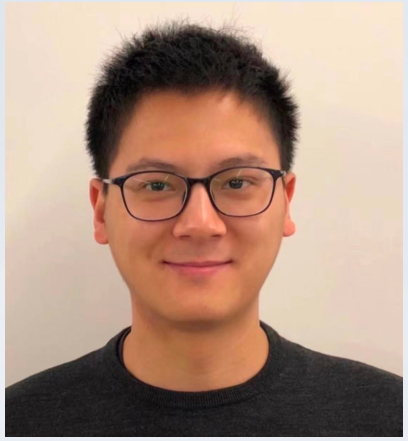
携程云平台基础设施变更管理实践

2024年4月

SREelite

https://sre-elite.com

个人介绍



刘芽

工作经历:

2015 年——至今,携程

个人简介:

- 2015年-2020年，参与并负责携程编译打包平台/GitlabCI
- 2021年，负责携程云平台基础设施运维

目录

CONTENT

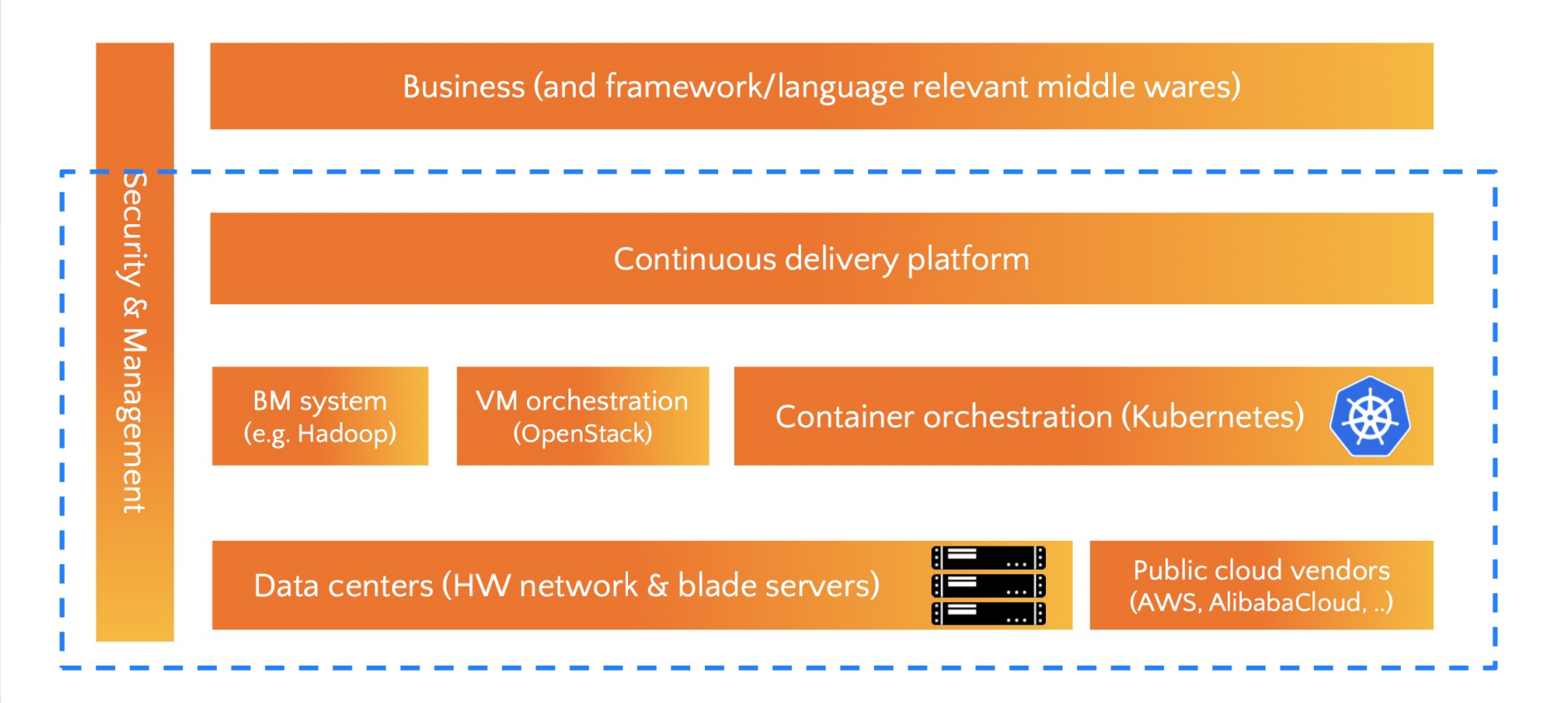
SREelite

01

携程云平台概览

SREelite

携程云平台概况



携程云平台 – Trip.com Cloud Platform

携程云平台概况

SREelite

计算资源

- Kubernetes
  - 20+ Clusters
  - < 5k nodes per Cluster
  - < 20w pods per Cluster
- OpenStack -> KubeVirt
  - 30k+ VM
  - 10k+ BM
- CDOS
  - Resource API

云原生网络

- OpenStack Neutron
  - VM IPAM
  - BM IPAM
- Cilium & EBPF
  - Container Networking
  - Cloud Native Security

分布式存储

- Object Storage
  - Ceph
  - S3
  - OSS
- Block Storage
  - Ceph rbd
  - Ebs
  - CSI
- FileSystem
  - JuiceFS

