

**Q. How does Amazon Prime video save 90% cost by moving from serverless to the monolith?**

Before we move on to understand how Amazon Prime video saves 90% cost by moving from serverless to monolith architecture. Lets first understand

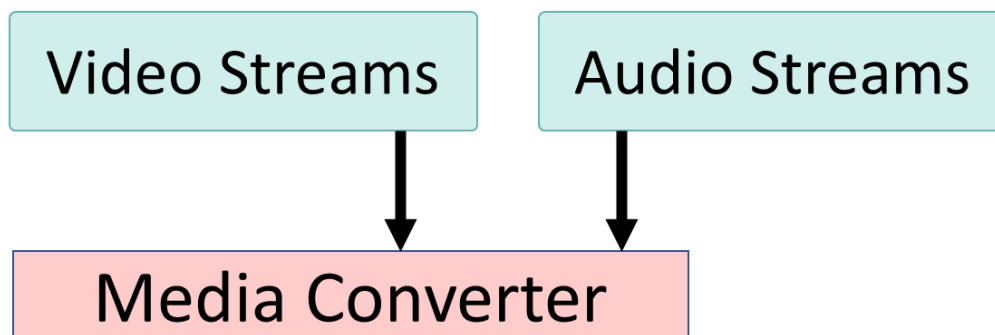
1. What was the initial design.
2. The problems with the initial design.

**Q. What was the initial design?**

Ans. Amazon Prime videos provide thousands of live streams to customers. They have a tool Defect Detector that takes audio and video streams. And sends a notification if there are quality issues (for example, block corruption or audio/video sync problems).

More inside view of design.

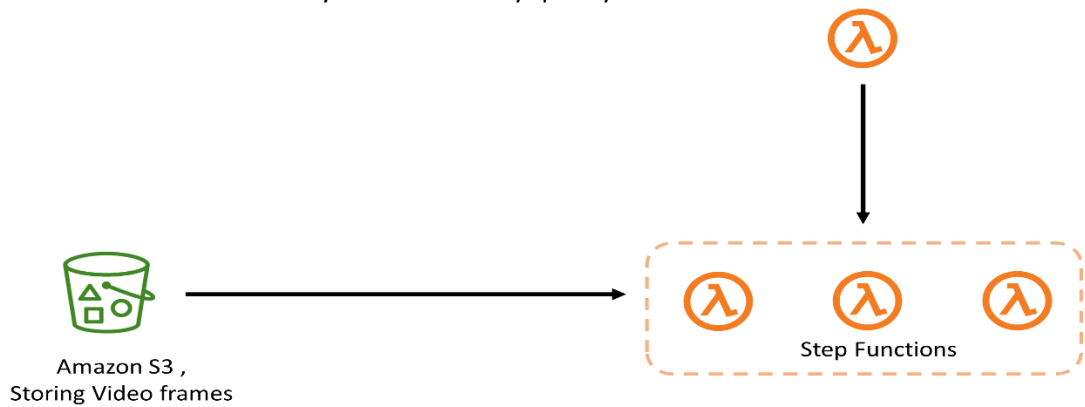
1. Video Quality Analysis Service takes video Streams and Audio Streams and sends the streams to Media Converter. As shown below.



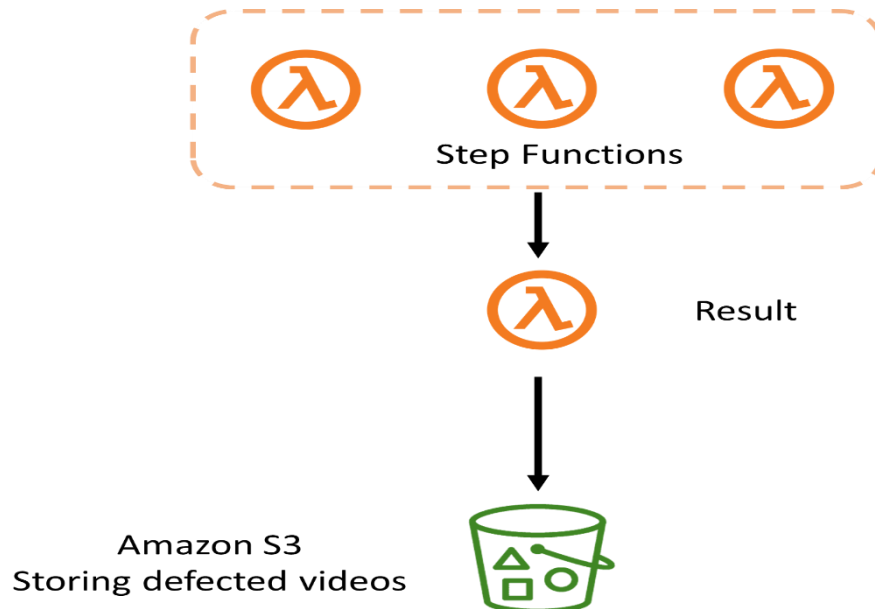
2. Media Converter combines audio and video streams to create video frames, once the frame is ready for a video the frame is saved in an amazon s3 bucket. And a message is sent to the defect detector orchestrator.



