

Video Processing Data Base

Andres Lozano

Octubre, 2025

Índice

Overview	3	Triggers	10
Objetivos	4	Stored Procedures	11
Situación problemática	5	Tablas	12
Solución	6	Métricas	28
Diagrama de Entidad Relación	7	Herramientas utilizadas	32
Vistas	8	Conclusión	33
Funciones	9		

Overview

Video Processing Pipeline es un sistema de base de datos diseñado para automatizar y gestionar la transcodificación de archivos multimedia a gran escala.

El sistema permite:

- Recibir videos en diversos formatos
- Procesarlos según diferentes perfiles de calidad
- Generar múltiples versiones optimizadas para distintos dispositivos y plataformas de streaming
- Gestionar clientes, usuarios, proyectos y cuotas de almacenamiento
- Mantener auditoría completa de todas las operaciones
- Notificar eventos relevantes automáticamente

Objetivos

Diseñar e implementar una base de datos relacional que permita:

- **Gestión de clientes y usuarios:** Sistema multiusuario con roles y permisos
- **Gestión de proyectos:** Organización de archivos por proyectos y clientes
- **Gestión de archivos fuente:** Almacenar metadatos de videos en múltiples ubicaciones
- **Control de trabajos de codificación:** Administrar cola de procesamiento con prioridades
- **Seguimiento de calidad:** Registrar métricas (PSNR, SSIM, VMAF)
- **Monitoreo de recursos:** Controlar utilización de servidores de procesamiento
- **Trazabilidad completa:** Desde archivo original hasta versiones finales
- **Auditoría:** Registro de todas las acciones realizadas
- **Notificaciones:** Alertas automáticas sobre eventos importantes

Situación Problemática

Las empresas de streaming y producción multimedia enfrentan desafíos al procesar grandes volúmenes de video:

Problemas sin sistema estructurado:

- Pérdida de trazabilidad en el procesamiento
- Duplicación de trabajo por falta de control de estado
- Ineficiencia en asignación de recursos computacionales
- Dificultad para medir calidad y rendimiento
- Falta de visibilidad en tiempos de procesamiento
- Ausencia de control sobre cuotas y facturación
- Imposibilidad de auditar acciones

