MuseBox

FPGA Machine Learning based system
for real-time AV Broadcasting applications

1. Introduction

Why Real-Time?

Definition

Real-time multimedia refers to applications in which multimedia data has to be delivered and rendered in real time; it can be broadly classified into interactive multimedia and streaming media.

- Work with live stream
  - Interactive or live applications need a «fast» response from the system (< video timing)
  - Speed required is dependent on multimedia quality
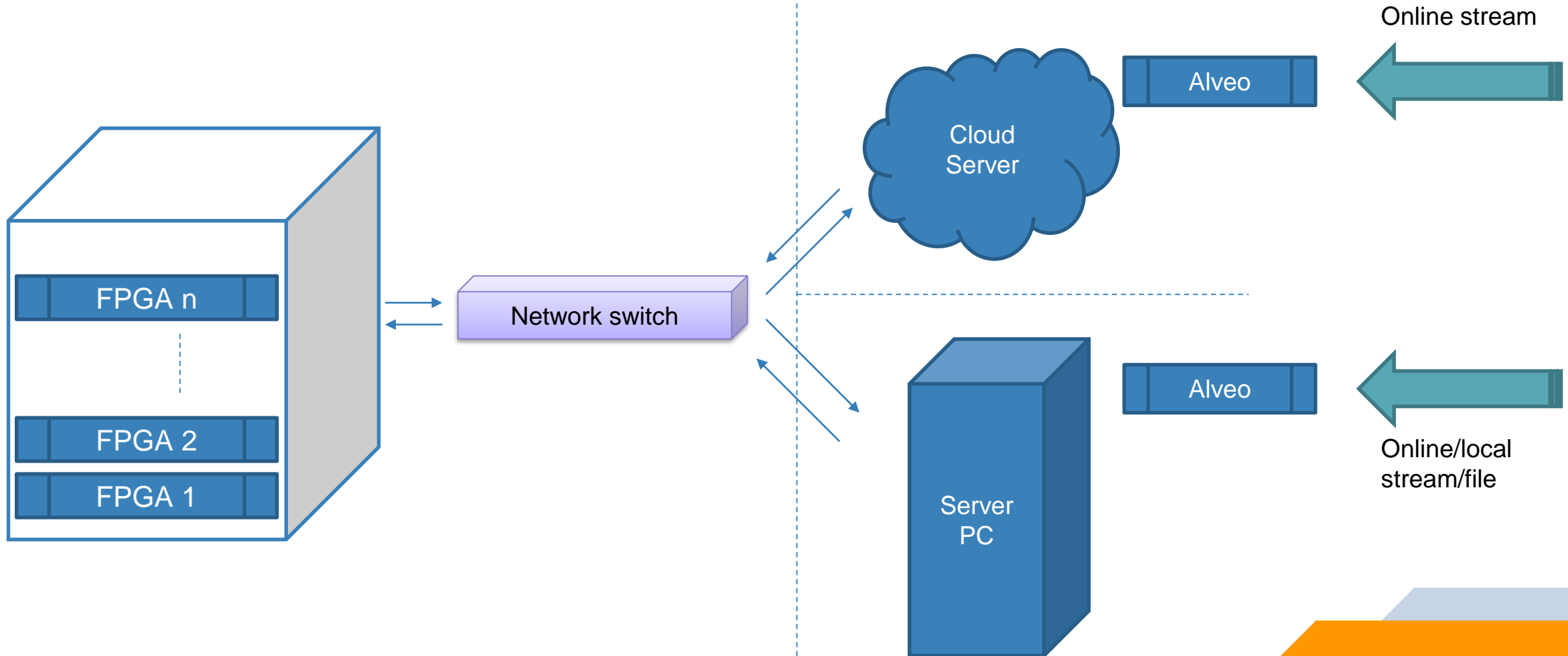 - Need a tradeoff between service cost and performances
- Work with local files
  - Big amount of files requires a very fast system
  - Files often not accessible outside the local network for legal reasons
  - Need a system for human timings

