Google Cloud Platform

Professional Cloud Architect Crash Course

By: Ryan Dymek

Who's Ryan Dymek?

- Over 20+ years Industry experience
- Pre-Cloud career path consisted of Infrastructure and Security focus
- Wrote my first program in BASIC in 1989, Learned C & C++, Pascal and Turbo Pascal in early '90s
- Started working on AWS in 2009, and GCP in 2016
- Have consulted and advised 20+ companies on the fortune 500 list

What I won't cover in this course

- Deep *individual product or service* details and understanding
- Automation code, scripting, etc. – Its not relevant for this exam!
- Easy stuff to find in documentation like what call or command may be used for something specific

If you are looking for a focus on GCP foundations, product and service details, consider taking my course titled:

Google Cloud Platform Associate Cloud Engineer Crash Course

What I WILL cover in this course

- Design principles – This is a design exam afterall!
- Exam testing strategies
- Complex scenarios & considerations
- Build a scenario, evaluate it, and improve
- Making your certification mean something - Don’t memorize, strive to *understand!*

Resources

Additional resources for this class can be found here. This link is revised regularly to provide additional, new, and revised resources

https://leandev.fyi/gcp-pca2020

Agenda – Day 1

- Segment 0: About the exam
- Segment 1: Planning and Designing a Cloud Architecture
- Segment 2: Solutions Infrastructure
- Segment 3: Analyzing and Optimizing Processes
- Wrap-Up

Agenda – Day 2

- Segment 4: Security and Compliance
- Segment 5: Implementing GCP
- Segment 6: Operational Reliability
- Segment 7: Prepping for your Exam
- Wrap-Up

Segment 0: About the Exam

Exam Details

Duration: 2 Hours

Fee: $200 USD

Exam Format: Multiple choice & Multiple Select

Exam Delivery Options:

- Online-Proctored
- Onsite-Proctored

Prerequisites: None

Recommended experience:

- 3+ Years Industry Experience
- 1+ Years designing and managing solutions using GCP

Exam Tips and Preparation

- Pace yourself – timing is important
- Run through the easiest, shortest questions first and then go back to the more complex questions when you have time (flag questions for follow-up)
- Always focus on the business requirements if the question provides them (pre-study the case studies)
- No partial credit for multiple answer questions!
- You can’t take notes during the exam, you must work in your head

Exam Tips and Preparation - Continued

- Some questions may not be scored, so don't let yourself get hung up on any single question, it *may* not even count – come back to it if you have time!
- Passing scores or percentages are not formerly defined by Google, but it is estimated by many to be around the 80% mark
- During your preparation, study your weakest topics early and often

Fundamental GCP Technologies and Methodologies to…

understand

Networking & Infrastructure

Question: Why are we even building a network infrastructure?

- VPC's, subnets, interfaces, firewalls and tags
- Cloud Interconnect, Dedicated Interconnect, Partner Interconnect, Direct Peering, Carrier Peering, CDN Interconnect, and Cloud VPN
- HTTP(s) and Network Load Balancing
- VPC Peering
- GCP *Projects* and their relationship to the infrastructure / VPC's
- Cloud CDN, Load Balancers, and Google Cloud Endpoints purpose in your infrastructure
- API Access in your Infrastructure
- Regions and Zones
- Edge Locations
- Cloud NAT
- BGP Basics (ASN's, eBGP vs iBGP, Neighbors, Route summary, etc.)

Datastores / Data Analytics

Questions: Database vs Storage? What's the difference? Which tool is right for the job?

- VM Persistent Disk vs Local SSD
- Cloud Storage and all associated storage classes
- Cloud SQL
- Cloud Spanner
- BigTable
- BigQuery
- Firestore
- Dataproc
- Dataflow
- Memorystore
- Pub/Sub
- Filestore

Compute

Question: Do we always need to run VM's?

- Compute Engine:
  - Virtual Machines
  - Images
  - Types
  - Sole-tenant
  - Interfaces
  - Firewalls
  - Tags
- App Engine Standard & App Engine Flexible
- Cloud Run
- Cloud Functions
- Google Kubernetes Engine (GKE)

Security

Questions: What security controls apply? When do I use a firewall and when do I use IAM?

