

Universidad de los Andes

Facultad de Ingeniería

**Desarrollo de Soluciones Cloud
Plan de pruebas de carga**

Grupo 1

Juan Sebastian Colmenares - 202312351

William Fernando Alarcón - 202423127

Jhon Mario Forero - 20123466

Contenido

E1-Plan de Análisis de Capacidad2

1. Descripción de la Aplicación.....2

2. Entorno de Pruebas.....2

3. Criterios de Aceptación2

4. Escenarios de Prueba3

5. Parámetros de Configuración8

6. Herramienta Seleccionada8

7. Escenarios Clave para Pruebas Futuras.....8

E1-Plan de Análisis de Capacidad

1. Descripción de la Aplicación

La aplicación consta de un backend en FastAPI y un frontend en Angular. El backend gestiona la carga, análisis y resumen de documentos con PostgreSQL y Redis. Se usa el modelo LLM local Mistral q4_k_m para el análisis de documentos. El frontend permite a los usuarios interactuar con la plataforma de análisis.

2. Entorno de Pruebas

Las pruebas se ejecutarán en una máquina local con las siguientes características:

- CPU: 16 núcleos
- Memoria RAM: 32GB
- Disco SSD: 512GB
- Sistema Operativo: macOS

3. Criterios de Aceptación

Los criterios de aceptación incluyen:

- Tiempo de respuesta óptimo para cargas estándar.
- Capacidad de procesamiento adecuada para múltiples solicitudes concurrentes.
- Uso eficiente de los recursos de hardware.

4. Escenarios de Prueba

Auth

Endpoint: /api/v1/auth/register (POST) - Registro de usuario

- Escenario de Prueba:

Validar el correcto funcionamiento del endpoint /api/v1/auth/register cuando se ejecutan solicitudes de forma concurrente.

- Pruebas:

- Simular múltiples solicitudes concurrentes (50-100 usuarios).
- Medir tiempos de respuesta en carga normal y carga alta.
- Evaluar utilización de CPU y memoria en el procesamiento.

- Criterios de Éxito:

- Respuesta en menos de 2 segundos para carga estándar.
- Correcto procesamiento del 99% de las solicitudes.

- Métricas Clave:

- Tiempo de respuesta promedio.
- Tasa de éxito de las solicitudes.

Endpoint: /api/v1/auth/login (POST) - Inicio de sesión

- Escenario de Prueba:

Validar el correcto funcionamiento del endpoint /api/v1/auth/login cuando se ejecutan solicitudes de forma concurrente.

- Pruebas:

- Simular múltiples solicitudes concurrentes (50-100 usuarios).
- Medir tiempos de respuesta en carga normal y carga alta.
- Evaluar utilización de CPU y memoria en el procesamiento.

- Criterios de Éxito:

- Respuesta en menos de 2 segundos para carga estándar.
- Correcto procesamiento del 99% de las solicitudes.

- Métricas Clave:

- Tiempo de respuesta promedio.

- Tasa de éxito de las solicitudes.

Documents

Endpoint: /api/v1/documents/upload (POST) - Carga de documentos

- Escenario de Prueba:

Validar el correcto funcionamiento del endpoint /api/v1/documents/upload cuando se ejecutan solicitudes de forma concurrente.

- Pruebas:

- Simular múltiples solicitudes concurrentes (50-100 usuarios).

- Medir tiempos de respuesta en carga normal y carga alta.

- Evaluar utilización de CPU y memoria en el procesamiento.

- Criterios de Éxito

- Respuesta en menos de 2 segundos para carga estándar.

- Correcto procesamiento del 99% de las solicitudes.

- Métricas Clave:

- Tiempo de respuesta promedio.

- Tasa de éxito de las solicitudes.

Endpoint: /api/v1/documents/listDocuments (GET) - Listar documentos del usuario

- Escenario de Prueba:

Validar el correcto funcionamiento del endpoint /api/v1/documents/listDocuments cuando se ejecutan solicitudes de forma concurrente.

- Pruebas:

- Simular múltiples solicitudes concurrentes (50-100 usuarios).

- Medir tiempos de respuesta en carga normal y carga alta.

- Evaluar utilización de CPU y memoria en el procesamiento.

- Criterios de Éxito:

- Respuesta en menos de 2 segundos para carga estándar.

- Correcto procesamiento del 99% de las solicitudes.

- Métricas Clave:

- Tiempo de respuesta promedio.