Scalable architecture
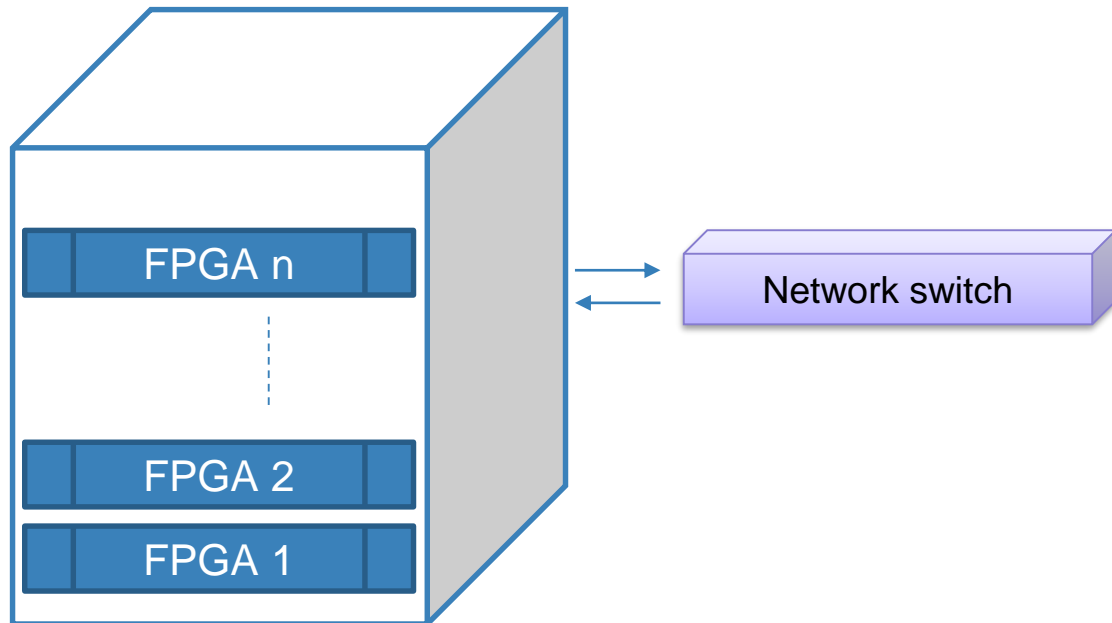
- The customer need a customizable architecture
  - ➢ Custom streams
  - ➢ Custom audio/video quality
  - ➢ Need for an upgrade
  - ➢ Output quality check
- Best tradeoff between system cost, maintenance cost and upgrade cost