- Identity and Access Management (IAM)
  - Primitive roles vs Custom Roles
  - Principle of least access
  - Service Accounts
  - Projects & Resource Hierarchy
- Firewalls
- Routing
- IP Addresses and Firewalls
- API Access Control
- IAM roles for billing

Costs

- Always consider data transfer!
- Operational overhead?
- Sustained Use & Committed Use Discounts
- Second, Minutes and Hours – Oh my!
- Per API Call? Too many calls and sloppy or inefficient code design
- Ryan's Exam Tip: The 5 cost elements
  - Storage, Network, Operations, Features, Service charge

Infrastructure or API Considerations

- Cost
- Operational Overhead
- Flexibility
- Fault Tolerance / High Availability
- Security Implications
- Scaling
- Tightly Coupled vs Decoupled
- Asynchronous vs Synchronous Processing

Exam Case Studies

- This is a scenario-based exam
- Review the GCP Exam Guide and learn the case studies
- You don’t know which case studies you *may* encounter on the exam
- Be familiar with the case studies to enhance your ability to answer questions quickly
- Don’t ignore these – spend quality time reviewing them

The nuances and minutia

- Some facts will be necessary to know but focus on design principles above all else

- Example 1: Consider an IOPS need when building a VM. Will *Persistent Disk Standard* meet the IOPS needs? Need *Persistent SSD* instead? Or could you solve the need with multiple disk & RAID?

- Example 2: Could you save money by moving to another storage class or disk type? Maybe even different storage altogether such as from persistent disk to cloud storage. Will it still meet the requirements?

Segment 1: Planning and Designing a Cloud Architecture

Its all about *design*

GCP Professional Cloud Architect certification:
Design, Plan & Optimize

GCP Associate Cloud Engineer certification:
Implementation, Deployment, and Management

Operations, Monitoring & Management

How do operations play into our designs?

Traditional IT designs vs operationally driven design

Cloud vs On-Prem: Similar but different

Site Reliability Engineering (SRE)

- Site Reliability Engineering book*
- *YOU* are the SRE!
- Error budgets and risks
  - SLA's, SLO's, & SLI's
  - Error *rates* vs uptime
  - GCP Operations (formerly Stackdriver) makes incorporating SLI's into your design simple

* Book is available online for free (See course resources) or paid print version

Recommended Enterprise Guidelines

- Understand the relationship between your organizational hierarchy and the resource hierarchy and the implications
- GCP *Project* and *Resource* structure
- Implement the principle of least access
- Centralize network control – Decouple network administration from project administration via use of Shared VPCs
- Realize network controls are a small fraction of your overall controls. Policies, roles, service accounts, and API endpoints account for most of your access control

Changing the “architecting” paradigm

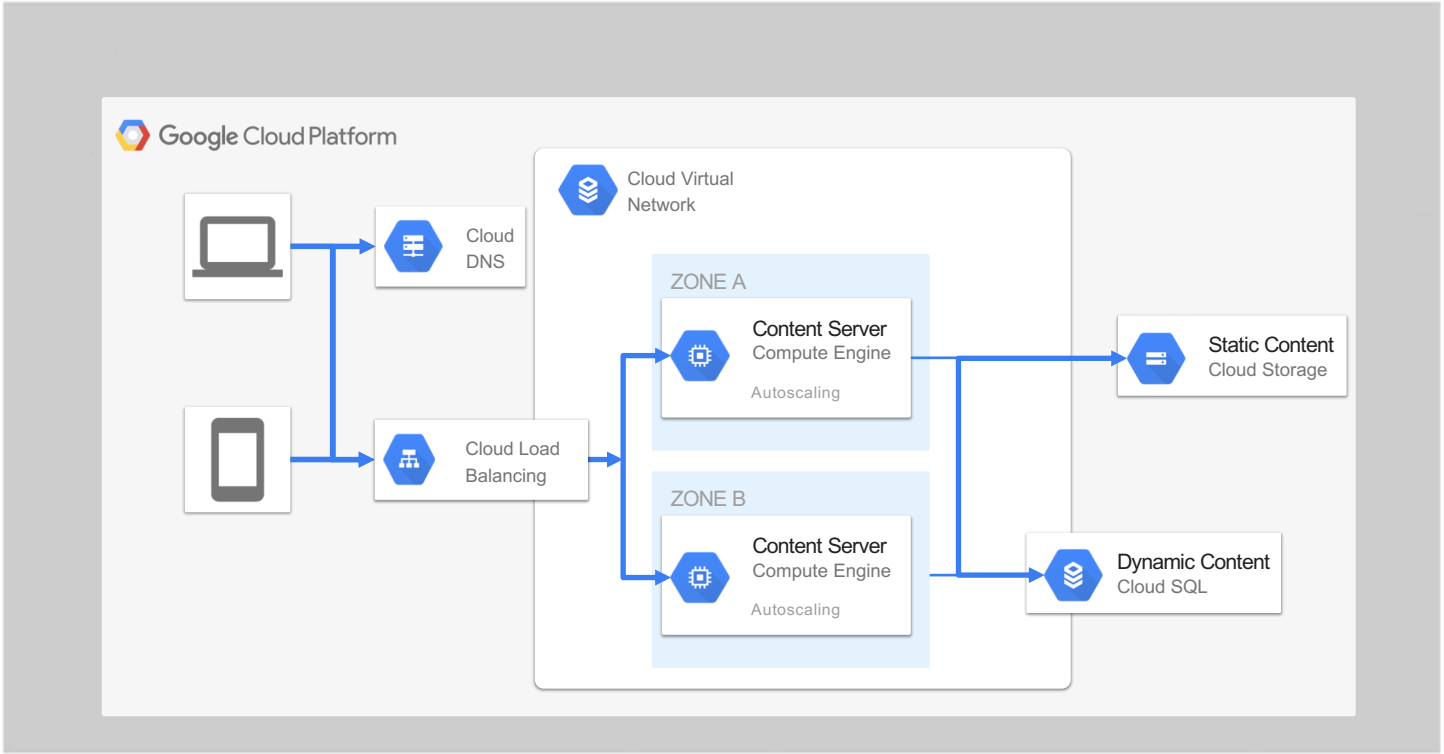
Questions to ask yourself:

- Who’s responsible for managing this?
- What layers of security apply?
- Architecture: Infrastructure, Software, Automation, Security, Data Analytics. In the end its all code!
- Costs: Per call, fixed, scaled, etc. Pay for what you use, not what you design or even create; decouple design, creation, and implementation

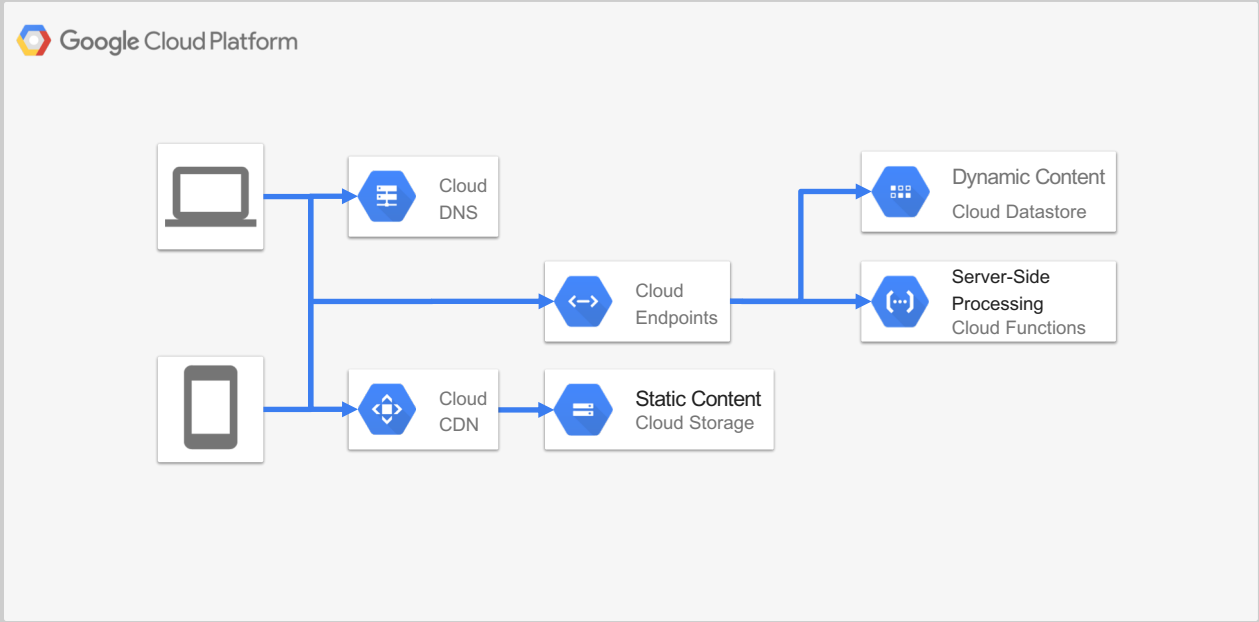
API First and Microservices

Consider the two following designs:

Solution #1: Infrastructure First



Solution #2: API First



CI/CD Pipelines

- Continuous Integration, Continuous Delivery, Continuous Deployment
- Development → Testing → Staging → Production
- Quality output, automated testing, and a “push” to production
- Minimize or eliminate human interaction with production systems
- Version control everything! Yes, even infrastructure!

