Deployment, monitoring and alerting, and incident response

Architecting with Google Cloud Platform: Design and Process

Last modified 2018-08-08

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

"Don't just design your service, design its operation."

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

https://pixabay.com/en/pressure-measurement-2176274/

Agenda

Deployment

Monitoring and alerting

Incident response

Photo service: Stabilization and operation

Design challenge #7: Monitoring and alerting

GCP Lab: Deployment Manager - Full Production

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Deployment

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Creating a launch checklist

Dependencies

- Shared infrastructure
- External and 3rd party

Plan for capacity

- Verify overload handling procedures

SPOFs ?

Security and access control

- Verify attack shields

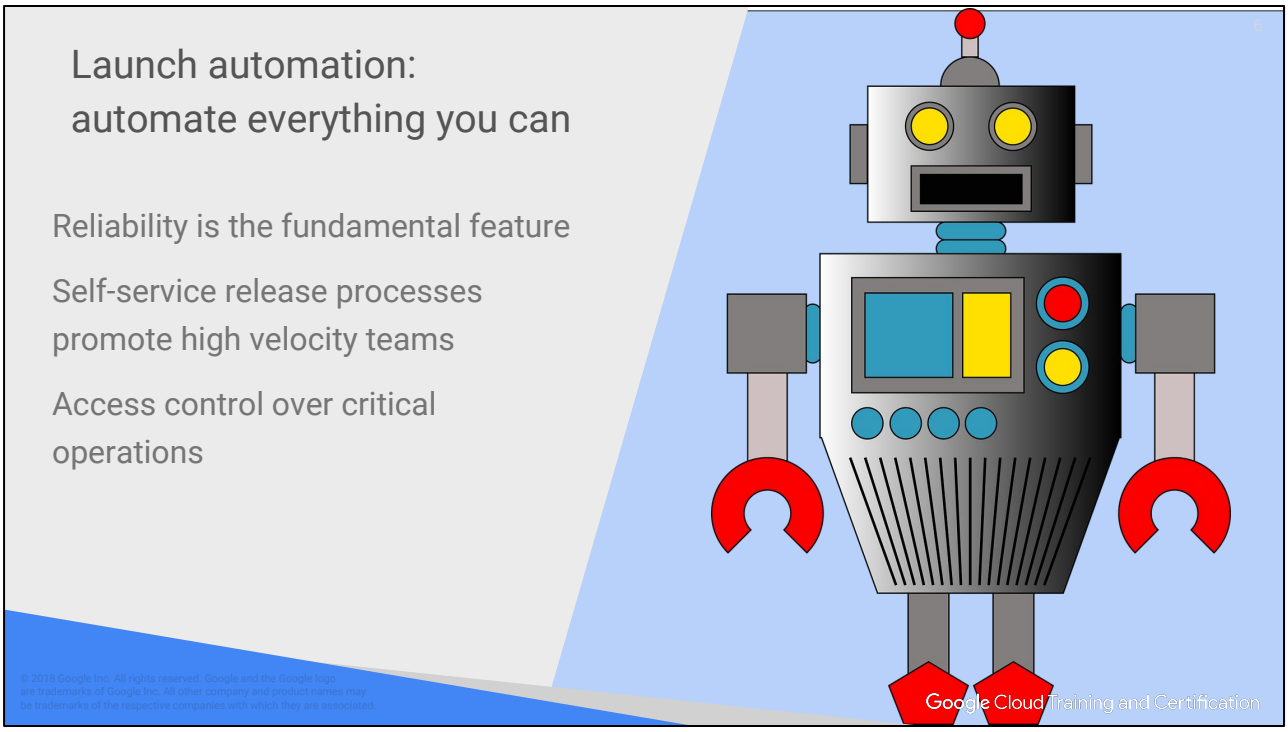
Rollout plan

- Gradual
- Staged / phased rollout
- Percent of users (0% to 100%)

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Not surprisingly, the launch checklist revisits many of the subjects already covered in this class.
Reference: SRE Book: Chapter 27 - Reliable Product Launches at Scale

https://pixabay.com/en/checklist-list-check-check-list-911840/



Reference: SRE Book: Chapter 7 - The Evolution of Automation at Google
Reference: SRE Book: Chapter 8 - Release Engineering

The main idea is that you should automate everything you can. Not just for efficiency, but for consistency, speed, quality, and measurability. Doing this at Google over many applications caused SREs to realize that the main feature of any application is 'reliability'. And launch and release automation has an influence over reliability.

Partial automation - Cases in which access control and human interaction required:

- Approving source code changes
- Determining release process activities
- Creating a new release
- Approvals
- Deploying the release
- Making changes to configurations

https://pixabay.com/en/auto-automatic-machine-robot-1294924/

Deployment Manager

Deployment Manager

Configuration

- Describes all the resources written in YAML syntax

Resources

- `resources:`
  - list of resources to create
    - list of resource properties

Template

- Building blocks of a configuration written in Python or Jinja2 and referenced in the YAML file