2. Architecture

Online stream

Alveo

Cloud
Server

Online/local
stream/file

Alveo

FPGA n

FPGA 2
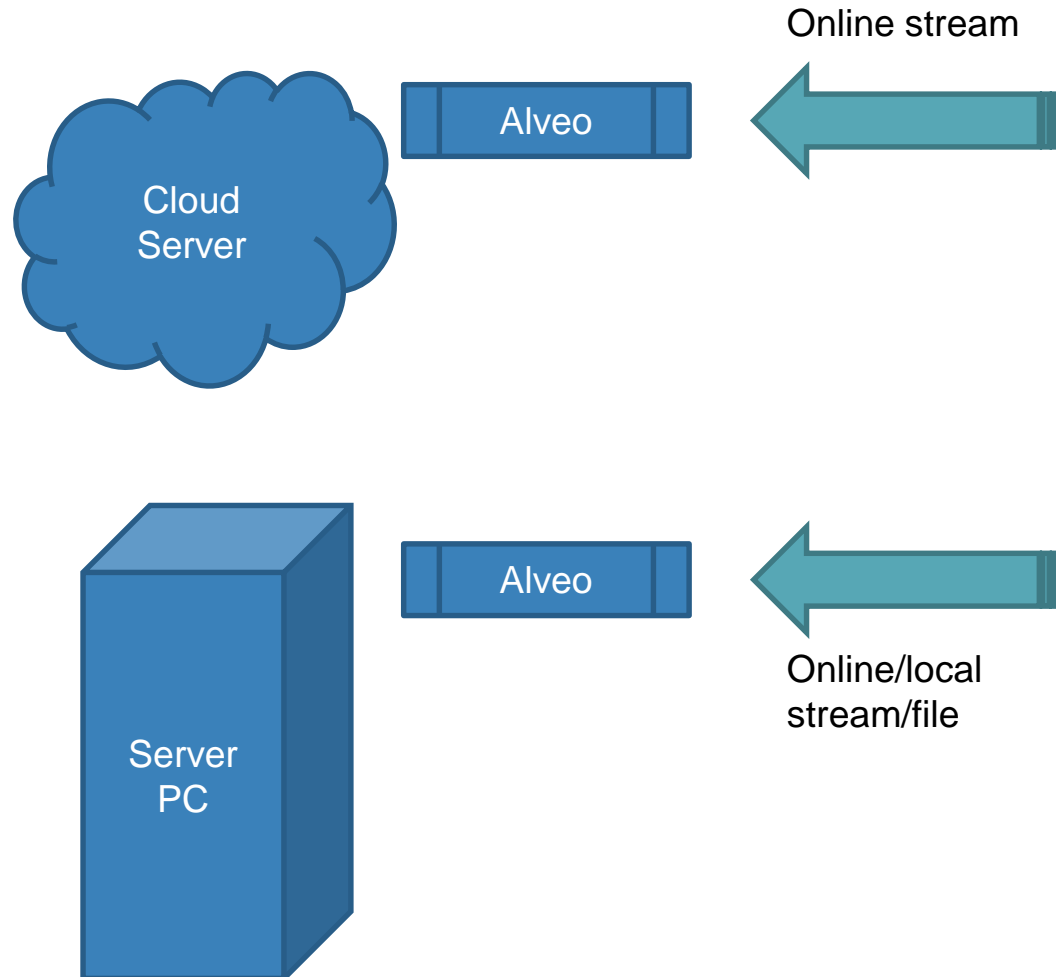
FPGA 1

Network switch

Server
PC

Hardware Architecture (cont.)

- FPGAs give a high-perfomances solution for multimedia applications
- SoC solution for FPGA-based systems gives a high scalable design thanks to microservices organization
- Easy to deploy, easy to customize
- Highly fault-tollerant

Hardware Architecture (cont.)
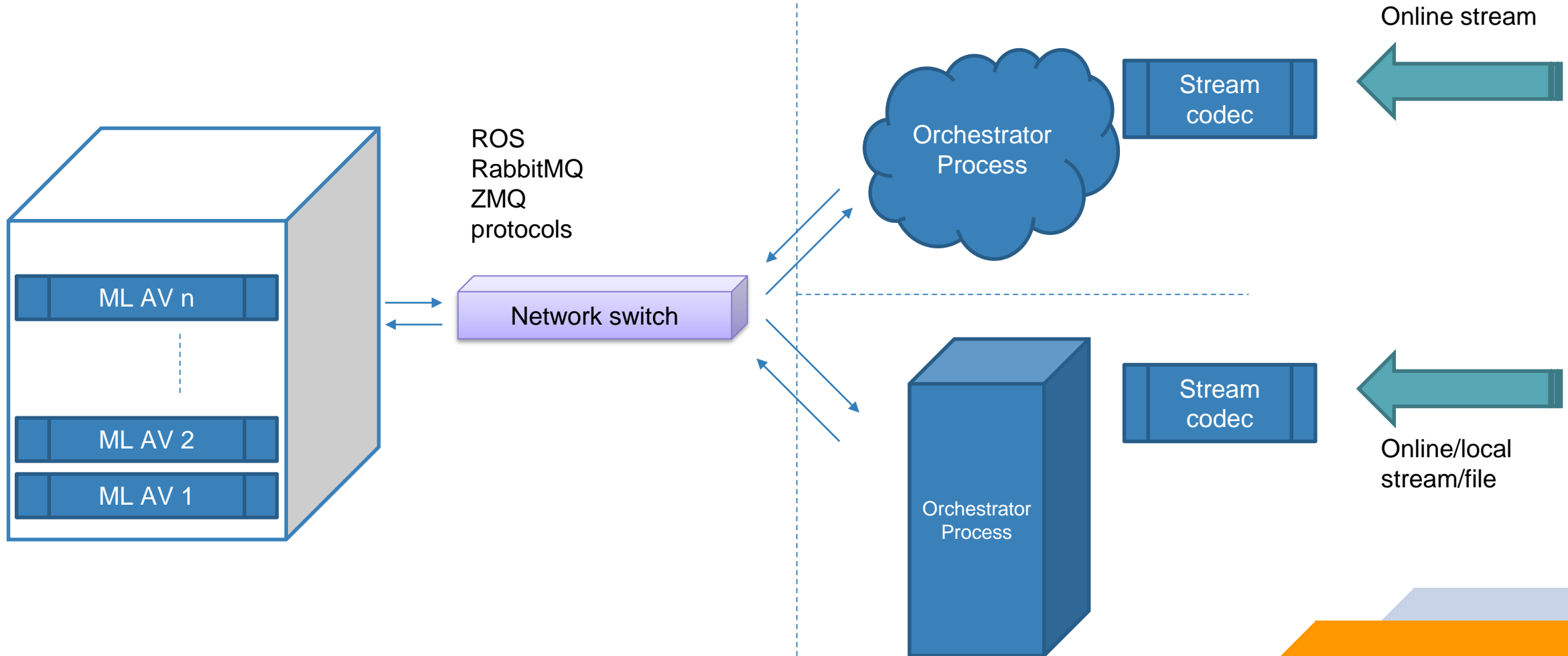
Online stream

