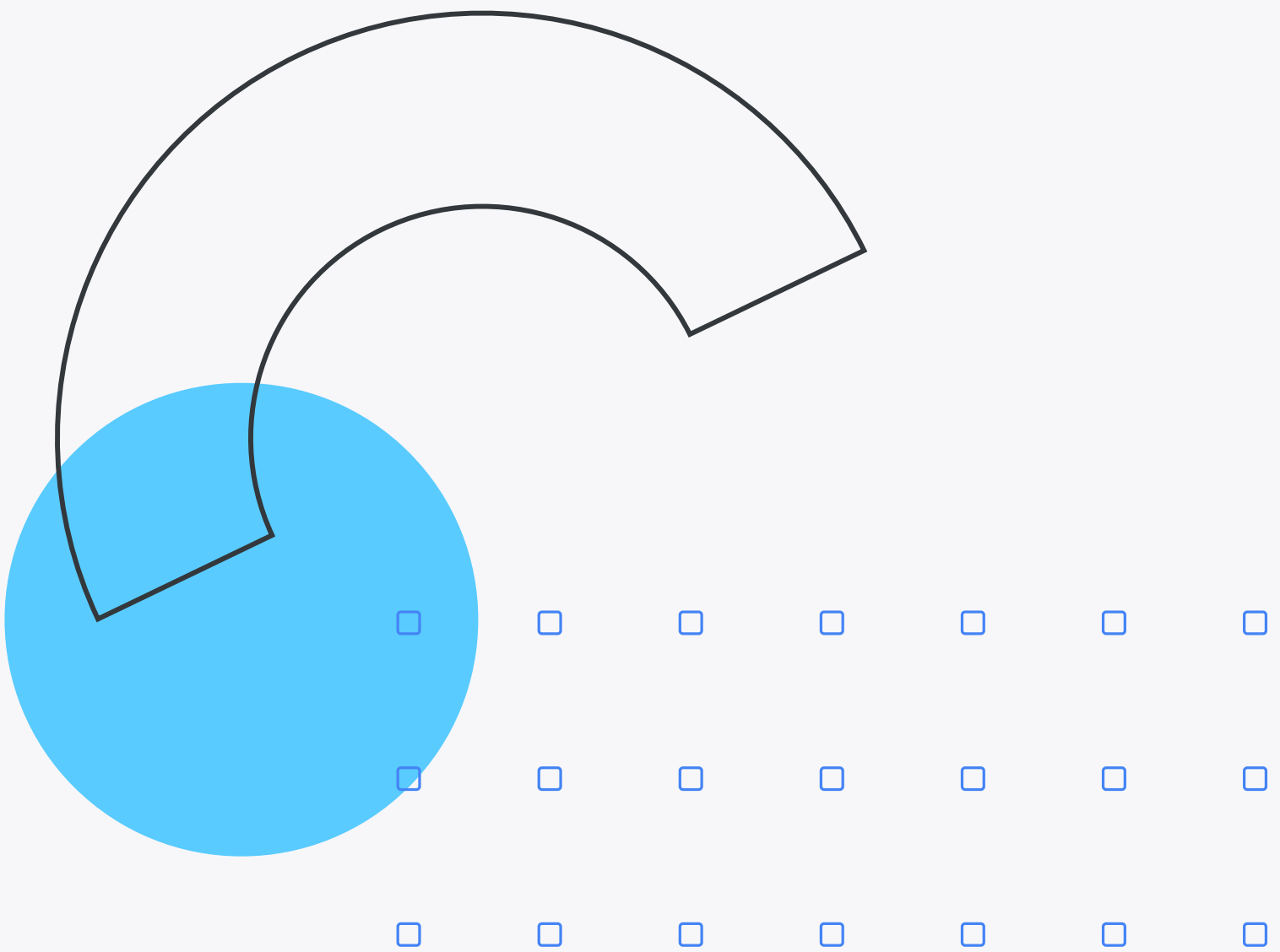


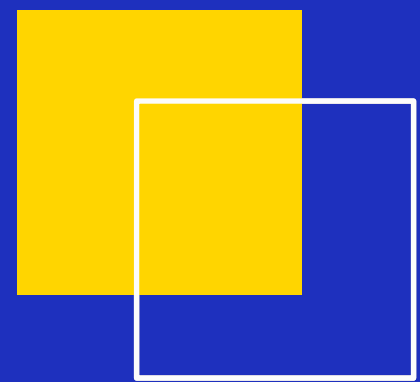
Databend

# Security on Cloud Data Warehouse

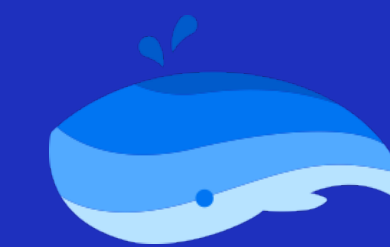
主讲人：李亚舟

2022.10



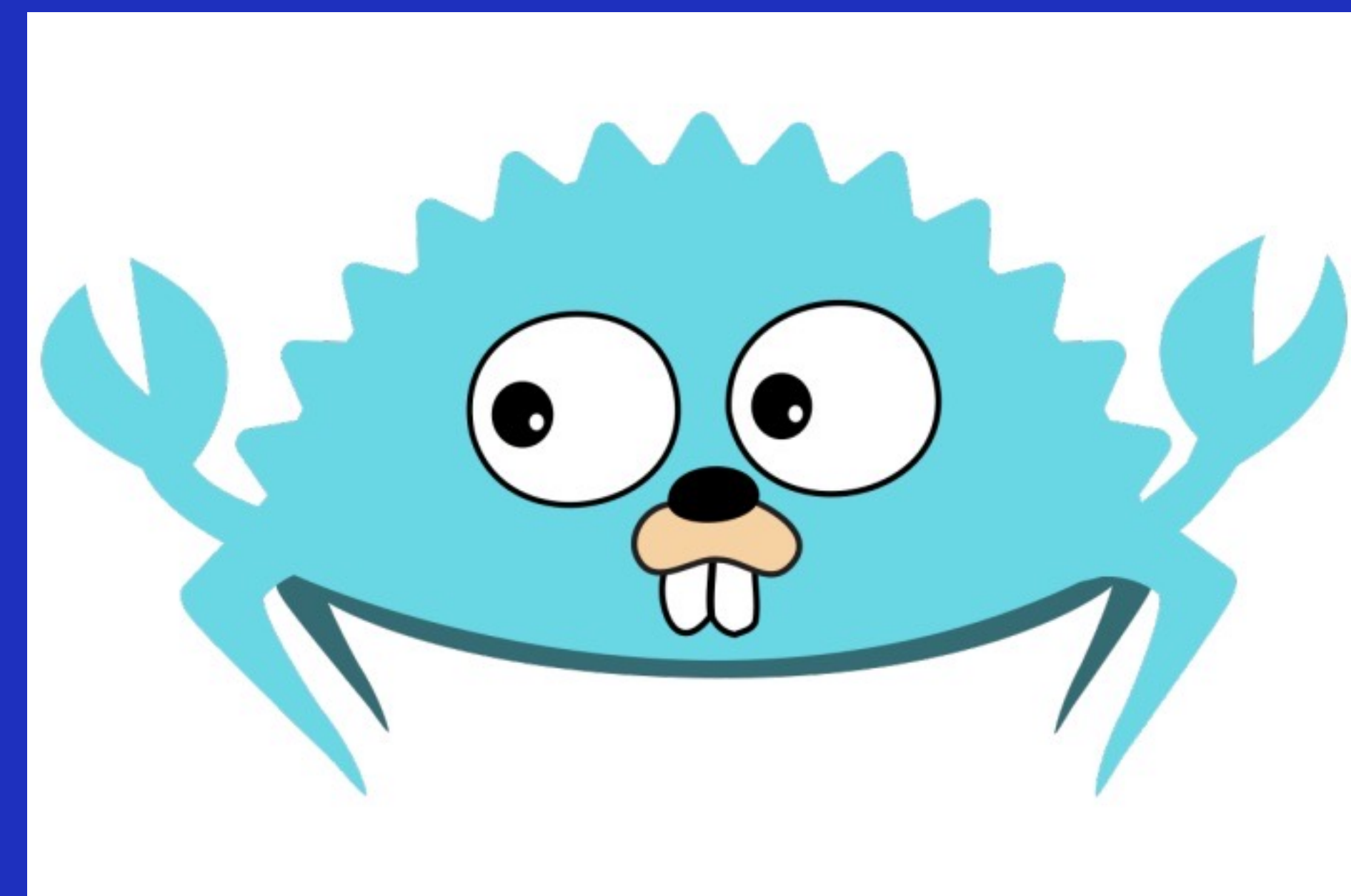


