



# How to Build a Healthy On-Call Culture

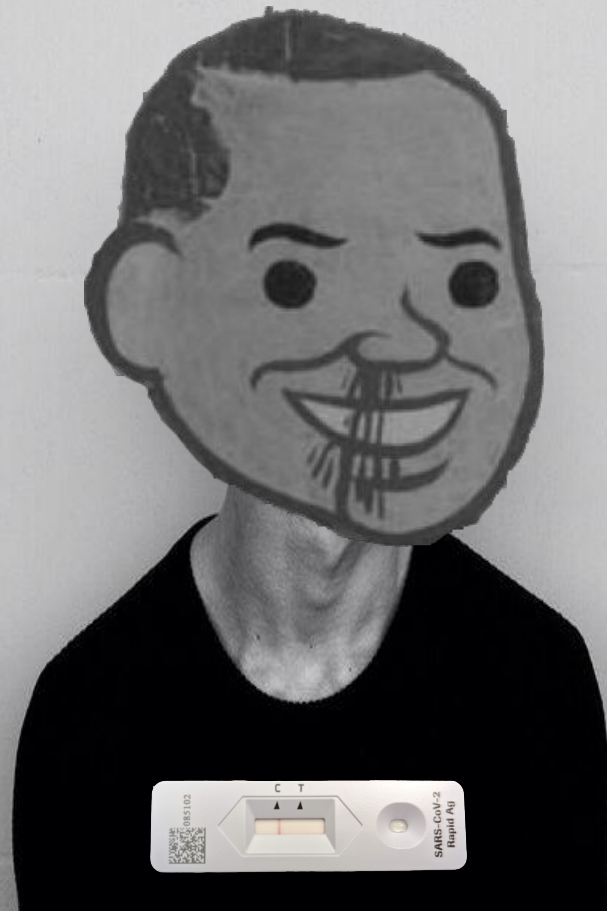
# Hello!

I am smalltown

MaiCoin Group Lead SRE

Taipei HashiCorp UG Organizer

AWS UG Taiwan Staff





**Organization Type**



**On Call**



**Monitoring**



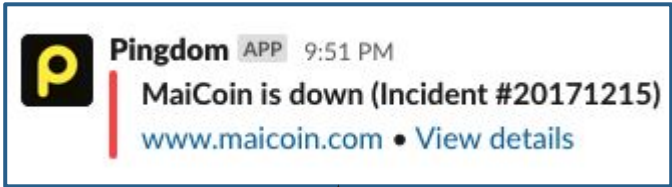
**Incident Response**



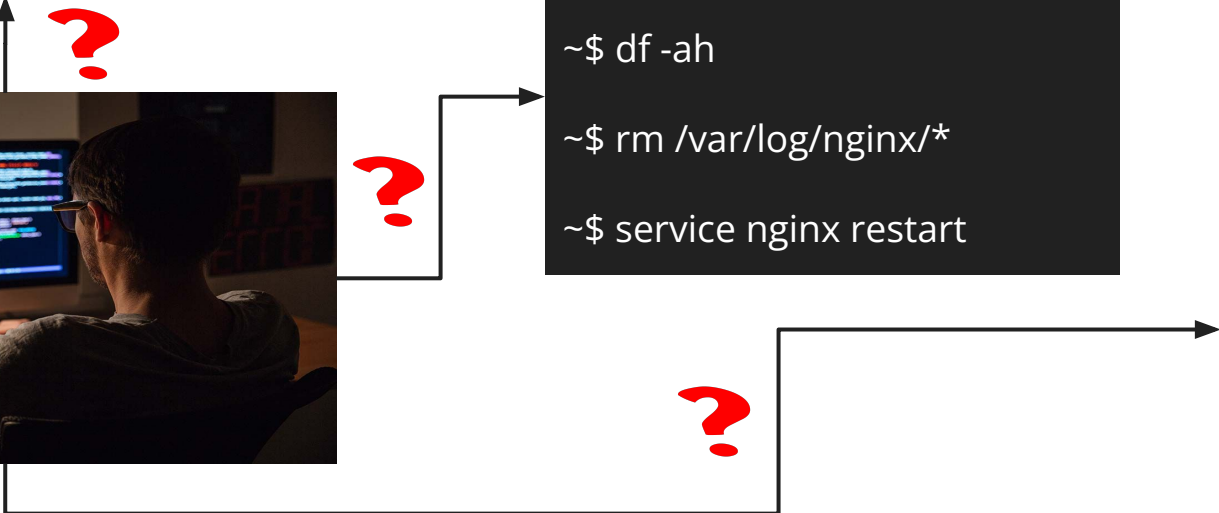
**Root Cause Analysis**



# Join MaiCoin First Week (End of 2017)



```
~$ ssh smalltown@172.0.10.1  
~$ df -ah  
~$ rm /var/log/nginx/*  
~$ service nginx restart
```





**Organization Type**



**On Call**



**Monitoring**



**Root Cause Analysis**



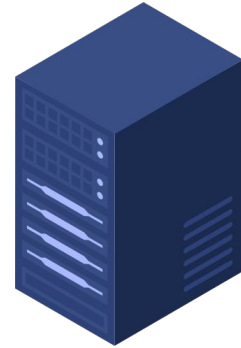
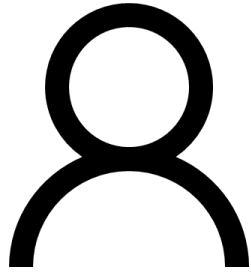
**Incident Response**



# Why Need Monitoring?

- Service **W/O Monitoring** Just Like a **Blind**
  - **Not Know** Service Usual Health Status
  - **No Notice** When Something Bad Will/Has Happen(ed)
  - **Not Investigate** Issue After Incident
- **No Monitoring** -> **No Measure** -> **No Quality**

# Common Monitoring Types



- End-To-End
  - Functionality
  - Performance
- Error Tracking

- Networking
  - Latency
  - Connectivity

- Infrastructure
  - CPU
  - Memory
  - Disk
  - ...