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

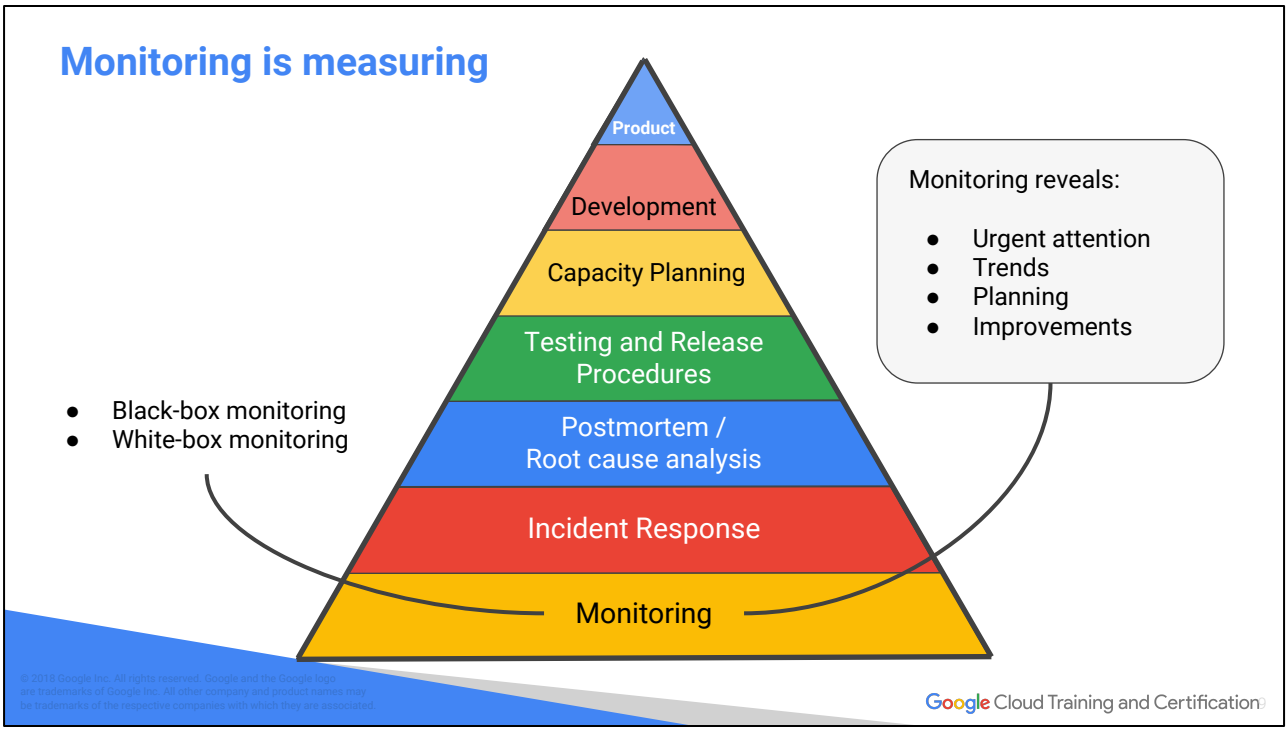
https://cloud.google.com/deployment-manager/docs/

A configuration describes all the resources you want for a single deployment. A configuration is a file written in YAML syntax that lists each of the resources you want to create and it's respective resource properties. A configuration must contain a resources: section followed by the list of resources to create.

Each time you update the configuration in Deployment Manager, it creates a new manifest, which is an object containing the state of the deployment.
A manifest is useful in troubleshooting Deployment Manager projects.

Monitoring and alerting

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.



Reference: SRE Book: Chapter 6 - Monitoring Distributed Systems
This is a model of the SRE Hierarchy. Notice that Monitoring is at the base of all other activities.

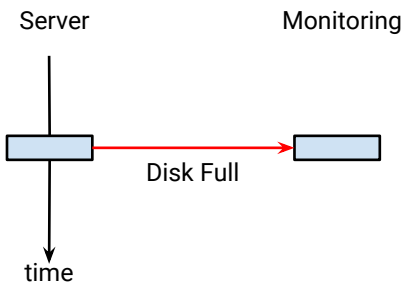
- Something needs immediate attention: "The website is down!"
- Monitor to reveal trends.
  - Something is wrong, but it is not urgent: "Average latency is 200ms higher than last week"
- Long term planning: "How much disk space will be needed for the next year"
- Make measurable improvements, prevent regression.: "If I try this new algorithm, what effect did that have on my performance"

One practical use of monitoring is debugging.
If an API Load balancer is flapping because of intermittent garbage collection, knowing that some Application Servers are having problems is not sufficient.
Which API Load Balancer is causing those problems? That could be determined by RPC error counts, perhaps, and the API LB's state (garbage collection run frequency).
Correlating multiple graphs can speed debugging considerably.

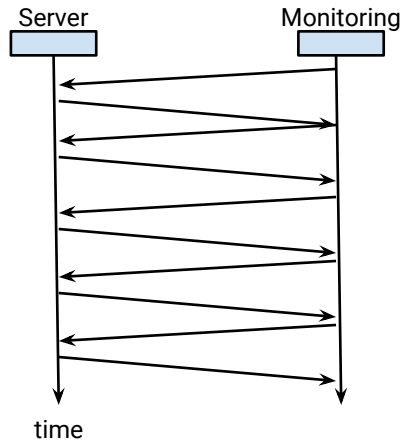
Complications in real systems: organic growth, loops, broad communications trees, different owners of parts of the system, horizontal and vertical boundaries/slices.

Push-based and pull-based metrics

Push



Pull



© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Push-based
Nothing happens until there is an event that notifies monitoring.
Benefits:

- Lower overhead
- Timely alerts
- Gets reliable data from short-lived processes (Example: worker nodes in map-reduce)

Issues:

- Relying on components to monitor themselves, when they may be unreliable, and when is is most important.
  - Example: alerts spooled to disk when disk is full
- Not useful for analysis of unusual, but not alert-worthy events or for long term trends

Pull-based
Monitoring system polls the server for internal state at intervals.
Benefits:

