deploying applications across multiple zones enables fault tolerance and high availability.

(Refer Slide Time: 03:39)



Behind the services provided by a Google cloud platform lie a huge range of GCP resources. Physical assets such as physical servers and hard disk drives and virtual resources such as virtual machines and containers, we manage these resources within our global data centers. As of mid 2019 GCP has expanded across 20 regions 61 zones and more than 200 countries and territories. This expansion will continue.

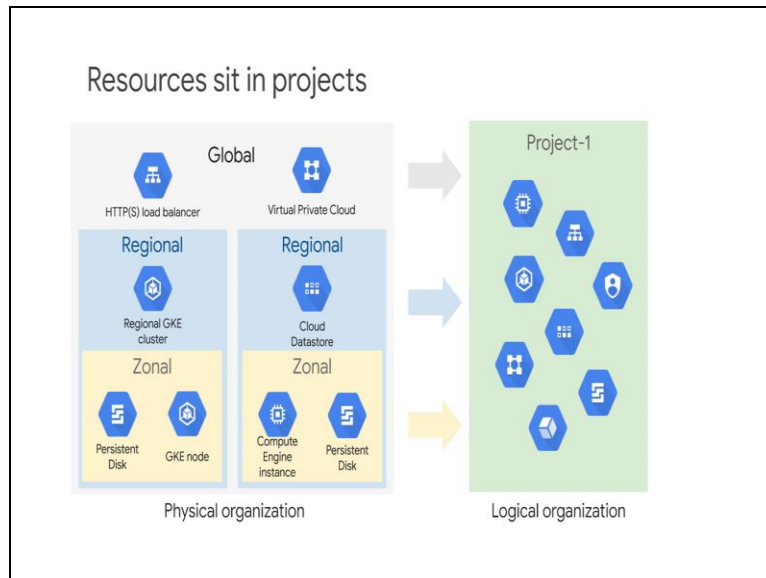
(Refer Slide Time: 04:17)

Zonal resources operate exclusively in a single zone

When you take advantage of GCP services and resources you get to specify those resources geographic locations. In many cases you can also specify whether you are doing so on a zonal level regional level or multi regional level. Zonal resources operate within a single zone which means that if a zone becomes unavailable, the resources won't be available either. A simple

example could be a compute engine virtual machine instance and it is persistent disks. GKE has a component called a node and these are zonal too.

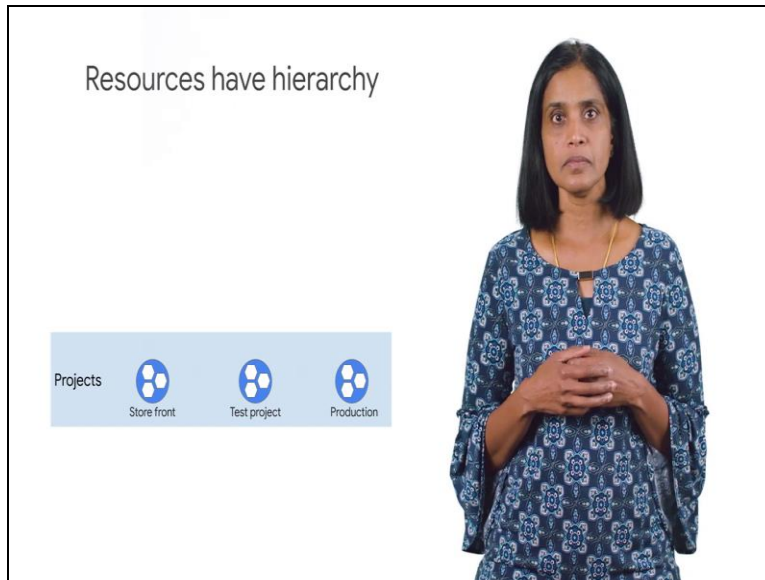
(Refer Slide Time: 05:00)



Regional resources operate across multiple zones but still within the same region. An application using these resources can be redundantly deployed to improve its availability. Finally global resources can be managed across multiple regions. These resources can further improve the availability of an application. Some examples of such resources include HTTP as load balancers and virtual private cloud networks.

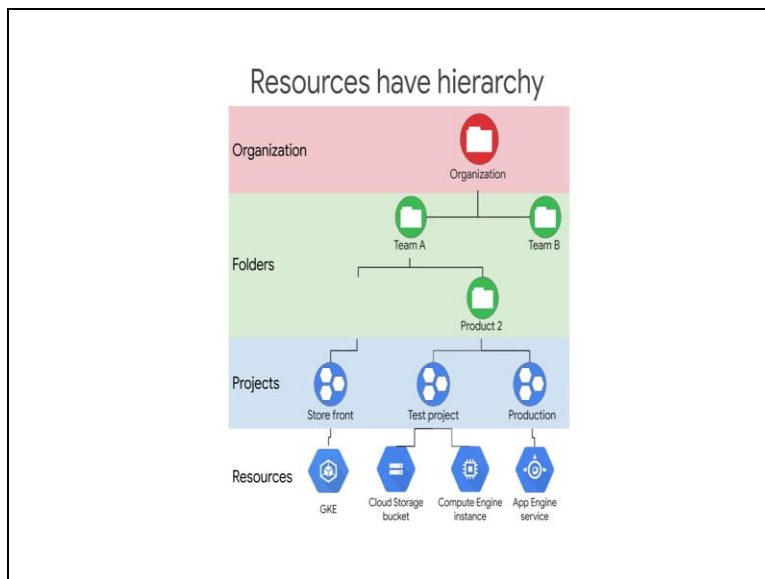
The GCP resources you use no matter where they reside must belong to a project. So, what is a project? A project is the base level organizing entity for creating and using resources and services and managing billing API's and permissions. Zones and regions physically organize the GCP resources you use. And projects logically organize them. Projects can be easily created managed deleted or even recovered from accidental deletions.

(Refer Slide Time: 06:17)



Each project is identified by a unique project ID and project number. You can name your project and apply labels for filtering. These labels are changeable but the project ID and project number remain fixed. Projects can belong to a folder which is another grouping mechanism.

(Refer Slide Time: 06:43)



You should use folders to reflect the hierarchy of your enterprise and apply policies at the right levels in your enterprise. You can nest folders inside folders, for example you can have a folder for each department and within each departments folder you can have subfolders for each of the teams that make it up. Each team's projects belong to its folder. A single organization owns the folders beneath it. An organization is the root node of a GCP resource hierarchy.

Although you are not required to have an organization to use GCP, organizations are very useful. Organizations let you set policies that apply throughout your enterprise. Also having an organization is required to use folders. The GCP resource are our key helps you manage resources across multiple departments and multiple teams within an organization. You can define an are our key that creates trust boundaries and resource isolation.

For example should members of your Human Resources team be able to delete running database servers and should your engineers be able to delete the database containing employee salaries? Probably not in either case. Cloud Identity and Access Management also called IAM lets you fine-tune access control to all the GCP resources you use. You define IAM policies that control user access to resources.