

智能语音V2.6部署指南

系统及软件要求

系统优化项

- 参数优化
- 禁用防火墙和SELinux
- 端口范围
- Docker运行时根目录

架构

- 概要图
- 服务清单

部署

- 配置项
- 授权服务
- 语音服务
- 模型准备
- 初始化
- 启动服务
- 自验证
- 停止服务
- 指定服务启动
- 指定服务停止
- 访问调用

维护

- 异常排查

系统及软件要求

指标	要求	自检命令
----	----	------

系统版本	CentOS 7.4+ 系列， 其它能支持docker的Linux系统理论上可以运行， 但可能存在不可预期的兼容性问题	uname
系统时间	CST 时区，北京时间，误差在1分钟内	date
CPU	32C+，支持avx2指令，主频2.5GHz+	grep avx2 /proc/cpuinfo
内存	128G，具体请查阅相应服务资源要求， asr或tts基础服务至少需要64G+， 具体需参考对应模型要求	free -g
磁盘	500G+	df -h
Docker	1.12.x + 或 17.06.0+	docker -v

注意：请部署实施前检查相关项是否满足要求，这是个基准配置，更多详情请参阅《部署架构与资源标准》。

友情提示：

1: Docker 通过yum在线安装会相对简单

docker官方安装指导：<https://docs.docker.com/install/linux/docker-ce/centos/#install-docker-ce>

2: Docker 离线安装请自己通过备份对应rpm安装。

3: 2.x架构均使用标准docker镜像交付，请自行了解所有docker相关技术和使用

系统优化项

注意：如下优化是必选项

参数优化

1. 文件描述符和进程数(线程数)等优化

文件路径： /etc/security/limits.conf

```
*      - nproc          65535
*      - sigpending     65535
*      - nofile         655350
```

文件路径： /etc/security/limits.d/20-nproc.conf 可能会为 例如90-nproc.conf，以实际为准

```
*      - nproc          65535
```

禁用防火墙和SELinux

```
systemctl stop firewalld.service      #停止firewall
systemctl disable firewalld.service   #禁止firewall开机启动
```

如果/etc/selinux/config中SELINUX=enforcing，需做如下修改并重启系统

修改/etc/selinux/config 文件
将SELINUX=enforcing 改为 SELINUX=disabled
重启系统生效

端口范围

修改临时分配端口范围

```
# 查看本机临时端口
sysctl net.ipv4.ip_local_port_range

# 修改到10000+起始，避免占用服务监听端口
sudo sysctl -w net.ipv4.ip_local_port_range="10000 64000"
```

Docker运行时根目录

注意: 由于docker的限制，如果在部署后再调整需要清理docker镜像和容器等后再重新配置并重启，建议在首次部署即考虑该项。

1. docker对磁盘的占用较大，请预先设置docker root dir为大磁盘路径，并重启查看是否生效

v17.05.0 以下版本配置方法

文件路径： /etc/docker.conf 或者 /etc/sysconfig/docker 请根据当前docker版本自行判断

```
--graph=/home/admin/docker
```

比如docker-ce 1.13版本修改docker root dir是按如下方法增加

```
# Modify these options if you want to change the way the docker daemon runs
OPTIONS='--selinux-enabled --log-driver=journald --signature-verification=false --graph=/home/admin/docker'
if [ -z "${DOCKER_CERT_PATH}" ]; then
    DOCKER_CERT_PATH=/etc/docker
fi
```

v17.05.0 以上版本配置方法

修改 /etc/docker/daemon.json ,如果不存在则创建一个

```
{  
  "data-root": "/new/path/to/docker-data"  
}
```

