

19 快速部署高可用的 ClickHouse 集群

ClickHouse on Amazon Web Services

GC-600-solution - ClickHouse on AWS

行业：通用

技术：Quickstart、大数据、快速部署、数据分析、智能湖仓



1. 总体介绍

方案背景 ClickHouse 是一款用于联机分析 (OLAP) 的开源列式数据库管理系统 (DBMS)，是目前国内非常流行的一种“实时数仓”的技术选型。ClickHouse 查询性能极快，每台服务器每秒钟可以处理数亿至数十亿行或者数十 GB 的数据，对于海量数据几乎可以实时生成报表。ClickHouse 的高性能源自以下特性：基于列式存储，通过 SQL 查询海量数据并实时生成分析报告；充分利用了所有可用的硬件优化技术，尽可能快地处理每个查询；向量化的查询执行引入了 SIMD 处理器指令和运行时代码生成技术；列式存储数据提高了 CPU 缓存的命中率。

适用客户 该方案适合有以下需求的客户：

- (1) 在针对实时数仓进行技术选型，希望在 Amazon Web Services 上使用 ClickHouse。
- (2) 在友商使用 ClickHouse 托管服务，需要迁移至 Amazon Web Services。
- (3) 对于海量数据查询，尤其是单表查询，有极致速度要求。ClickHouse 具有数据压缩的优势，能够与更小的数据库（如时间序列数据库）竞争并超越它们。

适用场景 该方案适用于多种与大数据、实时数据分析相关的场景，例如：电子商务中用户行为数据的记录和分析，广告行业中的数据分析，IT 运维监控分析，等等。

客户痛点 亚马逊云科技在数据仓库方面提供了 Amazon Redshift 和 Amazon EMR，但主要偏向离线分析场景。目前，暂时还没有 Amazon Web Services 服务很好地匹配“实时数仓”的需求。

2. 客户价值

- (1) 省事：极大地简化了集群部署配置的复杂度，减轻了客户监控和运维的工作量。
- (2) 经济：支持 X86 和 ARM 架构，提高硬件投入性价比。
- (3) 高质量：提供了部署高可用集群的最佳实践。

3. 解决方案概要

方案概要 该方案帮助客户快速在亚马逊云上部署高可用的 ClickHouse 集群，并与 Amazon S3 和 Amazon CloudWatch 等服务进行了集成。客户可以通过部署界面选择部署规模和配置参数。为了便于测试与使用部署好的 ClickHouse 集群，该方案以 OnTime 航班飞行公开数据集为例，对数据基于 Grafana 工具进行可视化，并提供统计分析报告。

方案特点 该方案帮助客户快速在亚马逊云上一键部署高可用的 ClickHouse 集群，支持中国区域和全球区域，有以下优点：

- (1) 与 Amazon Web Services 云服务预集成：该方案集成了 Amazon S3 实现分级存储，与 Amazon CloudWatch 集成实现日志采集和资源指标仪表盘，并提供 Grafana 图形化客户端。
- (2) 预置样例数据集：该方案在部署时可自动导入 OnTime 航班样例数据集，方便客户快速上手体验。
- (3) 支持 ARM 架构部署：该方案支持在采用 ARM 架构的 Amazon EC2 实例上部署，帮助客户节省硬件投入。