Alveo

Cloud
Server

CLOUD SERVER
- Easy to deploy and develop, it is the best solution for an online analyzer that works with an online stream
- It can record/parse streams without network problems
- It's a good choice for simple complexity application and for hardware costs
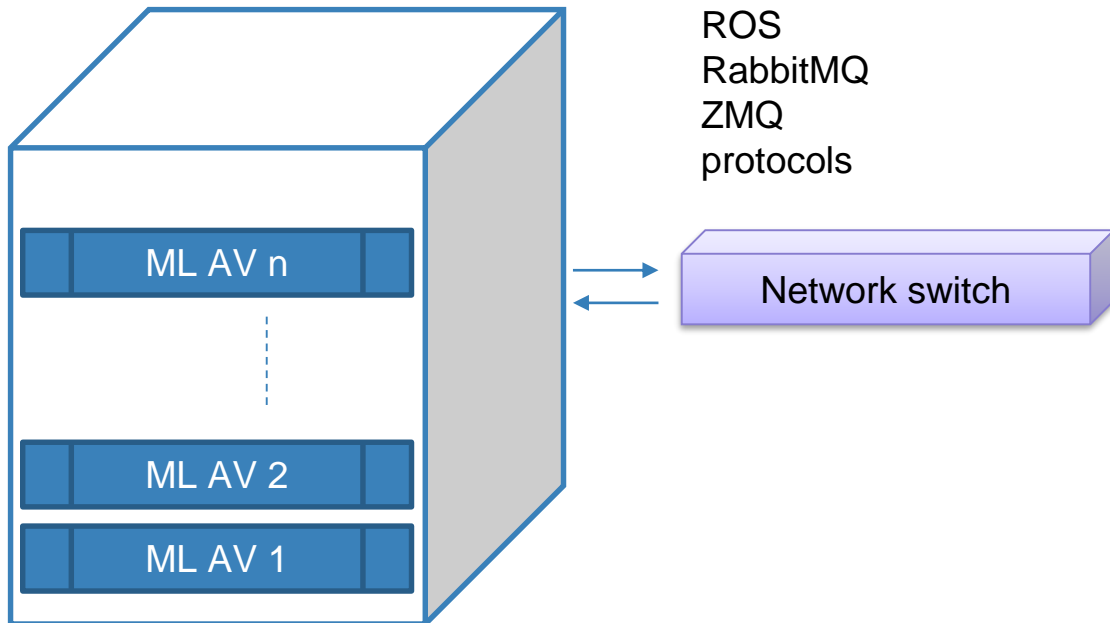
Alveo

