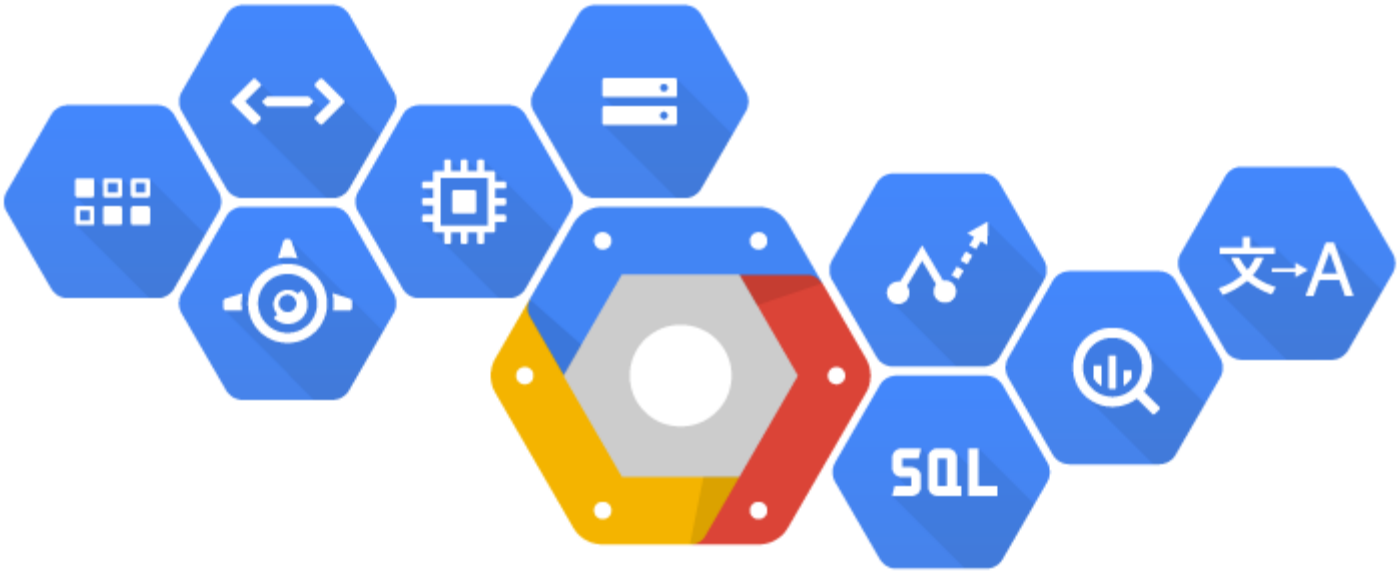
Introduction to Google Cloud Platform



Google Cloud Study Jam Session KCCITM

Agenda

- Why Google Cloud ?
- Infrastructure underpinning Google Cloud
- Components of Google Cloud
- Compute Services
- Networking Services
- Storage Service
- Big Data
- Machine Learning

Why Google Cloud ?

“Google Cloud is underpinned by the same infrastructure and innovation that powers Google products”

“Google has scaled seven products each of which has over a billion users each, every single day Google handles 1.4 petabytes of information in Gmail alone with 99.97% availability ”

“We are at the beginning of what’s possible with the cloud”

- Sundar Pichai (GCP Next 16 Keynote)

Why Google Cloud ?

Google's ability to build, organize, and operate a huge network of servers and fiber optic cables with an efficiency and speed that rocks physics on its heels.

This is what makes Google Google: its physical network, its thousands of fiber miles, and those many thousands of servers that, in aggregate, add up to the mother of all clouds"