- Tasa de éxito de las solicitudes.

Analysis

Endpoint: /api/v1/analysis/summarize/{document_id} (GET) - Resumir documento

- Escenario de Prueba:

Validar el correcto funcionamiento del endpoint /api/v1/analysis/summarize/{document_id} cuando se ejecutan solicitudes de forma concurrente.

- Pruebas:

- Simular múltiples solicitudes concurrentes (50-100 usuarios).

- Medir tiempos de respuesta en carga normal y carga alta.

- Evaluar utilización de CPU y memoria en el procesamiento.

- Criterios de Éxito:

- Respuesta en menos de 20 segundos para carga estándar.

- Correcto procesamiento del 99% de las solicitudes.

- Métricas Clave:

- Tiempo de respuesta promedio.

- Tasa de éxito de las solicitudes.

Endpoint: /api/v1/analysis/ask/{document_id} (POST) - Preguntas sobre documento

- Escenario de Prueba:

Validar el correcto funcionamiento del endpoint /api/v1/analysis/ask/{document_id} cuando se ejecutan solicitudes de forma concurrente.

- Pruebas:

- Simular múltiples solicitudes concurrentes (50-100 usuarios).

- Medir tiempos de respuesta en carga normal y carga alta.

- Evaluar utilización de CPU y memoria en el procesamiento.

- Criterios de Éxito:

- Respuesta en menos de 20 segundos para carga estándar.

- Correcto procesamiento del 99% de las solicitudes.

- Métricas Clave:

- Tiempo de respuesta promedio.

- Tasa de éxito de las solicitudes.

Conversations

Endpoint: */api/v1/conversations/list (GET)* - Listar conversaciones

- Escenario de Prueba:

Validar el correcto funcionamiento del endpoint */api/v1/conversations/list* cuando se ejecutan solicitudes de forma concurrente.

- Pruebas:

- Simular múltiples solicitudes concurrentes (50-100 usuarios).

- Medir tiempos de respuesta en carga normal y carga alta.

- Evaluar utilización de CPU y memoria en el procesamiento.

- Criterios de Éxito:

- Respuesta en menos de 2 segundos para carga estándar.

- Correcto procesamiento del 99% de las solicitudes.

- Métricas Clave:

- Tiempo de respuesta promedio.

- Tasa de éxito de las solicitudes.

Endpoint: */api/v1/conversations/byId/{conversation_id}* (GET) - Obtener detalle de conversación

- Escenario de Prueba:

Validar el correcto funcionamiento del endpoint `/api/v1/conversations/byId/{conversation_id}` cuando se ejecutan solicitudes de forma concurrente.

- Pruebas:

- Simular múltiples solicitudes concurrentes (50-100 usuarios).
- Medir tiempos de respuesta en carga normal y carga alta.
- Evaluar utilización de CPU y memoria en el procesamiento.

- Criterios de Éxito:

- Respuesta en menos de 2 segundos para carga estándar.
- Correcto procesamiento del 99% de las solicitudes.

- Métricas Clave:

- Tiempo de respuesta promedio.
- Tasa de éxito de las solicitudes.

Default

Endpoint: `/health (GET)` - Verificar estado del servicio

- Escenario de Prueba:

Validar el correcto funcionamiento del endpoint `/health` cuando se ejecutan solicitudes de forma concurrente.

- Pruebas:

- Simular múltiples solicitudes concurrentes (50-100 usuarios).
- Medir tiempos de respuesta en carga normal y carga alta.
- Evaluar utilización de CPU y memoria en el procesamiento.

- Criterios de Éxito:

- Respuesta en menos de 2 segundos para carga estándar.
- Correcto procesamiento del 99% de las solicitudes.

- Métricas Clave:

- Tiempo de respuesta promedio.

- Tasa de éxito de las solicitudes.

5. Parámetros de Configuración

Las pruebas utilizarán Apache JMeter con la siguiente configuración:

- Número de usuarios simulados: 50-100
- Duración de las pruebas: 10 minutos por escenario
- Monitoreo del uso de CPU y memoria.

6. Herramienta Seleccionada

Se usará Apache JMeter por su facilidad de configuración y métricas detalladas.

7. Escenarios Clave para Pruebas Futuras

- Evaluar el rendimiento del frontend ante múltiples solicitudes simultáneas.
- Analizar la eficiencia del backend en procesamiento por lotes.