- lots of data
- if the server doesn't respond, that's data too

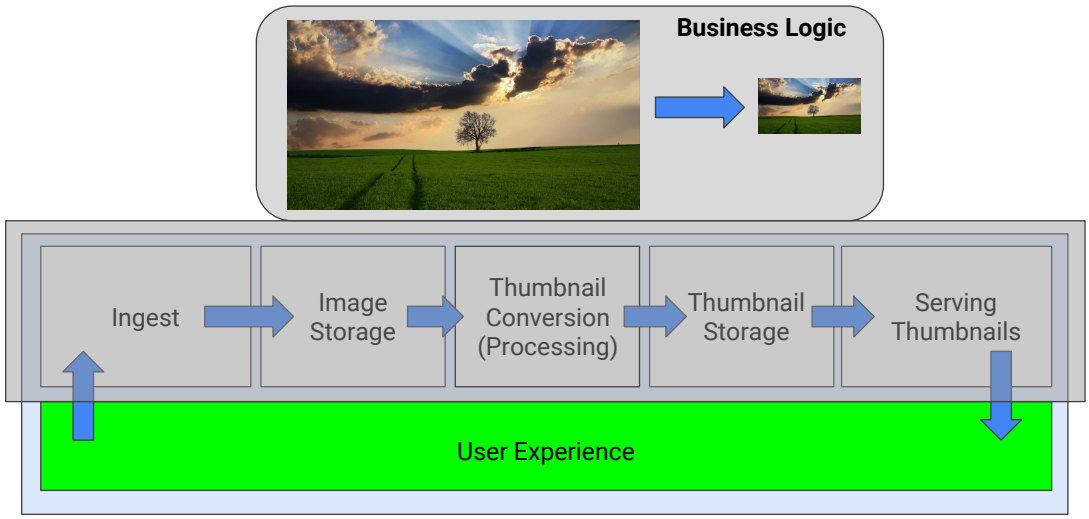
Issues:

- Invasive

- higher overhead / load
- if polling rate is wrong, important events could be missed between polls

It is important to use your SLOs and SLIs to poll at the optimal frequency (request rate per second).
Determine what it means and what action to take if there is no response or a delayed response.

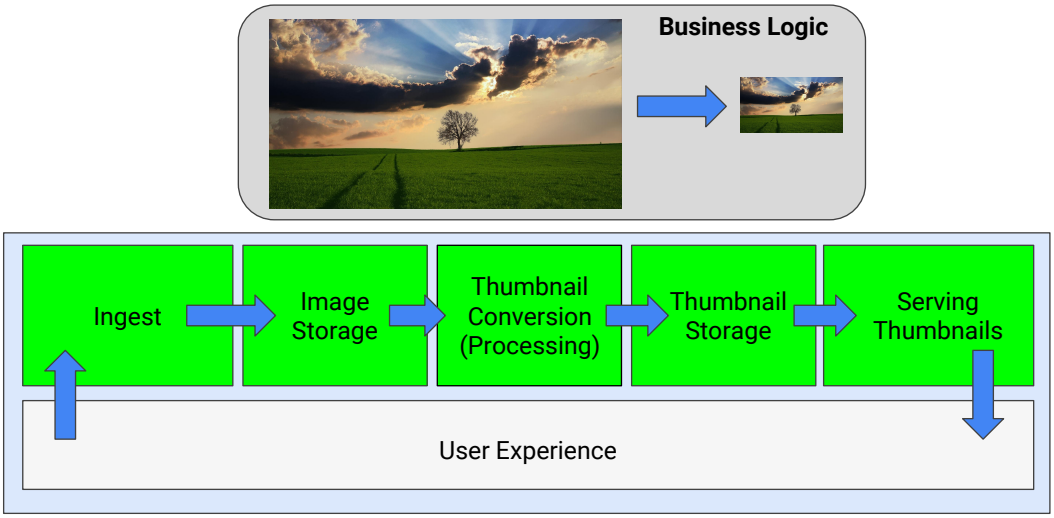
Black box monitoring applied to the sample service



© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

You are emulating a customer workflow. Monitor everything for troubleshooting, but alert on customer pain.

White box monitoring applied to the sample service



© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

White box monitoring is the internal system for troubleshooting, but alert on customer pain.

The Output of Monitoring Systems

Construct your monitoring strategy carefully.

A monitoring system should do more than trigger an alert based on a value or condition.

Monitoring systems should not rely on a human to decide when to react.

Acceptable output of a monitoring system:

- Alerts: A human must take action immediately
- Tickets: A human should take action, but the situation isn't yet urgent
- Logging: Diagnostic information only

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

A monitoring system should not rely on a human to determine if a condition necessitates action. Instead, the monitoring system should be able to differentiate urgent requirements from important ones, and then alert accordingly. To prevent humans from becoming accustomed to alerts that may not require urgent attention, the monitoring system should only alert a human when immediate action is required to maintain the operation of the service being monitored.

Alerting should be used for critical, time-sensitive notifications. For non-critical events that still require a human to react, use ticketing. For event information that is purely diagnostic in nature, log the information to a searchable solution such as Stackdriver.

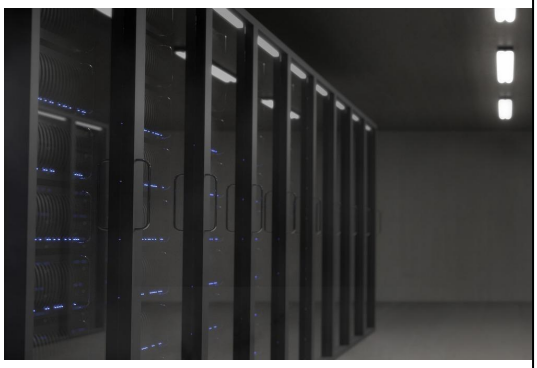
12-factor administration and operations design in GCP

Treat logs as event streams

- Stackdriver Logging

Run admin/management tasks as one-off processes

- Service Accounts
- Stackdriver
  - Monitor, Error Reporting, Tracing, Debugging