02

基础设施变更管理原则

基础设施变更管理原则

SREelite

导致变更失败的主要因素

- 人为因素
- 流程因素

- 我们把问题归为人的因素时，并不仅指某人有意的行动
- 我们定义的人为因素：没有按照规定以及标准操作流程（SOP）而导致的重大故障

三步法则

- 我是这个工作的合适人选吗？
- 我有能力执行这个任务吗？
- 我能把控整个任务吗？

SREelite

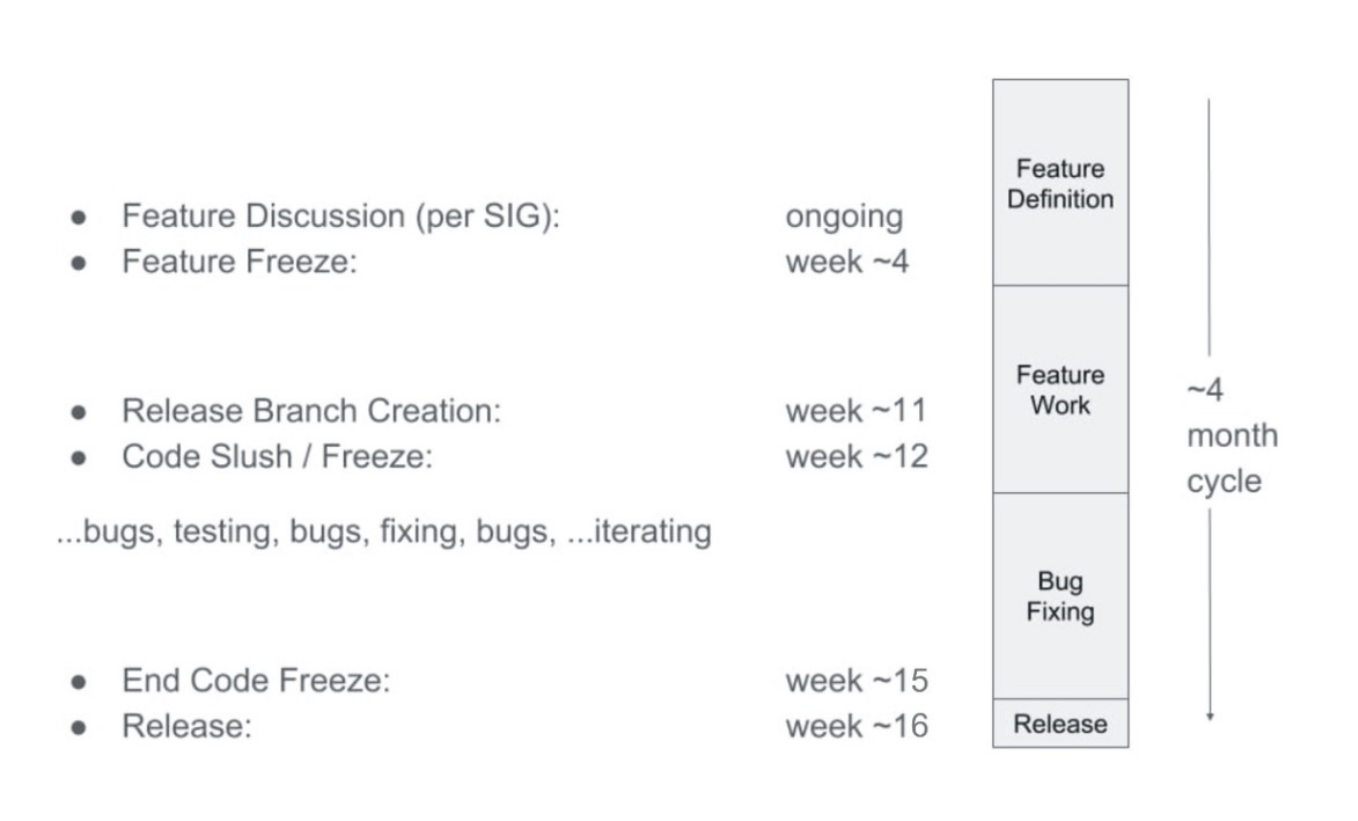
基础设施变更管理原则

- 计划性
- 灰度原则
- 可回滚原则
- 杜绝循环依赖
- 建立风险评估机制

基础设施变更管理原则 - 计划性

SREelite

- 与普通应用的变更相比，基础设施变更应该严格按照计划执行
- 典型例子: Kubernetes 4month 发布模式，一年3个固定版本



基础设施变更管理原则 - 灰度原则

SREelite

- 严格按照 Availability Zone (AZ) 定义，控制变更的灰度
- 不同 AZ 的服务，不共享一套基础设施
- 控制基础设施变更的爆炸半价
- 生产不同 AZ 的基础设施变更，时间上必须间隔 1 天
- “慢即是快”

基础设施变更管理原则 - 可回滚原则

SREelite

- 基础设施变更比较复杂，工程师容易忽略回滚策略