Online/local
stream/file

Server
PC

PC SERVER
- The best solution for a local stream or local archive
- Often the only solution for media that have copyright or legal constraints
- The best solution for high complexity application that requires a custom hardware
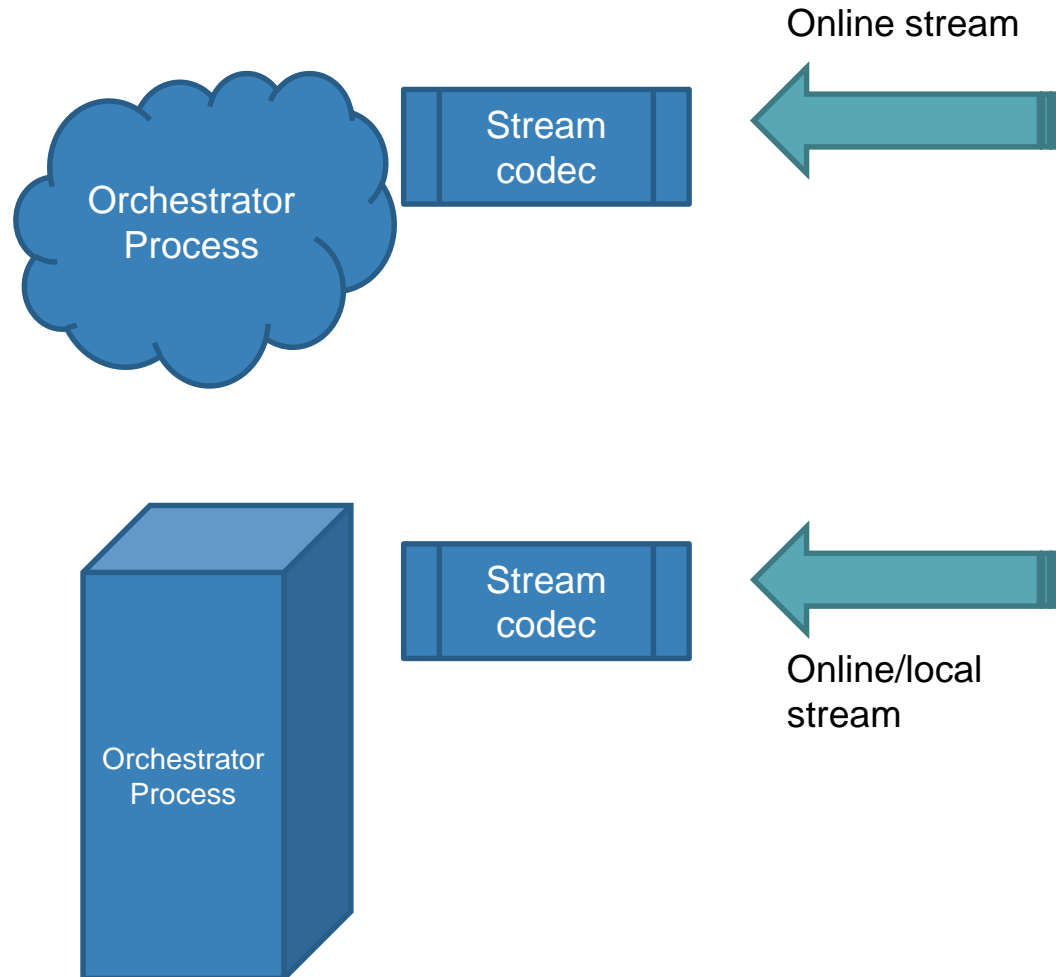
Software Architecture

ML AV n

ML AV 2

ML AV 1

ROS
RabbitMQ
ZMQ
protocols

Network switch

Orchestrator
Process

Orchestrator
Process

Stream
codec

Stream
codec

Online stream

Online/local
stream/file

Software Architecture (cont.)