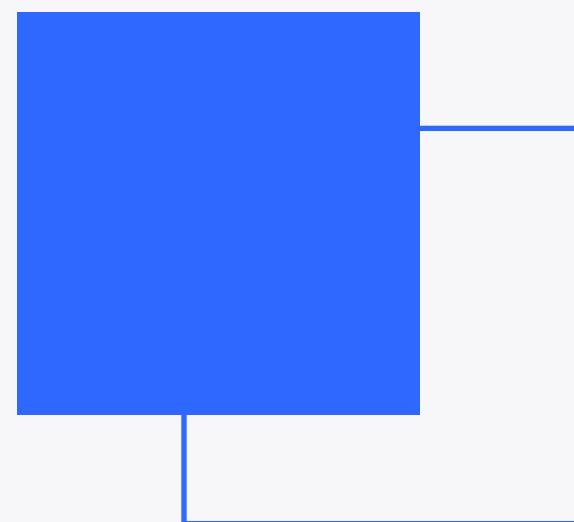
# 李亚舟



Databend

- Databend Cloud 技术负责人
- 曾负责猿辅导容器云架构、知乎数据库架构

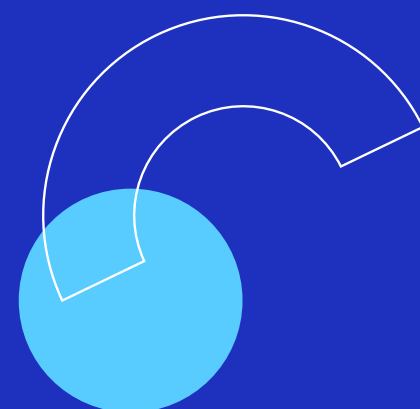
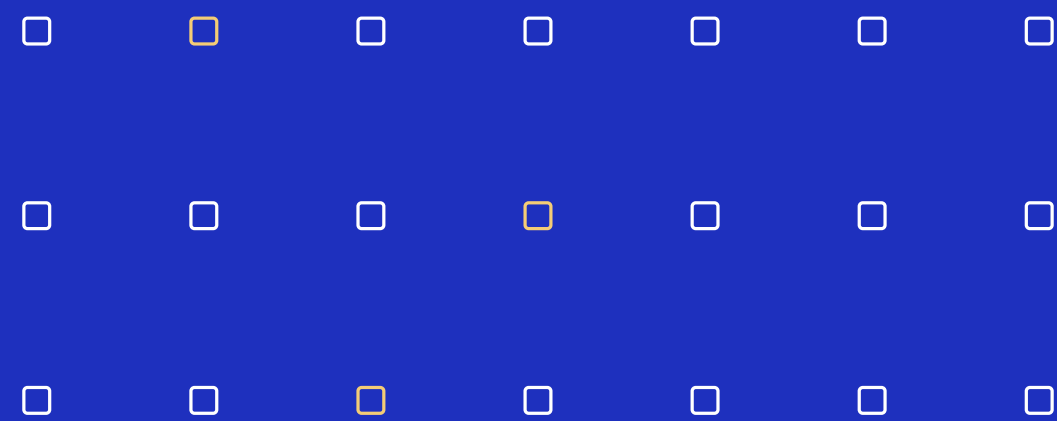




