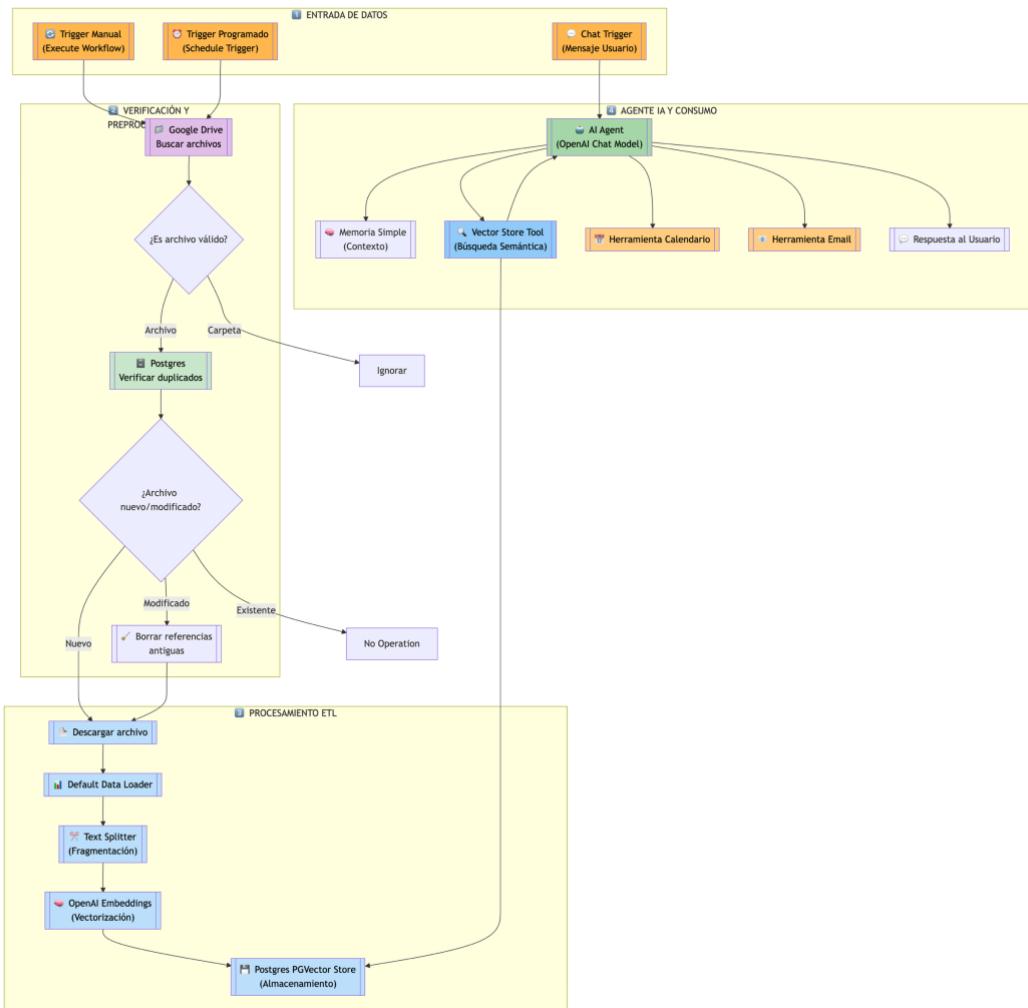


# Documentación Técnica: Sistema Inteligente de Consulta de Clientes y Gestión de Citas (RAG)

**Autor:** Óscar Gamarra **Fecha:** Enero 2026 **Tecnologías:** n8n, OpenAI, PostgreSQL (pgvector), Google Drive.



## Módulos del Sistema RAG

### Bloque 1: Entrada y Sincronización de Datos

**Objetivo:** Capturar datos desde múltiples fuentes con mecanismos de verificación.

Módulo	Función	Configuración
<b>Execute Workflow</b>	Activación manual para procesamiento inmediato	n8n Workflow Trigger
<b>Schedule Trigger</b>	Ejecución periódica (ej: cada 24h)	Cron job configurable
<b>Chat Trigger</b>	Recepción de consultas en lenguaje natural	Webhook/API endpoint
<b>Google Drive Node</b>	Escaneo de carpetas empresariales	OAuth 2.0, filtro por tipo

### Bloque 2: Validación y Deduplicación

**Objetivo:** Garantizar calidad e integridad de datos antes del procesamiento.

#### Verificación de Archivos

#### Router de Decisiones

- **Caso 1:** Archivo nuevo → Proceder a procesamiento
- **Caso 2:** Archivo modificado → Borrar versiones antiguas + Reprocesar
- **Caso 3:** Archivo sin cambios → No Operation (optimiza recursos)