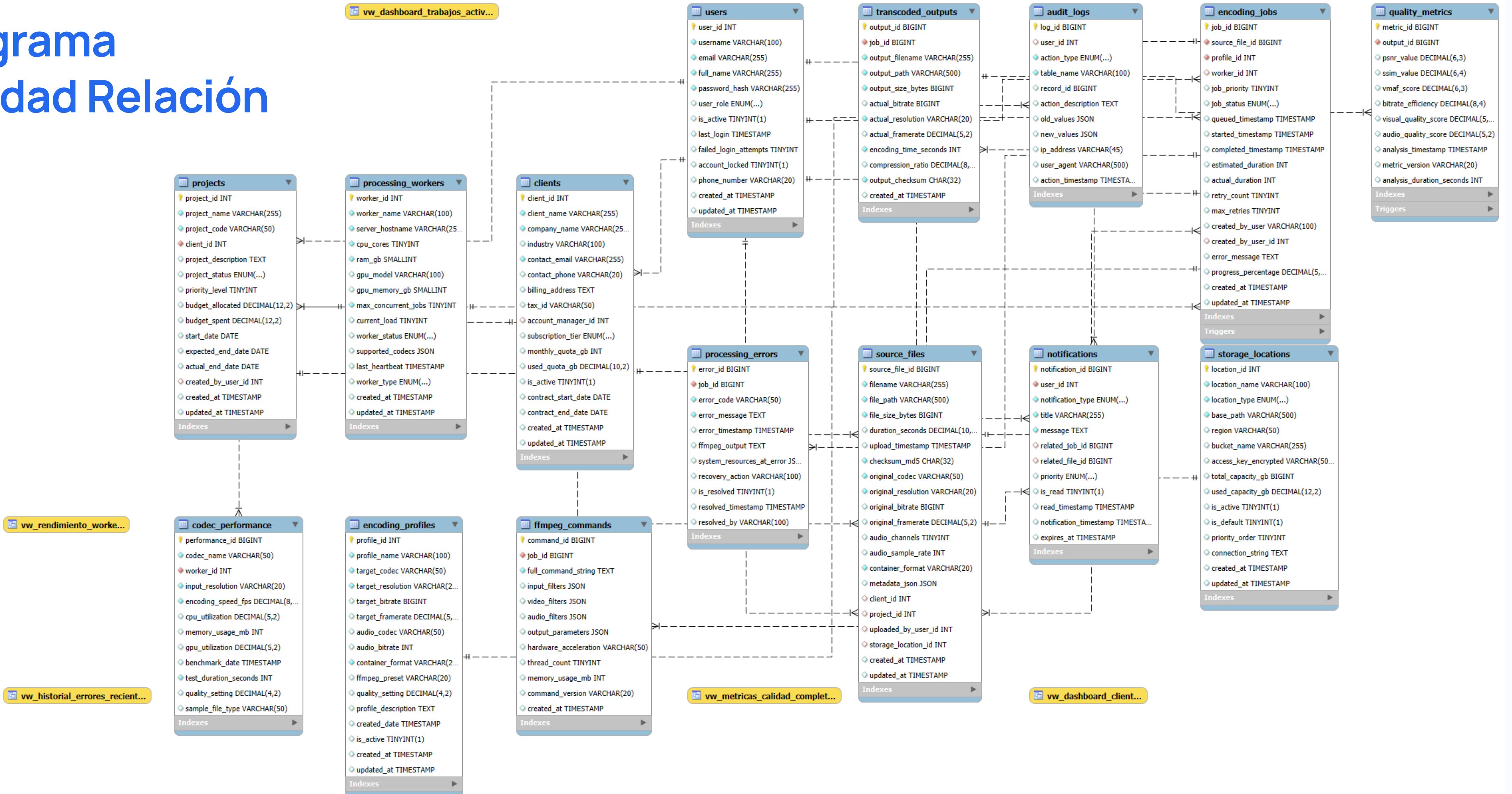


Solución

La base de datos implementada plantea la siguiente solución:

- **Centralizar información** del pipeline de procesamiento
- **Automatizar** gestión de colas con prioridades
- Proporcionar **métricas detalladas** de calidad
- **Optimizar** utilización de recursos
- **Gestionar** clientes, proyectos y cuotas
- Mantener **auditoría completa**
- **Notificar** eventos relevantes automáticamente

Diagrama Entidad Relación



Funcionalidades

Vistas

VW_DASHBOARD_TRABAJOS_ACTIVOS

Vista consolidada de trabajos en procesamiento para monitoreo en tiempo real.

VW_METRICAS_CALIDAD_COMPLETAS

Consolida métricas de calidad con información contextual para análisis comparativo.

VW_RENDIMIENTO_WORKERS

Analiza rendimiento y utilización de cada worker.

VW_HISTORIAL_ERRORES_RECIENTES

Monitorea errores recientes para identificar patrones.

VW_DASHBOARD_CLIENTES

Vista consolidada de actividad y uso por cliente para gestión de cuentas.

Funcionalidades

Funciones

FN_CALCULAR_EFICIENCIA_COMPRESION

Calcula el porcentaje de reducción de tamaño tras la compresión.

FN_OBTENER_TIEMPO_ESPERA_ESTIMADO

Estima tiempo de espera para trabajos en cola según prioridad.

FN_VERIFICAR_DISPONIBILIDAD_WORKER

Verifica si un worker está disponible para nuevos trabajos.

Funcionalidades

Triggers

TRG_VALIDAR_WORKER_ANTES_ASIGNACION

Valida que el worker asignado esté activo y tenga capacidad

TRG_ACTUALIZAR_CARGA_WORKER_COMPLETADO

Decrementa carga del worker cuando trabajo finaliza.

TRG_REGISTRAR_ERROR_TRABAJO_FALLIDO

Registra automáticamente error cuando trabajo falla.

Funcionalidades

Stored Procedures

SP_ASIGNAR_TRABAJO_A_WORKER

Asigna automáticamente un trabajo al worker más apropiado.

SP_REPORTE_RENDIMIENTO_PERIODO

Genera reporte completo de rendimiento del sistema.

SP_REINTENTAR_TRABAJOS_FALLIDOS

Reintenta automáticamente trabajos fallidos elegibles.

Tablas

Audit Logs

Trazabilidad completa de acciones en el sistema.

CAMPO	log_id	user_id	action_type	table_name	record_id	action_description	old_values
TIPO DE DATO	BIGINT	INT	ENUM	VARCHAR	BIGINT	TEXT	JSON
DESCRIPCIÓN	id del log	id del usuario	tipo de acción	nombre de la tabla	id del récord	descripción de la acción	valores pasados



CAMPO	new_values	ip_address	user_agent	action_timestamp
TIPO DE DATO	JSON	VARCHAR	VARCHAR	TIMESTAMP
DESCRIPCIÓN	valores nuevos	direccion ip	modelo explorador web	fecha y hora de accion

Clients

Administra empresas/organizaciones que usan el servicio.