# 目录

CONTENTS

- 什么是云数仓
- 云数仓的安全挑战
- 云上安全 101
- 云数仓的安全实践



# 什么是云数仓

# 什么是云数仓



弹性

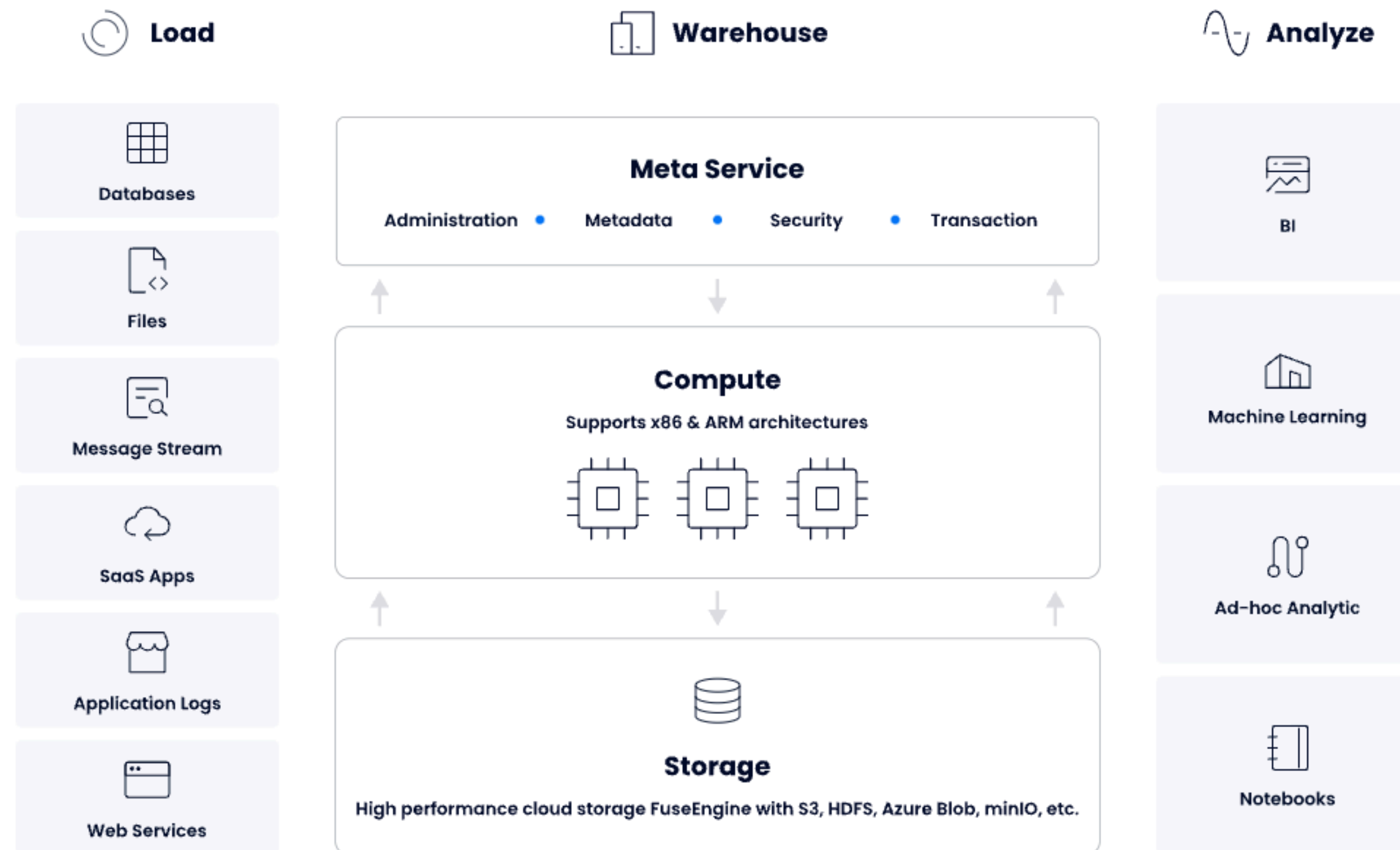
按需使用

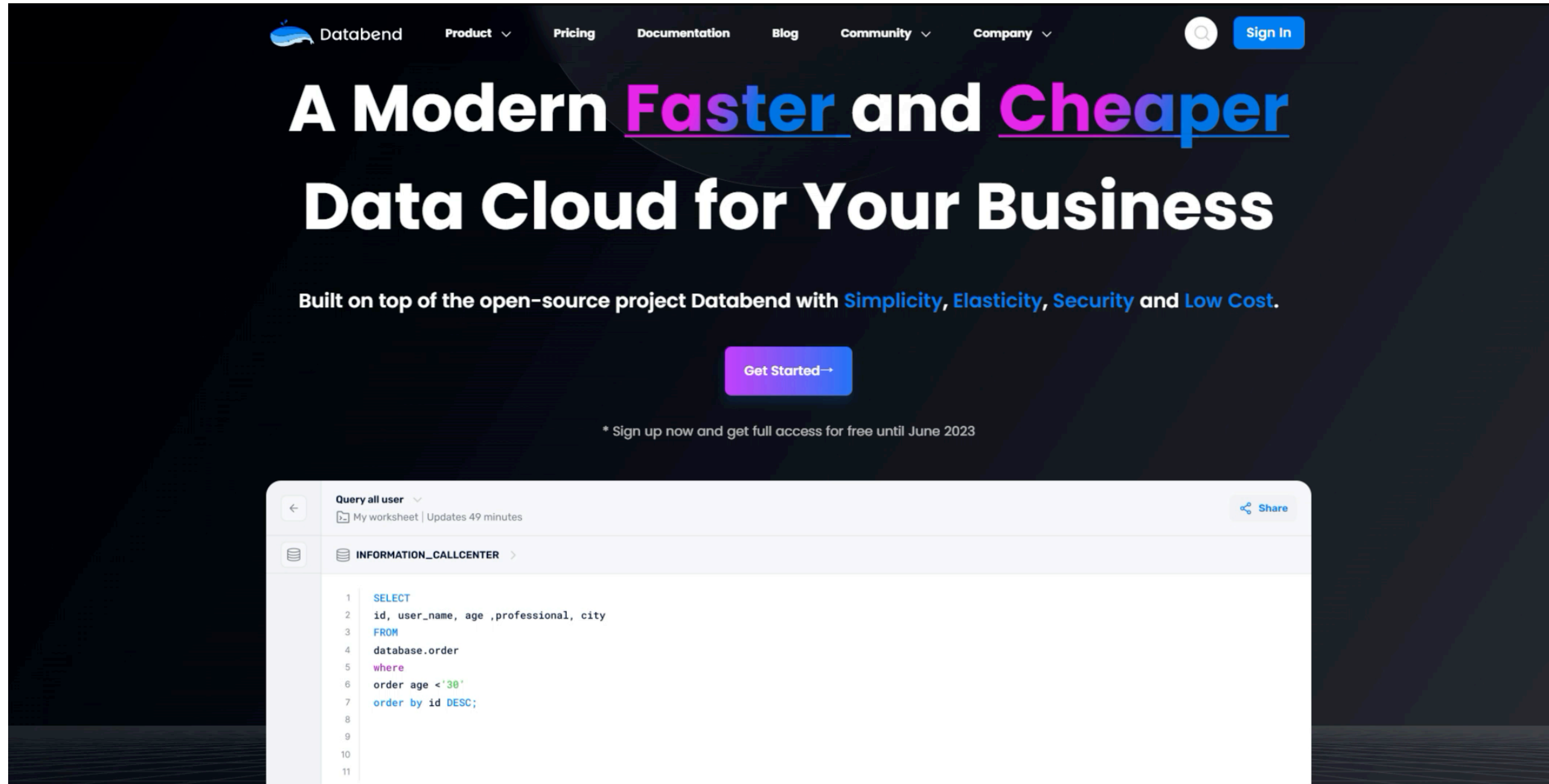
低成本

零运维

安全可靠

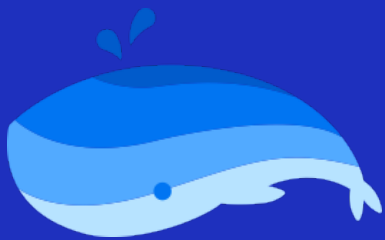
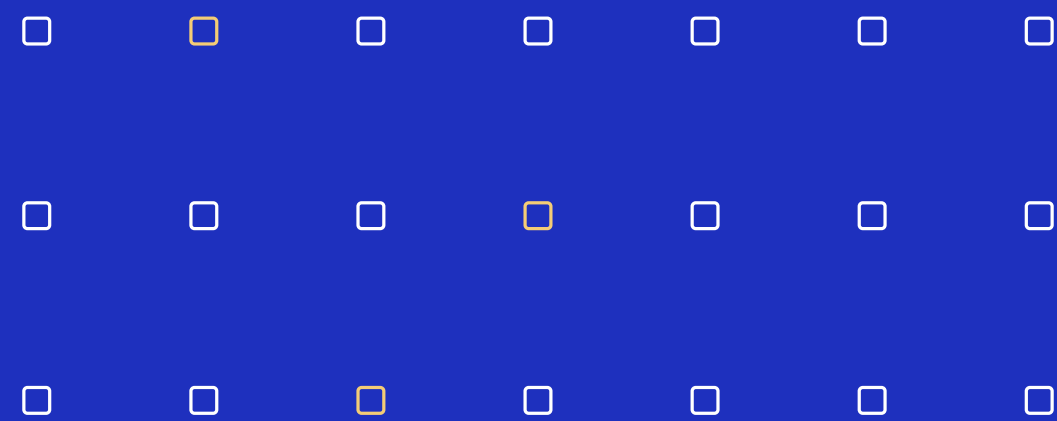
生态集成



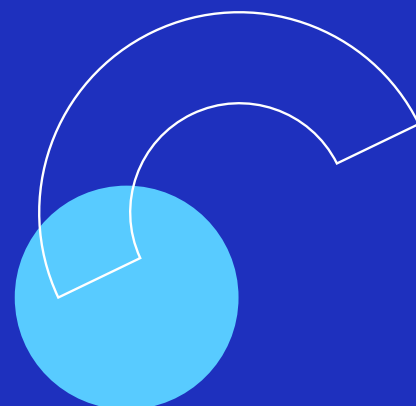


The screenshot shows the Databend Cloud website. The header includes the Databend logo, navigation links for Product, Pricing, Documentation, Blog, Community, and Company, a search icon, and a Sign In button. The main headline reads "A Modern Faster and Cheaper Data Cloud for Your Business". Below this, it states "Built on top of the open-source project Databend with Simplicity, Elasticity, Security and Low Cost." and features a "Get Started" button. A note at the bottom of the main section says "\* Sign up now and get full access for free until June 2023". The bottom of the screenshot shows a code editor interface with a SQL query: 

```
1 SELECT
2 id, user_name, age ,professional, city
3 FROM
4 database.order
5 where
6 order age <'30'
7 order by id DESC;
```