Migration & Hybrid Environments

- There is no “best” design
- “Best” is whatever meets the requirements in a way that aligns with your business
- Migrating may require “lift and shift”, or you may consider a full redesign and rewrite, or somewhere in the middle
- Consider all requirements before coming to conclusions
- Understand dependencies and limitations of the existing architecture
- Always remember what the “companies” focus is, and objective of the project

New Designs

- Opportunity to go microservices or serverless, but should you?
- Evaluate and understand the *risk of building your own infrastructure* as opposed to serverless
- Process flow – who manages what? Do you have to adhere to legacy methods?
- Cost models

Designing an Infrastructure

What to consider?

- Do actual IP addresses matter?
- Consider Projects, IAM, and VPC relationships
- Why create multiple VPC networks? For what purpose?
- Is “network” the only way to move data?
- Is network access the only way to manage resources?
- What about patching and other maintenance tasks, what network access is required, if any?
- What about API access, peering controls, and interconnect options?

Whiteboard

Let's design together!

Q & A

Segment 2:
Solution Infrastructure

Networking: Subnets

Let's talk subnets…

- What’s their purpose?
- Why have them?
- On-prem vs GCP:
  Small nuances make profound differences
- Scope: Regional

Networking: Routing

- Auto-Mode VPC networks vs Custom Mode
- Custom Routes: Static Routes, or routes propagated from your Cloud Routers
- Understand CIDR notation, RFC1918 IPv4 specifications, and IPv6 basic concepts
- Most specific route wins (like traditional routing)
- This area tends to be many individual's weakness on the exam

Networking: VPC Specifics

- Global scope, no specific region
- A Shared VPC can be used to keep a VPC Network in a common host project
- VPC Network Peering will work across or within Projects
- Connect a VPC to your datacenter via Cloud VPN or Cloud Interconnect options
- No Layer 2 (Broadcast or Multicast), only IPv4 Unicast (Layer 3) communications are supported
- IPv6 is supported at the global load balancer, *but not within the VPC*

Networking: Hybrid

Spend some time understanding the purpose of, and the differences between the following connectivity options:

- Cloud VPN, IKE (UDP 500) & IPSEC (ESP 50, NAT-T UDP 4500) Port Numbers and throughput (3 Gbps per tunnel)
- Cloud Interconnect
- Dedicated Interconnect (10Gbps or 100Gbps) options
  - Can be bonded – Up to 2x 100Gbps and 8x 10Gbps
- Partner Interconnect
- Direct Peering
- Carrier Peering
- CDN Interconnect

Designing Storage Systems

Consider the following when choosing storage:

- Access Patterns (How will this be used?)
- Size of the data (Bytes or Petabytes – Very different!)
- Rate of the data (Streaming, Batch, Archive)
- Throughput and Scaling (Rate + Size, and fluctuations)
- Structured vs. Unstructured (CSV or a Tweet)
- Resiliency Requirements (Cache vs Persistent Datastore)

Types of storage, what to choose?

- Could more than one storage solution be used?
  - E.g. Cloud SQL for transactional writes and Memorystore with Redis for sub-millisecond reads
- Analytics (OLAP) could (should?) be separate from transactional workloads (OLTP)
- Columnar vs Row Storage
- Object Storage vs File vs Block storage
- Databases as storage?

Columnar vs Record (Row) Storage

Query for average age

Typical Record oriented storage

id	f_name	l_name	age	country
1	Patrick	Debois	55	Belgium
2	Elon	Musk	48	USA
3	William	Inmon	75	USA
4	James	Dixon	52	USA
5	Thomas	Kurian	60	USA

Columnar Storage

1|2|3|4|5

Patrick|Elon|William|James|Thomas

Debois|Musk|Inmon|Dixon|Kurian

55|48|75|52|60

Belgium|USA|USA|USA|USA

So what solutions use Columnar Storage?

BigQuery, Parquet, ORC to name a few…

Compute

- Are compute options limited to just VMs?
- A few compute options…
  - Compute Engine (Virtual Machines)
  - Cloud Functions
  - Kubernetes
  - App Engine Standard / App Engine Flexible
  - Cloud Run

Infrastructure Provisioning

Some options for provisioning infrastructure…

System / OS Automation	Infrastructure Templates
Chef Puppet Ansible Saltstack	Terraform Deployment Manager

Kubernetes

What is Kubernetes?

Why & when would you use Kubernetes?

When would you NOT use Kubernetes?