ROS
RabbitMQ
ZMQ
protocols

ML AV n

ML AV 2

ML AV 1

Network switch

- Queue-based messages
  - With standard network protocol, it can be easily connected to existing systems
  - Useful for load balancing and stream sequencing
- Machine learning with FPGA
  - Vitis AI Model Zoo provides a big repository for standard deep neural network (Vgg, ResNet, InceptionNet, MobileNet, etc.)
  - AI SDK and DNNDK framework based gives a good opportunity to work with custom neural network created in tensorflow and caffee
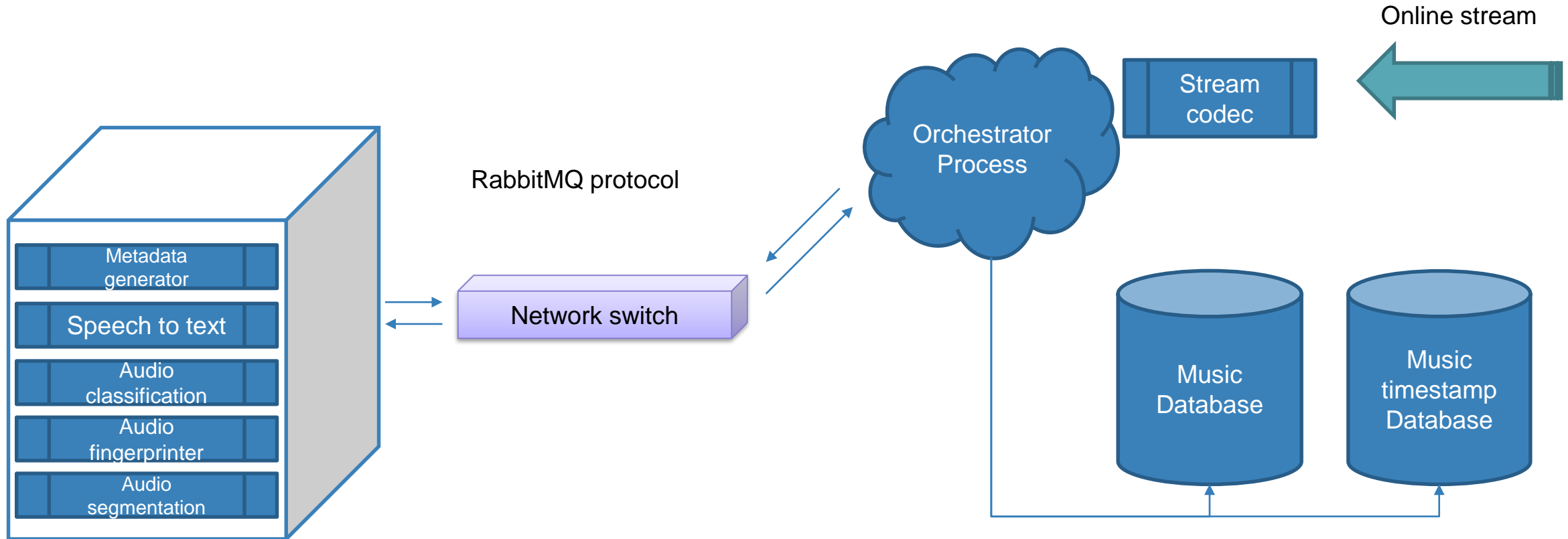
Software Architecture (cont.)

Orchestrator Process

Stream codec

Online stream

Orchestrator Process

Stream codec

Online/local stream

- ALVEO technology gives the best scalable solution for stream decoding/encoding and custom IP for software acceleration
- Often compressed streams require big resources in order to preserve latency constraints
- The system orchestrator gives the best solution in order to collect datas and apply custom data process
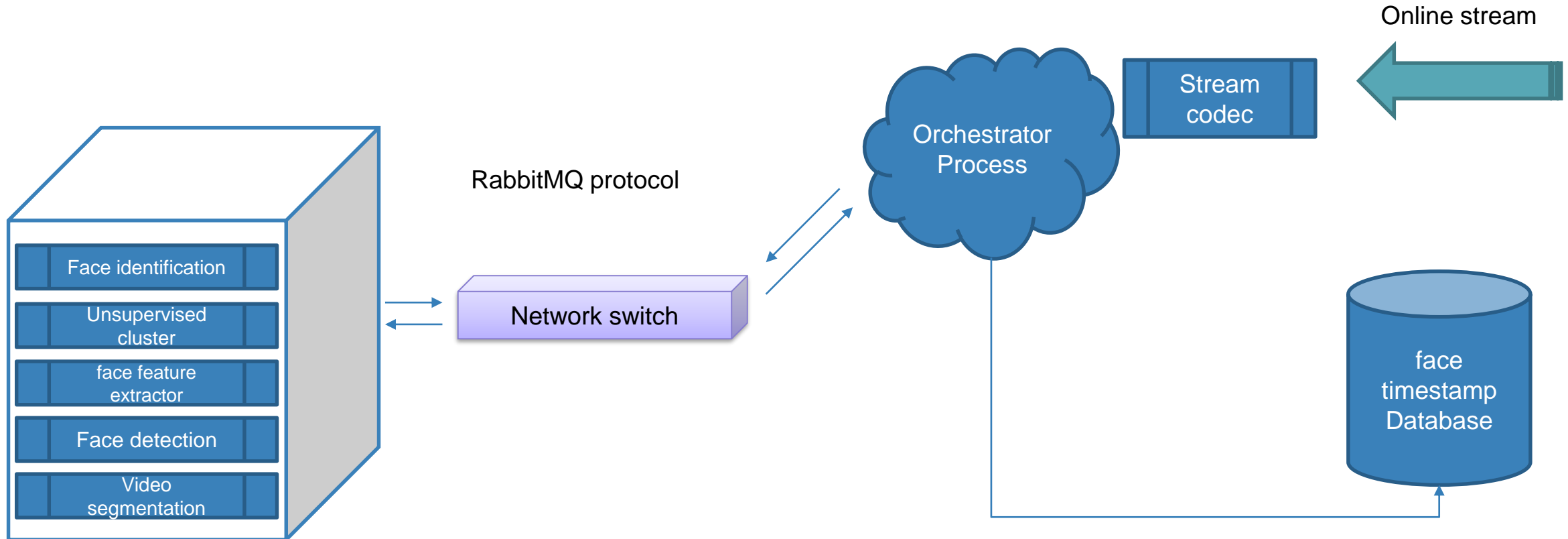- The orchestrator can be easily connected to a database for data collection

