NobleProg

GCP Fundamentals

Core Infrastructure

Presented by Bin Wu

The World's Local Training Provider

Outline

- Cloud Computing Introduction
  - IaaS
  - PaaS
- Cloud Services
  - Network
  - IAM
  - Utils
  - Computing & Storage
  - Big Data & Machine Learning

NobleProg

GCP Introduction

- Advantages
  - PAYG (Pay As You Go)
  - Capital expenses => Operating expenses
  - Low cost (Maintainability, Extensibility, Stability) => focus on Biz
  - Container service with Kubernetes
  - Machine learning with services encapsulation
  - More developer friendly
- Infrastructure as a Service
- Platform as a Service
- X as a Service

NobleProg

Network - Regions & Zones

NobleProg

Services - GCP Ecosystem

- Computing
  - VM, PaaS, Containers
- Storage
  - Block, Object, DB engine
- Management
  - Monitoring
- Security
  - IAM
- Development

NobleProg

GCP - Projects

- Services are associated with project
  - Track resource and quota usage
  - Enable billing
  - Manage permissions and credentials
  - Enable services and APIs
- Identify a project
  - Name, Number, ProjectID (aka Application ID)
- Interacting with projects using the web console or API

NobleProg

Lab - Sign up (20min)

- Account sign-up
- Create a project

NobleProg

Identity and Access Management

Who can do what things against on which resource.

- Authentication
  - Identify a user or a service
- Authorization
  - Define what a user or service can do
  - RBAC
- IAM
  - Service account (Json file)
  - User account (OAuth2.0)
  - API Key