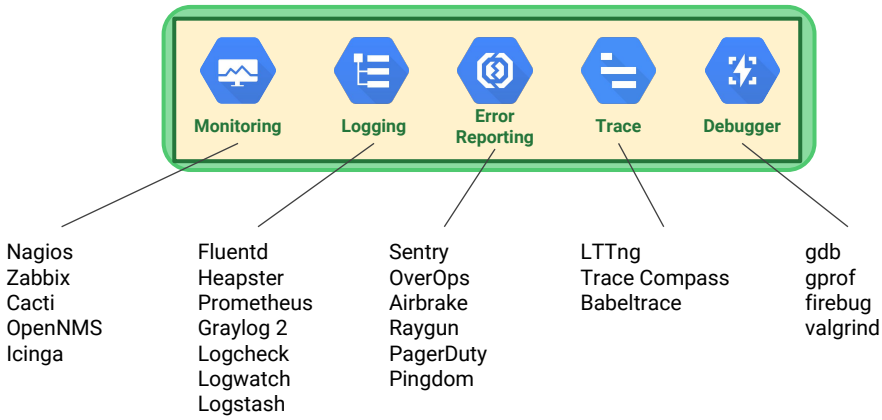
© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Logs: Log files track all kinds of information, from “app has started” to “critical error.” The idea here is not to view them as only forensic tools, but to use them in a monitoring and incident response service. Consolidate logs for convenience. Analyze logs for patterns for optimization or to identify infrequent errors. **Stackdriver Logging:** https://cloud.google.com/logging/

Admin Processes: Plan for running administrative tasks by providing access in the production environment with data cleanup on bad data that has been discovered. Don't run updates directly against a database. Don't run updates from a terminal window. Running analytics for reporting. Performing A/B test of new features. Administrative debugging.
Service Accounts: https://cloud.google.com/iam/docs/service-accounts
Stackdriver: https://cloud.google.com/stackdriver/

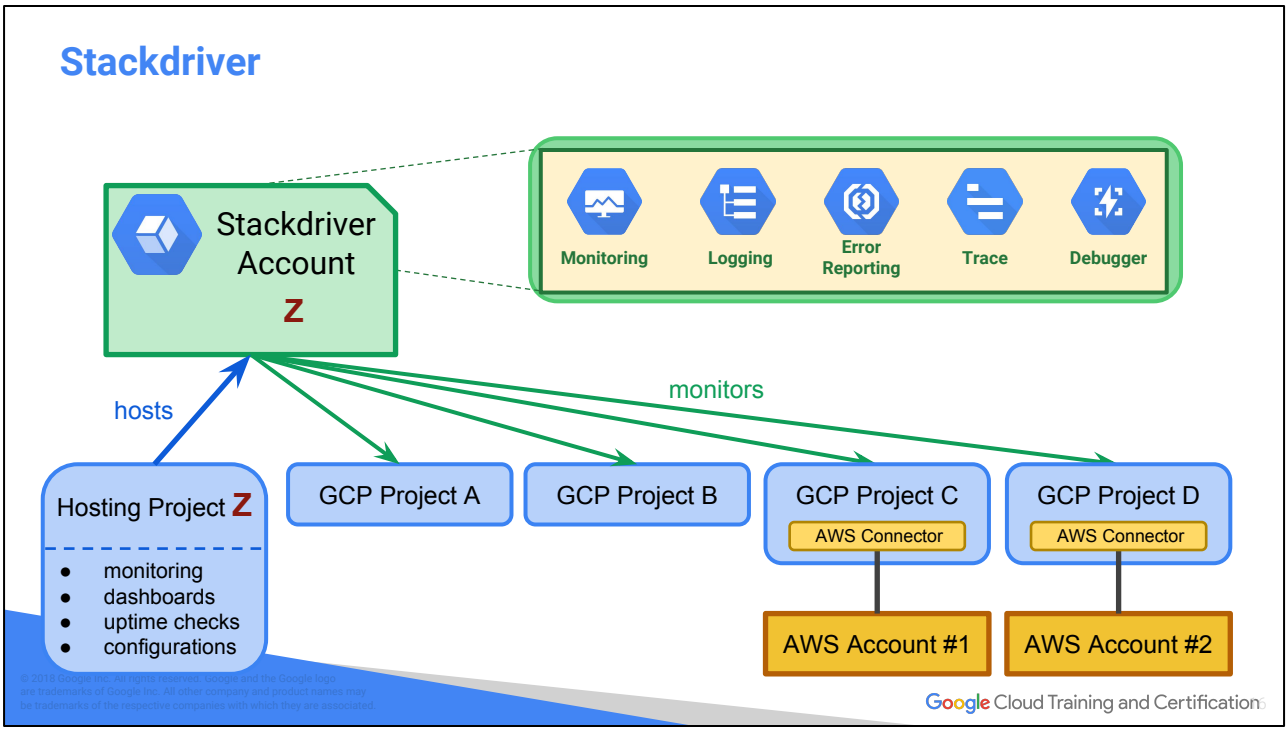
https://pixabay.com/en/server-space-the-server-room-dark-2160321/

On GCP Stackdriver unifies the tools you need



© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

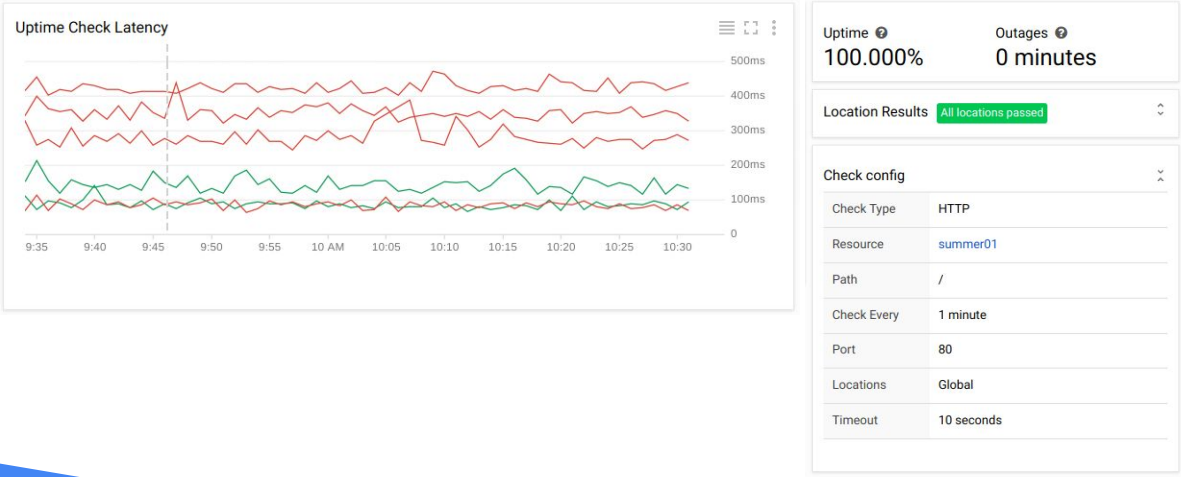
https://en.wikipedia.org/wiki/List_of_debuggers