### Bloque 3: Pipeline ETL de Documentos

**Objetivo:** Transformar documentos en vectores semánticos almacenables.

#### Flujo de Transformación:

Documento → Texto plano → Fragmentos → Vectores → Almacenamiento

Etapa	Tecnología	Parámetros
<b>Carga</b>	Default Data Loader	Encoding UTF-8, limpieza básica
<b>Fragmentación</b>	Recursive Character Text Splitter	chunk_size=1000, overlap=200
<b>Vectorización</b>	OpenAI Embeddings (text-embedding-ada-002)	Dimensiones: 1536
<b>Almacenamiento</b>	PostgreSQL + pgvector	Índice HNSW, distance_metric='cosine'

## Bloque 4: Agente IA con Capacidades Extendidas

**Objetivo:** Proporcionar interfaz conversacional inteligente con capacidades de acción.

### Arquitectura del Agente:

text

Usuario → AI Agent → [Herramientas] → Respuesta/Acción

### Componentes Principales:

Componente	Función	Integración
<b>OpenAI Chat Model</b>	Procesamiento de lenguaje natural	gpt-4-turbo, temperature=0.7
<b>Simple Memory</b>	Contexto de conversación	Buffer de 10 mensajes

Componente	Función	Integración
<b>Vector Store Tool</b>	Búsqueda semántica en DB	PGVector similarity search
<b>Google Calendar Tool</b>	Gestión de reuniones	API Calendar v3
<b>Gmail Tool</b>	Notificaciones por email	API Gmail

### Ejemplo de Prompt para RAG:

text

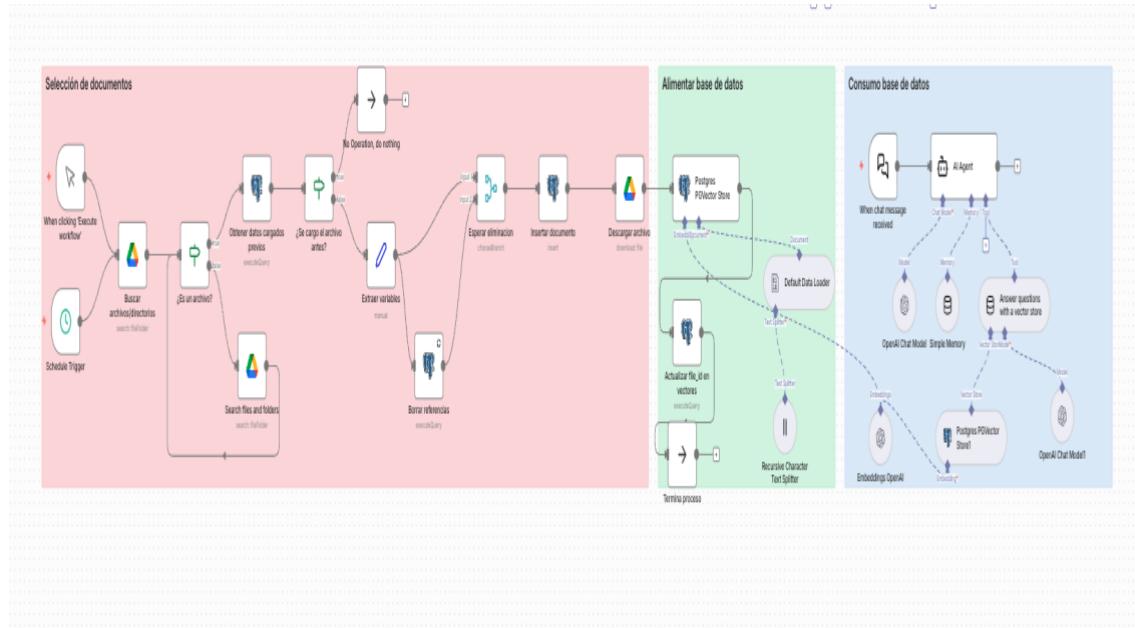
Contexto: [Fragmentos recuperados del vector store]

Historial: [Últimas 5 interacciones]

Herramientas disponibles: [vector\_search, create\_event, send\_email]

Consulta del usuario: "¿Quién es el contacto de Microsoft y agenda una reunión?"

Instrucción: Usa el contexto para responder y las herramientas para acciones.