NobleProg

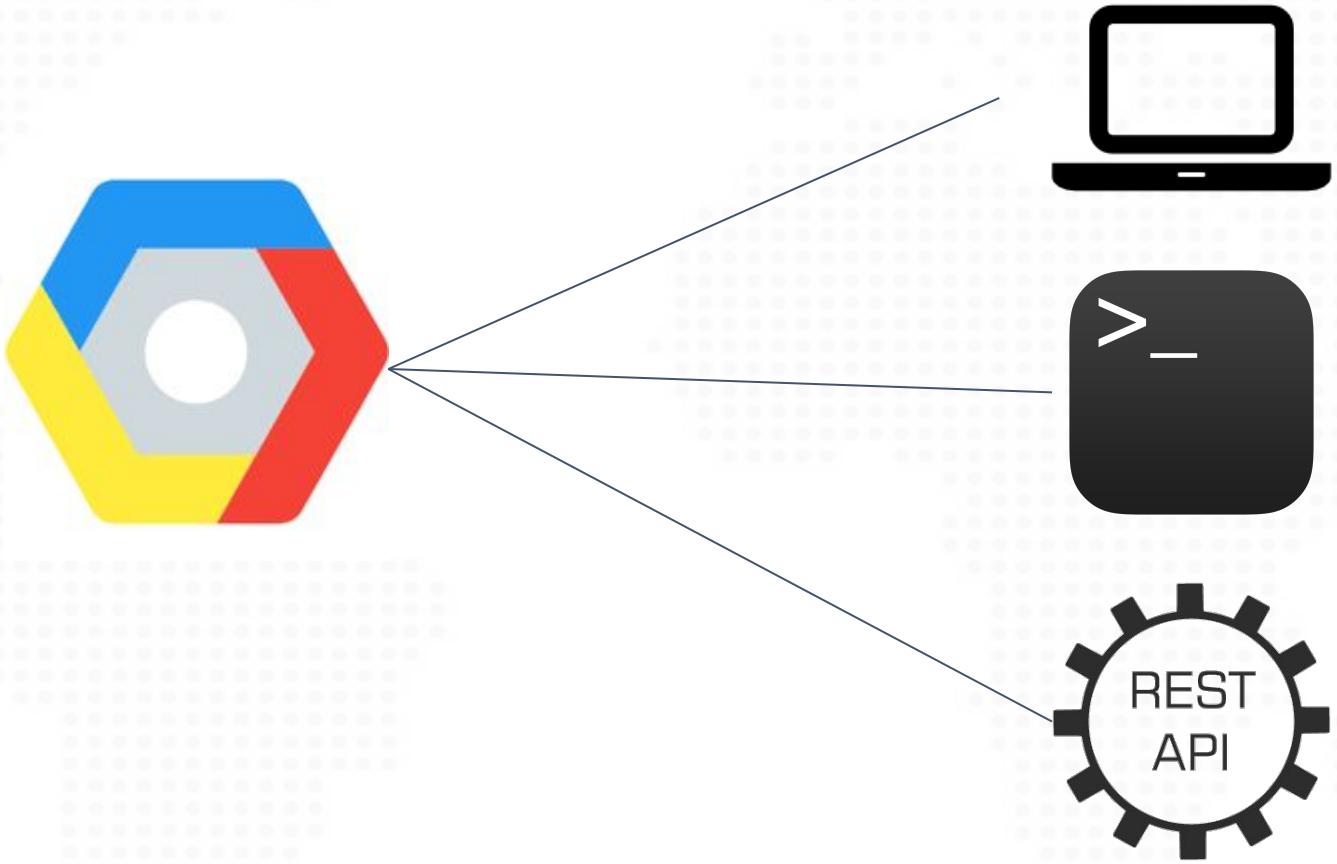
IAM - Primitive Roles

 **Owner**	 **Editor**	 **Viewer**	 Billing **admin**
Invite / remove members Delete project Add Editor rights	Deploy Applications Modify code Configure services Add viewer rights	Read-only Access	Manage billing Add / remove administrators

A project can have multiple owners, editors, viewers and billing administrators.

NobleProg

How to connect to GCP?



Cloud Console - Web UI

Cloud Shell (Command-line interface)

Cloud SDK / REST-based API

NobleProg

GCP Web Console

- Centralized console for all project data
- Developer tools
  - Source repo
  - Shell in browser
- Access to product APIs
- Manage projects and services

NobleProg

Google Cloud SDK

- SDK includes CLI tools for cloud platform products and services
  - gcloud, gsutil (cloud storage), bq (BigQuery)
- Available as docker image
- Available via Cloud Shell
  - Containerized version of cloud sdk running on compute engine instance

NobleProg

RESTful APIs

- Programmatic access to products and services
  - Typically JSON as an interchange format
  - OAuth2.0 protocol
- Enabled through the Google Cloud Platform Console
- Most APIs include daily quotas and rates (limits) that can be raised by request
  - Important to plan ahead to manage your required capacity
- Experiment with APIs Explorer

NobleProg

APIs Explorer

- URL: https://developers.google.com/apis-explorer
- It's an interactive tool that lets you easily try Google APIs using a browser
- With the APIs Explorer, you can:
  - Browser quickly through available APIs and versions.
  - See methods available for each API and what parameters they support along with inline documentation.
  - Execute requests for any method and see responses in real time.
  - Make authenticated and authorized API calls with ease.

NobleProg

Lab - Getting started (30min)

- Create VM
  - Web console
  - Command-line (in web console)
- Setup a dev environment based on your preference (PHP, java, etc.)
  - VM
  - Connect to MySQL
- Google cloud launcher

NobleProg

GCP PaaS (App Engine)

- A platform for building scalable web applications and mobile backends
- App Engine makes deployment, maintenance, and scalability easy. So you can focus on innovation and your business

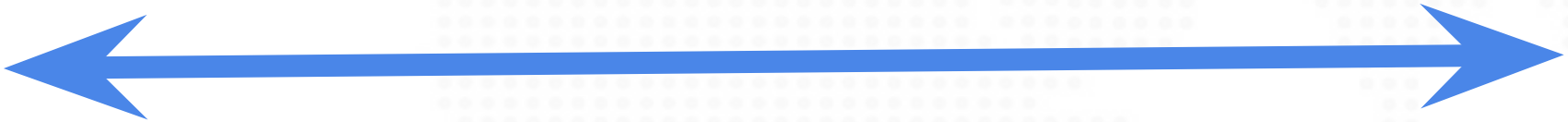
NobleProg

IaaS and PaaS

Compute Engine

App Engine

Towards managed infrastructure (DevOps)



Towards managed services (NoOps)

IaaS

Raw compute, storage and network
More granular control

PaaS

Present runtimes Java, Go, PHP, Python
...
Focus is application logic (Business)

Pay for what you allocated
More management overhead

Pay for what you use
Less management overhead

NobleProg

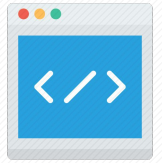
App Engine Standard Environment

- Managed runtimes for specific versions of Java, Python, PHP & Go
- Autoscale workloads to meet demand
- Free daily quota, usage based pricing
- SDKs for development, testing and deployment
- Need to conform to sandbox constraints:
  - No writing to local file system
  - Request timeouts at 60 seconds
  - Limit on 3rd-party software installations