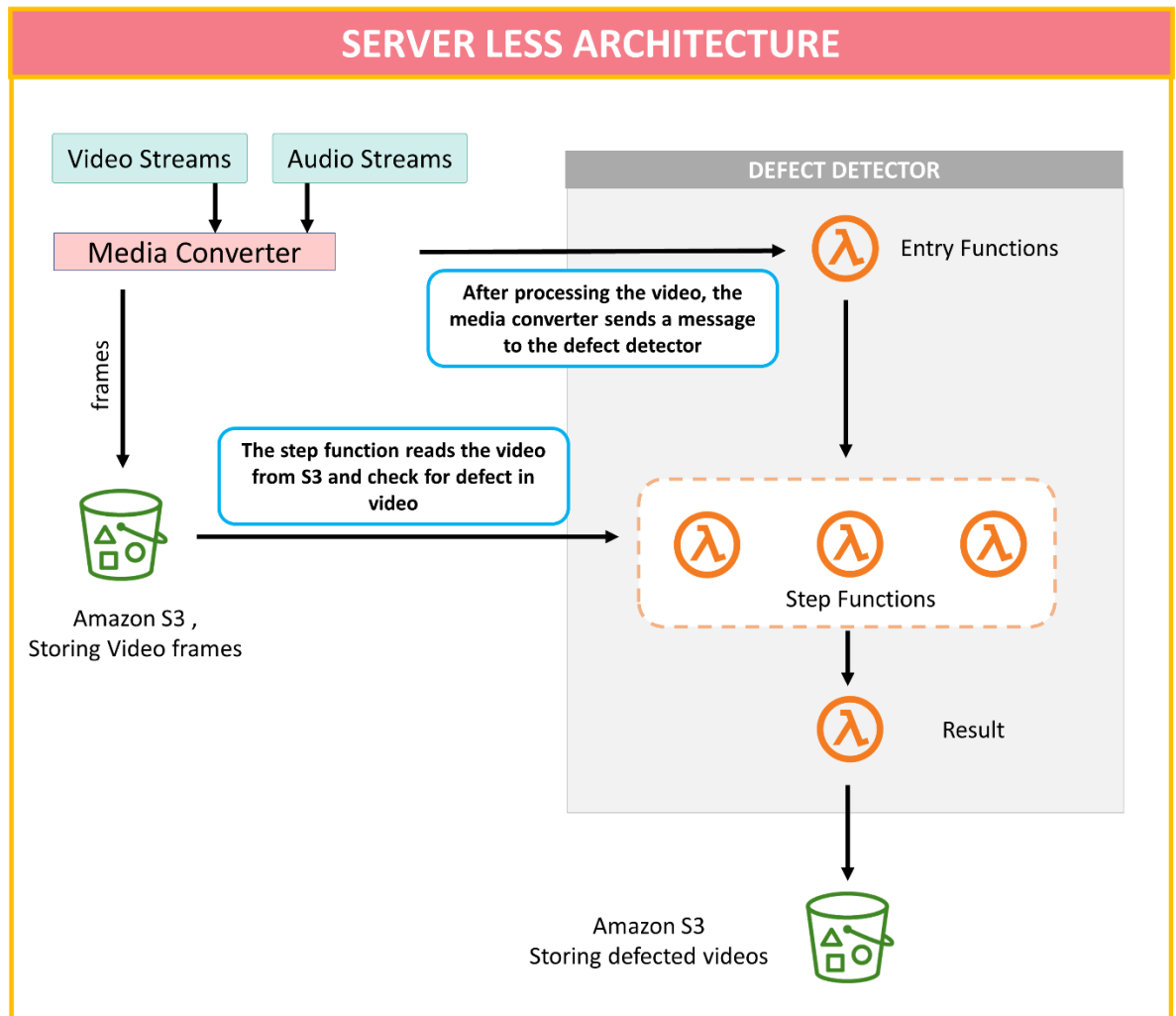
- Orchestrator sends a message to step functions to download the video frames from the Amazon S3 bucket and analyze if there is any quality issue.



- Finally if there are any quality issues a notification will be sent and a defective frame will be stored in the Amazon S3 bucket.



5. Below is the complete architecture.



**Q.) Why was the serverless or Defect detector initial design more expensive. ?**

Ans. Due to below two reasons the defect detector's initial design was more expensive:-

1. Increasing the number of step functions during horizontal scaling increases the cost. In the below image, you can see there are only three-step functions. If you increase the number of step functions to 5 to process more videos, the cost will also increase.