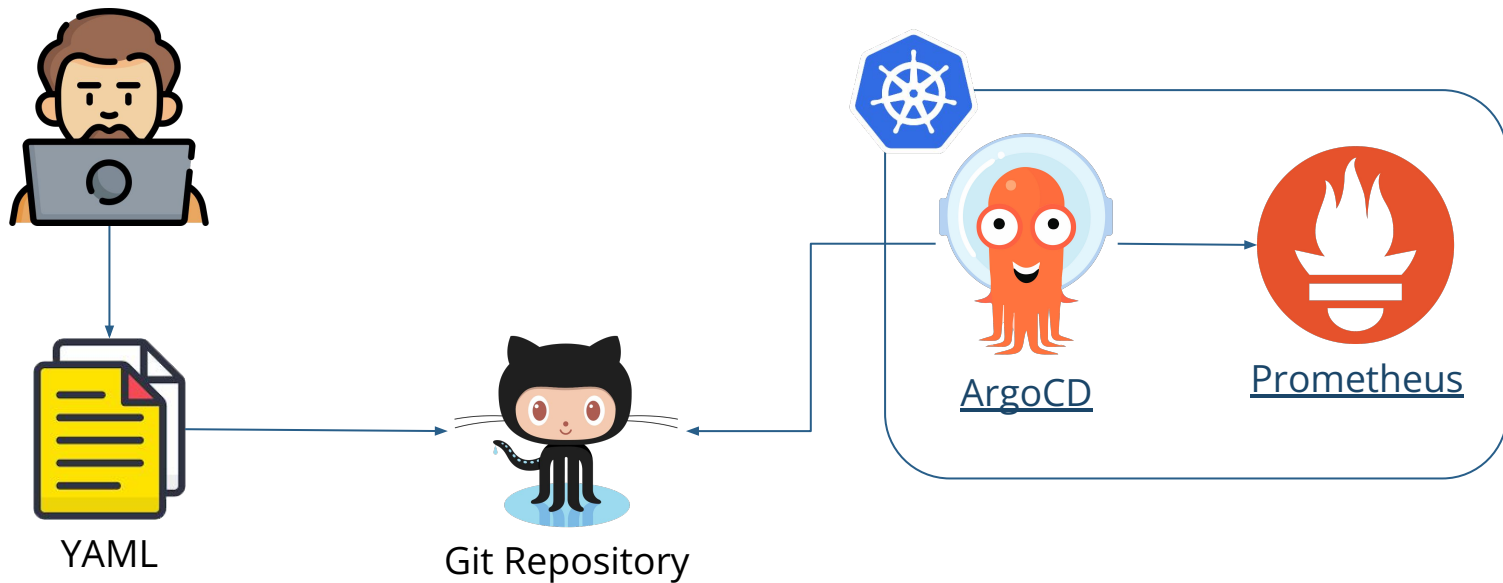
# External Monitoring



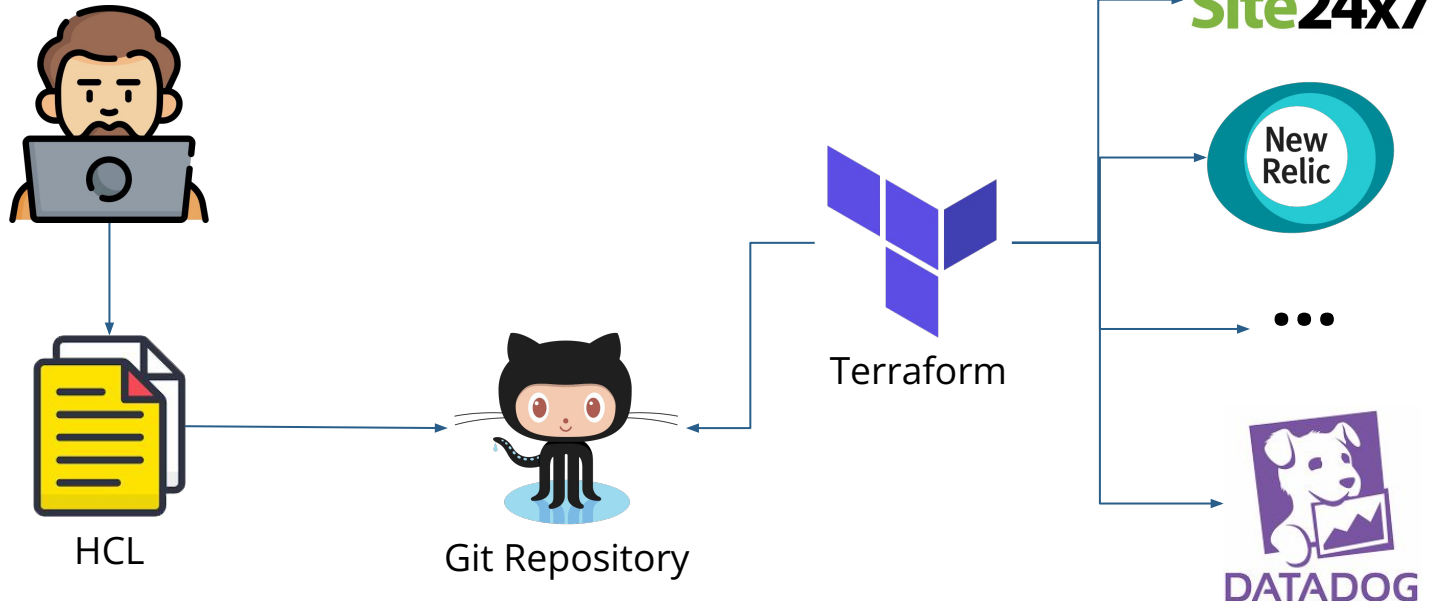
- DNS
- Certificate
- Geolocation (Global Ping)
  - Routing
  - Submarine Cable
- Content Delivery Network
- Web Application Firewall
- ...