3. Example applications

Music Recognizer

Online stream

Stream codec

Orchestrator Process

RabbitMQ protocol

Metadata generator

Speech to text

Audio classification

Audio fingerprinter

Audio segmentation

Network switch
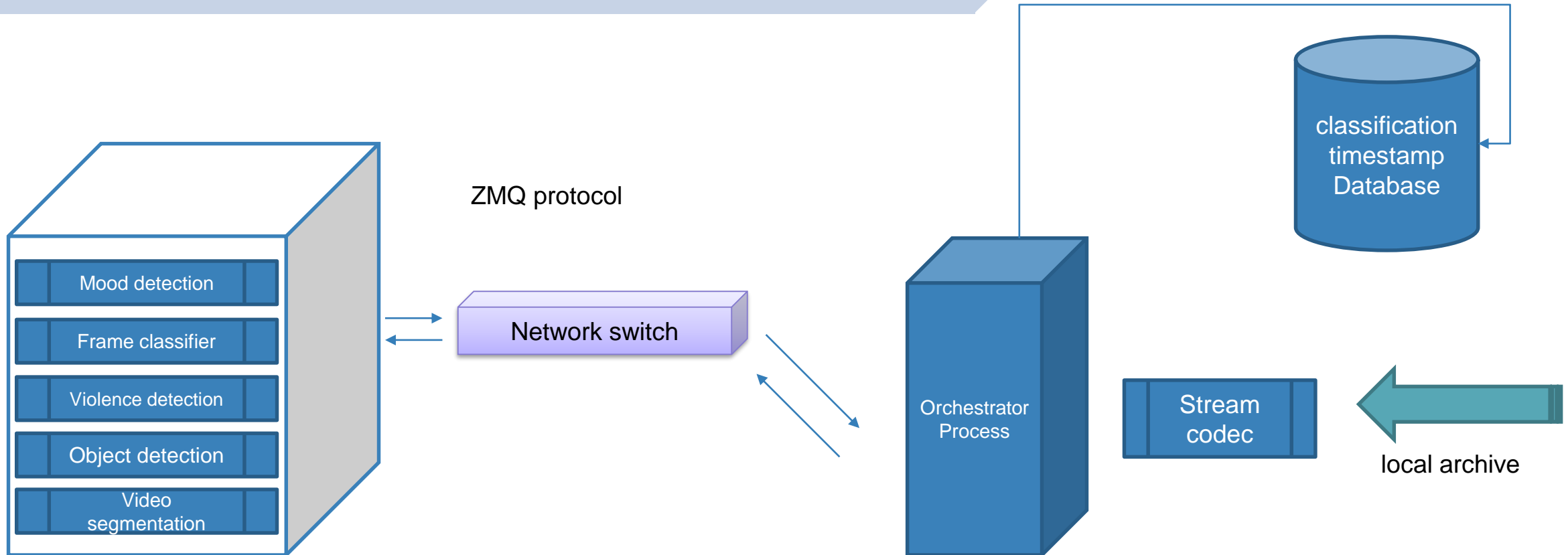
Music Database

Music timestamp Database

Possible final applications: copyright analyzer, collaborative systems, music recognition, etc.
Actually deployed on MRadio system that is an **Audio stream analyzer** used all over Italy for music reporting
(https://mradio.it/)