Databend



# 云数仓的安全挑战

## 数据隔离、数据加密

- 每个租户不可以访问其他租户的数据
- 所有落地的数据必须经过加密

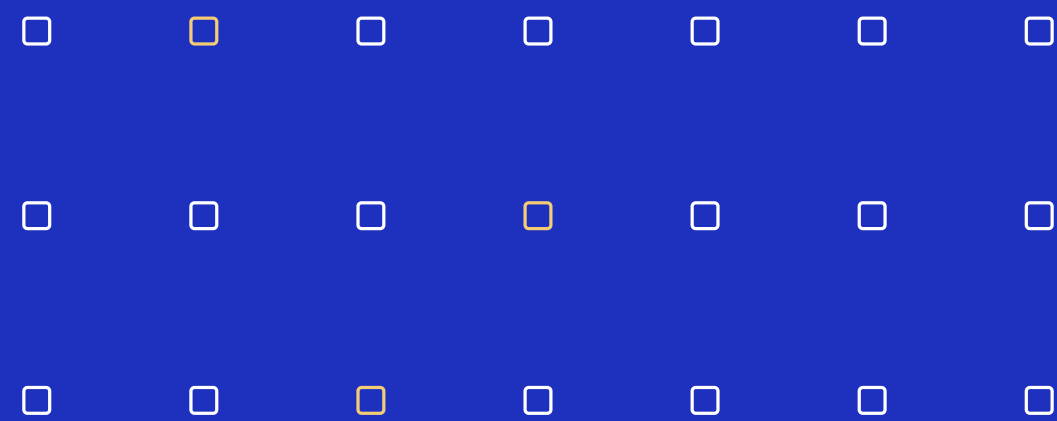
## 证书管理、Token 轮换

- 能够自动管理证书与 Token 轮换
- 避免基础设施中存在长生命周期的 AccessKey/SecretKey

## 合规认证

- 使 Infra 所有操作有审计与审批





# 云上安全 101

# 云比自建更安全？

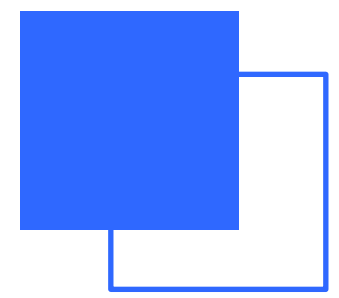


## 物理安全

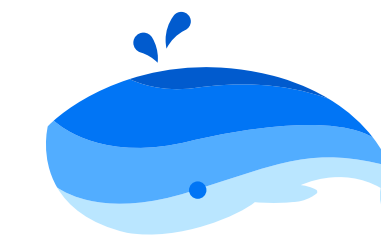
- 硬盘、内存条被人进机房偷了怎么办？

## 体系建设

- 建设审计、认证授权、KMS、加密等安全机制，可以站在云的肩膀上



# Shared Responsibility Model



Databend

## 云厂商负责云基础设施的安全性

- 包括设备的物理安全、确保不同用户之间的数据不会被其他租户访问到；

## 用户负责应用程序的安全

- 包括实现正确的访问控制、认证、加密逻辑



## 云上管理授权的中心

- 云上的所有资源都可以通过 IAM 进行访问控制
- 每条 IAM 规则大致上包含有 Resource、Action、Effect、Principal 几个参数

## 云上管理认证的中心

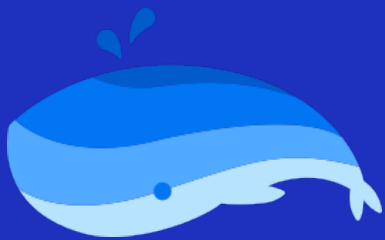
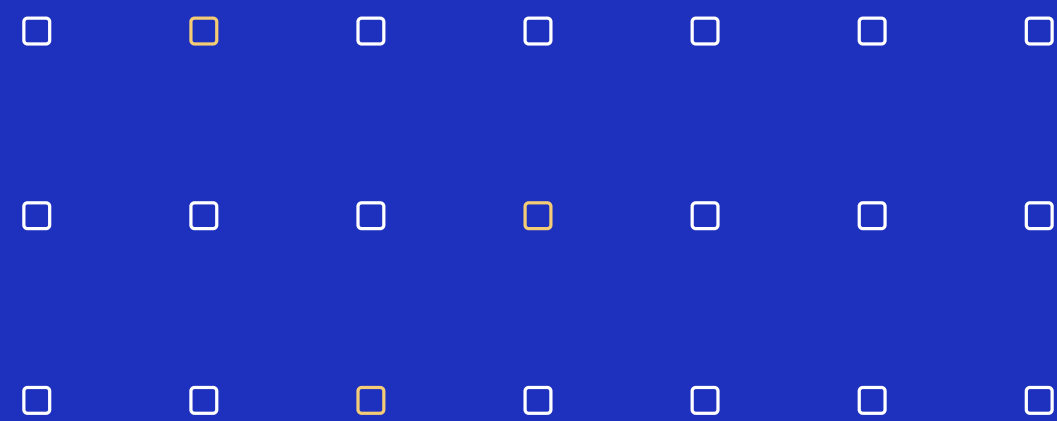
- STS 能够帮助认证的请求主体获取临时的 Security Token
- 应用程序可以根据这一 SecurityToken 去访问 AWS 的云服务
- 比如访问 s3 数据时，AWS 能够结合 IAM 策略，判断该身份是否允许读写特定的 S3 Object