# Monitoring As Code - Internal



# Monitoring As Code - External





**Organization Type**



**On Call**



**Incident Response**



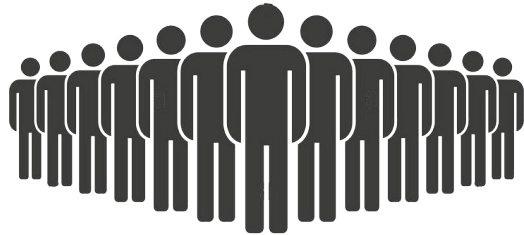
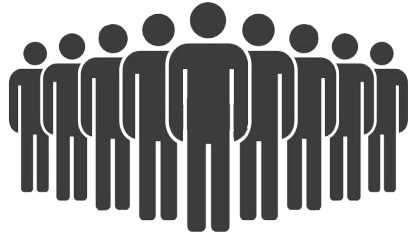
**Monitoring**



**Root Cause Analysis**



# Organization Scale



MaiCoin

# Organization Geographical Distribution



**VS**



# Organization Architecture - Traditional DC

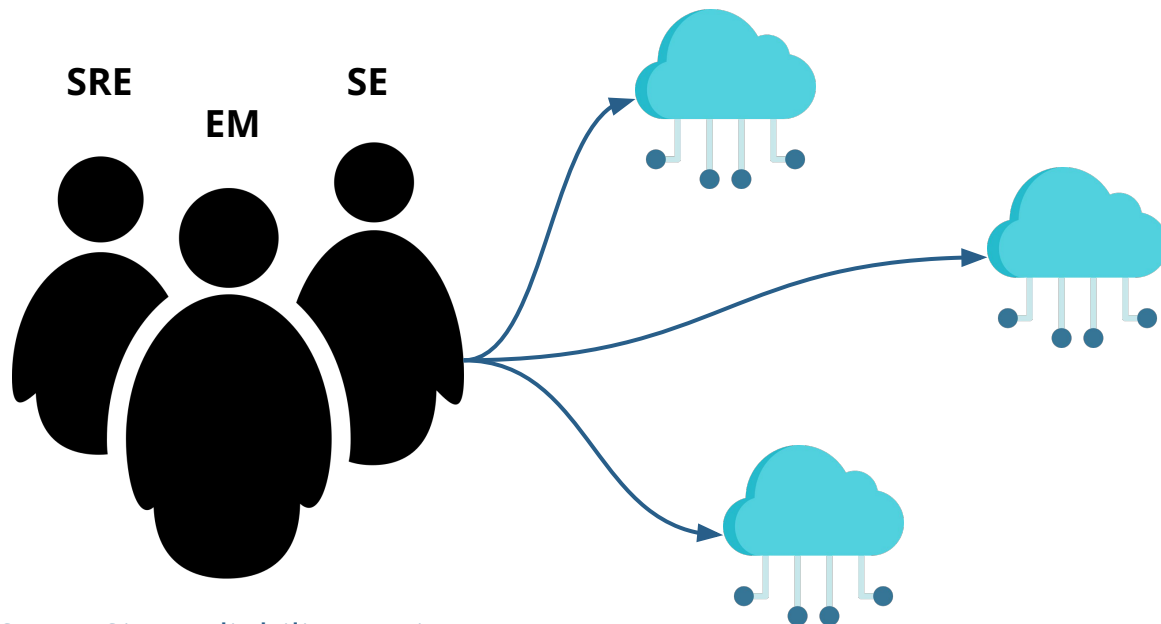


Development



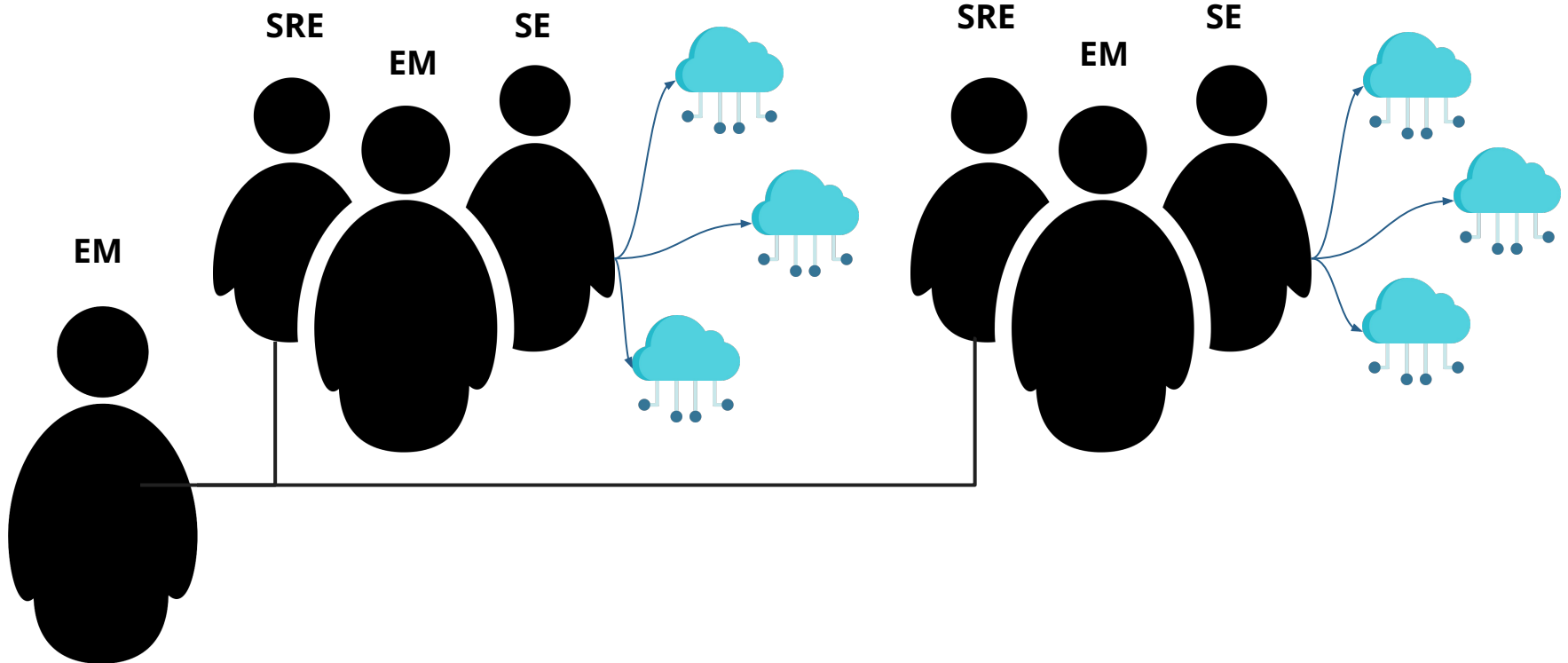
Operations

# Organization Architecture - DevOps



SRE = Site Reliability Engineer  
EM = Engineering Manager  
SE = Software Engineer