Monitoring and Logging work everywhere..

A Stackdriver Account is the root entity that holds monitoring and configuration information. You add projects into an Account to monitor them.
You can have multiple accounts to organize separate systems and applications. However, all items that you want to monitor together should be in the same account.
To access AWS account, you must configure a project in GCP to hold the AWS Connector. Each Account or Project can only be monitored by one Stackdriver Account at a time.
A special project called the "hosting" project contains monitoring, configurations, dashboards, uptime checks, and so forth. The Stackdriver Account acquires the name of the hosting account project (in the diagram, Project "Z" becomes Stackdriver Account "Z").
Stackdriver is available in two service tiers, "Basic" (free) and "Premium" which is charged by montiored resources in the account. AWS connection require the "Premium" service.

Uptime check details



© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

https://cloud.google.com/monitoring/alerts/uptime-checks

Uptime checks verify the availability of your service from locations around the world. You can use them as part of alerting policies or you can use the uptime checks overview to review their status.

Creating an Alert

Create new alerting policy

1 **Conditions**

Basic Conditions

HTTP check on instance summer01 — Edit Delete

Violates when: Uptime Check Health on Instance (GCE) summer01 fails

+ Add Another Condition

2 **Notifications** (optional)

When alerting policy violations occur, you will be notified via these channels. Learn more

Email — demo@example.com

+ Add Another Notification

3 **Documentation** (optional)

When email notifications are sent, they'll include any text entered here. This can convey useful information about the problem and ways to approach fixing it.

Edit | Preview — Markdown Formatting Help

 Main Server health check failed
+ Server named summer01 failed a Stackdriver uptime check
+ IP Address of the server is: 104.197.58.79

4 **Name this policy**

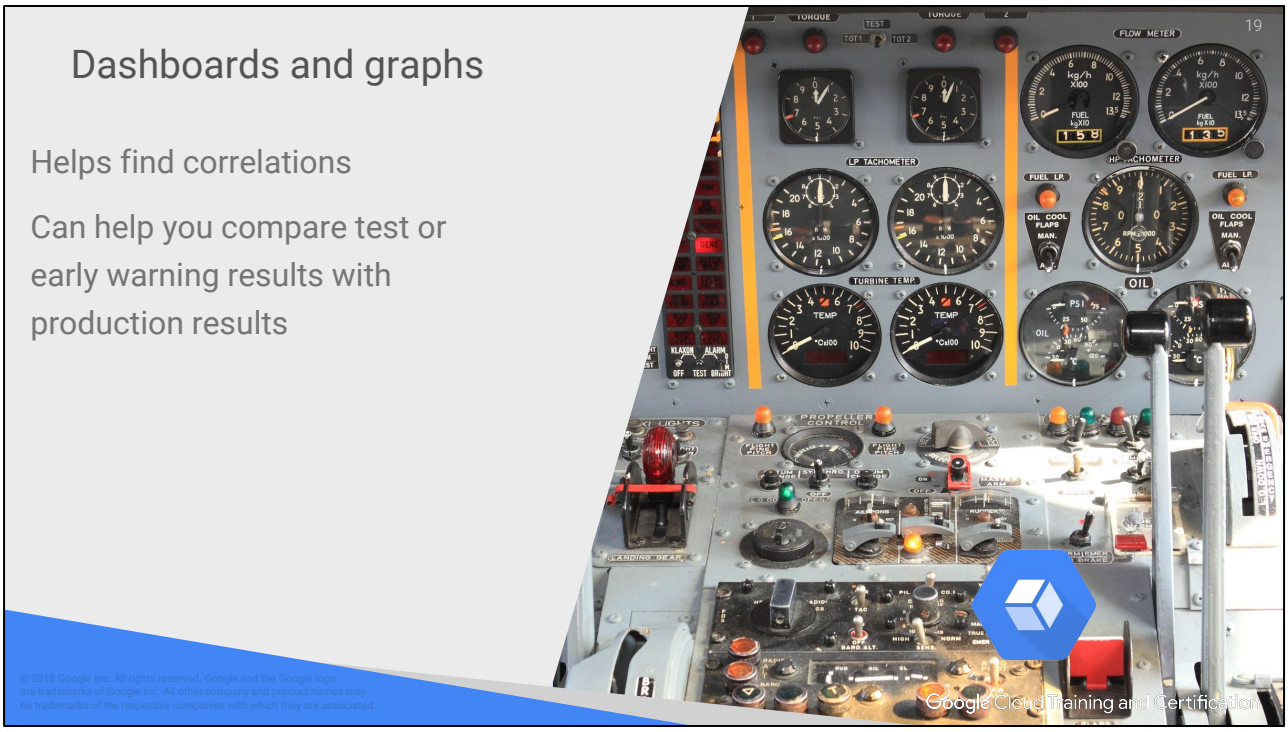
A policy's name is used in identifying which policies were triggered, as well as managing configurations of different policies.