NobleProg

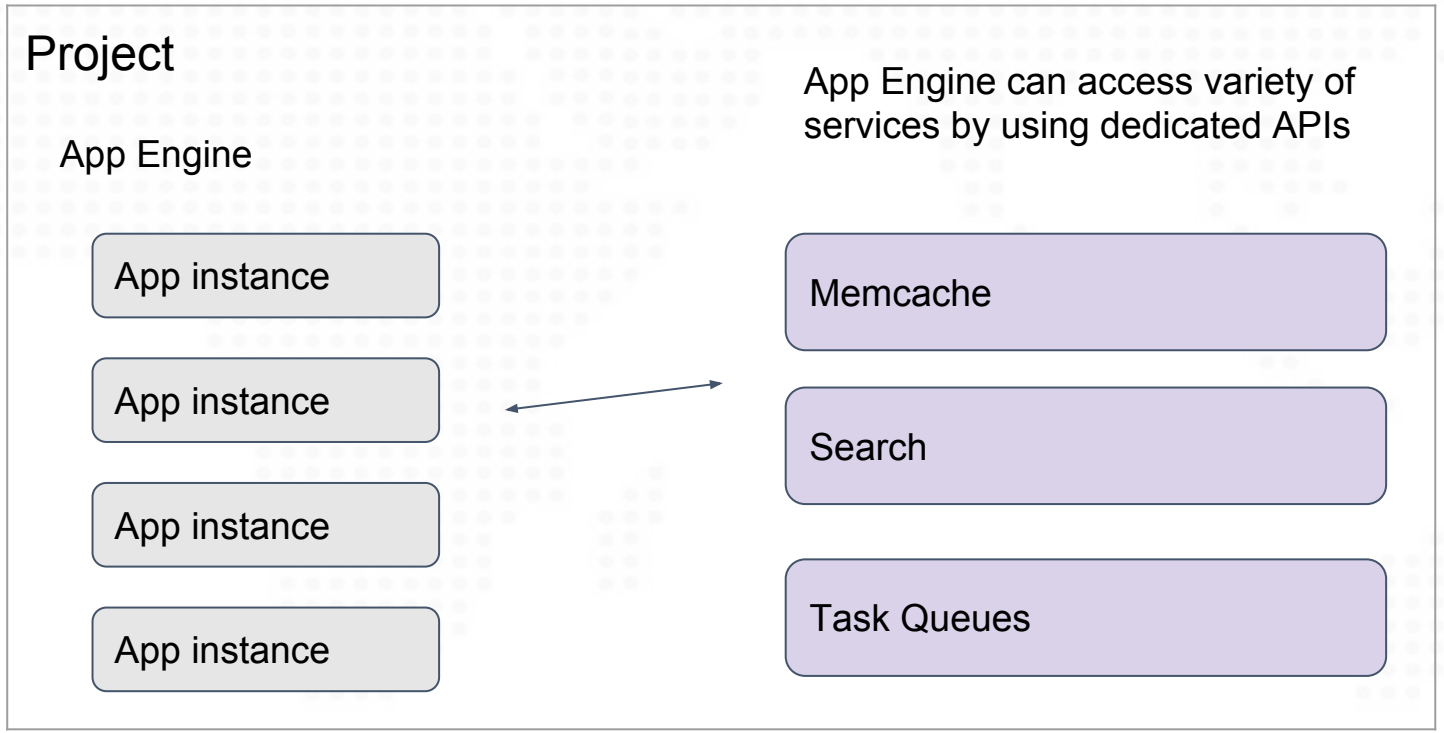
Sample App Engine workflow - Web App

1. Develop & test the web application locally



2. Use the SDK to deploy to App Engine

3. App Engine automatically scales & reliably serves your web application



NobleProg

App Engine Flexible Environment

- Build, deploy containerized apps with a click
- Standard runtimes - Python, Java, Go, Node.js
- Custom runtime support for any developing language that support HTTP requests
- Local development relies on Docker.