# Our Organization Architecture







**Organization Type**



**On Call**



**Incident Response**



**Monitoring**



**Root Cause Analysis**

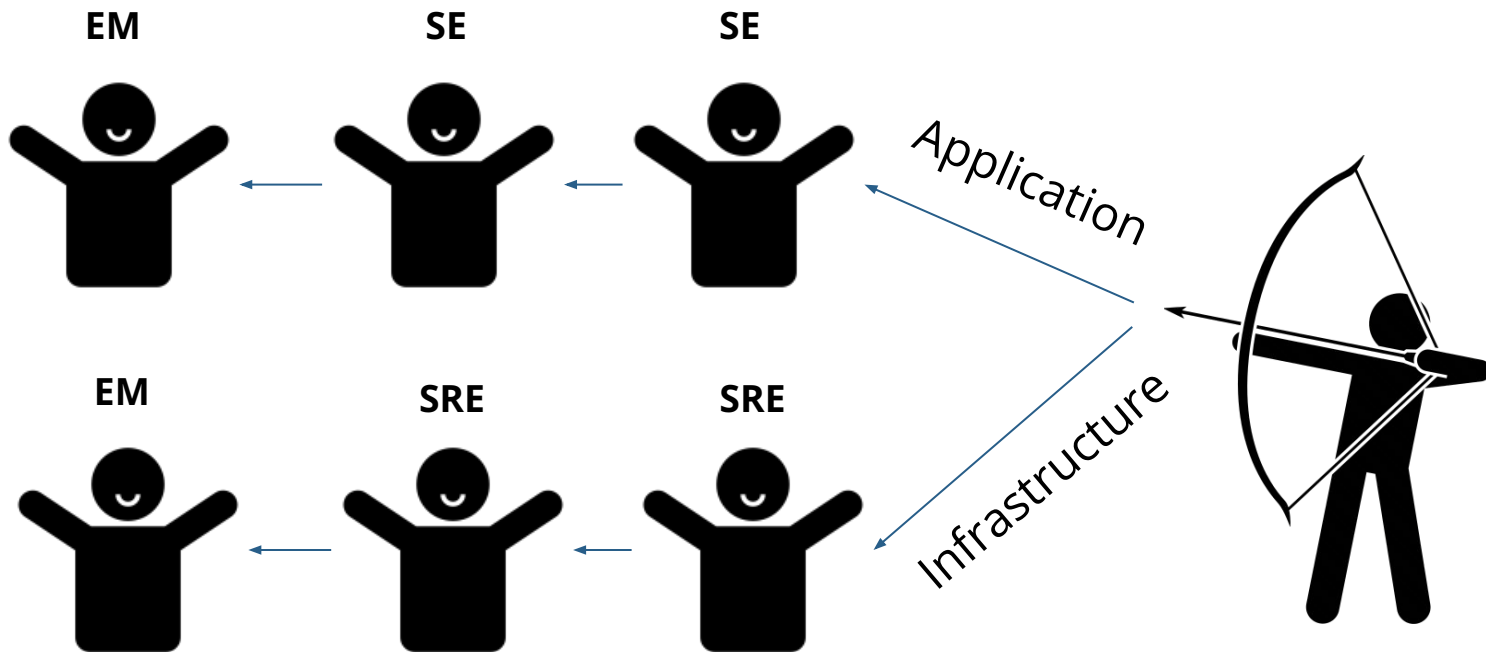


# On Call Engineer Responsibility

- Routine Operation Job
- Handle Incident