Uptime Check Policy

Save Policy | Cancel

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Documentation associated with Alert makes it easier to scale and distribute monitoring responsibilities.
PAGE 109 Chapter 10
Reference: SRE Book: Chapter 10 - Practical Alerting from Time-Series Data



Dashboards and graphs can help identify correlations when there are symptoms with no obvious cause.

Build dashboard and graph resources over time. For example, if you once spent a lot fime tracing a latency blip to a JVM garbage collection cycle, generate a graph of GC events for each server and graphs for end-user latency. Add an overlay of the two events to a dashboard. Next time it happens, even in a different service, you will spot it immediately.

Complicated dashboards can take longer to scan, especially when you are stressed. Comment out old graphs, when they are no longer useful.

If you have the resources, keep collecting the time series, so in a pinch, you can make the graph again.

https://pixabay.com/en/aircraft-fighter-cockpit-instrument-1367289/

Example of Stackdriver dashboard with charts

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Stackdriver charts are dynamic. So in this case, if you were to drag the horizontal line on the chart that reflects CPU utilization, the line that reflects bandwidth will move accordingly.

Install Stackdriver monitoring and logging agents

```
curl -O https://repo.stackdriver.com/stack-install.sh
sudo bash stack-install.sh --write-gcm

curl -sSO
https://dl.google.com/cloudagents/install-logging-agent.sh
sudo bash install-logging-agent.sh
```

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

https://cloud.google.com/monitoring/quickstart-lamp

Incident response

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

User trust **depends** on incident response

PROCESS

User trust depends on consistency and reliability.

To develop and strengthen trust:

- Minimize outages
- Transparency
- Respond and learn quickly

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud Training and Certification

Google learned that user trust depends on incident response. Users don't expect a perfect service. They are looking for a service provider that is transparent, responds quickly to issues, and learns from their mistakes.

Users trust you with their data, services, and apps. They depend on you to minimize outages and to keep their services running.
One of the ways to view trust is to think about how consistent and reliable your systems are over time.
The more consistent your service reliability over time, the more customer confidence in the product strengthens.

Transparency, in a business or governance context, is honesty and openness. It can also mean predictable. As a transparent computer program is one for which the output is entirely predictable from the input. Another way of describing this quality is to try to avoid surprising the customer, especially with bad news.

https://pixabay.com/en/trust-faith-encouragement-1418901/

Develop a **structured** incident response

PROCESS

Being prepared:

- Fast response
- Know what to do when

Consistency:

- Reduce duplication of effort
- Know who is doing what

SRE Culture

- Collaboration, communication
- Learning, and sharing
- Organized flow of information
- Organized customer notifications

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

The time to decide how to respond to incidents is not when they are happening.
SRE is Google's Site Reliability Engineering. It is a methodology for creating reliable global-scale cloud-based services.

https://pixabay.com/en/meeting-together-cooperation-1015316/

Fast response: Have pre-instrumented "EASY" buttons for responding to incidents

Good for incident response:

- Microservices architecture
- Continuous Integration / Continuous Development (CI/CD)

Monitoring dashboard

- For monitoring throughputs, loads, and consumption of resources

Alerting regimen

- Notify people before (or as soon as) users experience issues

Plans and tools for responding to issues

- A set of EASY buttons

https://pixabay.com/en/aircraft-fighter-cockpit-instrument-1367289/
https://pixabay.com/en/help-button-red-emergency-support-153094/
https://pixabay.com/en/phone-smartphone-mockup-hands-girl-2212961/
https://pixabay.com/en/checklist-clipboard-questionnaire-1622517/

Site Reliability Engineering

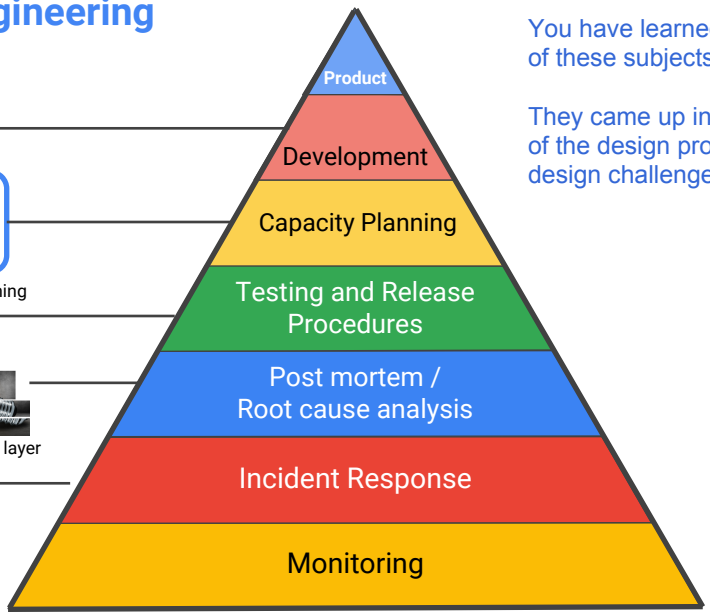
Business-logic layer

Capacity planning

Resiliency

Business-logic layer

Security



You have learned about most of these subjects already.

They came up in the context of the design process or in the design challenges.

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Review

Here is the list of process lessons covered throughout the class. Some of them came from costly mistakes that you won't have to repeat

These best practices set you up for success responding to incidents.

There are just a few practical items left to cover on the subject of incident response.

1. Create actionable alerts
2. Test before and during launch
3. Keep development simple, iterative, and aligned
4. Systematic logical troubleshooting
5. Collaboration and communication
6. Find and fix the real problems
7. Policy for writing postmortem reports
8. Prevent overload, make load testing real
9. Prepare the team using simulation sessions
10. Meetings matter for resiliency
11. Prepare the team for disasters by using drills
12. Have a plan for dealing with a major outage
13. User trust depends on incident response
14. Develop a structured incident response including a dashboard, alerting regimen, and plans and tools for responding to issues
15. Have a balanced approach to the launch process

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

12-factor "admin/management tasks as one-off processes"

Separate administration from application interfaces.
When a service has an issue, you don't want the diagnostic information or the access for fixing the problem to rely on the same systems that are in trouble.