- 测试环境需要验证回滚步骤是否可执行
- 部分极端情况，可以将新建集群作为回滚方案
  - 依赖快速构建环境的能力+数据备份+配置备份
- 典型的回滚策略
  - 代码和配置文件都需要版本管理，杜绝仅回滚镜像
  - 引入中间版本
  - 谨慎删除旧字段

基础设施变更管理原则 - 杜绝循环依赖

SREelite

- 核心基础设施间的依赖需要定期梳理
- 核心基础设施自愈能力
- 网络&虚拟化控制器，做到无外部依赖自启动
- 容易忽略的依赖项:
  - DNS
  - 分布式存储

基础设施变更管理原则 - 建立风险评估机制

SREelite

- 梳理基础设施管理服务，定义 P0 / P1 / P2服务标准
- 针对 P0业务，建立定期风险 review 机制，review 内容包括：
  - 故障爆炸半径
  - 业务架构
  - 业务部署规范
  - 业务故障预案
- 针对 P0 业务，进行故障演练，验证是否满足服务 P0 级定义，并建立常态化演练机制

03

基础设施变更管理实践

SREelite

https://sre-elite.com

基础设施变更计划 – 日常运维变更实施方法

SREelite

云组件部署机器调整、底层服务参数调优、DR调整、域名及访问入口调整、备份机制优化、日志轮转调整、内核参数变更等 IAAS 组件日常运维管理相关变更，适用于如下变更流程及变更原则:

1）生产环境计划内变更，需要在每周四的周会前提交 Release Plan，周会上review 变更计划。采用一票否决制，有任何不确定因素及反对意见的，变更搁置

2）变更计划需要有明确的回退步骤

3）对生产环境应用有一定概率产生网络中断等影响的变更，需要安排outage窗口、通过评审后实施

4）对发布系统有较大概率产生影响时间大于 15 分钟的变更，提前一天做好用户通知报备

基础设施变更模版

SREelite

- 变更计划由变更委员会 Review
- 关联 Review 通过且 Merged 状态的 Merge Request
- 详细的 Rollout & Rollback Plan
- 需要在发布窗口期内变更
- 实时通知 NOC 变更状态

@sys-ccb （注意：部分项目是 @ sys-dp）

产品

- 填写变更的项目名称（必填）

Changelog

- 填写变更内容（必填）

merge_request

- 填写 merge_requests 链接 （必填）

Migrations

- 根据是否需要做 migration 填写

关联系统

- 影响到的其它系统

发布环境

- 填写发布的环境

Rollout Plan

- 填写发布操作步骤 （必填）

Rollback Plan

- 填写回退步骤 （必填）

发布窗口

- 填写计划发布的时间

实际开始日期

- 填写实际开始发布的日期（必填）

实际结束日期

- 填写实际结束发布的日期（发布周期大于一天时，必填）

预计发布时间

- 20 分钟

发布状态

- [ ] 发布中
- [ ] 成功
- [ ] 回退

配置一致性巡检

- 配置不一致是万恶之源
- 变更工艺确保配置一致 & 巡检工具收口

基础设施变更计划 – 基础设施大版本升级

SREelite

- Cilium v1.11 升级 v1.12：差异分析，测试环境验证，灰度发布生产

升级注意事项

https://docs.cilium.io/en/v1.12/operations/upgrade/#id3

1. cilium-operator/cilium-agent 默认 Prometheus metrics 端口号发生了变化

- cilium-agent 9090->9962
- cilium-operator 6942->9963

2. 在 Azure IPAM 模式下，--azure-use-primary-address 默认是关闭的，即不使用 ENI primary ip，aws --aws-use-primary-address 也是一样的，需要注意开启

3. 公有云 operator 能够支持通过 instance-tags-filter 配置过滤 EC2 实例

4. cilium-operator pprof 相关的配置进行了重命名

- pprof -> operator-pprof
- pprof-port -> operator-pprof-port

5. kube-proxy-replacement 配置里的 probe 选项被标记成废弃，将会在 1.13 中移除

- https://docs.cilium.io/en/v1.12/gettingstarted/kubeproxy-free/#kube-proxy-hybrid-modes

- ETCD 版本升级 , v3.3 -> v3.4

- K8S Apiserver 版本升级, v1.13 -> v1.17 -> v1.19

IaC 实践: StackStorm & SaltStack

SREelite

SaltStack 管理云平台基础设施服务

- SaltStack Formula & Pillar 机制，标准化配置和敏感信息分离
- 不同 Region 不共用 SaltStack Master，缩小变更操作的爆炸半径
- Cilium Agent 不依赖 K8S DaemonSet, 用 salt + docker-compose 管理 Cilium 服务

StackStorm 固化操作流程

- 使用 StackStorm 提供的事件驱动和 runbook 技术栈，实践 IaC
- Chatbot & Stackstorm工作流对接
- 机器配置更新依赖 SaltStack 提供的能力
- 对 Stackstorm 执行情况进行深入分析，持续优化运维变更的 workflow