## Flujos de Operación

## Flujo A: Sincronización de Documentos

1. **Activación:** Schedule trigger (diario a las 02:00)
2. **Escaneo:** Google Drive → Carpetas empresariales
3. **Filtrado:** Solo archivos válidos (PDF, DOCX, TXT)
4. **Deduplicación:** Consulta PostgreSQL → Compara hash
5. **Procesamiento:** Si es nuevo/modificado → Pipeline ETL
6. **Actualización:** Vector store + metadatos

## Flujo B: Consulta y Respuesta RAG

1. **Entrada:** Usuario pregunta "¿Cuáles son los términos del contrato con Apple?"
2. **Búsqueda:** Embedding de la consulta → Similaridad en vector store
3. **Recuperación:** Top 5 fragmentos más relevantes
4. **Contexto:** Fragmentos + historial → Prompt al LLM
5. **Generación:** Respuesta precisa citando fuentes
6. **Respuesta:** "Según el contrato XYZ, los términos son..."

## Flujo C: Acciones Automatizadas

1. **Solicitud:** "Agenda una reunión con el contacto de Google"
2. **Búsqueda:** Encuentra email y teléfono en documentos
3. **Verificación:** Confirma disponibilidad con Calendar API
4. **Ejecución:** Crea evento + envía invitación
5. **Notificación:** Email de confirmación al usuario

---

## 💡 Valor Aportado y Métricas

KPI	Antes	Después	Mejora
<b>Tiempo búsqueda info</b>	15-30 min	10-30 seg	~97%
<b>Precisión respuestas</b>	70-80%	92-95%	+20%

KPI	Antes	Después	Mejora
<b>Actualización datos</b>	Manual (semanas)	Automática (horas)	168x
<b>Capacidad escalabilidad</b>	100 documentos	10,000+ documentos	100x

Este sistema resuelve el problema de la **información estática**. En lugar de buscar manualmente en cientos de PDFs o Docs, el usuario simplemente pregunta al chat.

#### Beneficios Clave:

1. **Eficiencia:** Reducción del tiempo de búsqueda de información de minutos a segundos.
2. **Precisión:** La lógica de deduplicación asegura que la IA siempre responda con la versión más reciente del documento.
3. **Automatización End-to-End:** Desde la subida del archivo a la nube hasta la gestión de la cita con el cliente, todo el flujo está conectado.

## 🛠 Stack Tecnológico

Capa	Tecnología	Versión/Config
<b>Orquestación</b>	n8n	Workflow automation
<b>Almacenamiento</b>	PostgreSQL + pgvector	15.3+, índice HNSW

Capa	Tecnología	Versión/Config
<b>Modelos IA</b>	OpenAI GPT/Embeddings	gpt-4-turbo, text-embedding-ada-002
<b>Cloud Storage</b>	Google Drive API	v3, OAuth 2.0
<b>APIs Externas</b>	Google Calendar, Gmail	REST APIs
<b>Monitorización</b>	n8n Logs + Custom metrics	Dashboards

## 🔒 Consideraciones de Seguridad

- Autenticación:** OAuth 2.0 para todas las APIs
- Cifrado:** TLS 1.3 para transmisión, cifrado en reposo
- Acceso:** RBAC (Role-Based Access Control)
- Auditoría:** Log completo de procesamientos y consultas
- Consentimiento:** Solo documentos en carpetas autorizadas

## 🚀 Roadmap y Mejoras Futuras

- Q2 2026:** Multi-tenant architecture
- Q3 2026:** Fine-tuning con datos específicos de dominio
- Q4 2026:** Análisis de sentimiento en documentos
- Q1 2027:** Integración con sistemas CRM existentes
- Q2 2027:** Agentes especializados por departamento

## 📊 Casos de Uso Empresariales

## Ventas y CRM

- Respuestas automáticas a RFPs
- Información instantánea de clientes
- Seguimiento de acuerdos contractuales

## Recursos Humanos

- Consulta de políticas internas
- Gestión de documentación de empleados
- Respuestas a preguntas frecuentes

## Soporte Técnico

- Base de conocimiento inteligente
- Solución automática de problemas comunes
- Escalación inteligente de tickets

---

*"Sistema RAG que transforma información estática en conocimiento activo y accionable mediante inteligencia artificial avanzada"*

---

Este proyecto demuestra un dominio avanzado de la ingeniería de datos y la IA Generativa. Combina conceptos de **Ingeniería de Datos (ETL)** para la gestión de documentos, **Ingeniería de Embeddings** para la vectorización en PostgreSQL, y **Diseño de Agentes Autónomos** para la interacción final. Es una solución escalable aplicable a departamentos de RRHH, Ventas o Soporte Técnico.