CAMPO	client_id	client_name	company_name	industry	contact_email	contact_phone	billing_address	tax_id	account_manager_id
TIPO DE DATO	INT	VARCHAR	VARCHAR	VARCHAR	VARCHAR	VARCHAR	TEXT	VARCHAR	INT
DESCRIPCIÓN	id del cliente	nombre del cliente	nombre de la empresa	industria	mail de contacto	numero de teléfono	direccion de cobro	tax id	manager id



CAMPO	subscription_tier	monthly_quota_gb	used_quota_gb	is_active	contract_start_date	contract_end_date	created_at	updated_at
TIPO DE DATO	ENUM	INT	DECIMAL	BOOLEAN	DATE	DATE	TIMESTAMP	TIMESTAMP
DESCRIPCIÓN	nivel de suscripción	uso mensual contratado	uso de la cuota contratada	cliente activo o no	fecha de inicio de contrato	fecha de finalización de contrato	fecha de creación	fecha de actualización

Codec Performance

Métricas de rendimiento de códecs en workers.

CAMPO	performance_id	codec_name	worker_id	input_resolution	encoding_speed_fps	cpu_utilization	memory_usage_mb
TIPO DE DATO	BIGINT	VARCHAR	INT	VARCHAR	DECIMAL	DECIMAL	INT
DESCRIPCIÓN	id de performance	nombre del codec	id del worker	resolución del input	velocidad de encoding	porcentaje de cpu utilizada	cantidad de memoria utilizada



CAMPO	gpu_utilization	benchmark_date	test_duration_seconds	quality_setting	sample_file_type
TIPO DE DATO	DECIMAL	TIMESTAMP	INT	DECIMAL	VARCHAR
DESCRIPCIÓN	uso de gpu	fecha y hora del benchmark	duración del test	seteo de calidad	genero del archivo

Encoding Jobs

Tabla central que gestiona trabajos de codificación.

CAMPO	job_id	source_file_id	profile_id	worker_id	job_priority	job_status	queued_timestamp	started_timestamp
TIPO DE DATO	BIGINT	BIGINT	INT	INT	TINYINT	ENUM	TIMESTAMP	TIMESTAMP
DESCRIPCIÓN	id del job	id del archivo source	id del perfil	id del worker	prioridad del job	status del job	fecha y hora del archivo en la cola	fecha y hora de comienzo del job

Encoding Profiles

Define perfiles de codificación disponibles.

CAMPO	profile_id	profile_name	target_codec	target_resolution	target_bitrate	target_framerate	audio_codec	audio_bitrate
TIPO DE DATO	INT	VARCHAR	VARCHAR	VARCHAR	BIGINT	DECIMAL	VARCHAR	INT
DESCRIPCIÓN	id del perfil del codec	nombre del perfil del codec	tipo de codec	resolución del codec	bitrate apuntado	framerate apuntado	codec de audio	bitrate de audio



CAMPO	container_format	ffmpeg_preset	quality_setting	profile_description	created_date	is_active	created_at	updated_at
TIPO DE DATO	VARCHAR	VARCHAR	DECIMAL	TEXT	TIMESTAMP	BOOLEAN	TIMESTAMP	TIMESTAMP
DESCRIPCIÓN	tipo de formato	preset de ffmpeg	seteo de calidad	descripción del perfil	fecha de creación	activo si/no	fecha y hora de creación	fecha y hora de actualización

Ffmpeg Commands

Comandos FFmpeg ejecutados para cada trabajo.

CAMPO	command_id	job_id	full_command_string	input_filters	video_filters	audio_filters
TIPO DE DATO	BIGINT	BIGINT	TEXT	JSON	JSON	JSON
DESCRIPCION	id del comando	id del job	comando completo de ffmpeg	filtros generales	filtros de video	filtros de audio



CAMPO	output_parameters	hardware_acceleration	thread_count	memory_usage_mb	command_version	created_at
TIPO DE DATO	JSON	VARCHAR	TINYINT	INT	VARCHAR	TIMESTAMP
DESCRIPCION	parametros de output	aceleración por hardware (nvec de nvidia)	cantidad de hilos	uso de memoria	version del comando	fecha y hora de creación

Notifications

Sistema de notificaciones para eventos relevantes.

CAMPO	notification_id	user_id	notification_type	title	message	related_job_id	related_file_id	priority
TIPO DE DATO	BIGINT	INT	ENUM	VARCHAR	TEXT	BIGINT	BIGINT	ENUM
DESCRIPCIÓN	id de la notificación	id del usuario	tipo de notificación	titulo	mensaje	relación job/id	relación file/id	prioridad



CAMPO	is_read	read_timestamp	notification_timestamp	expires_at
TIPO DE DATO	BOOLEAN	TIMESTAMP	TIMESTAMP	TIMESTAMP
DESCRIPCIÓN	leido si/no	fecha y hora de leído	fecha y hora de la notificación	fecha y hora de expiración

Processing Errors

Registro de errores durante procesamiento.