修改完成后，重启docker，systemctl restart docker

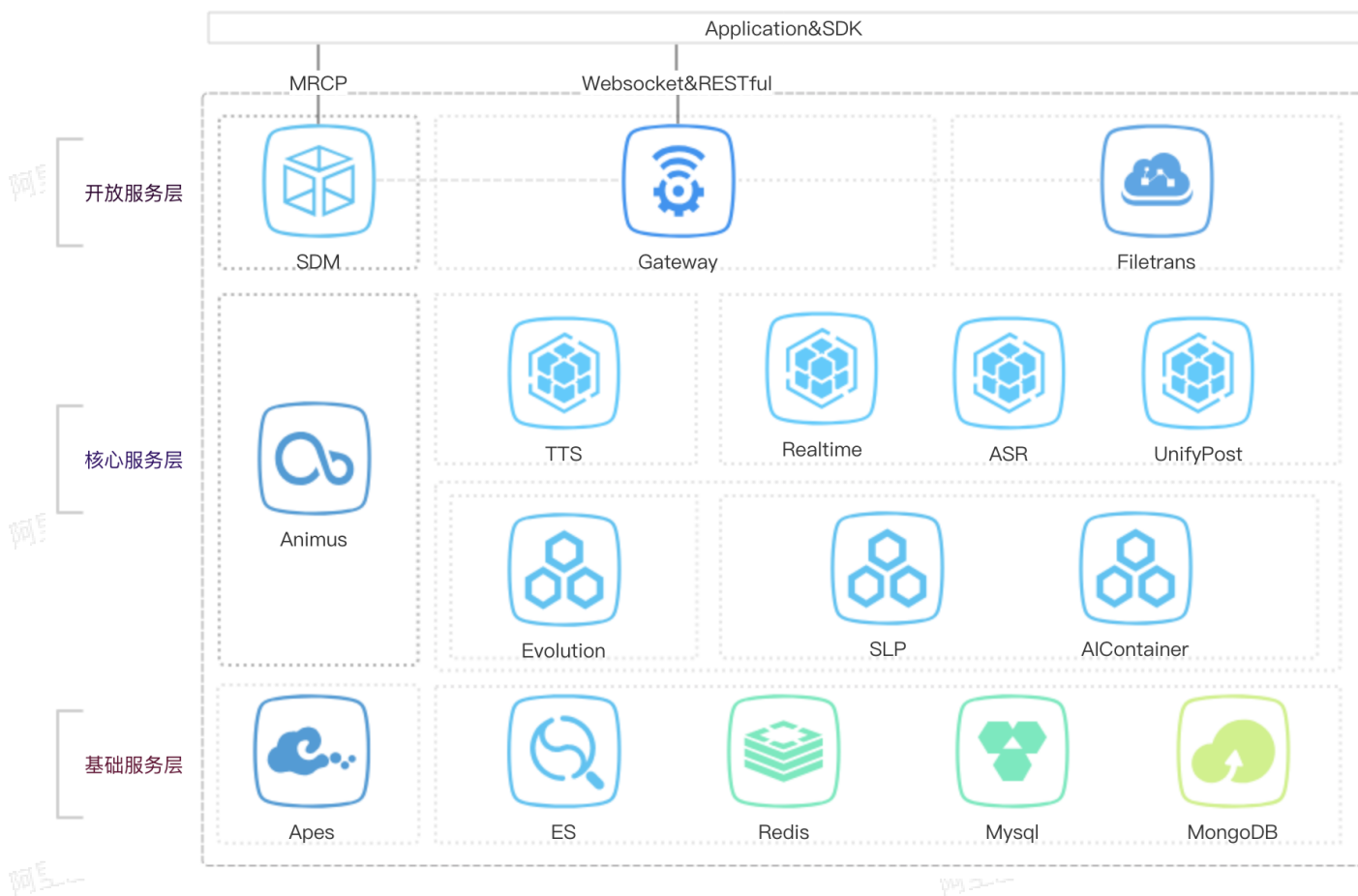
执行 docker info | grep "Docker Root Dir" 查看是否是设置的目录

架构

专有云V2启用全新架构，全新变化：

- 全系容器化，可运行在物理机，swarm，k8s等多种体系内，真正实现WORA(Write once, run anywhere)
- 全新交互协议，更健壮
- 自主研发的基础服务能力开放，如网格寻址，负载均衡等能力
- 自主管控运维监控能力开放
- 服务与模型松耦合
- 语音算法全系升级
- 定制化服务全面开放

概要图



服务清单

- 所有应用除管控设施外均以简要名代表，实际名会有nls或nls-cloud前缀，以下服务是按需提供。
- 详情请参考 《智能语音V2.X部署架构与资源标准》

部署

注意：

将提供的安装包放置到您需要部署运行的位置，推荐放置到 `/home/admin` 目录下，后续沟通均默认以此目录为基准。请保证该目录磁盘可用空间满足资源标准。

配置项

当前所有的开放配置项都在 `service/conf/nls.conf` 里，里面有所有中间件，apes等连接相关的配置，您可以按需修改。如果是单实例，您无需做任何配置，直接下一步吧。

授权服务

重要：

- Apes 必须预先部署和授权激活后才可以进行其他相关部署。

- Apes 提交序列号后未授权前不允许重启Apes本身，否则序列号将变更，对应授权码将失效。
- Apes 会限制运行实例 和 并发控制等。
- Apes 更新授权同新申请步骤一致。
- 该步骤只需在未授权前需要单独启动，授权通过后不再需要单独启动授权服务，其可以随着 start.sh 启动。
- 更多信息 请参考 《Apes部署指南》

```
chmod +x ./bin/*

# 单独提前启动授权服务
./bin/apes.sh start

# 获取序列号，请将序列号信息发送给商务同学进行授权，更新授权同如下步骤一致
./bin/apes.sh get

# 根据商务同学提供的授权码进行激活， 注意 未激活前请勿重启apes或者关机重启等操作，否则序列号
将变化
./bin/apes.sh put 您的授权码

./bin/apes.sh ctx 查看授权信息
```

语音服务

拷贝对应软件包到对应服务器，请选择放置到磁盘空间较大的路径下。

- 以下命令请以root权限 或者 sudo 超级权限执行
- 请保证端口不被占用，目前使用的为 80 443 3306 6379 7000-9000 27017