IaC 实践: 云基础组件变更管理 GItOps

SREelite

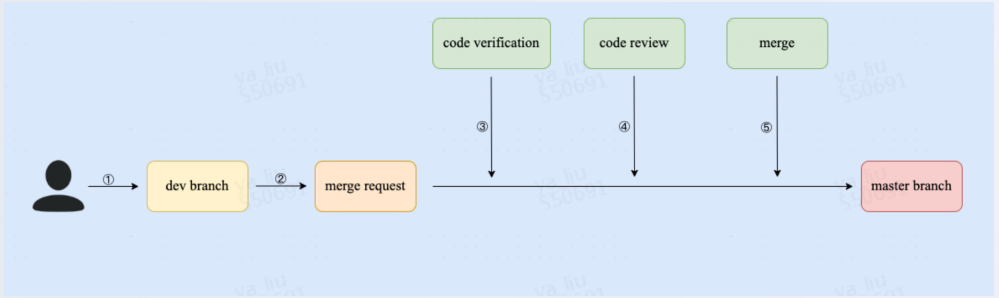
- Kustomize 来管理多个集群和环境的基础组件 Yaml 配置文件
- Git Workflow（Tag/ MergeRequest）进行版本控制
- 基于 GitlabCI，CI/CD Pipeline 将基础组件部署到集群中

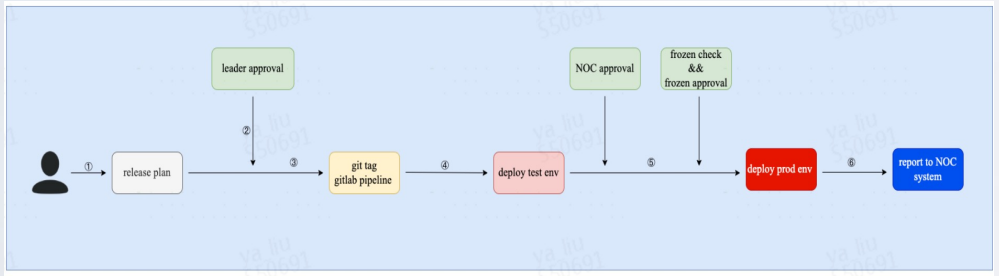
Layout structure

The layout of configs as below

```
- base
-   | appnameA
-   | appnameB
-   | appnameC
-   ...
- overlays
-   | clusterA
-       | appnameA
-       | appnameB
-   | clusterB
-       | appnameA
-       | appnameB
-       | appnameC
-   | clusterC
-       | appnameC
-   ...
```