## 帮助保存密钥

- 在数据加密、签名中，最担心的问题莫过于密钥泄露
- KMS 服务可以在认证身份后，通过外部不可见的密钥对数据进行加解密，全程无法得到密钥的明文

## 落地数据的加密

- 表示云厂商中的存储在落盘时需要经过加密，不然就存在有盘失窃时将数据窃走的风险
- 云上的所有数据都需要开启 Encrypt At Rest



Databend



# 云数仓的安全实践

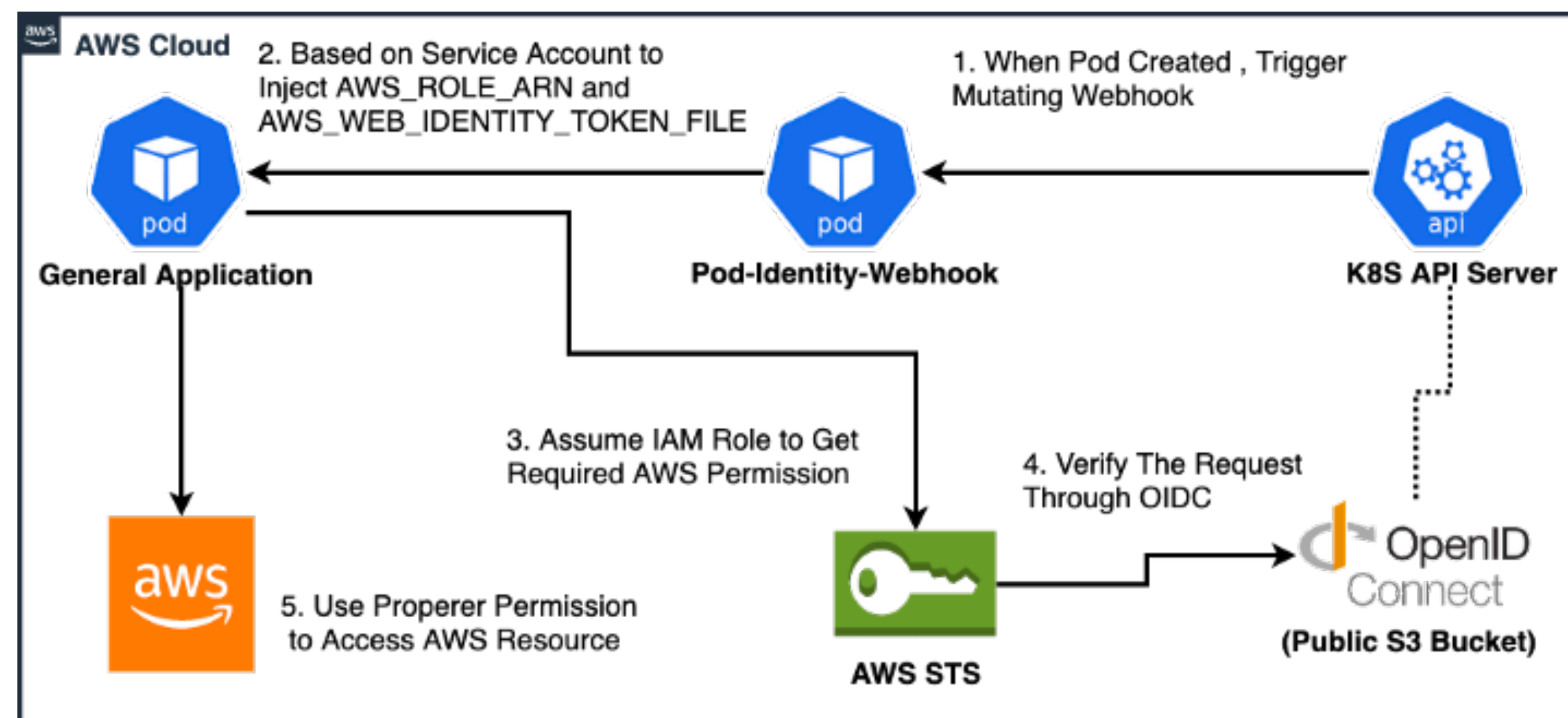


## Make an architecture secure by default

- Use an IaC tool (like Pulumi or Terraform)
- Use a Monorepo, with Github Action to automate every Infra changes
- Use an auto rotated identity token & certificates
- Use encrypt at rest
- Use the IAM policy with least privilege for each tenant
- Use a password-less infra access plane

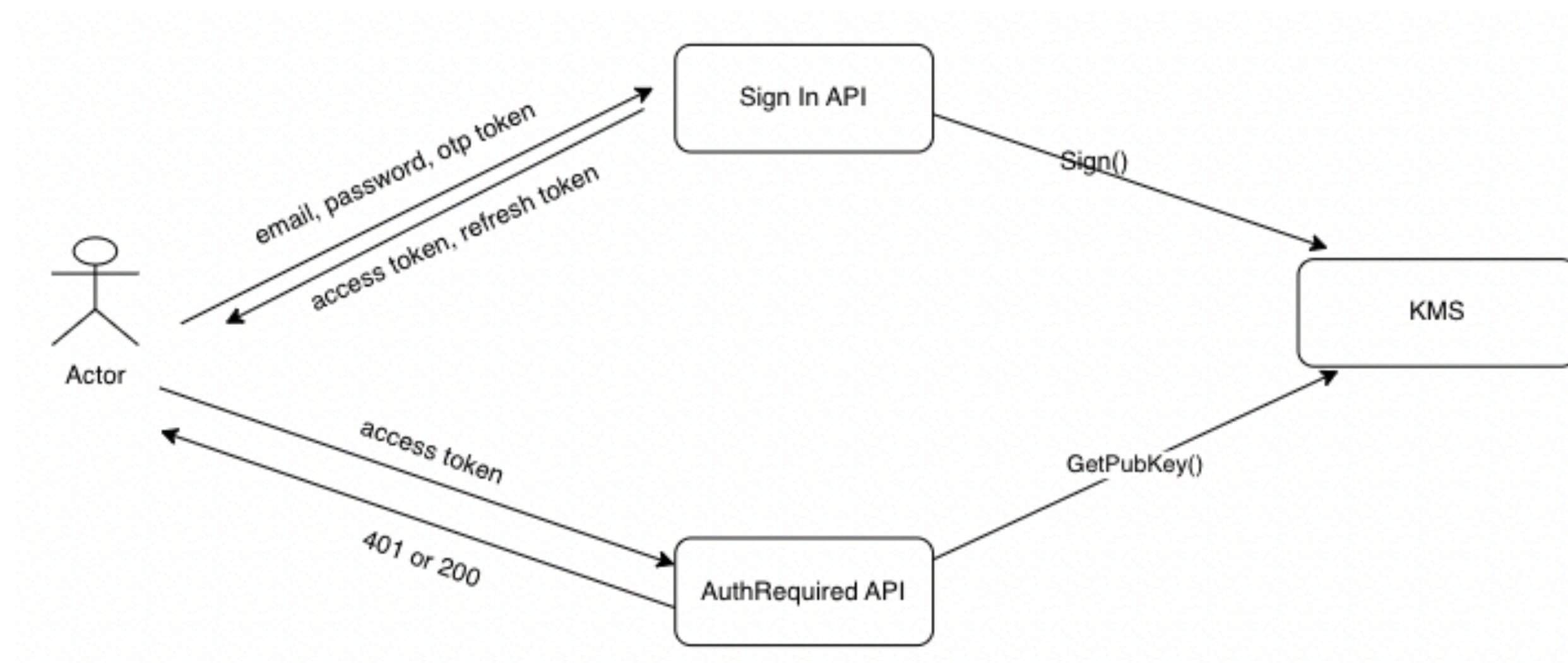
## 将 k8s 的 Service Account 与 AWS 的 Role 进行关联