Q & A

Segment 3: Analyzing and Optimizing Processes

Discussion

Many technical problems can be solved…

…through process

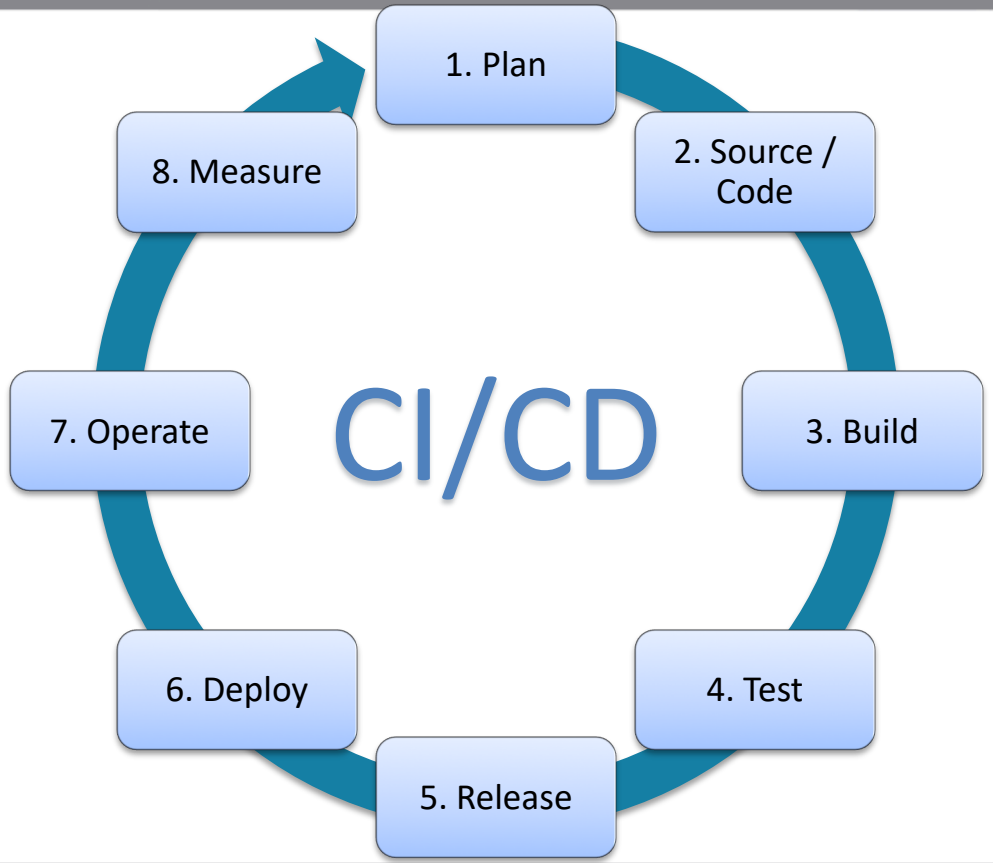
Software Development Lifecycle

&

Continuous Integration / Continuous Deployment

Projects, VPC's, or Shared VPCs?

The CI/CD Pipeline



Value Streams, Visibility, Change Management

What are Value Streams?

Improve work visibility

Evaluate & review my provided or changed resources

Testing, Validating & Troubleshooting

- Dev / Test / Prod parity
- Templates
- Private Catalogs
- Testing Tools (i.e. Selenium, WireMock, Locust, Etc.)
- Understand the need to distribute your test systems
- Isolate testing; Same subnet or VPC, different subnet or VPC, over VPN, Inside, Outside, etc.
- Perform tests in parallel and independently when possible; A failed test should not impact another test

Infrastructure as Code & Private Catalog

- Understand how Deployment Manager works
- How should templates be managed at scale in an organization?
- Sharing templates
- “Modular” templates
- Private catalog with sample or custom templates and share across projects or within a project
- Governance & Compliance

Business Continuity

Business Continuity should address the following:

- Replication vs Backups or Snapshots
- Game Day Exercises – Don’t just test in test, test in production!
- Recovery Time Objective (RTO) vs Recovery Point Objective (RPO)
- Cost Factors
- SLO and SLA factors on Business Continuity
- High Availability (HA) vs Fault Tolerance (FT) vs Disaster Recovery (DR)
- Do the resources for DR already need to be running? It depends.
- Chaos Engineering

Cost Exercise: Compare these datastores

Which is *most* cost effective?