NobleProg

App Engine Flexible vs. Standard Env

	Standard Environment	Flexible Environment
Instance startup	Milliseconds	Minutes
SSH access	No	Yes (not default)
Scaling	Manual, basic, automatic	Manual, automatic
Write to local disk	No	Yes (Ephemeral)
Support for 3rd party binaries	No	Yes
Network access	Via App Engine services	Yes
Customizable stack	No	Yes

NobleProg

Google Cloud Endpoints

- Build your own API running on App Engine standard
- Expose your APIs using a RESTful interface
- Includes support for OAuth2.0 authorization
- Generate client libraries.

NobleProg

Google Cloud Datastore

- Database designed for application backends
- NoSQL store for billions of rows
- Schemaless access, no need to think about underlying data structure
- Local development tools
- Automatic scaling and fully managed
- Built-in redundancy

Lab - App Engine and Datastore (30min)

Build a web app in either Python/Java/Node.js etc. on your choice with simple CRUD operations

Persists data in cloud datastore

NobleProg

Google Cloud Storage

- High performance, internet-scale, immutable BLOB (binary large object) storage
- **NOT** a file system
- Simple administration and does not require capacity management.
- Data encryption in-flight and at rest
- All storage classes accessed through the same APIs.

NobleProg

Cloud Storage Classes

Storage classes	Features	Best for
Multi-Regional (Standard)	• 99.95% availability SLA* • Geo-redundant	Storing data that is frequently accessed ("hot" objects) around the world, such as serving website content, streaming videos, or gaming and mobile applications.
Regional (Standard)	• 99.9% availability SLA* • Lower cost per GB stored • Data stored in a narrow geographic region	Storing frequently accessed in the same region as your Google Cloud DataProc or Google Compute Engine instances that use it, such as for data analytics.
Nearline	• 99.0% availability SLA* • Very low cost per GB stored • Data retrieval costs • Higher per-operation costs • 30-day minimum storage duration	Data you do not expect to access frequently (i.e., no more than once per month). Ideal for back-up and serving long-tail multimedia content.
Coldline	• 99.0% availability SLA* • Lowest cost per GB stored • Data retrieval costs • Higher per-operation costs • 90-day minimum storage duration	Data you expect to access infrequently (i.e., no more than once per year). Typically this is for disaster recovery, or data that is archived and may or may not be needed at some future time.

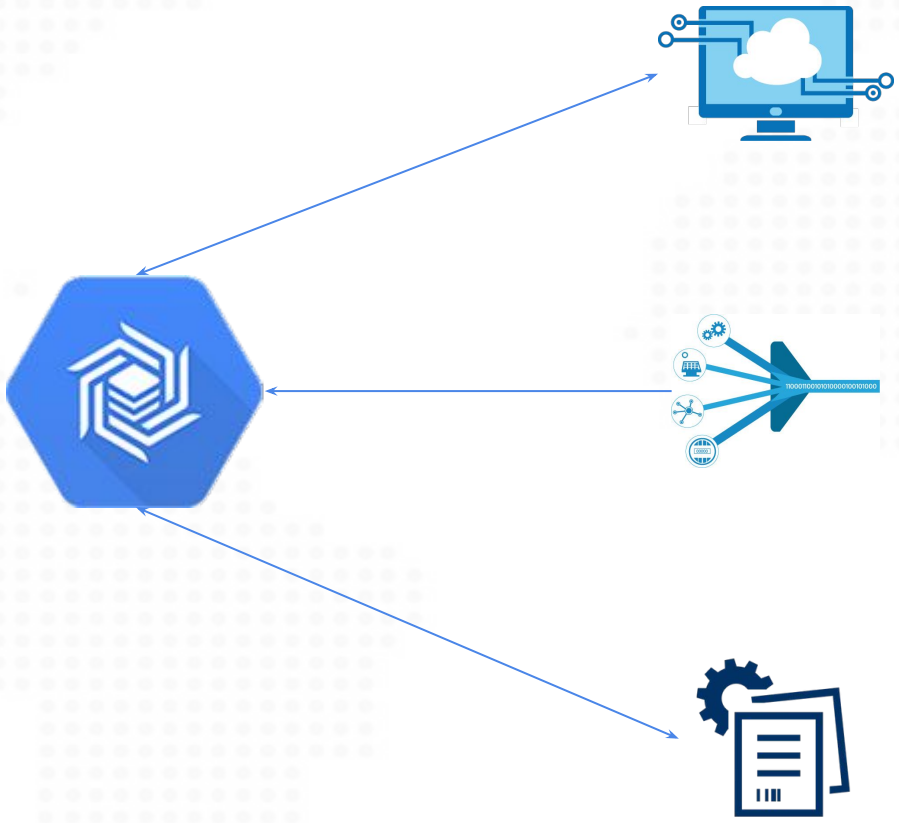
NobleProg

Google Cloud Bigtable

- Fully managed NoSQL, wide-column DB service for heavy workload application from TB to PB
- Integrated
  - Access using HBase API
  - Native compatibility with big data, hadoop eco
- Protected
  - Replicated
  - Data encryption in-flight
- Proven
  - Drives major apps such as Google Analytics and Gmail