CAMPO	error_id	job_id	error_code	error_message	error_timestamp	ffmpeg_output	system_resources_at_error
TIPO DE DATO	BIGINT	BIGINT	VARCHAR	TEXT	TIMESTAMP	TEXT	JSON
DESCRIPCIÓN	id del error	id del job	codigo del error	mensaje del error	fecha y hora del error	output que da ffmpeg	cpu/espacio en el disco/ memoria al momento del error



CAMPO	recovery_action	is_resolved	resolved_timestamp	resolved_by
TIPO DE DATO	VARCHAR	BOOLEAN	TIMESTAMP	TIMESTAMP
DESCRIPCIÓN	mensaje con los reintentos o acciones que tomo el usuario	terminado si/no	fecha y hora de terminado	terminado por que usuario

Processing Workers

Servidores disponibles para procesamiento de video.

CAMPO	worker_id	worker_name	server_hostname	cpu_cores	ram_gb	gpu_model	gpu_memory_gb	max_concurrent_jobs
TIPO DE DATO	INT	VARCHAR	VARCHAR	TINYINT	SMALLINT	VARCHAR	SMALLINT	TINYINT
DESCRIPCIÓN	id del worker	nombre del worker	nombre del servidor	cantidad de cores cpu	cantidad de ram	model de gpu	memoria de gpu	máximos jobs concurrentes

CAMPO	current_load	worker_status	supported_codecs	last_heartbeat	worker_type	created_at	ended_at
TIPO DE DATO	TINYINT	ENUM	JSON	TIMESTAMP	ENUM	TIMESTAMP	TIMESTAMP
DESCRIPCIÓN	cantidad de carga en el momento	status del worker	codecs soportados	ultimo heartbeat	tipo de worker	fecha y hora de creación	fecha y hora de finalización

Projects

Agrupa archivos de video organizados por cliente.

CAMPO	project_id	project_name	project_code	client_id	project_description	project_status	priority_level	budget_allocated
TIPO DE DATO	INT	VARCHAR	VARCHAR	INT	TEXT	ENUM	TINYINT	DECIMAL
DESCRIPCIÓN	id del proyecto	nombre del proyecto	código del proyecto	id del cliente	descripción del proyecto	status del proyecto	nivel de prioridad	presupuesto



CAMPO	budget_spent	start_date	expected_end_date	actual_end_date	created_by_user_id	created_at	updated_at
TIPO DE DATO	DECIMAL	DATE	DATE	DATE	INT	TIMESTAMP	TIMESTAMP
DESCRIPCIÓN	presupuesto gastado	fecha de inicio	fecha de finalización esperada	fecha de finalización real	usuario creador	fecha y hora de creación	fecha y hora de actualización

Quality Metrics

Métricas de calidad calculadas para cada video.

CAMPO	metric_id	output_id	psnr_value	ssim_value	vmaf_score	bitrate_efficiency	visual_quality_score
TIPO DE DATO	BIGINT	BIGINT	DECIMAL	DECIMAL	DECIMAL	DECIMAL	DECIMAL
DESCRIPCIÓN	id de métrica	id del output	valor psnr	valor ssim	valor vmaf	eficacia del bitrate	puntaje de calidad de video



CAMPO	audio_quality_score	analysis_timestamp	metric_version	analysis_duration_seconds
TIPO DE DATO	DECIMAL	TIMESTAMP	VARCHAR	INT
DESCRIPCIÓN	puntaje de calidad de audio	fecha y hora del análisis	versión de la métrica	duración del análisis

Source Files

Almacena información de archivos de video originales.

CAMPO	source_file_id	filename	file_path	file_size_bytes	duration_seconds	upload_timestamp	checksum_md5	original_codec
TIPO DE DATO	BIGINT	VARCHAR	VARCHAR	BIGINT	DECIMAL	TIMESTAMP	CHAR	VARCHAR
DESCRIPCIÓN	id del archivo source	nombre del archivo	ruta del archivo	tamaño del archivo	duración del archivo	fecha y hora de carga	hash	codec original

CAMPO	original_resolution	original_bitrate	original_framerate	audio_channels	audio_sample_rate	container_format	metadata_json
TIPO DE DATO	VARCHAR	BIGINT	DECIMAL	TINYINT	INT	VARCHAR	JSON
DESCRIPCIÓN	resolución original	bitrate original	framerate original	cantidad de canales de audio	sample rate del audio	formato del archivo	metadata

CAMPO	client_id	project_id	uploaded_by_user_id	storage_location_id	created_at	updated_at
TIPO DE DATO	INT	INT	INT	INT	TIMESTAMP	TIMESTAMP
DESCRIPCIÓN	id del cliente	id del proyecto	usuario que subió el archivo	lugar donde se guarda el archivo	fecha y hora de creacion	fecha y hora de actualización

Storage Locations

Gestiona ubicaciones de almacenamiento (cloud y local).