代码分层结构





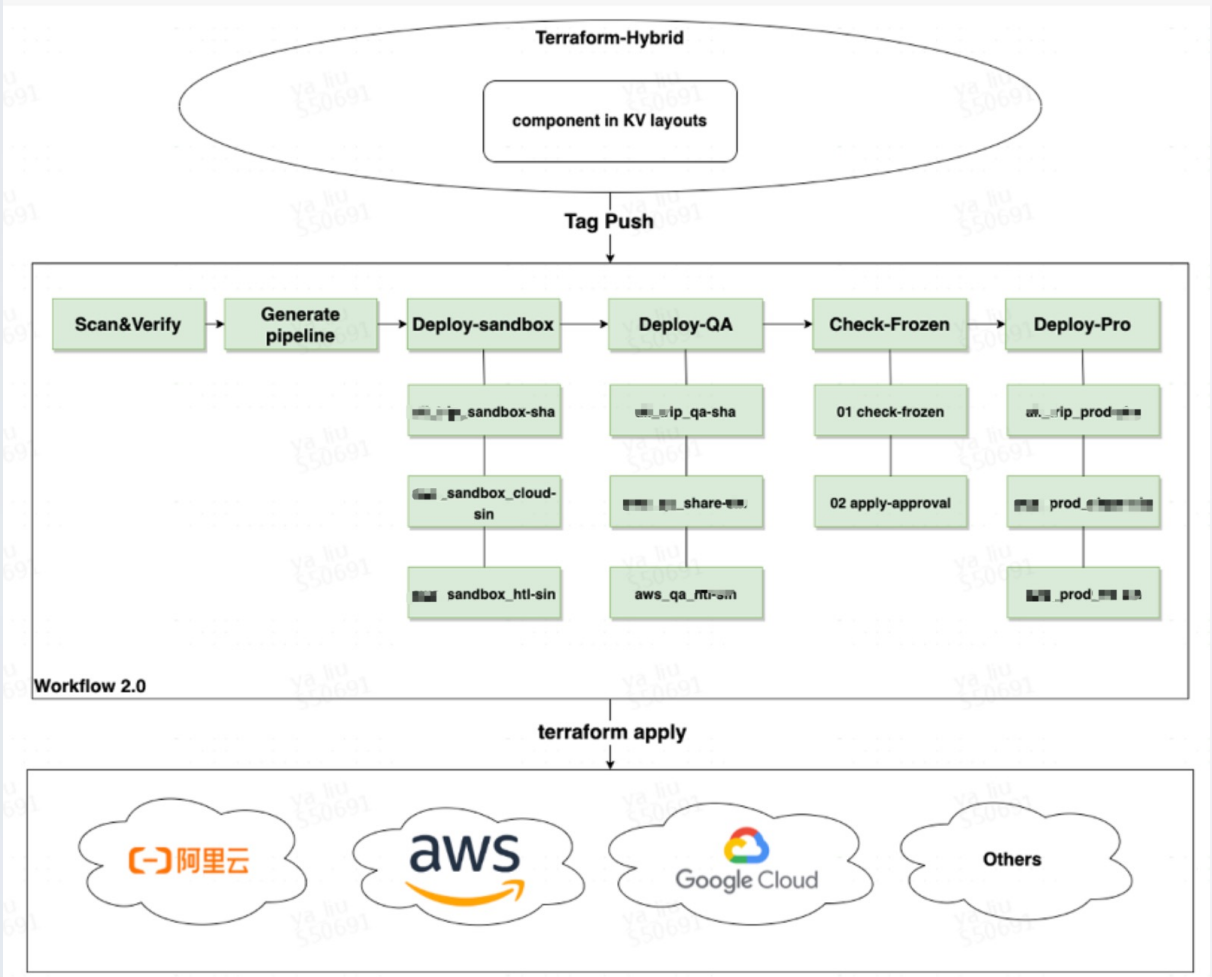
CI/CD 流程

IaC 实践: Terraform

SREelite

```
├── DOMAIN_OWNERS
├── config
│   ├── ali.yaml
│   ├── aws.yaml
│   └── gcp.yaml
├── deploy
│   └── provider
│       ├── aws
│       │   ├── component
│       │   │   └── component1
│       │   └── account
│       │       └── account1
│       │           ├── component
│       │           │   ├── component1
│       │           │   └── component2
│       │           └── region
│       │               ├── region1
│       │               │   └── component
│       │               │       ├── component1
│       │               │       └── component2
│       │               ├── region2
│       │               │   └── component
│       │               │       └── component1
│       │               └── region3
│       │                   └── az
│       │                       ├── az1
│       │                       │   └── component
│       │                       │       └── component1
│       │                       └── az2
│
├── modules
│   └── provider
│       └── aws
│           └── module1
```