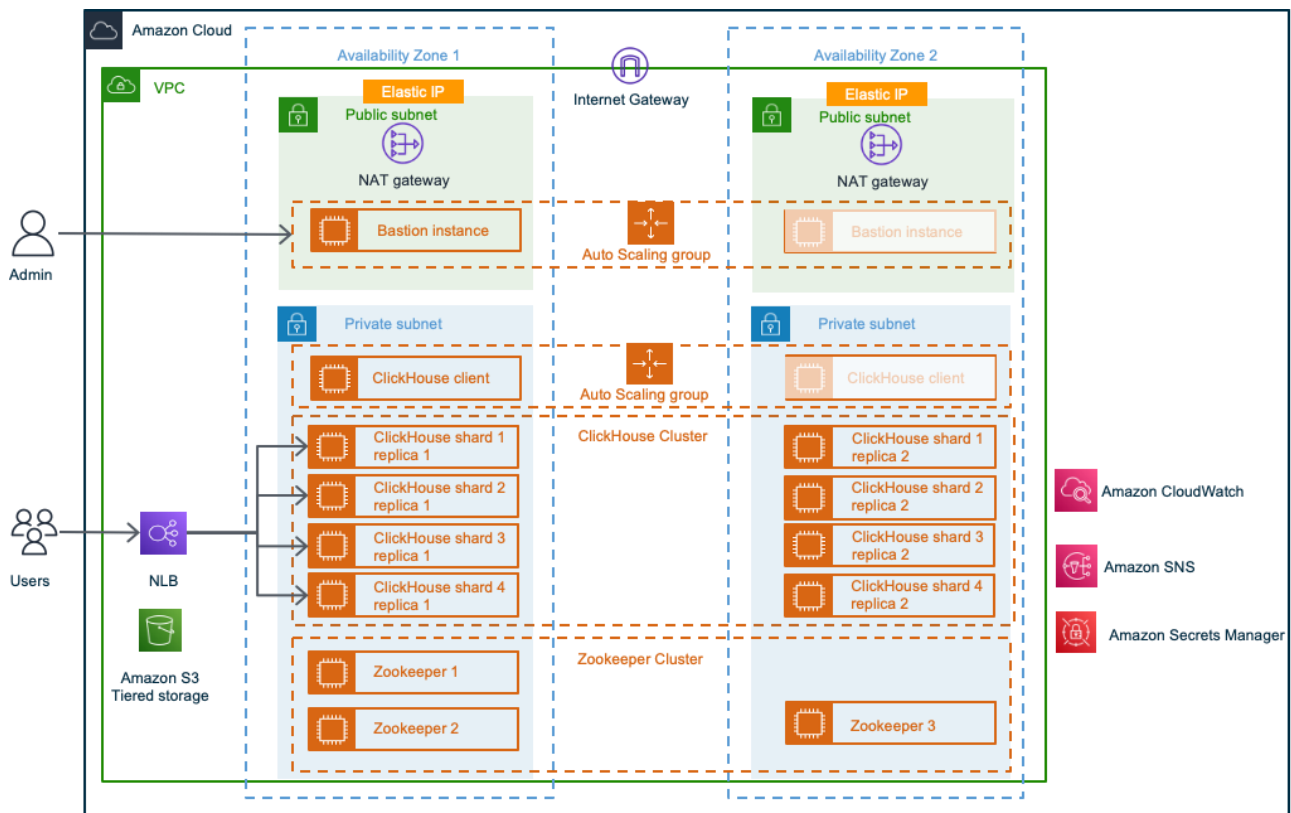
部署前提

- (1) 确保客户账户下的 Amazon Elastic IP、VPC、EC2、Auto Scaling Group、NAT Gateway、EBS 等服务没有达到使用上限。如果某一项已经达到上限，那么请向后台提交工单申请提高限额。
- (2) 如果客户在部署区域内没有密钥对，请在 Amazon EC2 控制台的导航窗格中，依次选择密钥对，创建密钥对，选择默认的 pem 格式，输入名称，然后选择创建秘钥。请妥善保管好此密钥，若丢失则将无法访问实例。

客户资源/能力评估 客户需要具备有关 SQL 的基本的知识和操作能力。

成本因素 该方案的成本主要为 Amazon EC2、Amazon EBS、Amazon S3 和 Amazon CloudWatch 服务的使用费用，以及两个在可用区之间复制数据时产生的费用，使用量与集群节点数、存储数据量和访问负载相关。

4. 架构图



主要服务： Amazon CloudWatch、Amazon EC2、Amazon S3、Amazon Secrets Manager、Amazon SNS、ClickHouse、NAT gateway

5. 资源和帮助

官网 (中国) » <https://www.amazonaws.cn/solutions/clickhouse-on-aws>

部署手册 (中文版) » https://s3.cn-north-1.amazonaws.com.cn/aws-dam-prod/china/Solutions/clickhouse-on-aws/Deployment_Guide_ClickHouse_on_AWS.pdf

Workshop » <https://catalog.us-east-1.prod.workshops.aws/v2/workshops/22377ab3-e5d8-496d-b8be-20a63668daf0/en-US>

部署手

GitHub » <https://github.com/aws-quickstart/quickstart-clickhouse-cluster>

PartnerCast: 快速构建高可用 ClickHouse 集群 » <https://explore.skillbuilder.aws/learn/course/internal/view/elearning/9817/aws-partnercast-kuai-su-gou-jian-gao-ke-yong-clickhouse-ji-qun-technical>

博客: 基于 Amazon EC2 快速部署高可用 ClickHouse » <https://aws.amazon.com/cn/blogs/china/based-on-amazon-ec2-rapid-deployment-of-high-availability-clickhouse/>

ClickHouse 官网 » <https://clickhouse.tech/>

ClickHouse 特性 » <https://clickhouse.tech/docs/zh/introduction/distinctive-features/>

Campaign ID » GC-600-solution - ClickHouse on AWS

产品经理 苏璠 (fansu@amazon.com)

项目负责人 乔伟 (wqqiao@amazon.com)

推广负责人 张濮 (puzhang@amazon.com)

SO8019 