CAMPO	location_id	location_name	location_type	base_path	region	bucket_name	access_key_encrypted	total_capacity_gb
TIPO DE DATO	INT	VARCHAR	ENUM	VARCHAR	VARCHAR	VARCHAR	VARCHAR	BIGINT
DESCRIPCIÓN	id del bucket de guardado	nombre del bucket	tipo de bucket	path del bucket	region	nombre del bucket	llave de acceso	capacidad total



CAMPO	used_capacity_gb	is_active	is_default	priority_order	connection_string	created_at	updated_at
TIPO DE DATO	DECIMAL	BOOLEAN	BOOLEAN	TINYINT	TEXT	TIMESTAMP	TIMESTAMP
DESCRIPCIÓN	capacidad en uso	activo si/no	default si/no	orden de prioridad	string de conexion	fecha y hora de creacion	fecha y hora de actualización

Transcoded Outputs

Archivos generados tras la codificación.

CAMPO	output_id	job_id	output_filename	output_path	output_size_bytes	actual_bitrate	actual_resolution
TIPO DE DATO	BIGINT	BIGINT	VARCHAR	VARCHAR	BIGINT	BIGINT	VARCHAR
DESCRIPCIÓN	id del output	id del job	nombre del output	ruta del output	tamaño del output	bitrate final	resolución final



CAMPO	actual_framerate	encoding_time_seconds	compression_ratio	output_checksum	created_at
TIPO DE DATO	DECIMAL	INT	DECIMAL	CHAR	TIMESTAMP
DESCRIPCIÓN	framerate final	tiempo de encoding	ratio de compresión	output hash	fecha y hora de creacion

Users

Gestiona los usuarios que interactúan con el sistema

CAMPO	<code>user_id</code>	<code>username</code>	<code>email</code>	<code>full_name</code>	<code>password_hash</code>	<code>user_role</code>	<code>is_active</code>	<code>last_login</code>
TIPO DE DATO	INT	VARCHAR	VARCHAR	VARCHAR	VARCHAR	ENUM	BOOLEAN	TIMESTAMP
DESCRIPCIÓN	id del usuario	nombre	email	nombre completo	hash de la contraseña	rol del usuario	activo si/no	fecha y hora ultimo login

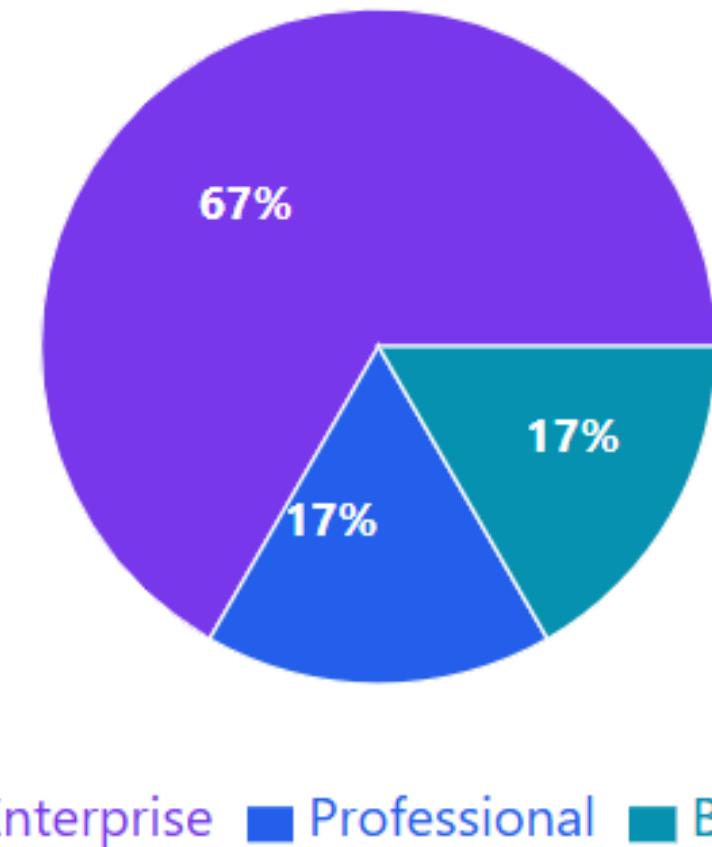


CAMPO	failed_login_attempts	account_locked	phone_number	created_at	updated_at
TIPO DE DATO	TINYINT	BOOLEAN	VARCHAR	TIMESTAMP	TIMESTAMP
DESCRIPCIÓN	intentos fallidos de login	cuenta bloqueada si/no	numero de teléfono	fecha y hora de creación	fecha y hora de actualización