- Runbook Refine/Writing
- Weekly On Call Report



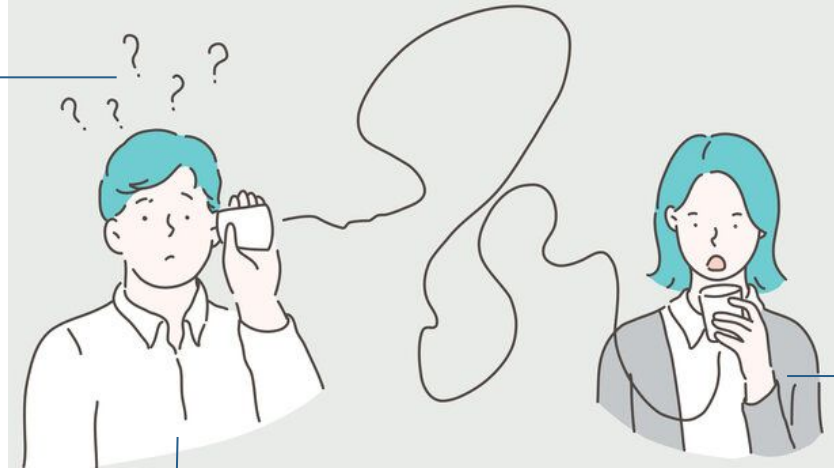
# On Call Model - Two Tracks



Escalation Path

# Notification Problem

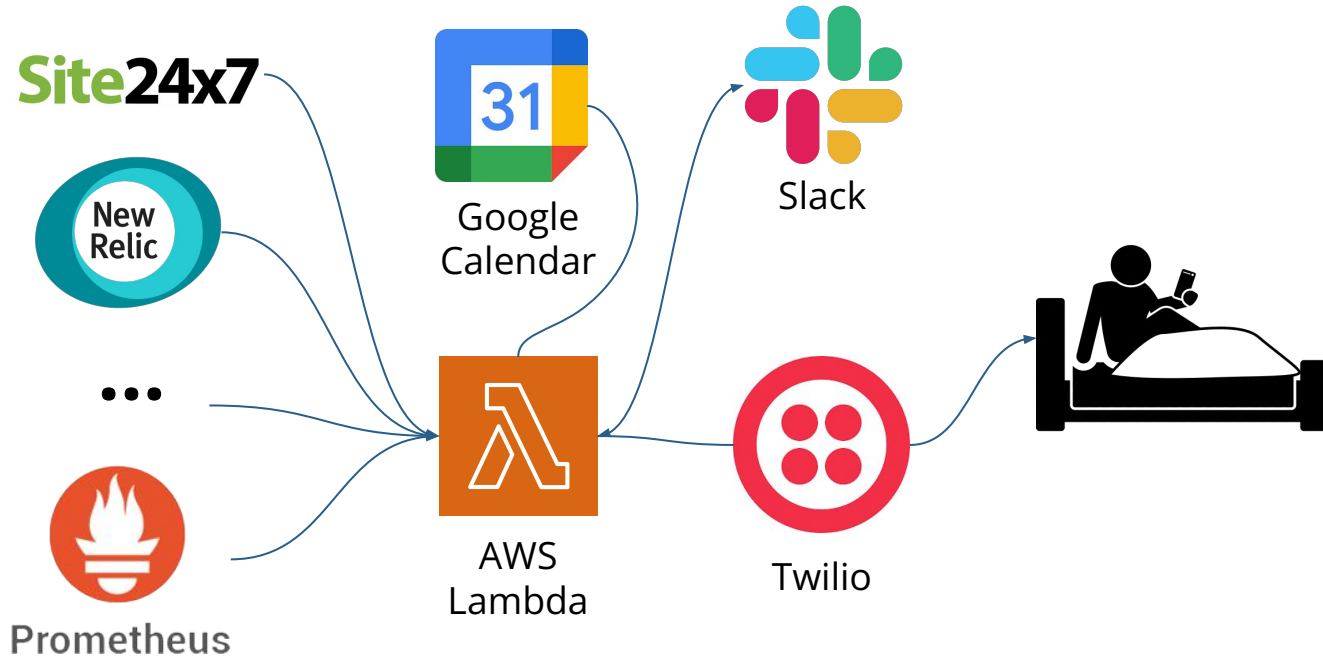
Voice Call Notification  
User Experience