2. The data transfer between step functions is expensive. In the below image in some scenarios, data has to be passed from one step function to another. Because one step function might have the logic to process corrupted blocks and the other step function will

check for audio video sync issues

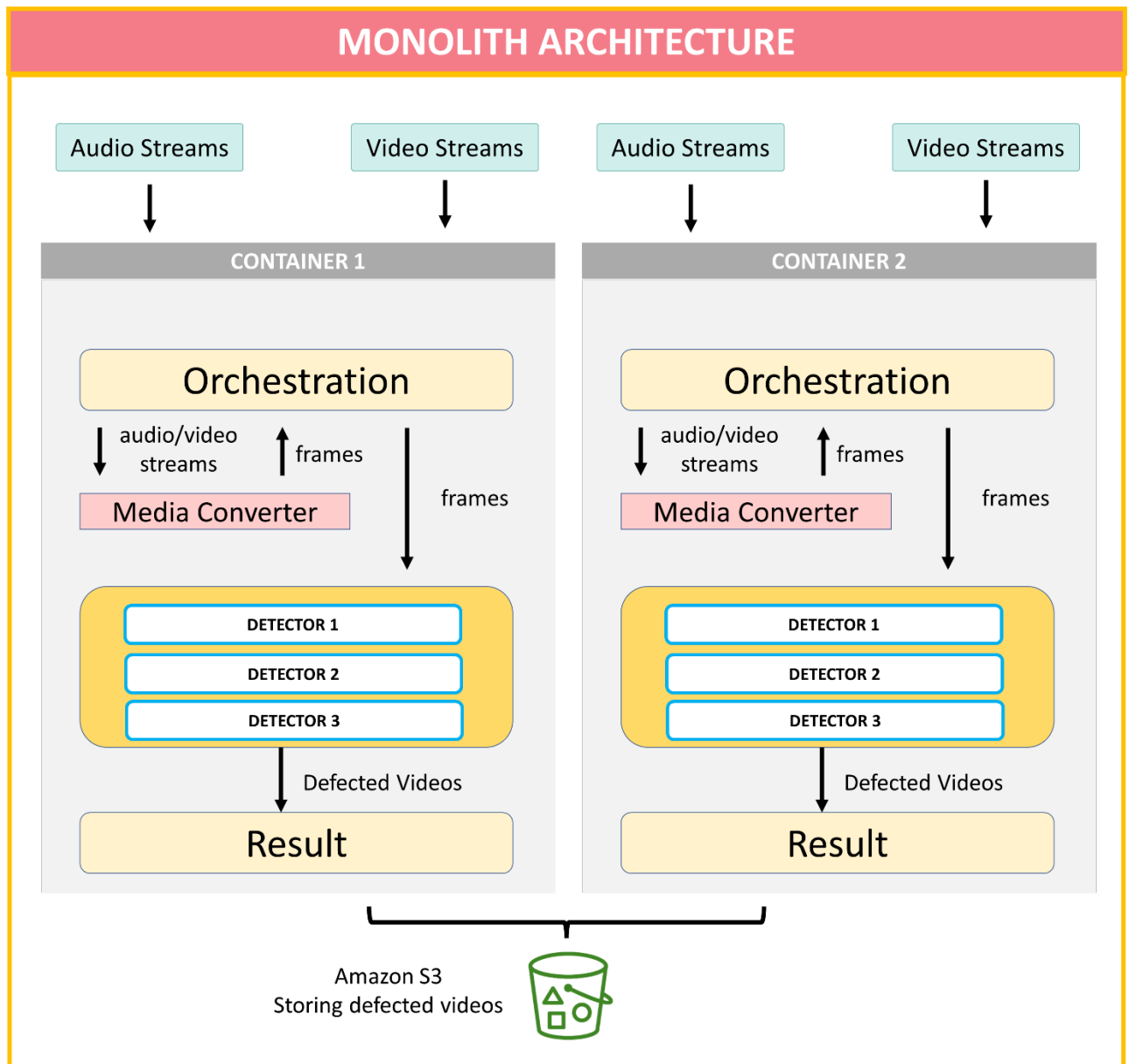


3. Making a lot of calls to an amazon s3 bucket, is expensive. In the below image when every calls are made to the amazon s3 bucket to download the videos, leads to more network calls, and more network calls costs more.



**Q.) How improved architecture work?**

Ans. Audio and video streams are sent to the orchestrator, the orchestrator will run in a container. The orchestrator will send the audio/video streams to the media converter and the media converter will send the video frames back to the orchestrator. Then orchestrator will send the video frames to defect detectors, defect detectors will check for video quality issues, and if there are any video quality issues a notification will be sent and the corrupted frames will be stored in Amazon s3. Check the below image to learn more.



**Q. How does Improved architecture save 90% cost by moving from serverless to the monolith?**

Ans: Below are the reasons improved architecture helps in saving 90% of costs.

1. Reducing costly network calls: As all the orchestrators, media converters, and defect detector works inside one container only reduces the costly network call to amazon s3 bucket. As in new architectures, the whole video frames are stored in the memory of the container only.
2. Horizontal scaling of step functions is more expensive: In the old architecture to analyze more amount of videos at a single point in time the number of step functions needs to be increased which leads to more costs. But in new architecture to handle large scale more parallel containers can be spined up. Horizontal scaling of containers is less expensive than horizontal scaling of step functions.