- 配合 IAM 实现租户间的数据强隔离



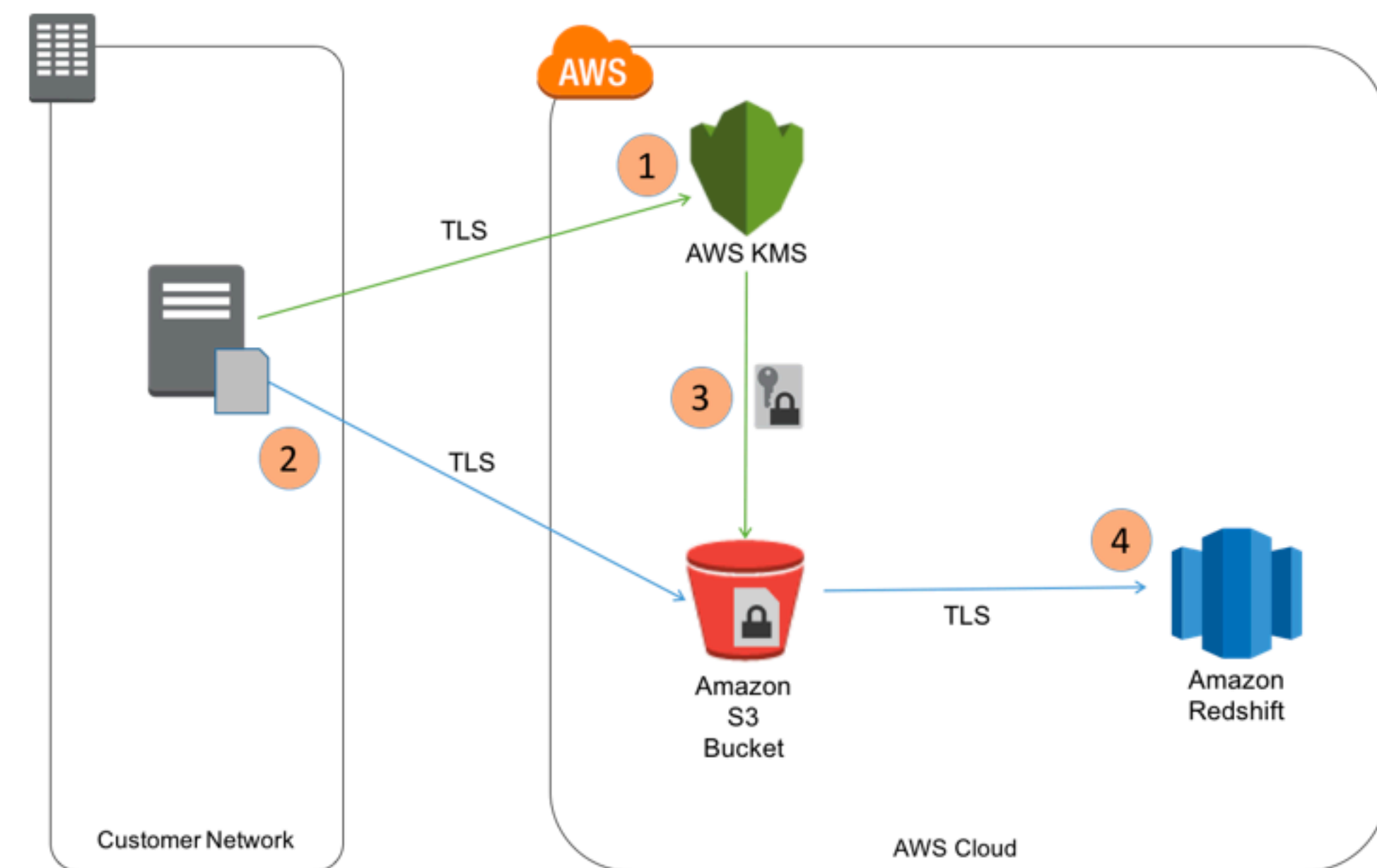
## 通过 KMS 对 JWT 进行签名

- 通过对外不可见的密钥进行签名
- 可以自动轮换密钥



## 服务端加密

- 通过 KMS 保证落入 s3 的所有数据均已进行加密

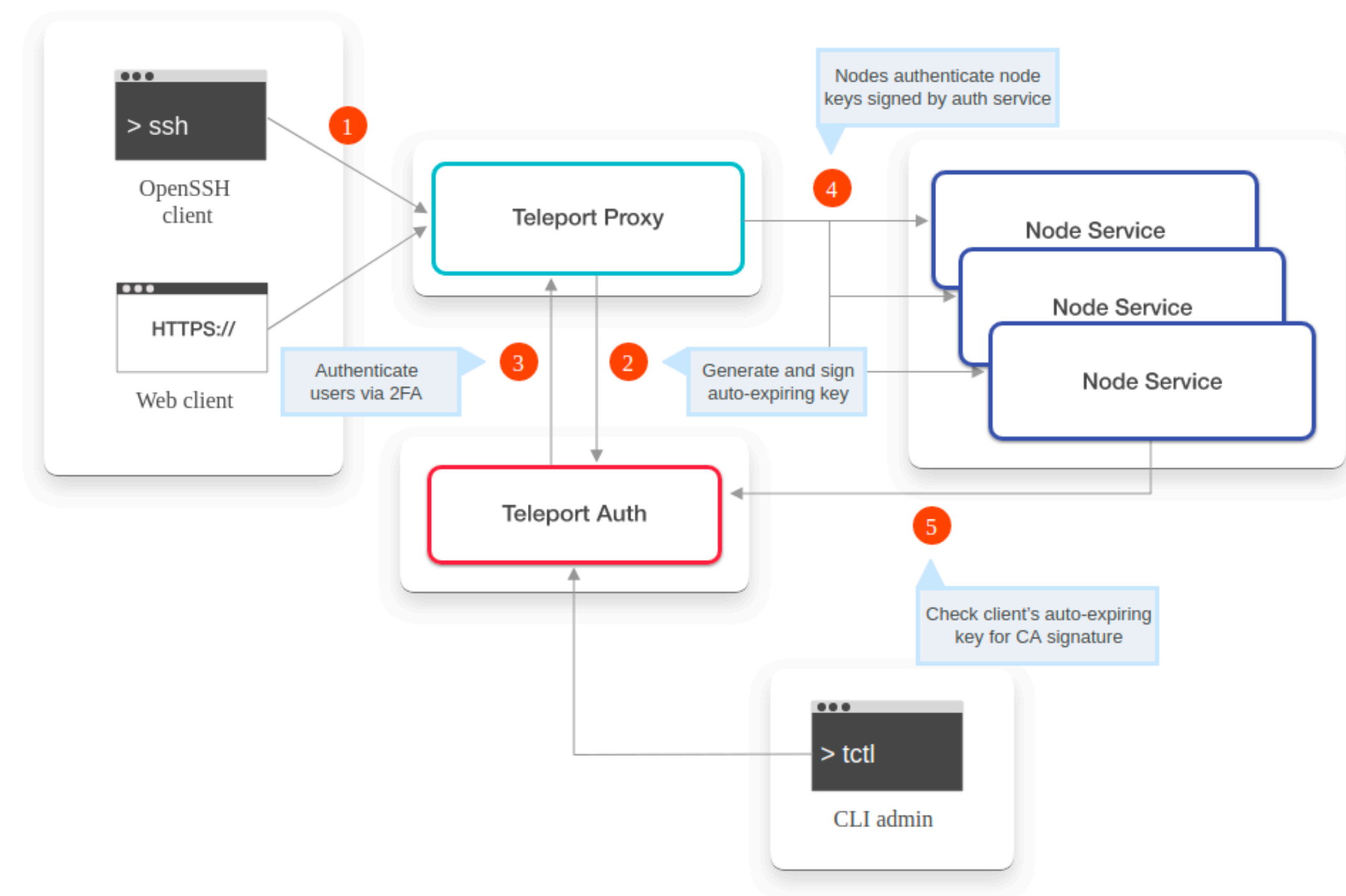


## 授权 Databend Cloud 访问自己的 Bucket

- CONNECTION = (AWS\_ROLE\_ARN=<role\_arn>)
