Architecture Report

KTH Group 2

May 30, 2017

Component Overview

The different components of the system can be found in Figure 1, where there are three principal components:

1. Workload generator
2. Service
3. Monitor

Architecture of the Application

A schematic representation of the architecture can be found in Figure 2. The clients connect to the webserver, which is a nginx webserver, and the webserver produces a job message which it puts in the distributed queue, RabbitMQ. A job message consists of:

1. Client ID, for the webserver to identify the client.
2. Additional information about the request.

A VM consumes a message from the queue, and sends a confirmation message to the client, via the webserver acting as a reverse proxy server now. The client gets the confirmation and an URI to where the client can use the HTTP POST command to send his input video. The converter gets the input video, process the video, and sends it back to the client, via the webserver. This process is depicted in Figure 3 shows the sequence diagram, which consists of the following messages:

1. The client requests a video conversion from the frontend.
2. The frontend requests the necesary resources from the backend.
3. The backend reserves the necessary resources.
4. The backend acknowledges by providing a resource block.

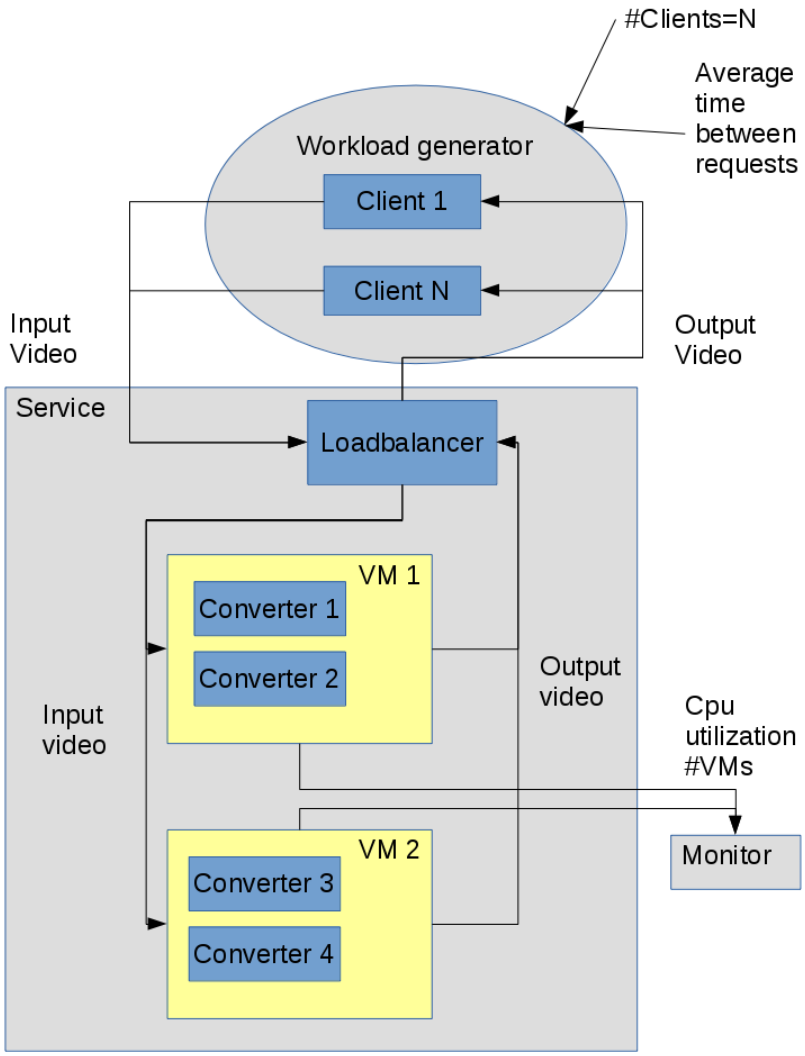

Figure 1: Graphical representation of the requirements and the components.

5. The frontend acknowledges that the video can be converted.
6. The client can start transferring the source video.
7. The backend converts the video to the desired format.