Google Cloud Storage

vs

Firestore (Formerly Datastore)

vs

CloudSQL

vs

Cloud Pub/Sub

Review Cost Structures

Ryan's 5 factors of cost:

- Storage
- Network
- Operations
- Features
- Service charge

Cost Scenarios – Not including free

- Scenario 1:
  - Near Real Time Data
  - 600 Writes Per Second (~1,555,200,000 IO/Mo)
  - 300 Bytes Average Write size (~466 GB/Mo)
  - OLAP and OLTP reads of data required
- Scenario 2:
  - Batch Style Data Ingestion
  - 10 Writes Per Second (~25,920,000 IO/Mo)
  - 10 **MiB** Average Write Size (259.2 TiB/Mo)
  - OLAP and OLTP reads of data required

Disclaimer

(Do your own analysis, this is not a substitute for due diligence!)

Scenario 1

	Cloud Storage	Firestore	CloudSQL	Cloud Pub/Sub
OLAP	Not Directly	✓	Limited	Not Directly
OLTP*	Not Directly	✓	✓	Not Directly
Storage Cost	@$0.026/GB = $12.12/Mo	@$0.108/GB = $50.33/Mo	@$0.34/GB (HA) = $158.44/Mo**	@$40/TiB = $37.32 / Mo
Transfer Charges	No Inbound*	No Inbound*	No Inbound*	No Inbound*
Operational Charges	@$0.05/10,000 = $7,776/Mo	@0.108/100k = $1680/Mo	db-n1-standard-4 = $197.25/Mo	None*
Features	None*	None*	None*	None*
Service Charges	None*	None*	None*	None*
Approximate Total	$7788.12/Mo	$1730.33/Mo	$355.69***	$37.32 + DataFlow + Cloud Storage

Scenario 2

	Cloud Storage	Firestore	CloudSQL	Cloud Pub/Sub
OLAP	Not Directly	✓		Not Directly
OLTP*	Not Directly	✓		Not Directly
Storage Cost	@$0.026/GB = $6739.20/Mo	@$0.108/GB = $27,993.60/Mo	Not Possible	@$40/TiB = $10,368/Mo
Transfer Charges	No Inbound*	No Inbound*		No Inbound*
Operational Charges	@$0.05/10,000 = $129.60/Mo	@0.108/100k = $28/Mo		None*
Features	None*	None*		None*
Service Charges	None*	None*		None*
Approximate Total	$6869/Mo	$28,000/Mo	Scale exceeds solution	$10,368/Mo + DataFlow + Cloud Storage

Q & A

Thank you!

- This concludes the end of day 1
- Its been an honor and a pleasure to share my knowledge with you
- See you for the next session tomorrow

Google Cloud Platform

Professional Cloud Architect Crash Course

By: Ryan Dymek

Welcome to Day 2!

Agenda – Day 2

- Segment 4: Security and Compliance
- Segment 5: Implementing GCP
- Segment 6: Operational Reliability
- Segment 7: Prepping for your Exam
- Wrap-Up

Segment 4:
Security & Compliance

Ever hear of the AAA? (hint, not the Automotive Club!)

Authentication

Prove that you are who you say you are

Google Account!

Authorization

What do you have access to?

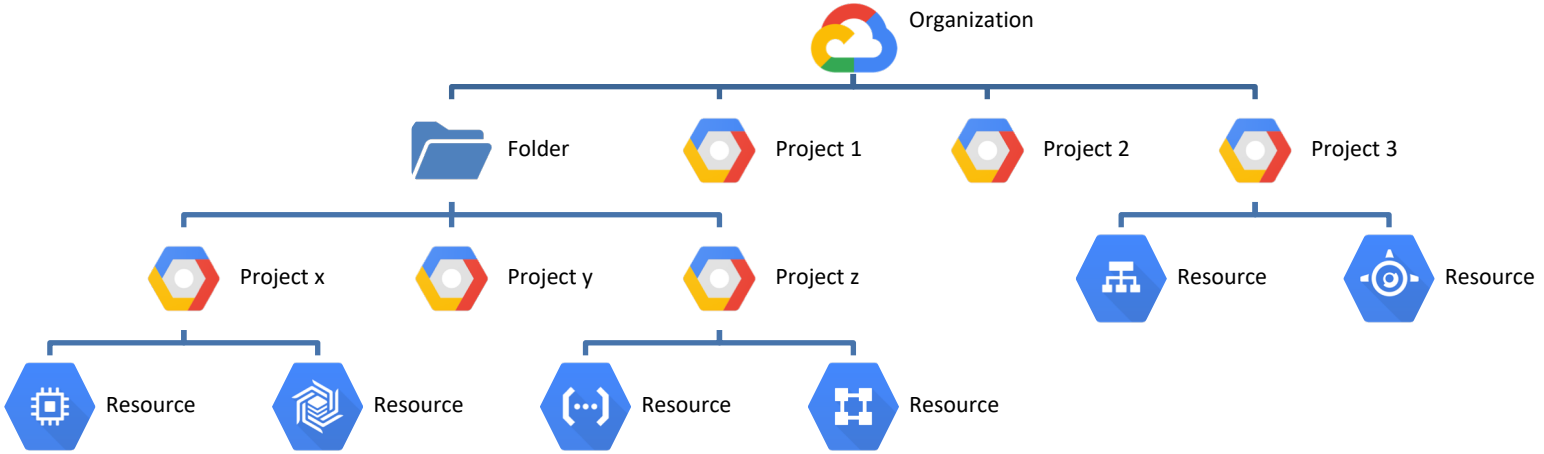
Roles – Organization, Folders, and Projects