NobleProg

Bigtable Access Pattern



Application API

Data can be read from and written to Cloud Bigtable through a data service layer like: Managed VMs, the HBase REST Server, a Java Server using HBase client. Typically this will be to serve data to applications, dashboards and data services.

Streaming

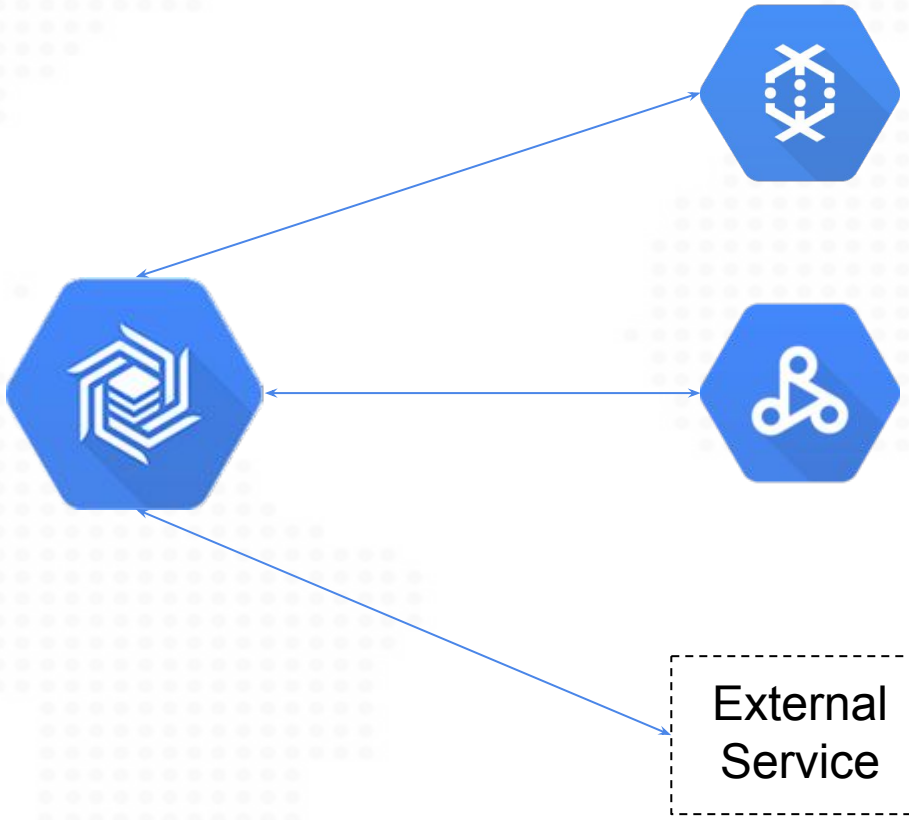
Data can be streamed in (written event by event) through a variety of popular stream processing frameworks like: Cloud Dataflow Streaming, Spark Streaming, Storm.

Batch Processing

Data can be read from and written to Cloud Bigtable through batch process like: Hadoop MapReduce, Dataflow, Spark. Often, summarized or newly calculated data is written back to Cloud Bigtable or to a downstream database.

NobleProg

Cloud Bigtable Integration



Google Cloud Dataflow

Use Cloud Dataflow connector for Bigtable for batch and streaming operations in pipelines.

Google Cloud Dataproc

Use Bigtable HBase client to integrate Hadoop jobs with Cloud Dataproc

On-premises, cloud-based Hadoop

Use Bigtable HBase client to integrate with Hadoop clusters

NobleProg

Google Cloud SQL

- Google-managed MySQL RDB in the cloud
- Pay-per-use model
- REST API for management
- Affordability and performance
- Google security
- Vertical scaling (read and write)
- Horizontal scaling (read)
- Seamless integration with App Engine and Compute Engine

NobleProg

Cloud SQL Integration

		
Cloud SQL can be used with App Engine using standard drivers like Connect/J for Java or MySQLdb for Python. App Engine applications are authorized to access CLoud SQL, and the instance can be configured to follow one application.	Compute Engine instances can be authorized to access Cloud SQL instances using an external IP address. Cloud SQL instances can be configured with a preferred zone - to stay close to the Compute Engine infrastructure.	Cloud SQL can be used with external applications and clients by authorizing IP addresses or networks using CIDR notation. Standard tools like MySQL Workbench can be used to administer databases. External read replicas can be configured.