Terraform 目录结构

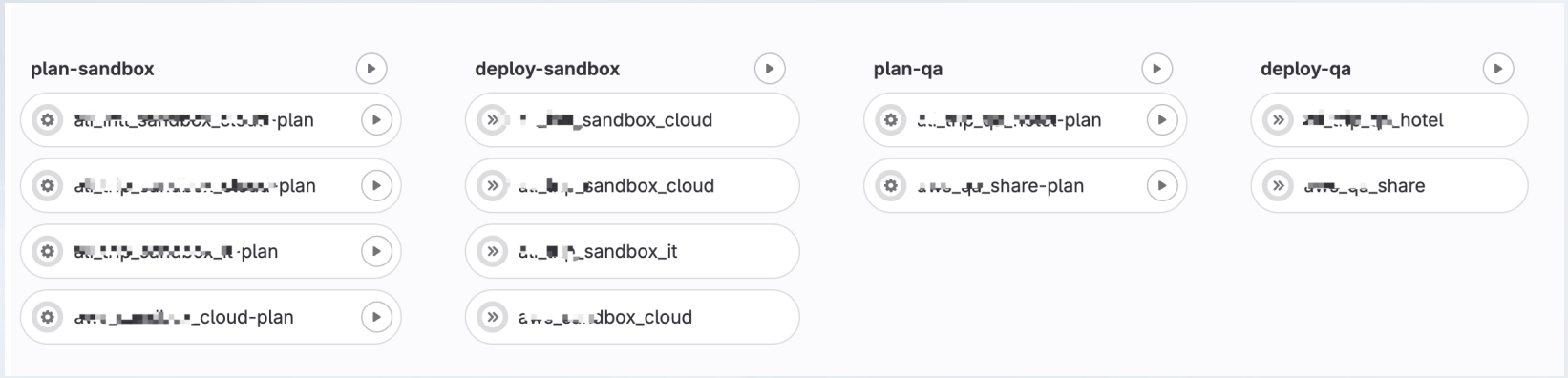


Terraform 部署流程

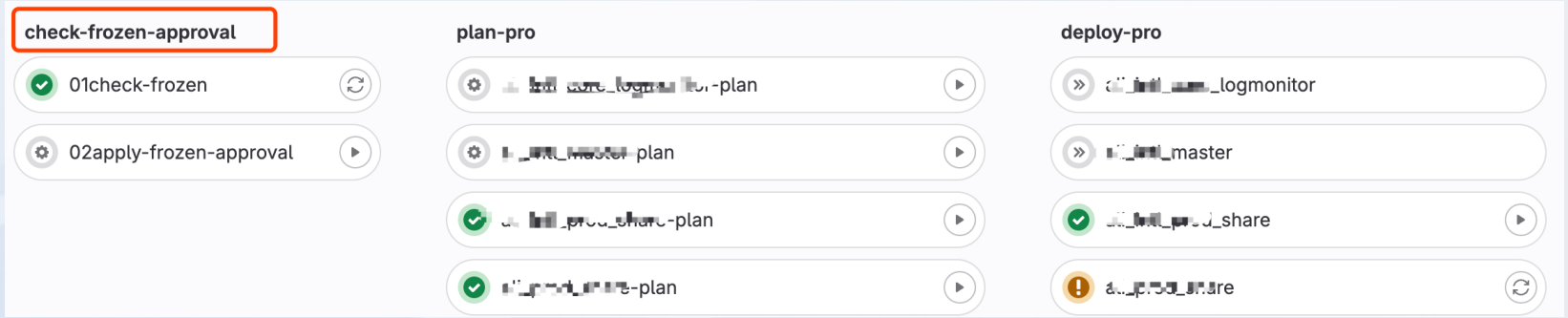
Terraform Release Pipeline

SREelite

Code Check



Deploy QA



Deploy PRO

04

基础设施变更管理思考及优化方向

思考及优化方向

SREelite

- 切记灰度，切忌不做回滚方案，面向灾难设计架构
- AI + ChatBot 可以承担更多的日常工作，对于变更可能造成的风险进行智能化提示
- 对于核心组件定期 Review，主动升级，避免跨多个大版本的升级操作
- 不畏惧基础设施变更的前提:
  - 具备持续变更能力
  - 对于基础设施有深入的研究和理解
  - 实践出真知