Communication  
Between Teams

Arrange On Call Engineer Schedule

# Notification System Architecture



# Notification System - Rotation

## On Call - Week 17

活動 工作 提醒

🕒 2月 15日 (星期二) - 2月 22日 (星期二)

全天

不重複 ▾

安排時間

---

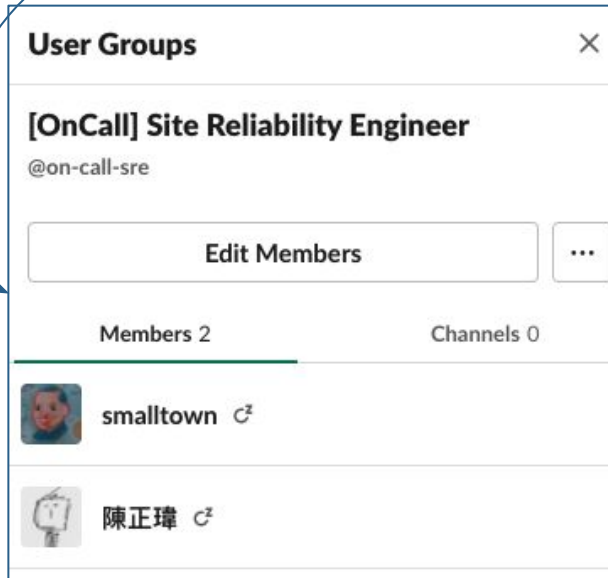
👤 新增邀請對象

-  smalltown  
主辦人
-  陳正璋 \*  
可不出席

Do

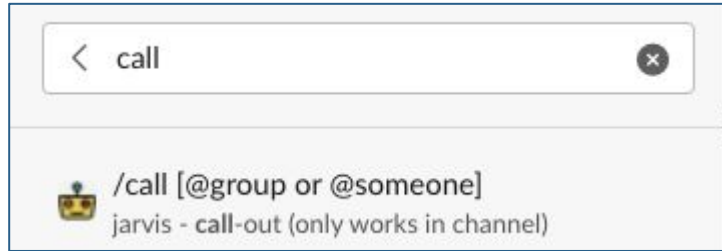
- On Call Model
- Rotation Frequency
- Fail to Answer a Page

# Notification System - Communication

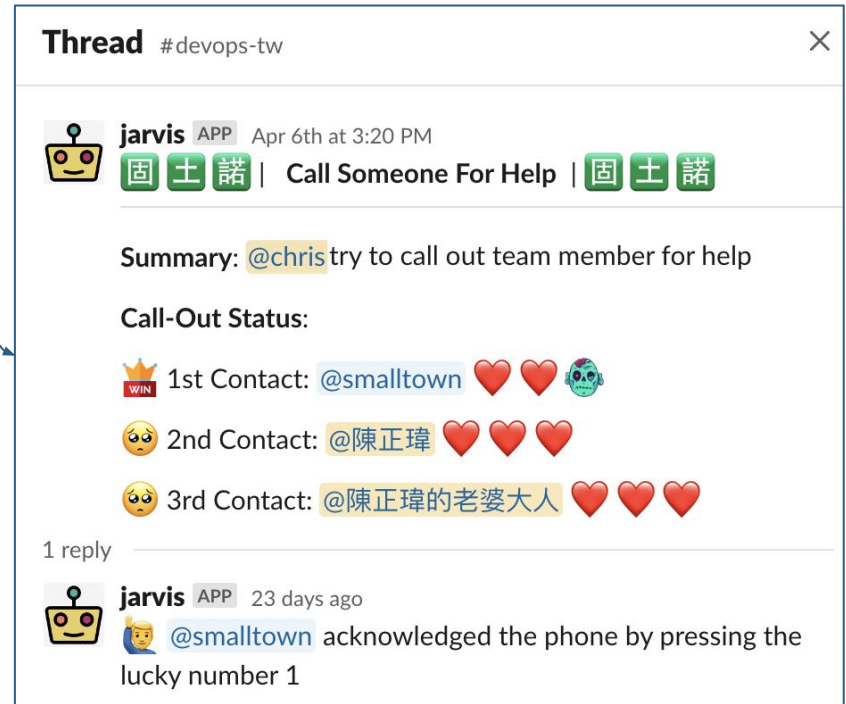


- Slack User Group Break the Communication Cap
- Check Document -> Check Calendar -> Tag Group Name

# Notification System - Effectiveness



- Contacts Phone Book -> Slack Command W/ User Group/Name
- Short Phone Call Notice -> Phone Call W/ Validation
- Notification Visibility -> Detail Status in Slack







**Organization Type**



**On Call**



**Monitoring**



**Root Cause Analysis**



**Incident Response**



# Aware Incident Before Customer



**MaiCoin** 14 小時 · 已讀

🚩 平台公告 | 4/29 系統維修

因系統維護，4/29(五) 10:30~11:00 接收與發送有高機率延遲，建議您避免於此期間操作出入金。造成不便，敬請見諒。

MaiCoin

## 最新公告

平台最新重要資訊

The screenshot shows a social media post from MaiCoin. The post header includes the MaiCoin logo, the name 'MaiCoin', and a verified status. Below the header, the post content is in Chinese, starting with a yellow flag icon and the text '平台公告 | 4/29 系統維修'. The main text of the post explains a system maintenance period on 4/29 (Friday) from 10:30 to 11:00, noting that there will be a high probability of delays in receiving and sending transactions, and advises users to avoid transactions during this time. The post is set against a red background with a sunburst pattern. At the bottom of the post, there is a white line-art illustration of an astronaut holding a megaphone, with the text '最新公告' (Latest Announcement) and '平台最新重要資訊' (Platform's latest important information) overlaid on the background.