- 通过 Assume Role, 允许 Databend Cloud 的 Warehouse 临时访问自己的云账号中的数据, 不需要保存 AK/SK

## Teleport

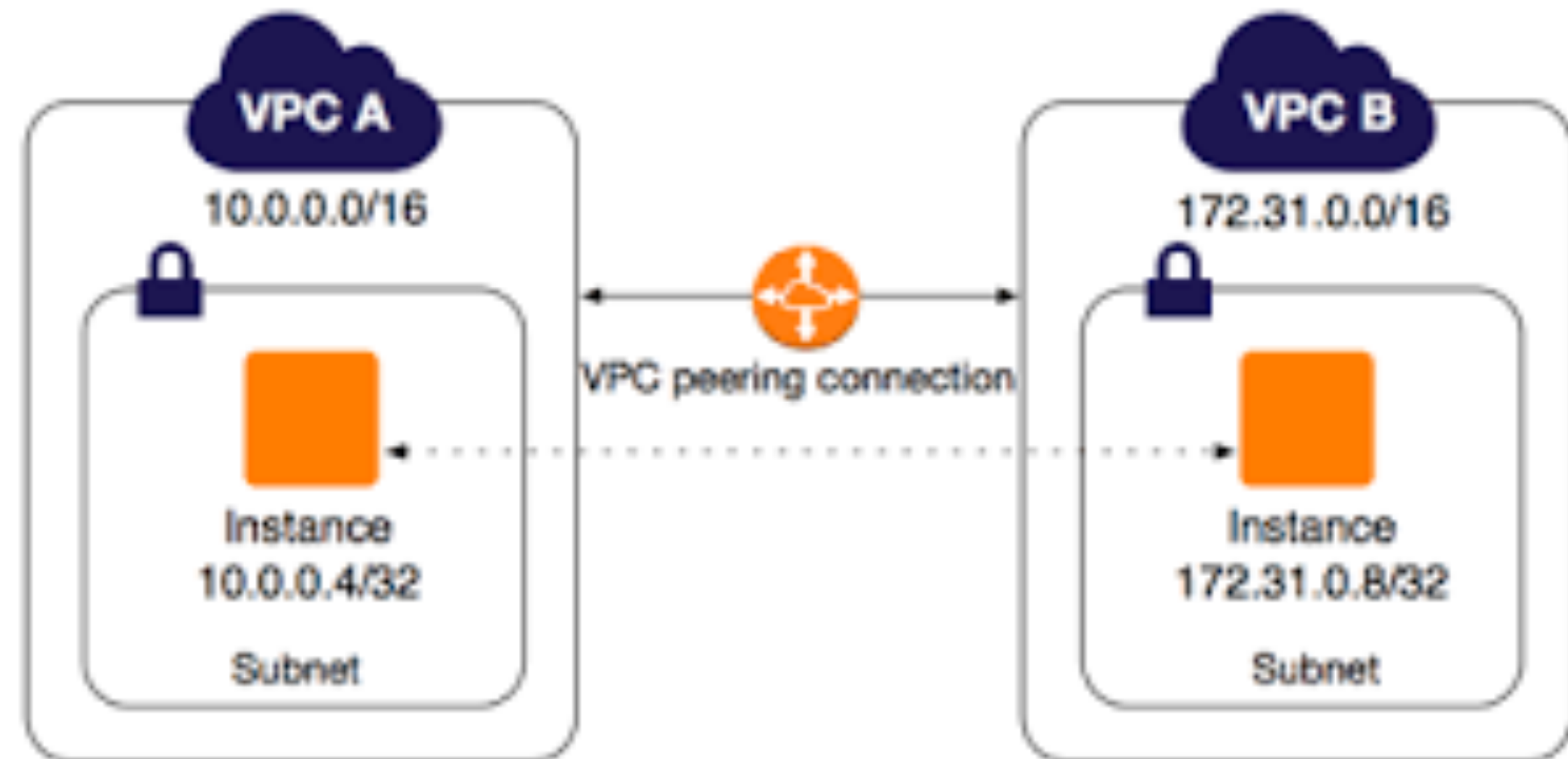
- 在登陆内网时，teleport 可以与认证 Provider 关联得到内网的访问权限，并签发 Short lived certificate 对应到一个 K8s 的 Role
- 对敏感操作需要更高权限的 Role 则需要通过审批

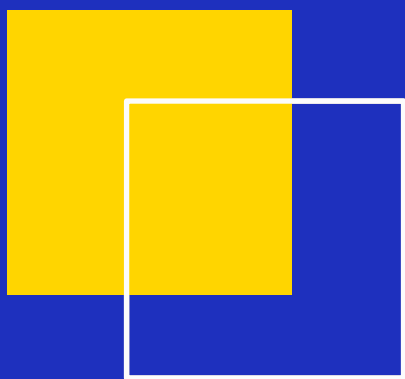




## 限制网络访问

- 限制租户只能通过 VPC 才能访问数据





# THANKS!