Signs of control plane functions

- Makes system decisions
- Originates and consumes system messages for coordination; signalling
  - System configuration, operation, management, or state information

Signs of data plane functions

- Receives and transmits information that is valuable and meaningful to the user
- Implements application logic and performs transformations on user data
- Forwards the results to the user

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Also called in-band and out-of-band communications.

Control Plane and Data Plane are networking concepts that can be applied as a useful analogy to system design.
Information and logic that controls the application should be separated from and treated differently from information and logic that is the subject of the application.

When a service has an issue, you don't want the diagnostic information or the access for fixing the problem to be relying on the same systems that are in trouble.

https://learningnetwork.cisco.com/thread/33735
http://sdntutorials.com/difference-between-control-plane-and-data-plane/
https://en.wikipedia.org/wiki/Forwarding_plane
https://pixabay.com/en/network-cables-rj45-patch-499792/

In the actual Virtual Private Cloud (VPC) network inside the Google Network, the Control Plane is concealed and the Data Plane is a standardized service.

Build a **playbook** based on alerts

PROCESS

For each alert:

- Severity and impact (useful for triaging multiple alerts)
- Verify that the problem is still occurring (and isn't a monitoring issue)
- Diagnose
  - Potential causes of problem and how to investigate them
- Fix the problem
  - Immediate mitigation
  - Long-term solution
- Post-fix (possibilities)
  - Do a post-fix cleanup
  - Escalate to another team

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud Training and Certification

At Google, a playbook is another term for an operations manual.
For every alert the monitoring system can generate, you should have a corresponding section in your operations manual to explain how to handle the event. The sections in your operations manual should be standardized to make the manual easier to read, and it should be written to the knowledge level of a newly onboarded employee. Don't make assumptions regarding any company-specific knowledge.

There are many resources online for building SECURITY incident response playbooks.
These can serve as a starting point and background for a broader set of playbooks for your system.
https://www.nist.gov/publications/computer-security-incident-handling-guide
https://www.incidentresponse.com/playbooks/improper-computer-usage
https://www.demisto.com/how-to-build-incident-response-playbooks/

https://pixabay.com/en/chalkboard-quote-1927332/
https://pixabay.com/en/book-read-literature-old-learn-1740515/

Create "easy buttons"

PROCESS

Use microservices with RESTful APIs

Use Google Cloud APIs

Drain a service for maintenance

- Load balancer controls
  - Connection draining
- Instance group / server controls
  - Pause scaling and shut down servers

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud Training and Certification

https://cloud.google.com/compute/docs/load-balancing/enabling-connection-draining

A few methods to consider:

- Use a microservices design leveraging Cloud Functions
- Use a bastion host with a service account
- Use Google Private Access (covered in the Security module)

https://pixabay.com/en/help-button-red-emergency-support-153094/

Balance interrupt-driven work and project-driven work

PROCESS

On-call rotation: funnel interrupts to one, nominated person

- The oncall makes sure that issues are fixed
- Lets everyone else work uninterrupted
- If nobody is oncall then everybody is oncall
- The rotation schedule makes sure the oncall doesn't burn out

Reduce noise

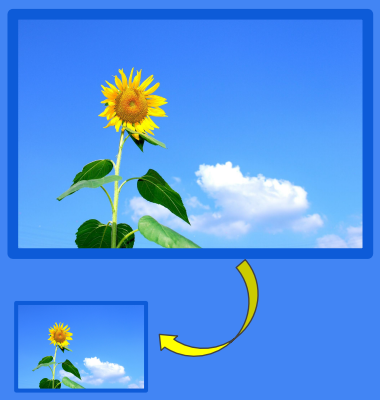
- Prune noisy alerting
- Automate anything that can be automated
- Reduce low-value work

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Balancing interrupt-driven work with project-work is difficult. Reduce the amount of low-value work the team does as much as possible - even if this involves some short-term pain.
Human effort does not scale well: your service should be robust and self-service as much as possible.

https://pixabay.com/en/safety-helmet-construction-hat-295057/
https://pixabay.com/en/helmet-firefighter-fire-equipment-2074951/

Stabilization and operation



3-6 months of stabilization of the solution

How do you keep the service running?

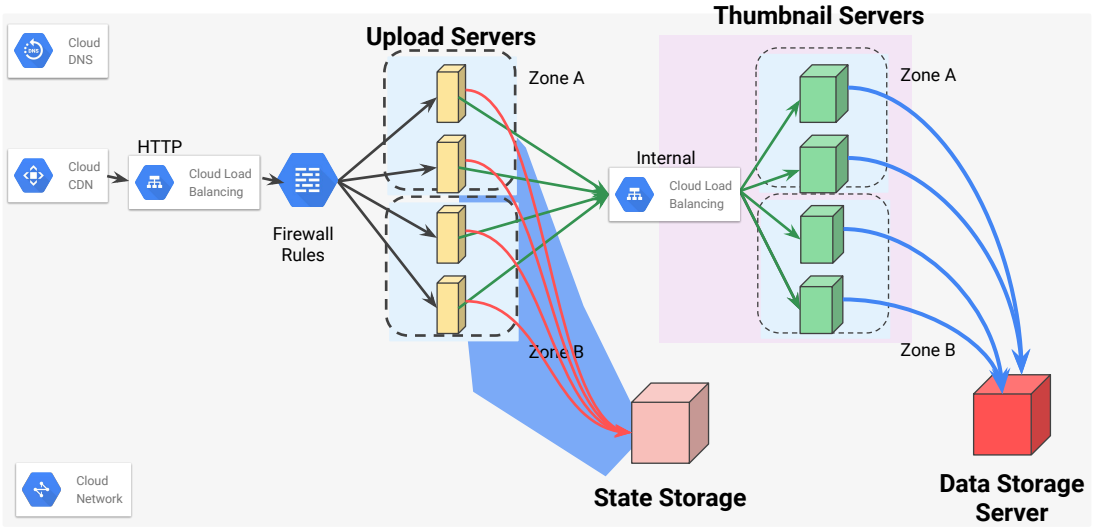
What to watch?