- Wired

Google's Network Infrastructure

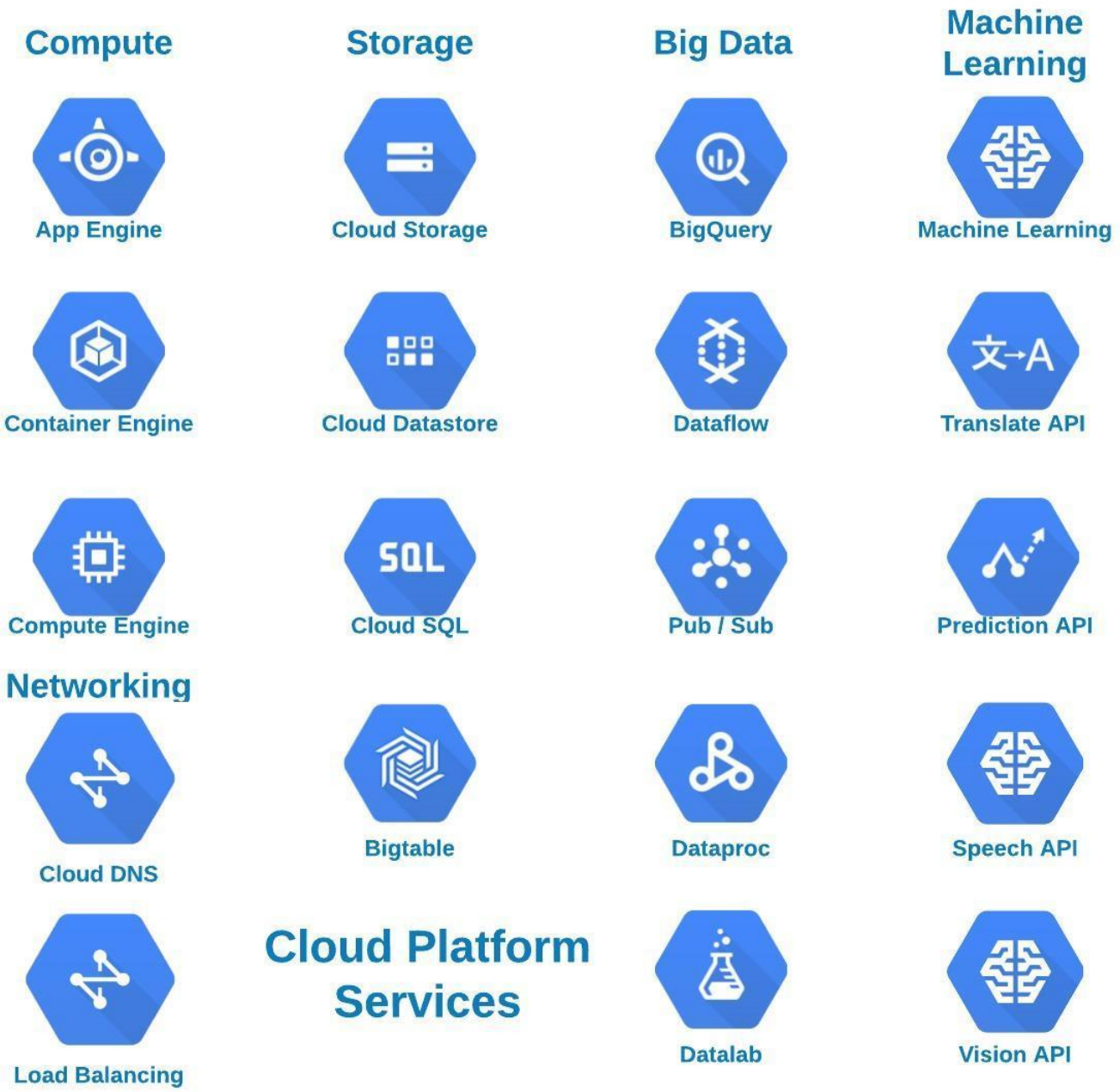
Global, meshed fiber backbone network interconnecting data centers with 70+ Edge points of presence in 33 countries with elements within ISP and access networks

Read More at:

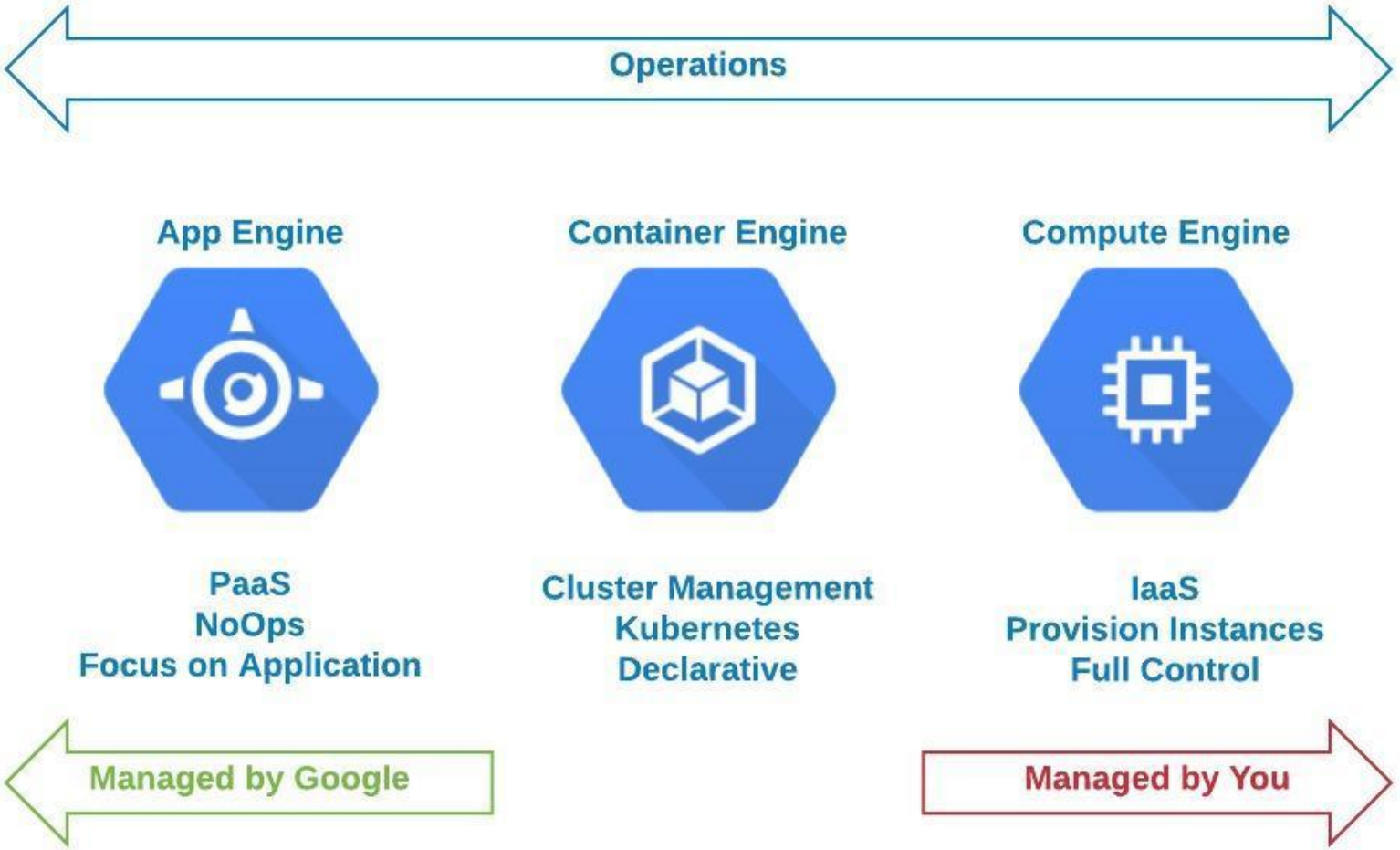
https://peering.google.com/#/infrastructure

https://cloudplatform.googleblog.com/2015/06/A-Look-Inside-Googles-Data-Center-Networks.html

http://www.wired.com/2015/06/google-reveals-secret-gear-connects-online-empire/


Compute
Storage
Big Data
Machine Learning
App Engine
Cloud Storage
BigQuery
Machine Learning
Container Engine
Cloud Datastore
Dataflow
Translate API
Compute Engine
Cloud SQL
Pub / Sub
Prediction API
Networking
Cloud DNS
Bigtable
Dataproc
Speech API
Load Balancing
Cloud Platform Services
Datalab
Vision API

Compute Services



Compute Engine

- Configurable Custom Machine Types
- Live migration
- Up to 2 GBPS networking between VMs
- Instance metadata and startup scripts
- HTTP(s) and Network load balancing
- APIs for auto-scaling and group management
- Sub-Hourly billing, Automatic sustained use discount
- Preemptible VMs (Spot Instances)

Container Engine

- Kubernetes based Container orchestration
- Uses underlying Compute Engine resources
- Declarative syntax for orchestration and scheduling Docker containers
- Managed Logging, Monitoring, and Scaling

App Engine

- Managed runtime for Java, Go, Python, & PHP
- Local SDK for developing, testing and deployment
- Auto-scaling based on demand
- Free daily quota, usage based billing
- 60s Request timeout
- Can’t write to local filesystem
- Limits on third party software

Load Balancing

- HTTP(S) and Network Load Balancing
- HTTP(S) Load balancing and auto-scaling across Compute Engine Regions
- Single Anycast external IP, simplifies DNS setup
- No pre-warming required, scales to 1 million+ QPS
- Policy based Auto-scaling of Instance groups
- Network Load balancing for TCP and UDP traffic within a Compute Engine Region
- Only healthy instances handle traffic

Cloud DNS

- Fully managed, Scalable and Highly Available DNS
- 100% availability SLA
- Programmatically manage zones and records with RESTful API
- Powered by global network of Anycast name servers
- Managed zones for projects
- Cost effective pricing tiers

Cloud Storage

- Highly scalable immutable object /blob store
- Standard variant (HA & low latency)
- Durable Reduced Availability variant (Reduced availability)
- Nearline Storage for archiving, backup and DR (~3s response)
- No capacity planning required
- All options accessed through the same API
- Can be mounted as file system using GCS Fuse