NobleProg

Choose a Storage

NobleProg

Storage Specs

PRODUCT	DESCRIPTION	GOOD FOR	COMMON WORKLOADS
Persistent Disk	Fully-managed, price-performant block storage that is suitable for virtual machines and containers.	• Block storage for Google Compute Engine and Google Container Engine • Snapshots for data backup	• Disks for virtual machines • Sharing read-only data across multiple virtual machines • Rapid, durable backups of running virtual machines
Google Cloud Storage	A scalable, fully-managed, highly reliable, and cost-efficient object / blob store.	• Images, pictures, and videos • Objects and blobs • Unstructured data	• Storing and streaming multimedia • Storage for custom data analytics pipelines • Archive, backup, and disaster recovery
Google Cloud Bigtable	A scalable, fully-managed NoSQL wide-column database that is suitable for both real-time access and analytics workloads.	• Low-latency read/write access • High-throughput analytics • Native time series support	• IoT, finance, adtech • Personalization, recommendations • Monitoring • Geospatial datasets • Graphs
Google Cloud Datastore	A scalable, fully-managed NoSQL document database for your web and mobile applications.	• Semi-structured application data • Hierarchical data • Durable key-value data	• User profiles • Product catalogs • Game state

NobleProg

Storage Specs cont'd

PRODUCT	DESCRIPTION	GOOD FOR	COMMON WORKLOADS
Google Cloud SQL	A fully-managed MySQL and PostgreSQL database service that is built on the strength and reliability of Google's infrastructure.	• Web frameworks • Structured data • OLTP workloads	• Websites, blogs, and content management systems (CMS) • Business Intelligence (BI) applications • ERP, CRM, and eCommerce applications • Geospatial applications
Google Cloud Spanner	Mission-critical, relational database service with transactional consistency, global scale and high availability.	• Mission-critical applications • High transactions • Scale + Consistency requirements	• Adtech • Financial services • Global supply chain • Retail
Google BigQuery	A scalable, fully-managed Enterprise Data Warehouse (EDW) with SQL and fast response times.	• OLAP workloads up to petabyte-scale • Big Data exploration and processing • Reporting via Business Intelligence (BI) tools	• Analytical reporting on large data • Data Science and advanced analyses • Big Data processing using SQL
Google Drive	A collaborative space for storing, sharing, and editing files, including Google Docs.	• End-user interaction with docs and files • Collaborative creation and editing • Syncing files between cloud and local devices	• Access files from anywhere through web, apps, and sync clients • Create and work on documents with coworkers • Backup photos and media

NobleProg

Lab - Cloud Storage

Extend your previous webapp to persist and load blob files

E.g.
Images
Text files
Etc.

NobleProg

Google Container Engine

- Virtualization at the operating system layer
- Separate operating system from application code and dependencies
- Isolates individual processes
- Popular implementations include Docker and rkt

NobleProg

Why Containers

- Support consistency across development, testing, and production environment
- Loose coupling between operating system and application layers
- Much simpler to migrate workloads between on-premises and cloud environments
- Support agile development and operations

Kubernetes ('k8s')

- Open source container cluster orchestration system
  - Automates deployment, scaling and operations for container cluster
- Based on Google's exp over 10+ years
- Built for multi-cloud world
  - Public, private and hybrid

NobleProg

K8s features

- Workload portability
  - Run in many environments, across cloud providers
  - Implementation is open and modular
- Rolling updates
  - Upgrade application with zero downtime
- Autoscaling
  - Automatically adapt to changes in workload
- Persistent storage
  - Abstracts details of how storage is provided from how it is consumed
- Load balancing
  - External IP address routes traffic to correct port

NobleProg

Google Container Engine

- Fully managed cluster management and orchestration system for running containers
  - Based on K8s
  - Uses Compute Engine instances and resources
- Complimentary services:
  - Google Cloud Container Builder: Create Docker container images from app code in Google Cloud Storage
  - Google Container Registry: Secure, private Docker image storage