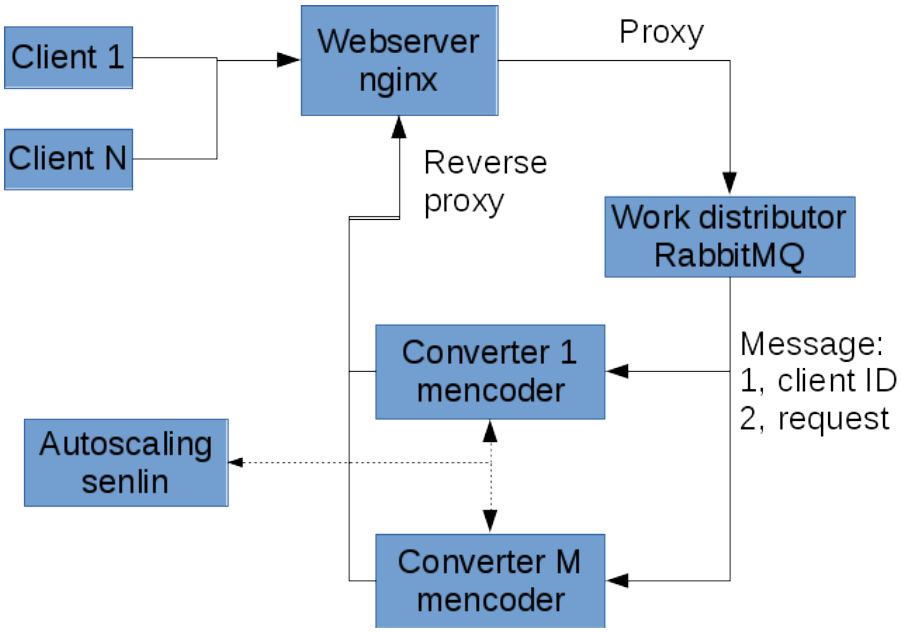

Figure 2: Architecture of a video converting application.

8. The resulting video file is transfered back to the client.

The OpenStack Senlin module surveys and monitors the load on the converter nodes. Senlin can use different policies to scale up and down the number of running vm. The senlin module would also act as the monitor in the component scheme.

The architecture have the following properties:

- No files are stored in the service.
- The requests are relayed from the webserver to the work distributor, a distributed message queue.
- A job consists of three parts:
  1. Send the video from the client to the worker, through proxy server.
  2. The worker process the video.
  3. The worker sends back the converted video to the client, through proxy server
- The senlin module scales the number of running VMs, based on predefined policies.

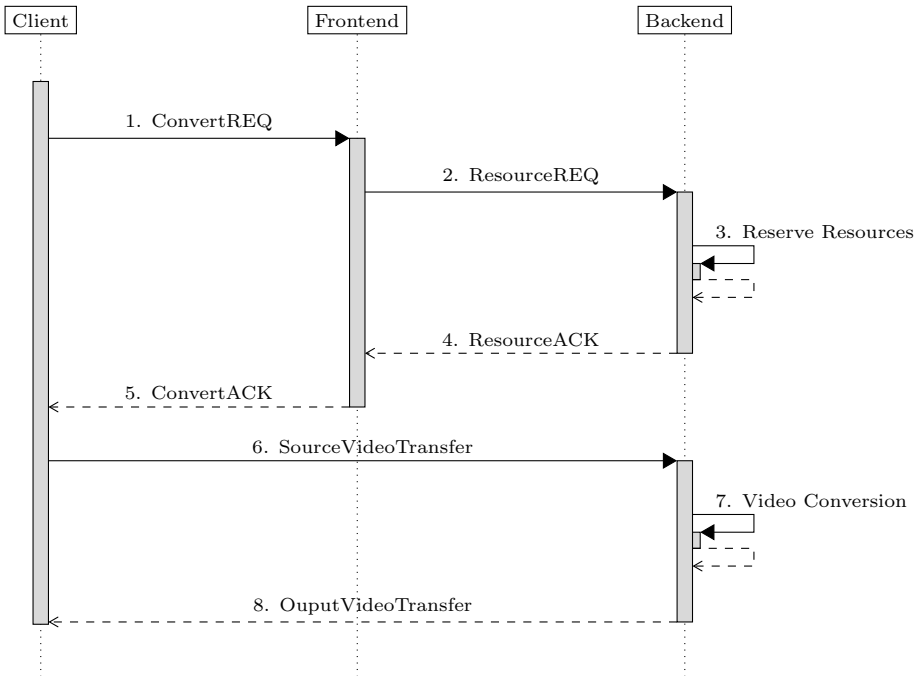

Figure 3:

Exposed Interfaces

The following interfaces are implemented:

1. REST HTTP API between client and webserver.
2. HTTP XML between nginx distributed queue (STOMP).
3. Openstack between converter VM and senlin.

Technology Stack

1. PyFLASK
2. Pyrequest
3. nginx
4. RabbitMQ
5. Senlin
6. mencoder
7. Python wrapper for mencoder

Scalability and Robustness

- Three vms are assumed to be always-available:
    1. webserver
    2. work distributor
    3. autoscaling
- The converter node VMs can be scaled up or down depending on the need.