Q&A

SREelite

https://sre-elite.com

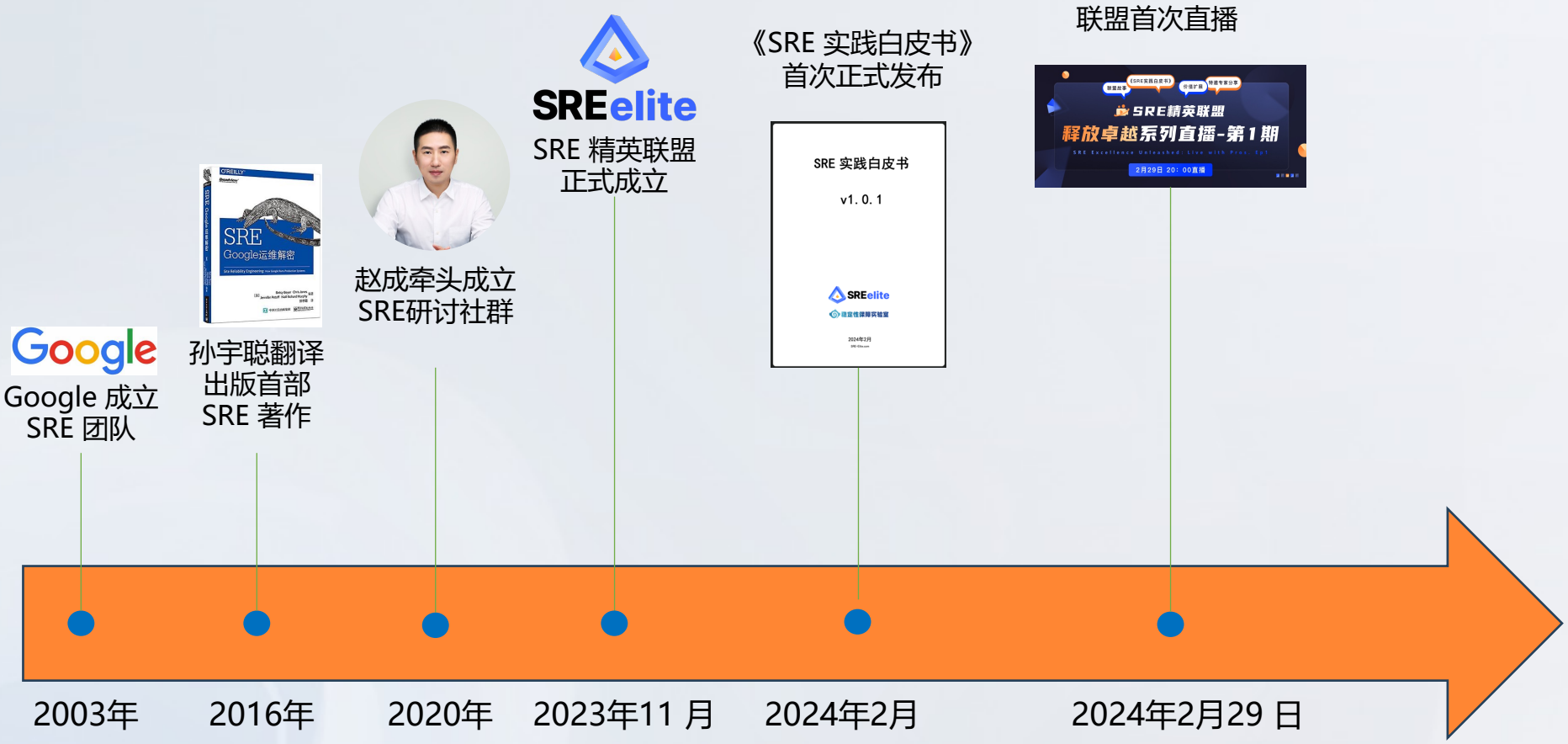
附录

供参考

https://sre-elite.com

“SRE精英联盟”概述

SREelite



《SRE实践白皮书》

SREelite

- 经历数年，20 多位一线专家协作编写。
- 扫码下载 v1.0.1。版本持续更新迭代中。
- 在官网 https://sre-elite.com/notice/下载最新版。

联系我们

SREelite

公众号

视频号

B 站

YouTube