# Incident Handle Terminology



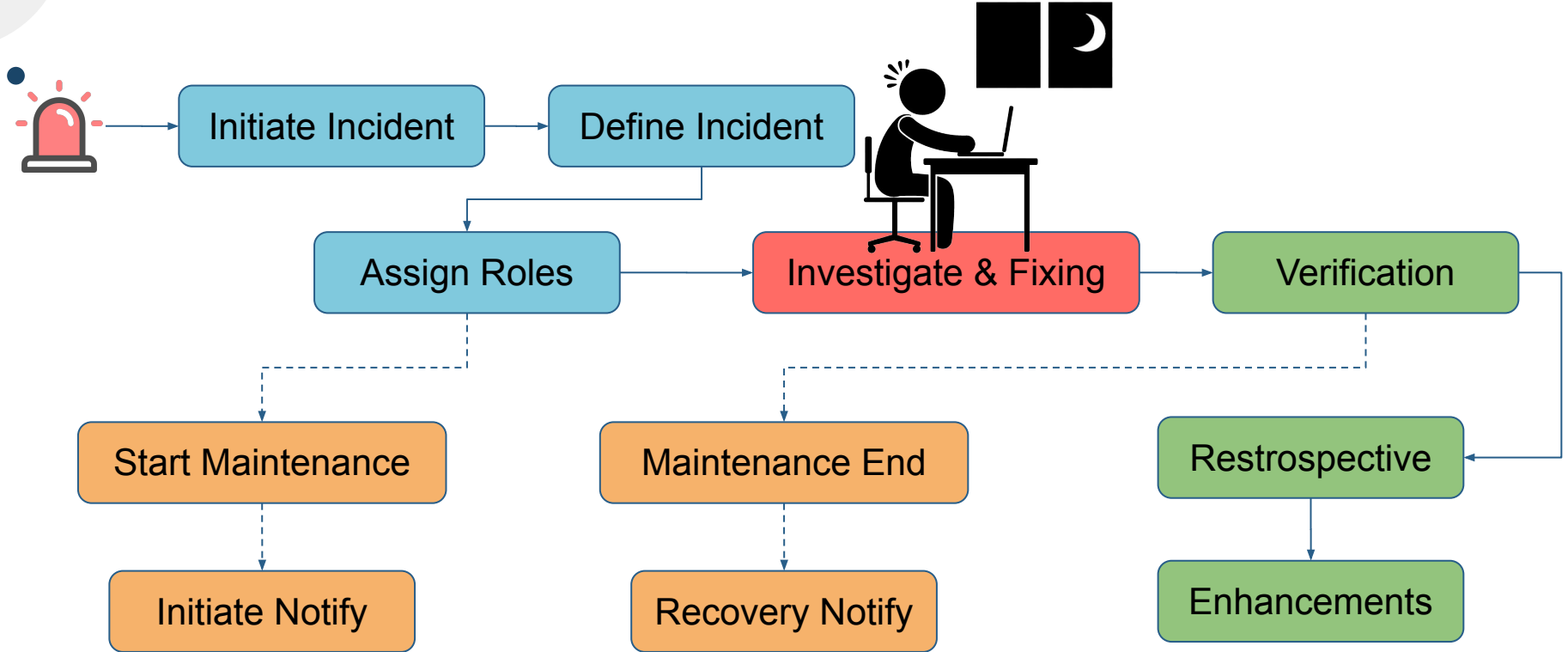
## Role

- Incident Commander
- Tech Lead
- Communication Lead
- Engineering Manager

## Incident Level

- S3
- S2
- S1
- S0

# Incident Handle Process



# Incident Visibility - Internal

## #incident-20220324-max-ethusdt-orderbook-incorrect

@Trouble Solver created this channel on March 24th. This is the very beginning of the #incident-20220324-max-ethusdt-orderbook-incorrect channel.

[Add description](#) [Add people](#) [Send emails to channel](#)

Thursday, March 24th ▾



**Trouble Solver** APP 11:37 PM

joined #incident-20220324-max-ethusdt-orderbook-incorrect along with 43 others.



**Trouble Solver** APP 11:37 PM

 Support Commands:

- `/119 update` can update incident attributes.
- `/119 status` can check current incident status.
- `/119 history` will display all incident history.