NobleProg

Google Container Engine - cont'd

- Uses a declarative syntax to manage applications
  - Declare desired application configuration, Container Engine implements, manages
- Decouples operational, development concerns
- Manages and maintains
  - Logging, health management, monitoring
- Easily update Kubernetes versions as they are released

Deploying apps: Container Engine vs. App Engine

	Container Engine	**App Engine Standard**	**App Engine Flexible**
Language support	Any	Java, Python, Go & PHP	Any
Service model	Hybrid	PaaS	PaaS
Primary use case	Container-based workloads	Web and mobile applications	Web and mobile applications, container-based workloads

NobleProg

Lab - Container Engine

Deploy an application with google container engine

https://codelabs.developers.google.com/codelabs/cloud-hello-kubernetes

NobleProg

Google Compute Engine

- Run large-scale workloads on virtual machines hosted on Google's infrastructure
- Robust networking features
  - Default, custom networks
  - Firewall rules
  - Regional HTTP(s) load balancing
  - Network load balancing
  - Subnetworks

NobleProg

Google Compute Engine - cont'd

- High CPU, high RAM, standard and shared-core machine types
- Persistent disks
  - Standard HDD, SSD, local SSD
  - Snapshots
- Resize disks, migrate instances with no downtime
- Instance metadata and startup scripts

NobleProg

Google Compute Engine - cont'd

- Advanced APIs for auto-scaling and instance group management
- Innovative pricing
  - **Per-minute** billing, sustained use discounts
  - Preemptible instances
  - High throughput to storage at no extra cost
  - Custom machine types - Only pay for the hardware you need

NobleProg

Lab - Compute Engine

Deploy a webapp in Compute Engine

NobleProg

Google Cloud Networking Overview

- Cloud Interconnect
- Cloud VPN
- Cloud DNS
- Load Balancing
- Cloud CDN

NobleProg

Google Cloud Interconnect

Carrier Interconnect	Direct Peering	CDN Interconnect
Enterprise-grade connections provided by carrier service providers	Connect your business directly to Google	Allows select CDN providers to establish direct interconnect links with Google's edge network at various locations

NobleProg

Google Cloud VPN

- Securely connect your network to Google Cloud Platform using IPSec VPN connection
  - Encrypts traffic over the Internet
- Google Cloud Router supports dynamic routing between Google Cloud Platform and your network

NobleProg

Google Cloud DNS

- Highly available and scalable DNS
  - Translates domain names into IP addresses
- Create managed zones, then add, edit, delete DNS records
  - Programmatically manage zones and records using RESTful API or command-line interface

NobleProg

Google Cloud Load Balancing

- HTTP(s) load balancing
  - Balance HTTP-based traffic across multiple Compute Engine regions
  - Global, external IP address routes traffic
  - Scalable, requires no pre-warming and provides resilience, fault tolerance
- TCP/SSL and UDP (network) load balancing
  - Spread TCP/SSL and UDP traffic over pool of instances within a Compute Engine region
  - Ensures only healthy instances handle traffic
  - Scalable, requires no pre-warming

NobleProg

Google Cloud CDN (Content Delivery Network)

- Use Google's globally distributed edge caches to cache http(s) load balanced content far closer to your users than your instances
  - Faster delivery of content to users while reducing costs
- Cloud CDN uses caches at network locations to store responses generated by instances

NobleProg

Operations and Tools

NobleProg

Google Stackdriver (Beta)

- Integrated monitoring, logging, diagnostics
- Works across GCP and AWS
- Open source agents, integration
- Powerful data, analytics tools
- Collaborations with PagerDuty, BMC, Splunk, others

NobleProg

Google Stackdriver (Beta) - cont'd

Monitoring

- Platform, system and application metrics
- Uptime / health checks
- Dashboards and alerts

Trace

- Latency reporting and sampling
- Per-URL latency and statistics

Logging

- Platform, systems and application logs
- Log search, view and filtering
- Log-based metrics

Error Reporting

- Error notifications
- Error dashboard

Debugger

- Debug applications

NobleProg

Google Cloud Deployment Manager

- Infrastructure management service
- Create a .yaml template describing your environment and use deployment manager to create resources
- Provides repeatable deployments