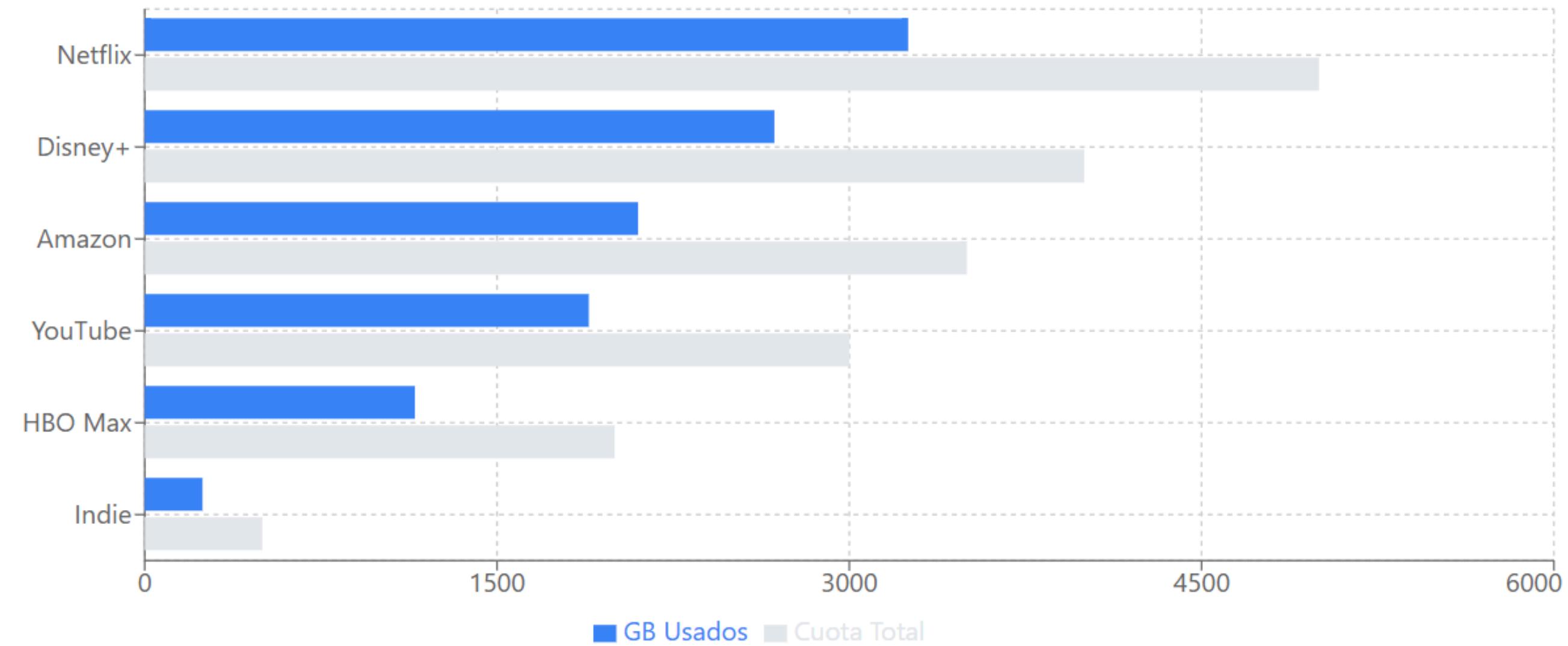
Métricas

Suscripciones y Clientes

Distribución por nivel de suscripción

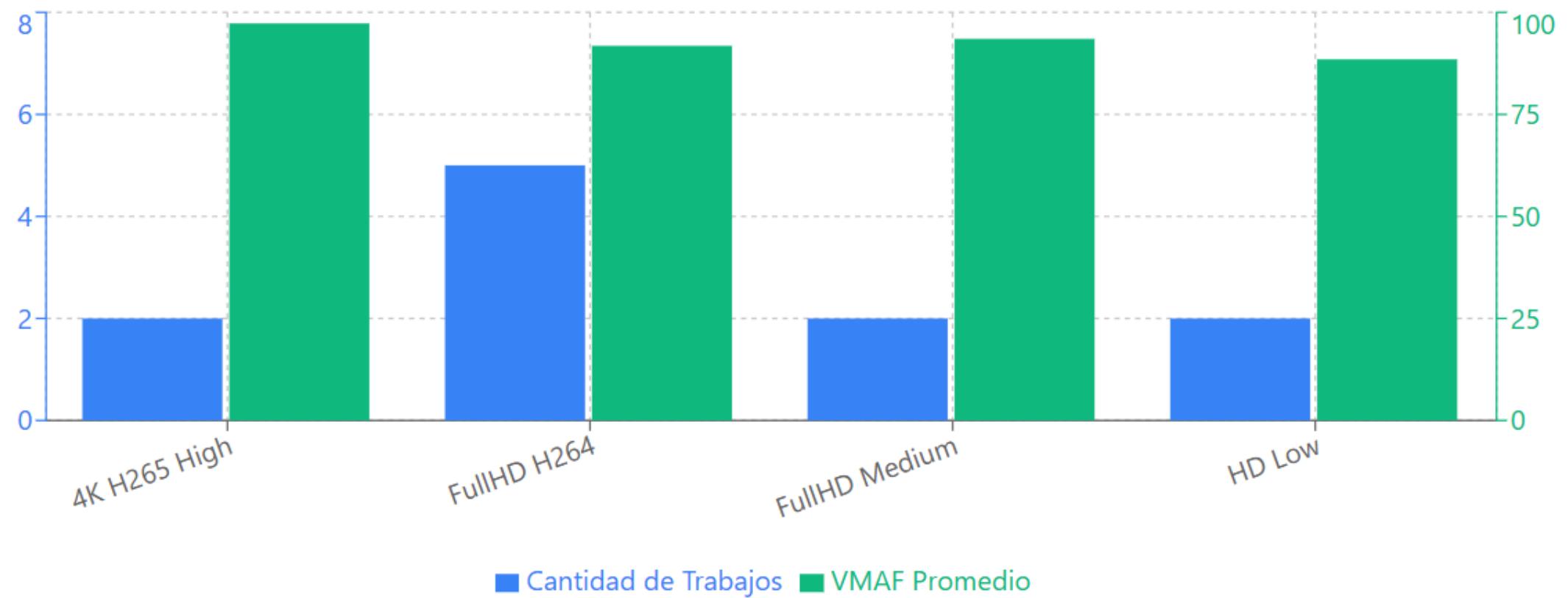


Uso de cuota por cliente (GB)