模型准备

安装包内一般内置默认模型,路径如下，其它模型类似存放在对应目录下：

类别	相对目录
ASR模型目录	service/resource/asr/default
TTS模型目录	service/resource/tts/default

- 注意： TTS服务内部已经带有一份基础模型，可以实际运行，如需升级或替换模型，才需考虑该步骤。
- 注意： 更新完模型需要重启服务

关于模型更新：备份该default目录，将新模型目录命名为default

```
cd service/resource/asr
mv default 你的备份目录/asr_20181101_bak
mv 新模型目录 default

# 重启服务
./bin/stop.sh
./bin/start.sh
```

以下操作默认在安装包内的 service 目录

初始化

执行如下命令导入镜像，如后续需要更新升级，同样导入相应镜像即可

```
chmod +x ./bin/*
./bin/init.sh
```

可以通过 docker images 查看镜像是否被加载。

- 镜像是可以按需加载的，比如只需要部署tts服务，则 ./bin/init.sh tts 其会自动加载tts相关的服务

启动服务

一键启动所有服务，用户可以按需定制启动相应服务

- 注意：docker运行后会对映射目录权限进行修改，请提前查看读写权限

```
./bin/start.sh
```

- 可以通过 docker ps 查看服务运行的状态
- 所有日志 在 bin的上层目录logs下面

重点关注：

- 应用理论上在五分钟内即可启动成功，asr或tts 服务由于需要加载模型，启动所需要的时间 根据模型，并发路数和您的机器配置有直接关系，具体启动详情可以关注日志

服务	重点关注日志
asr	logs/nls-cloud-asr/alizr.log

tts

logs/nls-tts/application.log

- 所有数据被持久化在 data 目录下，包括 mysql, redis等，请勿随意删除

自验证

服务启动后，需要运行自验证程序，来确保服务已经正常。

- 请按需选择，您未部署的功能请无须关注

查看当前安装服务端口状态是否异常，如出现异常，请排查具体服务的日志

```
sh ./bin/status.sh
```

基于sdk的demo进行基础功能测试，如出现异常，请排查具体服务的日志

```
sh ./demo/demo.sh
```

停止服务

一键停止所有服务

```
./bin/stop.sh
```

指定服务启动

如果您需要按需指定服务运行，可以进行如下操作（请在您已经知晓相关架构和流程细节后再自行按需定制操作）

```
./bin/start.sh 服务名1(比如nls-cloud-gateway) 服务名2 等
```

指定服务停止

如果您需要按需指定服务运行，可以进行如下操作（请在您已经知晓相关架构和流程细节后再自行按需定制操作）

```
./bin/stop.sh 服务名1
```

访问调用

目前gateway基于websocket协议和sdk交互，已经直接提供REATful接口，sdk使用请查阅sdk相关文档

功能	连接地址
一句话识别&实时语音转写&语音合成	ws://gateway所在IP:8101/ws/v1
一句话识别RESTful	http://gateway所在IP:8101/stream/v1/asr
语音合成RESTful	http://gateway所在IP:8101/stream/v1/tts
录音文件识别RESTful	http://gateway所在IP:8101/stream/v1/filetrans
MRCP协议调用ASR&TTS	sip地址： nls-cloud-sdm的ip, sip端口是7010

- 鉴权：专有云可忽略此概念，阿里云官网关于此功能的描述对于专有云无效；
appkey：专有云可忽略此概念，[appkey 参数传入固定字符串"default" 即可](#)；
access token：专有云可忽略此概念，[token参数传入固定字符串"default" 即可](#)；

维护

- 为了更便捷快速排查和定位问题，客户对于专有云部署的环境需要了解基本的异常定位方法，最基本原则：除了看现象，还要看日志。
- 当前版本所有应用日志都定向到了同一个目录下，启动时有日志所在目录提示，客户可以根据上述概要图分析日志，如若有必要，可以提供打包的日志进行分析。
- 如果存在服务启动等异常，且可能导致故障无法排查恢复时，请卸载所有容器服务，并删除运行时数据或者直接更换目录重新部署

异常排查

- 请查看硬件配置和系统版本是否达到标准，比如内存，硬盘，cpu，是否centos7.2+等，查看局域网络是否异常，比如存在ip冲突，dns配置不当，路由不当等
- 查看各个docker 服务是否都启动正常，docker ps多次，看看运行时间是否如期，如果容器服务有异常，会退出并重新启动，可以根据 docker logs 容器id 来查看该容器的stdout日志
- 查看对应服务进程是否存在 ps -ef | grep 进程关键字，几次，看看进程id是否不变
- 查看对应服务端口是否存在，telnet 127.0.0.1 对应端口 ,并确保端口未被其他应用占用:sudo netstat -pant | grep 端口
- 查看对应服务日志，日志按标准存储在 logs/对应服务目录下

access.log	记录该应用核心请求的扼要信息，比如请求参数，延时，返回消息等
application.log	记录该应用的所有日志
error.log	记录该应用的错误信息