When to worry?

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

https://pixabay.com/en/summer-sunflower-flowers-sky-cloud-368224/

Monitoring and Alerting



© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

What monitoring and alerting to set up for the service?

What if there is a loop?

What if there are runaway consumption of resources?

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

YOUR TURN

Design challenge #7

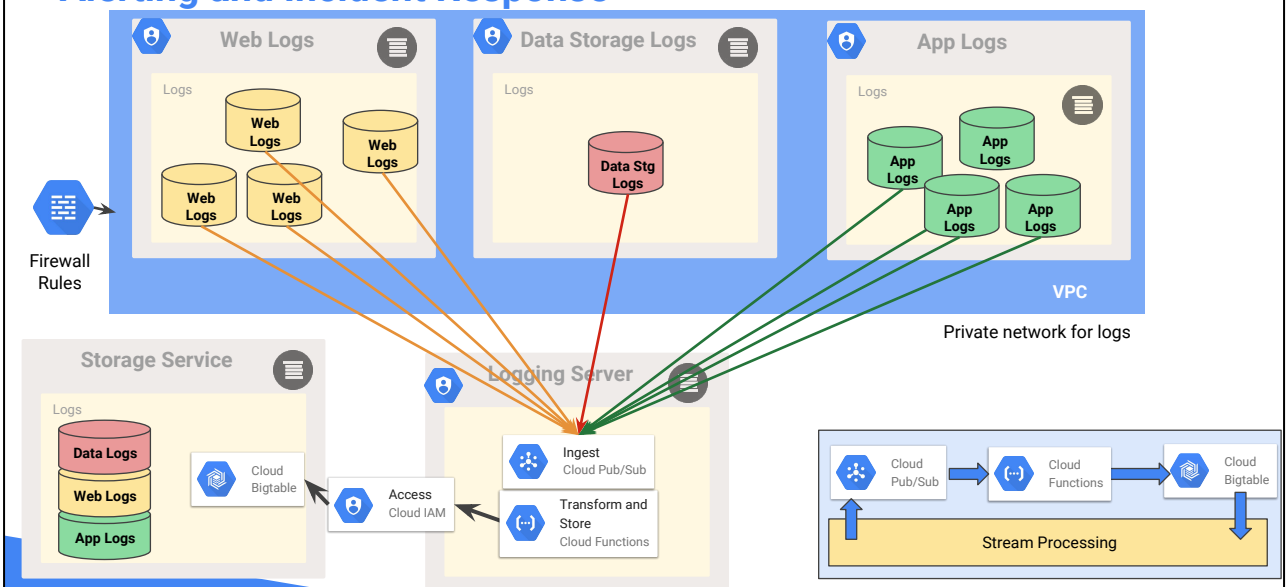
Monitoring and alerting

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud Training and Certification

https://pixabay.com/en/the-strategy-win-champion-1080527/

Alerting and Incident Response



© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Take a few minutes to design your solution

Problem: The logging service is not stable. It needs monitoring and alerting added. Consider what you think would be important to monitor and under what conditions alerts notifications should be sent. Make a list of three monitoring and alerting items you would put in place to support stabilizing the solution.

There are multiple designs possible depending on your assumptions. Your solution may be better than the one shown. The point of this exercise is to "think about the design" to develop your architecting skills.

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

What monitoring and alerting to set up for the logs?

Logs no longer being generated/recorded (watchdog).

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Reference Architectures Online

Google Cloud Platform Solutions

https://cloud.google.com/solutions/

Tutorials and Solutions

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

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud Platform

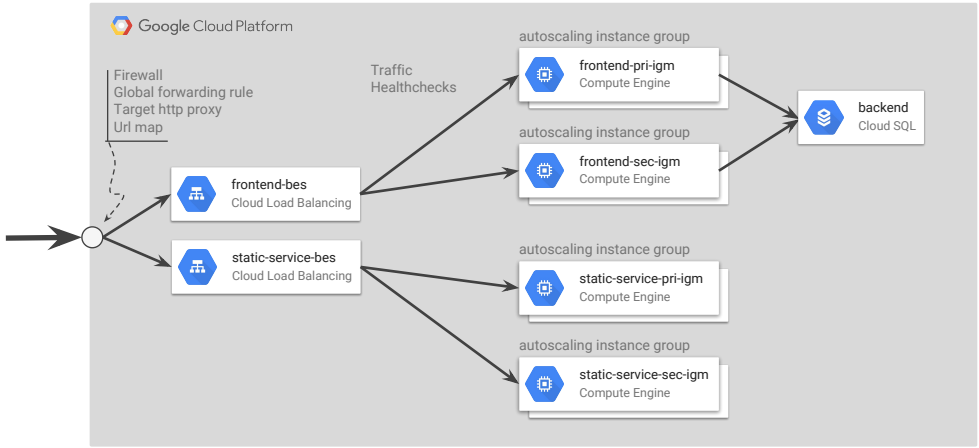
GCP Lab

Deployment Manager - Full Production

Lab 4: A complex and full production deployment with Cloud SQL server, static and active autoscaling. Start it up and then apply Stackdriver monitoring, uptime checks, alerting and create a dashboard with charts. (Logbook application).

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Lab deployment



© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud Training and Certification

© 2018 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.