NobleProg

Google Cloud Source Repositories

- Fully-featured Git repositories hosted on Google Cloud Platform
- Supports collaborative development of cloud applications
- Includes:
  - Source code editor
  - Integration with Stackdriver debugger

NobleProg

Google Cloud Functions

- Create single purpose functions that responses to events without a server or runtime
  - Events like: New instance created, file added to cloud storage etc.
- Written in JavaScript, execute in managed Node.js environment on Google Cloud Platform

NobleProg

Big Data & Machine Learning

NobleProg

BigQuery

- Fully-managed analytics data warehouse
  - Provides near real-time interactive analysis of massive datasets (hundreds of TBs)
- Query using SQL-like syntax
- Zero administration for performance and scale - NoOps
- Runs on Google's fully managed, secure, high-performance infrastructure
  - Compute and storage are separated with a petabit, high-speed network in between
  - Only pay for storage, processing used
- Automatic discount for long term data storage

NobleProg

Google Cloud Pub/Sub

- Scalable, reliable messaging system for Google Cloud Platform and beyond
- Supports many-to-many asynchronous messaging
- Includes support for offline consumers
- Based on proven Google technology
- Integrates with Cloud Dataflow for data processing pipelines

NobleProg

Google Cloud Pub/Sub - cont'd

- Uses push/pull subscriptions to topics
- Use cases:
  - Building block for data ingestion in Dataflow, Internet of Things (IoT), Marketing Analysis
  - Foundation for Dataflow streaming
  - Push notifications for cloud-based applications
  - Connect applications across Google Cloud Platform (push/pull between Compute Engine and App Engine)

Google Cloud Dataflow

- Managed service for executing scalable and reliable data pipelines
- Write code once and get batch and streaming
  - Transforming-based programming model
- Clusters are sized for you
- Processes data using Compute Engine instances

NobleProg

Google Cloud Dataflow - cont'd

- Integrates with GCP services like Cloud Storage, Cloud pub/sub, BigQuery, BigTable
- Open source Java and Python SDKs
- Use cases
  - ETL (extract/transform/load) pipeline to move, filter, enrich, shape data
  - Data analysis - batch computing or continuous computation using streaming
  - Orchestration - create pipelines that coordinate services, including external services

NobleProg

Google Cloud Dataproc

- Fast, easy, managed way to run Hadoop and Spark/Hive/Pig on Google Cloud Platform
- Benefit from cloud integration
  - Cloud storage
  - Stackdriver
- Customize and configure clusters using initialization actions

Google Cloud Dataproc - cont'd

- Create cluster in 90 sec or less
- Dataproc clusters billed minute-by-minute
  - Save money using preemptible instances for batch processing
- Scale cluster up and down even when jobs are running
- Developer tools
  - RESTful API
  - Integration with Google Cloud SDK

NobleProg

Google Cloud Dataproc - cont'd

- Use cases
  - Easily migrate on-premise Hadoop jobs to the cloud
  - Quickly analyze data (i.e. logs) stored in Cloud Storage - create cluster in less than 90s then delete it immediately
  - Use Spark/SparkSQL to quickly perform data mining and analysis
  - Use Spark Machine Learning Libraries (MLlib) to run, e.g. classification, algorithmes

Google Cloud Datalab

- Interactive tool for large-scale data exploration, transformation, analysis, visualization
  - Analyze data in BigQuery, Compute Engine and Cloud Storage using Python, SQL and JavaScript
  - Easily deploy transformation, analysis models to BigQuery
- Integrated, open source
  - Runs on Google App Engine
  - Build on Jupyter (formerly IPython)
  - Use Google Charts or matplotlib for easily visualizations
- Code, documentation, results, visualization in intuitive notebook format

NobleProg

Machine Learning Services

- TensorFlow
  - Open source tool to build and run mainly neural network models
  - Wide platform support: CPU or GPU; mobile, server or cloud
  - Developed by researchers and engineers of Google Brain
- Cloud ML - fully managed ML service
  - Faster training, better accuracy versus competing systems
  - Familiar notebook-based developer experience
  - Optimized for Google infrastructure, integrates with BigQuery and Cloud Storage
- ML APIs
  - Vision: identify objects, landmarks, text, explicit content detection
  - Translation: includes language detection
  - Speech: stream results in real-time, detects 80 languages

NobleProg

Lab - BigQuery

In this lab, you load a CSV file into a BigQuery table. After loading the data, you query it using the BigQuery web user interface, the CLI, and the BigQuery shell.

NobleProg