**\*\*Notes:** These commands are only available in incident channels.

 Checkout MAX/MaiCoin emergency maintenance SOP if needed: [SOP Docs](#)

[s0] max-ethusdt-orderbook-incorrect

**Stage:** Init incident

**Roles:**

IC (Incident Commander): @smalltown TL (Tech Lead): Unassigned

CL (Communications Lead): Unassigned EM (Engineering Manager): Unassigned

**Status:**

ETHUSDT orderbook 不正確

↓ Latest messages

# Incident Visibility - External

- Service Health Page
- Mobile App Push Notification
- Facebook, Telegram, Twitter... Customer/Vendor Communication Channel





**Organization Type**



**On Call**



**Incident Response**



**Monitoring**



**Root Cause Analysis**



# What is Root Cause Analysis (RCA)?

- A Systematic Process for Identifying “Root Causes” of Problems or Events and an Approach for Responding to Them
  - What Happened
  - How it Happened
  - Why it Happened...so
  - Actions for Preventing Reoccurrence are Developed



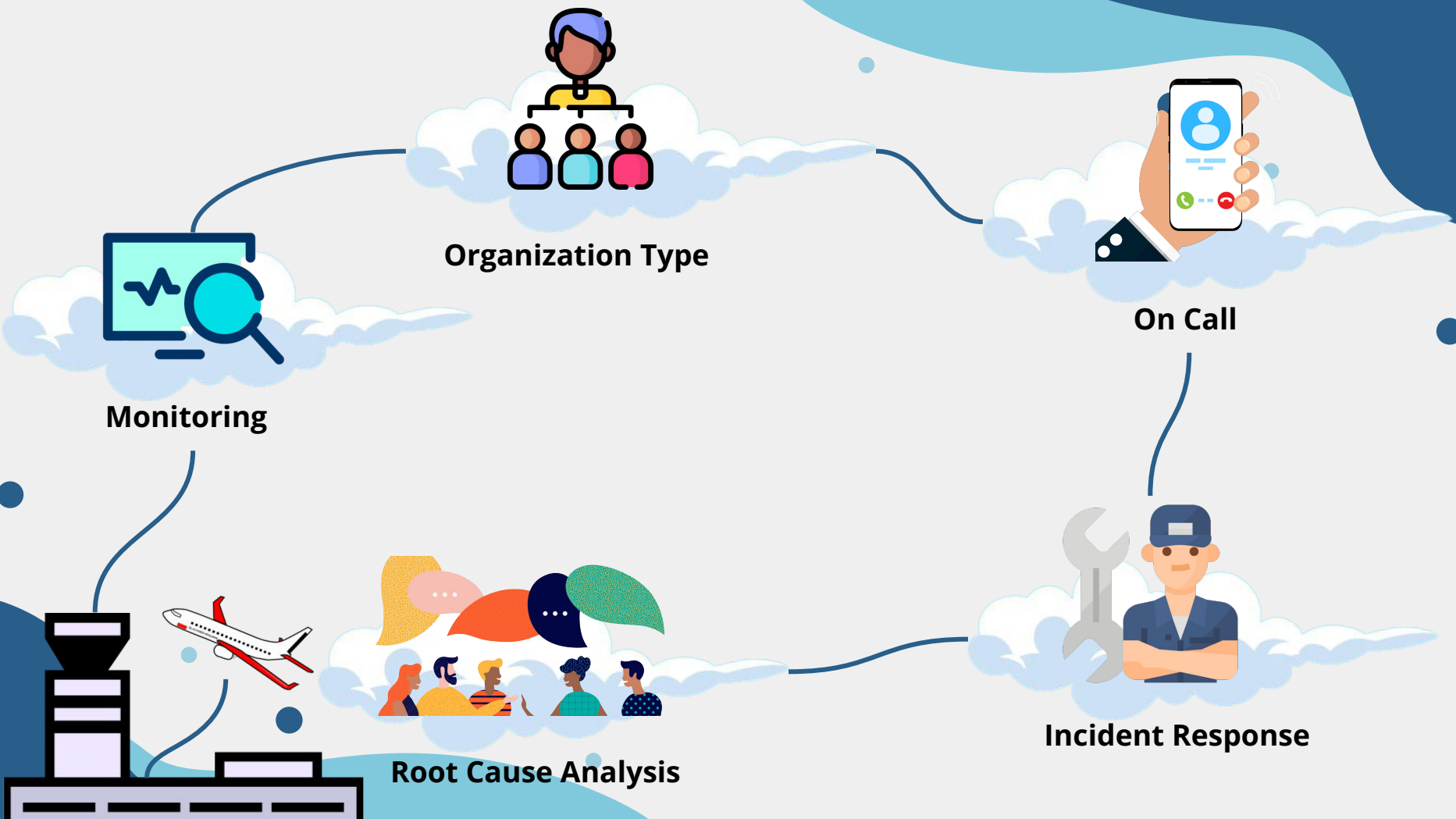


# How to Prepare RCA?

- Incident Timeline
  - 2022/04/29 16:00 Receive Alert from Prometheus
  - 2022/04/29 16:03 Start Incident Response
  - ...
- Findings/Root Cause
  - The HTTP Status Code Return 400
  - Finding TLS Certificate Expired
- Follow-up/Corrective Action
  - Monitoring All TLS Certificate **[Ticket: ID-1234][Owner: smalltown][Status: On-Going]**



# Blameless Culture



**Monitoring**



**Organization Type**



**On Call**



**Incident Response**



**Root Cause Analysis**

# THANKS!

**ANY QUESTIONS?**

**You can find me at my office:**

- MicroService Engineer
- Backend Engineer
- Frontend Engineer
- ...