Accounting

Where's the proof?

Audit Logs, Application Logs, VPC Flow Logs

Resource Hierarchy



IAM: Roles (Authorization)

- Primitive Roles
  - Legacy
  - Broad levels of access
  - Existed before the current IAM implementation
  - Owner, Editor, and Viewer roles
- Predefined Roles
  - Granular access for a specific service provided and managed by Google
- Custom Roles
  - Ability to fully use the *principle of least access**
  - Preferred but requires more operational overhead

Permissions vs Roles vs Policies

What is a permission?

May *normally,* but not always, correspond with a REST method. Ultimately a permission is the direct access itself to a specific “call”

What is a Role?

A collection of permissions

What is a Policy?

A policy “binds” members to a Role. May contain certain conditions that must be met.

IAM: Policies

- Bindings
  - Members - Who?
  - Role – What permissions?
  - Condition – Under what circumstances?
- AuditConfig – Configure audit logging for the policy (Accounting)
- Metadata
  - Etag – Concurrency control for policy consistency
  - Version – Schema version definition

IAM: Service Account

- Used by an application or VM instance, not a person
- Identified by its (generated) email address, unique to the account
- Do not contain passwords and cannot login via browser or cookies
- Associated with RSA key-pairs used for authentication to Google
- Cloud IAM *Permissions* can be granted to users for impersonation
- *Not* members of your G Suite Domain

Data Protection

Considerations for Data Protection

- Data classification
- Encryption & Key Management – Who controls the keys?
- Encryption in *Transit* vs Encryption at *Rest*
- Access controls – IAM, Firewalls, etc
- Auditing – Logging, Monitoring, etc.
- Compliance controls – PII, PCI, HIPAA, FIPS, etc.

Security Tooling

- Cloud Armor
  - DDoS Protection
  - Web Application Firewall (WAF)
  - Additional Geo-Based access controls
  - Custom Layer 7 Rules / Filtering policies
- Packet Mirroring and integration with partners of your choice
- Cloud Security Command Center (Cloud SCC)

Segment 5: Implementing GCP

API’s

What are API’s after all, and why are they important?

Marketing CRM vs Sales CRM – No API

Non-API example

(Whiteboard)

Marketing CRM vs Sales CRM – API

API example

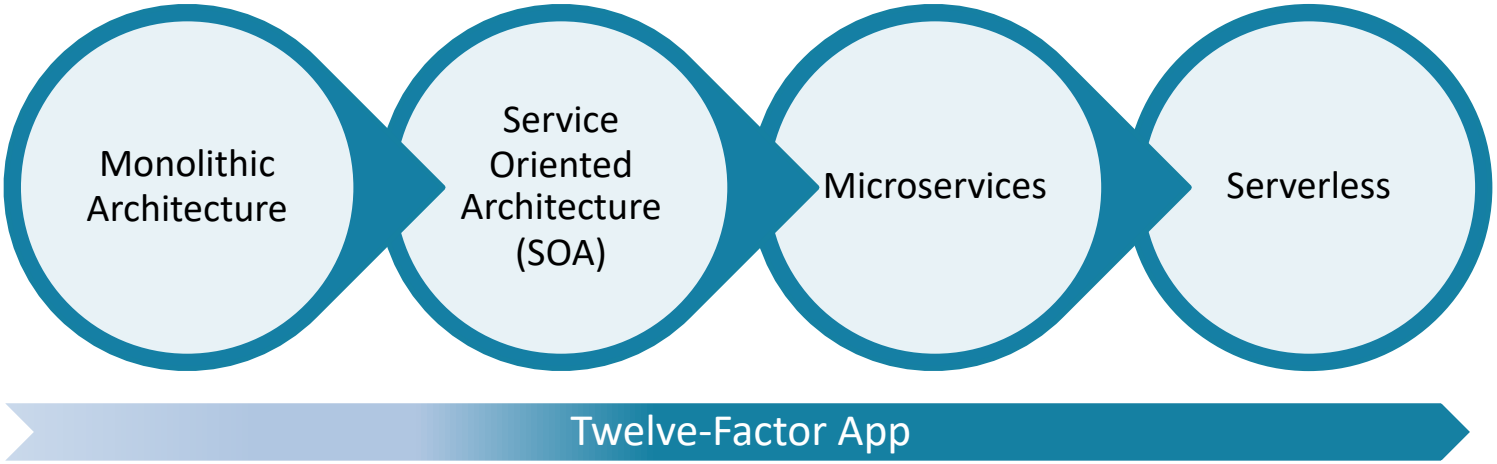
(Whiteboard)

