My Zeus: A Product Architecture and Implementation Report

# Executive Summary

MyZeus is an ecosystem designed to unlock the immense, untapped potential within video data. Our mission is to transform passive video from a static record into a dynamic, queryable, and actionable resource. We will achieve this through a synergistic suite of two products: the Aether Cloud, a powerful infrastructure-as-a-service (IaaS) for comprehensive video analysis, and the Argus Sentinel, a sophisticated application that leverages Aether to provide users with unprecedented capabilities in intelligent monitoring, surveillance, and agentic automation. MyZeus will serve as the central nervous system for the next generation of visually-aware AI applications.

# Part 1: The Aether Cloud - Foundational Vision Intelligence

## 1.1 Vision and Purpose

The Aether Cloud is the engine of the MyZeus ecosystem. It is a highly scalable, robust, and secure cloud platform built on Amazon Web Services (AWS), designed to ingest, process, index, and comprehend video at an industrial scale. It is not merely storage; it is an active intelligence layer. Developers and businesses will connect to Aether via a comprehensive API to power their own applications, while our flagship application, Argus Sentinel, will run on its core.

A screenshot of a cloud computing system

AI-generated content may be incorrect.

## 1.2 Core Architecture on AWS

Ingestion & Storage:

1. AWS S3 (Simple Storage Service): The primary data lake for storing raw video files. It provides durability, scalability, and cost-effective storage.
2. AWS Kinesis Video Streams: For ingesting and processing real-time video streams from thousands of connected cameras simultaneously.

Compute & Processing:

1. Amazon EC2 (Elastic Compute Cloud): GPU-enabled instances (P4d, G5 series) will be used for the heavy-lifting of AI model inference and video transcoding.
2. AWS Lambda: For event-driven processing. For example, a Lambda function can be triggered upon a new video upload to S3 to kick off the analysis pipeline.
3. AWS Batch: To manage and execute large-scale, asynchronous batch processing jobs, such as re-indexing an entire archive with a new AI model.

Database & Indexing:  
Amazon Aurora/RDS: A relational database to store structured metadata: user information, camera details, file locations, timestamps, and job statuses.  
Amazon OpenSearch Service (or a dedicated Vector Database): This is the core of our search capability. It will store and index text from transcriptions (for keyword search) and, most importantly, the high-dimensional vector embeddings generated by our AI models for powerful semantic search.

API & Security:  
Amazon API Gateway: To create, manage, and secure our RESTful and WebSocket APIs for both internal and external use.  
AWS Cognito & IAM: For robust user authentication, authorization, and fine-grained access control to ensure data privacy and security.

## 1.3 The Video Intelligence Pipeline

When a video enters the Aether Cloud, it undergoes a multi-stage, automated analysis pipeline:

Ingestion & Normalization: The video (either an uploaded file or a live stream) is received. AWS Elemental MediaConvert is used to transcode it into a standardized format and resolution for consistent processing.  
Scene Detection & Frame Analysis: The video is segmented into logical scenes and shots. Keyframes are extracted at regular intervals or when significant visual changes occur.  
Multimodal Feature Extraction (The AI Core): This is where the deep analysis occurs. A suite of specialized AI models runs in parallel:  
Audio Processing: The audio track is separated. A state-of-the-art transcription model (e.g., a fine-tuned Whisper) converts speech to text. Speaker diarization identifies who is speaking and when.  
Object & Entity Recognition: Models like YOLO (You Only Look Once) or DETR are used to identify and tag thousands of objects, people, and animals within each frame.  
Action Recognition: Spatiotemporal models (e.g., Timesformer) analyze sequences of frames to identify complex actions and events, such as 'a person is running,' 'a car is turning left,' or 'a baby is waking up.'  
Facial Recognition & Attribution (Opt-In): For private libraries, users can opt-in to create a catalog of individuals. The system can then identify and tag these people across their entire video collection.  
Text-in-Video (OCR): Optical Character Recognition models extract and index any text that appears visually, such as on signs, whiteboards, or computer screens.  
Vectorization & Indexing: The outputs from all AI models are converted into vector embeddings using multimodal models like CLIP. These vectors, which represent the semantic meaning of the content, are stored in our vector database. All other metadata (transcripts, object tags, timestamps) is sent to Aurora and OpenSearch.

# Part 2: The Argus Sentinel - Applied Intelligent Monitoring

## 2.1 Product Vision

Argus Sentinel is the command center for MyZeus users. It is a sophisticated, cross-platform desktop and mobile application that provides a user-friendly interface to the immense power of the Aether Cloud. With Argus, users can connect their video sources, define monitoring tasks, assign AI agents, and search their visual world with simple, natural language. It is designed for everyone from a parent monitoring a nursery to a scientist automating data collection to a factory manager overseeing an assembly line.

## 2.2 Key Features & System Design

Unified Dashboard: A central hub to view all connected camera streams, recent events, and agent statuses in real-time.  
Effortless Camera Connectivity:  
IP Cameras: Natively supports common protocols like RTSP and ONVIF, allowing users to connect most modern security cameras by simply entering their IP address and credentials.  
Desktop Client: A lightweight application that can capture a user's screen, a specific window, or a connected webcam. This client will leverage core components from the OBS (Open Broadcaster Software) framework for its robust and efficient screen/video capture capabilities, which then securely streams the feed to the Aether Cloud.  
Mobile App: The MyZeus mobile app can use the phone's camera as a live video source, perfect for on-the-fly monitoring tasks like the workout analysis example.

A computer monitor on a desk

AI-generated content may be incorrect.

**The Agentic Workflow Builder**: This is the core interactive feature. It's a no-code/low-code interface where users define tasks for their AI agents. It uses a simple 'If-This-Then-That' logic.

A computer screen with a hand on a mouse

AI-generated content may be incorrect.

Trigger (The 'If'): A user can select from a library of triggers or define their own. Examples: 'If a person is detected,' 'If the sound of crying is heard for more than 30 seconds,' 'If the text 'Build Complete' appears on screen.'  
Action (The 'That'): Define what happens when the trigger condition is met. Examples: 'Send a push notification with a snapshot,' 'Send an email to my team,' 'Log the event to a Google Sheet,' 'Trigger a webhook to an external system.'  
The Agentic System in Practice:  
A user creates a task, e.g., 'Alert me when my 3D print is finished.'  
The user connects their webcam pointing at the printer via the Argus desktop client.  
In the Workflow Builder, they set the trigger: 'If the text on the printer's LCD screen contains the words 'Print Complete'.'  
They set the action: 'Send a push notification to my phone.'  
Upon saving, Argus instructs the Aether Cloud to deploy a persistent, asynchronous 'agent' (a lightweight monitoring process). This agent continuously analyzes the Kinesis Video Stream from that specific webcam, running the OCR model on the relevant portion of the frames. When the text is detected, it executes the defined action via AWS SNS (Simple Notification Service).