Cloud Datastore

- NoSQL database that can scale to billions of rows
- Fully managed service
- Automatically handles Sharding and Replication
- Support for ACID transactions, SQL like queries
- Fast and Highly Scalable
- Local development tools
- Access from anywhere through a RESTful Interface
- Free daily quota

Cloud Bigtable

- Massively scalable NoSQL
- For large workload applications - Terabytes to petabytes of data
- Low latency and high throughput
- Accessed using HBase API
- Native compatibility with Hadoop ecosystem
- Replicated storage
- Role based ACLs
- Encryption of in-flight and at rest data
- Used by Google Analytics and Gmail

Cloud SQL

- Managed MySQL
- Packages and Pay-per-use billing
- Second generation Cloud SQL is currently in Beta
- Vertical scaling for read and write
- Horizontal scaling for read
- Seamless integration with App Engine, and Compute Engine
- Data is automatically encrypted
- Automatic failover for high availability

Big Data Services (Fully Managed)

BigQuery
Analytics data warehouse
Stream data at 100,000 rows per second

Dataflow
Stream and Batch processing of data
Unified programming model

Pub/Sub
Scalable & Reliable enterprise messaging middleware

Dataproc
Managed Hadoop, Spark, Pig and Hive at affordable pricing

BigQuery

- Fully managed petabyte scale analytics data warehouse
- Near real-time interactive analysis of massive datasets
- Based on columnar structure for performance
- SQL like syntax for querying
- Scale storage and compute separately
- Pay for storage and compute used
- Benefit from integration points developed by partners

Dataflow

- Unified programming model for developing and executing scalable and reliable data pipelines
- Support for ETL, Analytics, Real-time computation, and Process orchestration
- Processes data using Compute Engine instances
- Open Source Java SDK for developing custom extensions
- Benefit from integration developed by GCP partners

Dataproc

- Fully managed Hadoop, Spark, Pig, and Hive
- Dataproc clusters can be resized at anytime, even when the jobs are running
- Clusters are billed minute-by-minute
- Clusters can use preemptible instances to further reduce cost
- Restful API and integration with Google Cloud SDK
- Easy to move existing ETL pipelines without redevelopment

Cloud Pub/sub

- Scalable and reliable messaging middleware
- Based on proven Google technologies
- Guaranteed “at least once” delivery with low latency
- Supports both pull and push delivery
- Fully managed and global by design taking advantage of all GCP regions
- Includes support for offline consumers

Cloud Datalab

- Interactive tool for large scale exploratory data analysis and visualization
- Based on Jupyter notebook (IPython)
- Code, documentation, results and visualizations all in notebook format
- Runs on Google App Engine
- Python, SQL, and JavaScript for data analysis
- Google charts or matplotlib for visualization
- Easy to deploy transformation, analysis models to BigQuery

Cloud Machine Learning

- Cloud Machine Learning is currently in Alpha
- Fully managed large scale Machine Learning Platform
- Fully managed and Integrated with Cloud Storage and BigQuery
- Uses open source TensorFlow framework that powers Google Photos, and Cloud Speech API
- Integrated with Cloud Dataflow for pre-processing
- Google has built custom Tensor Processing Units for efficiently running Machine Learning
- http://venturebeat.com/2016/05/18/google-is-bringing-custom-tensor-processing-units-to-its-public-cloud/
- http://www.infoworld.com/article/3072569/cloud-computing/googles-cloud-strategy-becomes-clearer-with-tensorflow.html

Translate API

- Simple API for translating an arbitrary string in to any supported language
- Programmatically detect a document's language
- Support for dozens of languages
- Highly Scalable high quality translation
- Supports Python, Java, Go and etc
- You can try it out from API Explorer
- Usage and billing calculated per million characters
- We can try it on APIs Explorer

Prediction API

- Predicts trends based on historical data
- Use cases:
  - Categorizing emails as spam or non-spam
  - Product recommendations
  - Assessing whether posted comments have positive or negative sentiment
- Data replicated using Cloud Storage
- Fast & Reliable (Most queries take less than 200 ms)
- RESTFul API is available for many popular languages

Cloud Vision API

- Image analysis based on powerful machine learning models
- Ability to classify images in to thousands of categories
- Detect individual objects and faces within the image
- API improves over time by building on insights
- Detect different types of inappropriate content
- Analyze emotional facial attributes
- Object Character Recognition to detect text with automatic language identification

Cloud Speech API

- Currently in Alpha
- Audio to text powered by neural network models
- Recognizes over 80 languages and variants
- Ability to filter inappropriate content
- Return partial results in real time as and when they become available
- Built-in noise elimination for a variety of environments
- API improves over time by building on insights

What Next

GCP Blog

https://cloudplatform.googleblog.com/

GCP Docs

https://cloud.google.com/docs/