API Best Practices

- The Twelve-Factor App
- Language Agnostic
- Do not pass database queries (SQL Statements, etc)
- Scrub for SQL Injection
- 1 Method per call
- OpenAPI v2 (formerly Swagger spec) – Industry standard for defining REST APIs

Architecture Exercise

Microservices



The 5 R's of Cloud Migration

Rehost	"Lift and Shift"
Refactor	Perform cloud infrastructure changes while maintaining similar/familiar code & frameworks.
Revise	Modification of existing code + rehosting or refactoring
Rebuild	Discard the code for an existing application, re-architect, and rebuild the PaaS
Replace	Utilization of new commercial software, and/or serverless technologies; 100% cloud native.

Interaction with GCP

- The exam – where to focus your study time
- GCP Tooling:
  - Cloud SDK
    - gcloud
    - gsutil
    - bq
    - Kubectl
  - Google Cloud Shell

Discussion

Build vs Buy

Segment 6: Operational Reliability

Monitor, Logging & Alerting

- Stackdriver has been renamed to *Operations*
- Collect metrics, logs and traces from Operations
- Built-in Dashboards and custom options
- Query and Analyze
- Setup alerts
- Works for resources in, and outside of GCP

Operations

- Logs Router – Route how/where your logs are delivered
- Error reporting – Aggregate and analyze errors
- Cloud Monitoring – Monitor and alert
- Cloud Monitoring Dashboards
- Service Monitoring – Topology and Context graphs
- Cloud Trace – Latency sampling and reporting
- Cloud Debugger – Inspect the state of your application at any code location without slowing requests
- Cloud Profiler – Continuous profiling of resource consumption
- Cloud Audit Logs – Near real-time User activity and visibility

CI/CD, Deployment & Release Management

- Deployment Manager
  - Structure your templates around operational needs
  - Decoupled units of deployments
  - Isolated by App, department, security needs, etc
  - Consider use of CI/CD pipelines with separate environments (e.g. Dev, Test, Prod)
  - Variables and
  - JSON vs YAML
  - Configuration -> Templates -> Resources

Support & QA

Considerations when you are building and designing a solution:

- How would you support, operate, monitor, and audit this solution?
- What is your QC/QA *process*? (tools can be leveraged differently depending on your process)
- Understand GCP Support tiers/structure
- Testing and automated approvals
- How different would the design be if we could leverage DevOps with CI/CD pipelines as opposed to manual operations?

Segment 7:
Prepping for your exam

First lets review what we've covered

- Segment 0: What to expect on the exam
- Segment 1: Planning and Designing a Cloud Architecture - Everything is related to this!
- Segment 2: Solution Infrastructure – VPC and related components
- Segment 3: Analyzing and Optimizing Process
- Segment 4: Security & Compliance
- Segment 5: Implementing GCP
- Segment 6: Operational Reliability
- …. And this… Segment 7: Prepping for the exam!

Practice

- Qwiklabs Quests for this Exam:
  - Google Cloud Essentials
  - Cloud Architecture
- Your own GCP Account – use free services, and possibly be willing to pay for some
  - Set Billing Alerts & Budgets to avoid surprises
  - Review Billing Reports to learn about costs
  - Understand how you pay for things
  - Understand the free tier

Create designs for yourself

How do I know if I’m ready for the exam?

Consider another cert!

- Get a 2 for 1 on your study efforts!
  **You'll still have to pay for 2 exams though, sorry ☹
- Google Cloud Platform Associate Cloud Engineer Crash Course – I have an upcoming course! Look at the schedule and reserve your spot
- Studying for the Cloud Engineer can assist in the Professional Cloud Architect Certification and vice versa
- Consider taking both the same day!
- It helps to architect and design if you know the products at hand
- The Associate Cloud Engineer is not necessarily easier, but involves less design and more hands-on

Q & A

Lets Connect!

https://www.linkedin.com/in/ryandymek

Thank You!