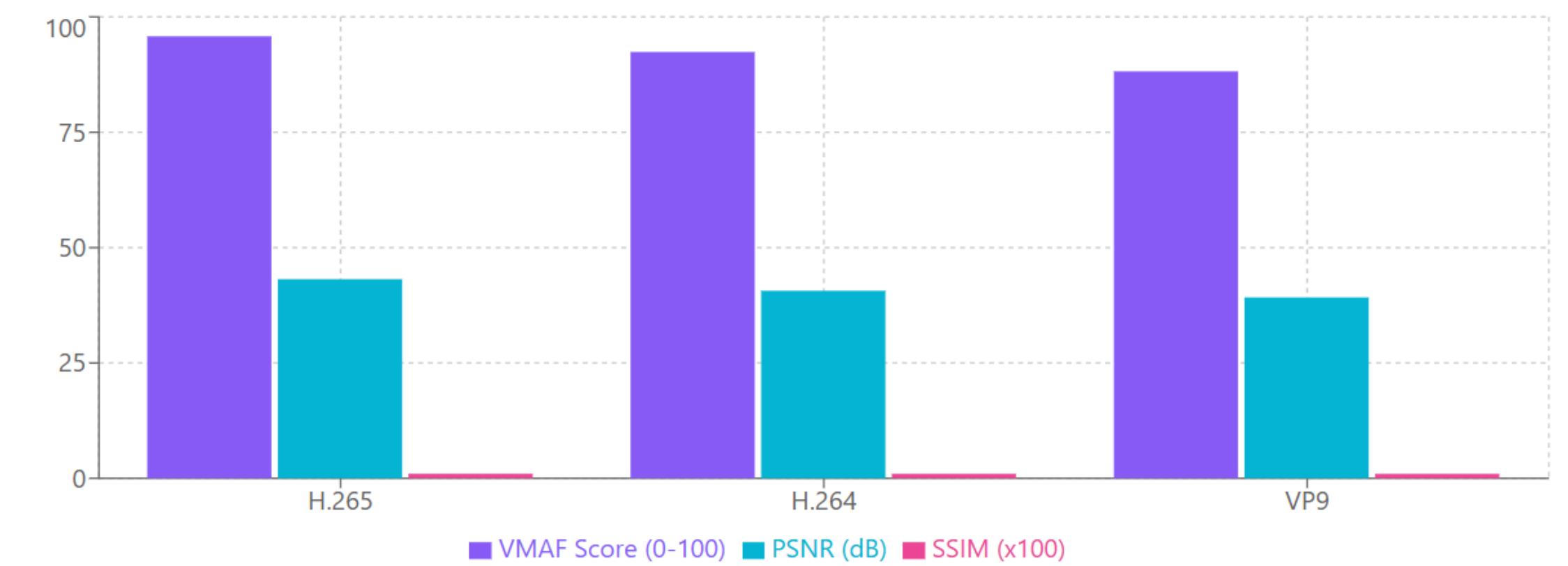


Codecs

Trabajos por perfil de codificación

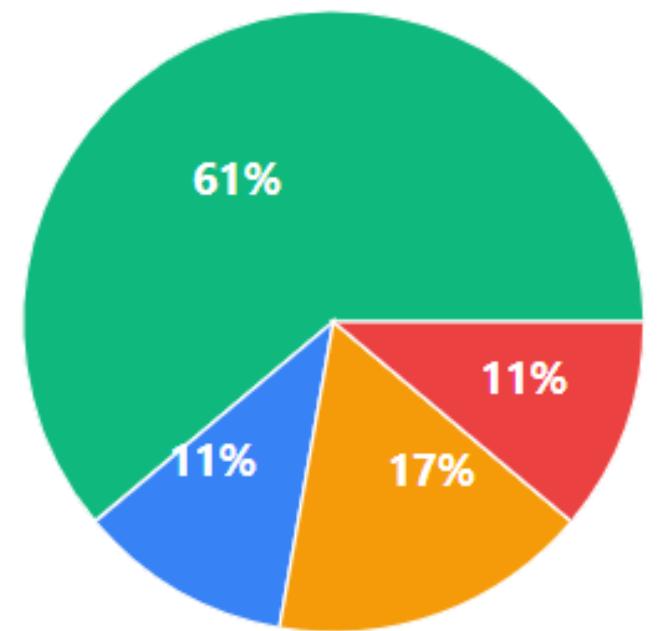


Calidad promedio por códec



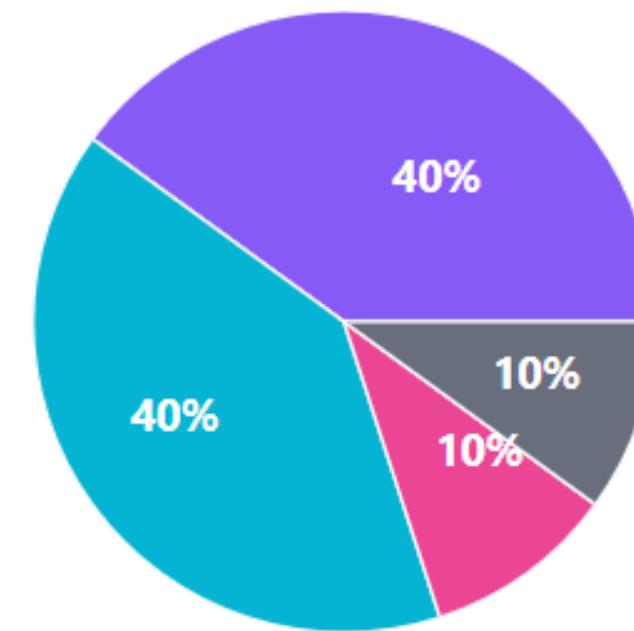
Distribución y Estado de los Workers

Distribución de estados de trabajos



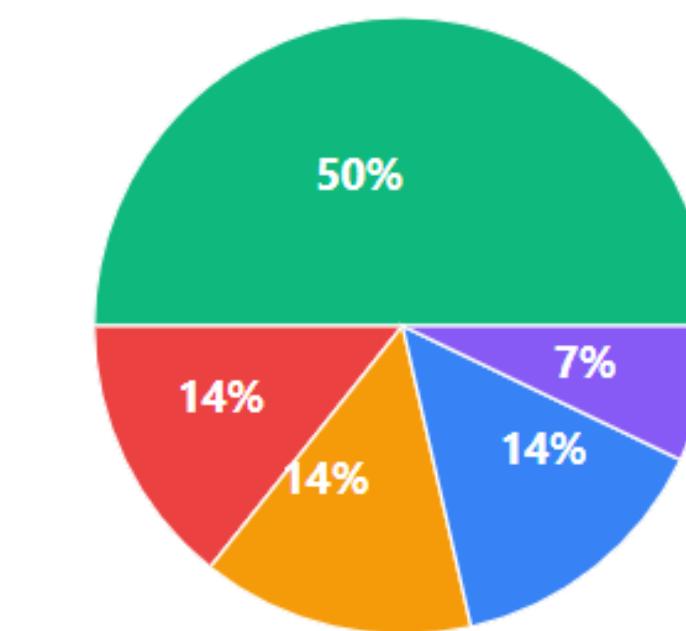
■ Completados ■ En Procesamiento ■ En Cola ■ Fallidos

Distribución por tipo de Worker



■ GPU Workers ■ CPU Workers ■ Hybrid Workers
■ En Mantenimiento

Notificaciones por tipo



■ Job Completed ■ Job Failed ■ Quota Warning
■ System Alert ■ Quality Issue

Herramientas utilizadas



MySQL Workbench para el diseño y testeo de la db



Claude para la generación de gráficos de métricas



Figma para la presentación

Conclusión

El sistema Video Processing Pipeline proporciona una **solución integral para gestión de procesamiento de video**.

CARACTERÍSTICAS PRINCIPALES:

- 15 tablas diseñadas
- 5 vistas para reporting
- 3 funciones útiles
- 3 stored procedures automatizados
- 4 triggers inteligentes
- Trazabilidad completa
- Gestión multiusuario
- Control de calidad con métricas estándar
- Optimización de recursos
- Auditoría completa
- Sistema de notificaciones

El sistema está listo para producción y puede manejar cargas empresariales.

¡Gracias!



Proyecto Final | Andrés Lozano

Octubre 2025

andyxinvierno@gmail.com