

Supplementary Material

Execution scenarios

λ -*CLOVR* is based on lambda architecture, and the execution scenarios undergo two types: the Speed Execution Scenario (Speed Layer) and the Batch Execution Scenario (Batch Layer). Both scenarios use Serving Layer for data processing. The λ -*CLOVR* modules are distributed in the cloud across various types of clusters, each of which can be scaled out as needed. Fig. S1 illustrates these execution scenarios, and detailed explanations are provided in the subsequent subsections.

Speed Execution Scenario

After subscribing to a video stream data source, the Speed Execution Scenario progresses through several stages, utilizing distinct λ -*CLOVR* components. These modules are distributed across cloud clusters of five different computing types, as illustrated in Fig. S1.

The VSAPS interfaces with external video stream sources, hosted by the Video Stream Acquisition Cluster. Video data streams are captured, converted into messages, grouped into micro-batches, compressed, and loaded onto the corresponding Kafka topic (UID_VDS) in the Distributed Message Broker Cluster. This cluster buffers the video data streams and IR within the distributed messaging system, maintaining two types of topics per registered user: UID_VDS and UID_VDS_IR.

The Video Stream Persistence Cluster uses VBDL services to load mini-batches from the UID_VDS topics, processing, encoding, and extracting metadata from the video data. The processed video and its metadata are then stored in the Video Big Data Store and Structured Metastore, respectively. The Video Stream Analytics Consumer Cluster processes the video streams from the UID_VDS topic in near real-time, using VBDAL services to analyze and compute deep features, which are then sent to the UID_VDS_IR topic in the Distributed Message Broker Cluster. The Feature Persistence Cluster continuously reads the IR from the UID_VDS_IR topic, encoding and aggregating them before indexing them in the DEFI.

Batch Execution Scenario

The Passive Execution life cycle consists of two types of clusters: the Raw Batch Video Data Service Cluster and the Batch Video Analytics Consumer Cluster. Users can upload batch datasets and models to the Raw Batch Video Data Service Cluster. The Batch Video Acquisition Service handles the uploaded batch datasets, which are processed by the Video Processor after being buffered. This processing extracts metadata from the batch datasets, which are then stored in the Video Big Data Store and the Structured Metastore, respectively.

The Batch Video Analytics Consumer Cluster loads instances of deep video processing services and performs offline video analysis. Once the IR is computed, it is indexed in the DEFI for future use.

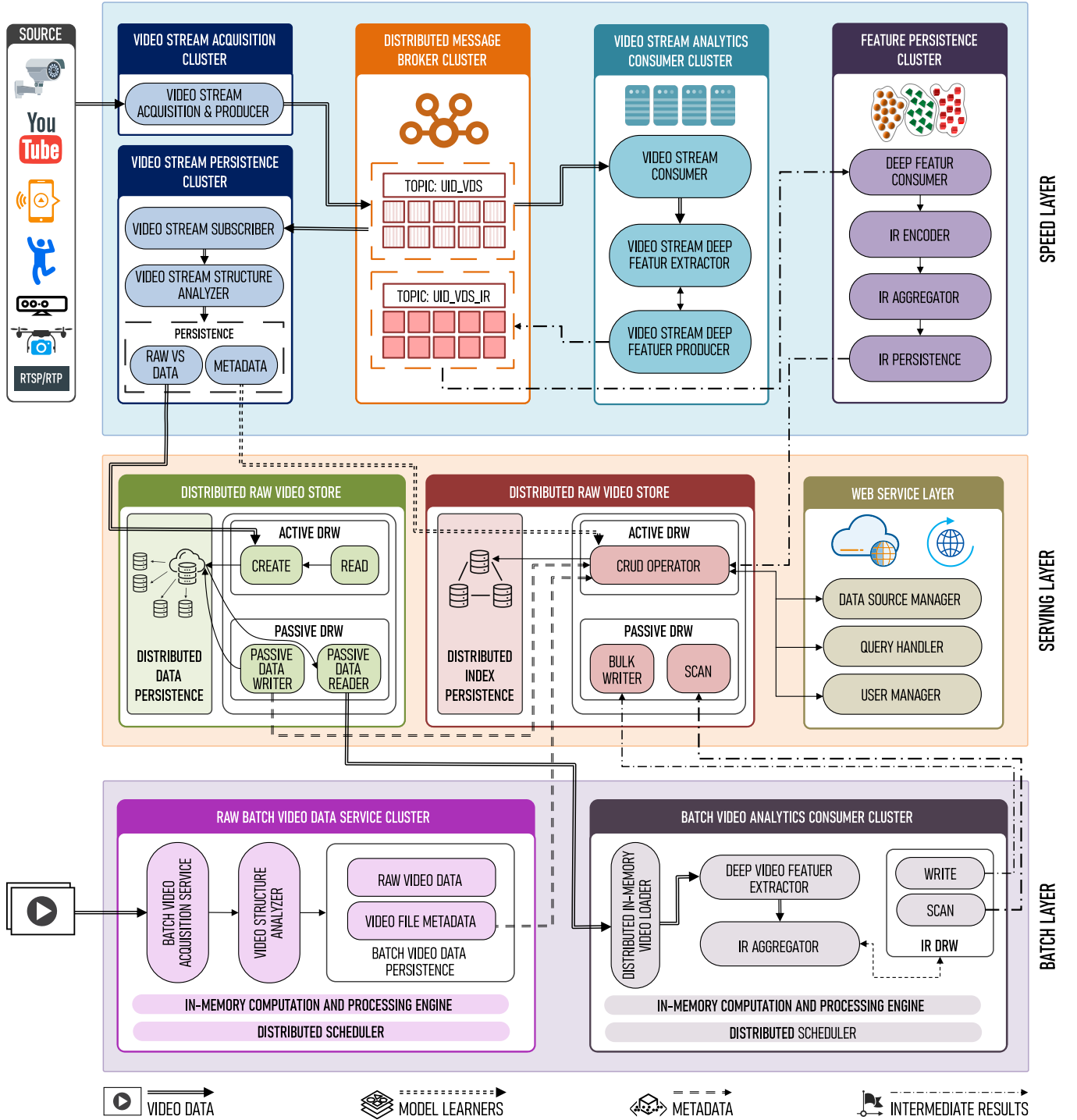


Fig. S1. λ -CLOVR Execution Scenarios under Lambda Architecture.