People Identification

Online stream

Stream codec

Orchestrator Process

RabbitMQ protocol

Network switch

Face identification

Unsupervised cluster

face feature extractor

Face detection

Video segmentation

face timestamp Database

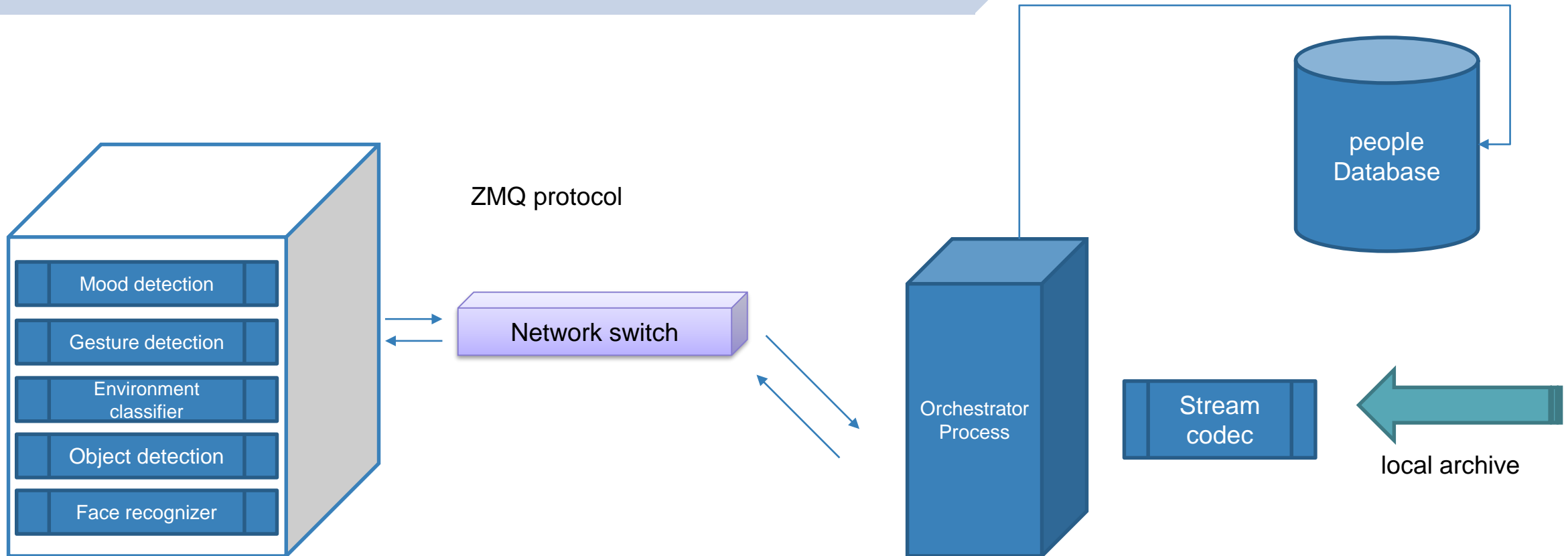Possible final applications: crowd identification, celebrity recognition, audience analysis, etc.

Local Video Classifier

Mood detection

Frame classifier

Violence detection

Object detection

Video segmentation

ZMQ protocol

Network switch

Orchestrator Process

classification timestamp Database

Stream codec

local archive

Possible final applications : video classification, audience analysis, sport analysis, censorship system, etc.

Healtcare

ZMQ protocol

Mood detection

Gesture detection

Environment classifier

Object detection

Face recognizer

Network switch

Orchestrator Process

people Database

Stream codec

local archive

Possible final applications : elderly support, emotional systems, pet robot, reception and presentation robot, etc.

Thanks

MakarenaLabs SRL
staff@makarenalabs.com