PIP INSTALL REQUIREMENTS

```
PS E:\EG_INNOVATIONS_TASK> cd .\rag_application\
PS E:\EG_INNOVATIONS_TASK\rag_application> py -3.11 -m venv venv

PS E:\EG_INNOVATIONS_TASK\rag_application> venv/Scripts/activate

⟨ (venv) PS E:\EG_INNOVATIONS_TASK\rag_application> pip install -r requirements.txt

Collecting faiss_cpu==1.12.0 (from -r requirements.txt (line 1))

Using cached faiss_cpu-1.12.0-cp311-cp311-win_amd64.whl.metadata (5.2 kB)

Collecting fastapi==0.119.0 (from -r requirements.txt (line 2))

Using cached fastapi=0.119.0-py3-none-any.whl.metadata (28 kB)

Collecting httpx==0.28.1 (from -r requirements.txt (line 3))

Using cached httpx-0.28.1-py3-none-any.whl.metadata (7.1 kB)
```

Start the application

Q1 CHAT BOT

UPLOAD PDF

REQUEST



Terminal Process

```
SECTION OF THE PROPERTY OF THE
```

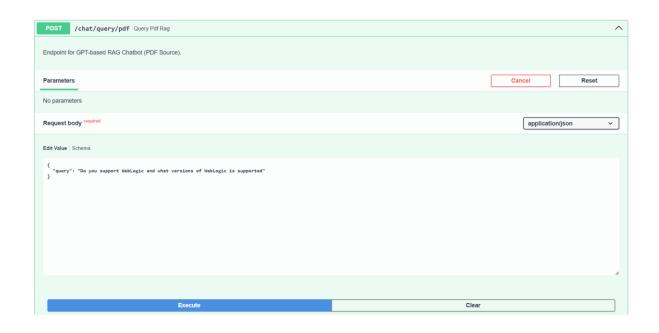
Output:



Answering PDF

Q1) Do you support WebLogic and what versions of WebLogic is supported

Answer: Yes, we support Oracle WebLogic. The supported versions are 5.1, 6.x, 7, 8.x, 9.x, 10.x, 11G, 12.x, and 14c.



```
INFO: 127.0.0.1:04887 - POSI /Cnat/upload/pur HITP/1.1 200 OK

9025-10-20 15:27:02 - rag_app - INFO - Controller: Received PDF query: Do you support WebLogic and what versions of WebLogic is supported

9025-10-20 15:27:02 - rag_app - INFO - PDF RAG: Executing Retrieve phase for query: Do you support WebLogic and what versions of WebLogic is supported

9025-10-20 15:27:03 - rag_app - INFO - Retrieved 5 relevant chunks from index.

9025-10-20 15:27:03 - rag_app - INFO - GeminiRAG: Preparing prompt for PDF query.

9025-10-20 15:27:07 - rag_app - INFO - Gemini API call succeeded. Response received.

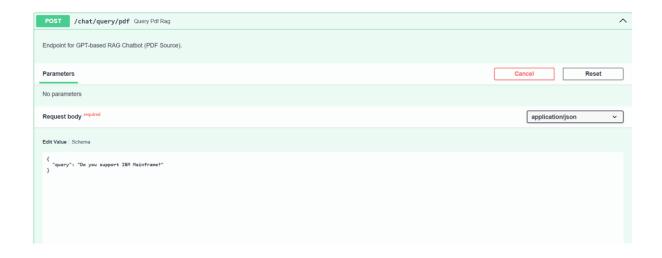
9025-10-20 15:27:07 - rag_app - INFO - PDF RAG response successfully generated via Gemini.

916: 127.0.0.1:61083 - "POST /chat/query/pdf HTTP/1.1" 200 OK
```



Q2) Do you support IBM Mainframe?

Answer: No.

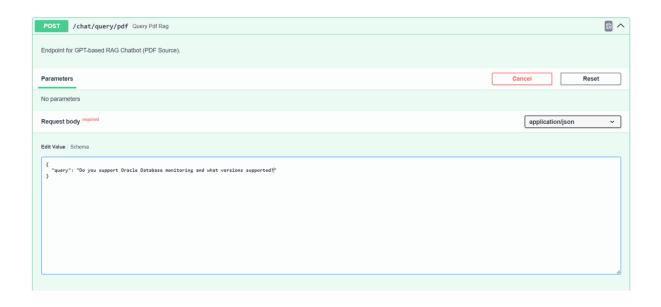


```
2025-10-20 15:28:33 - rag_app - INFO - Controller: Received PDF query: Do you support IBM Mainframe?
2025-10-20 15:28:33 - rag_app - INFO - PDF RAG: Executing Retrieve phase for query: Do you support IBM Mainframe?
2025-10-20 15:28:33 - rag_app - INFO - Retrieved 5 relevant chunks from index.
2025-10-20 15:28:33 - rag_app - INFO - GeminiRAG: Preparing prompt for PDF query.
2025-10-20 15:28:38 - rag_app - INFO - Gemini API call succeeded. Response received.
2025-10-20 15:28:38 - rag_app - INFO - PDF RAG response successfully generated via Gemini.
INFO: 127.0.0.1:61800 - "POST /chat/query/pdf HTTP/1.1" 200 OK
```



Q3) Do you support Oracle Database monitoring and what versions supported?

Answer) Yes, Oracle Database monitoring is supported. The versions supported are 7, 8, 9, 10G, 11G, 12c (including multi-tenant setup), 19c, and 21c.



```
2025-10-20 15:30:37 - rag_app - INFO - Controller: Received PDF query: Do you support Oracle Database monitoring and what versions supported?
2025-10-20 15:30:37 - rag_app - INFO - PDF RAG: Executing Retrieve phase for query: Do you support Oracle Database monitoring and what versions supported?
2025-10-20 15:30:37 - rag_app - INFO - Retrieved 5 relevant chunks from index.
2025-10-20 15:30:37 - rag_app - INFO - GeminiRAG: Preparing prompt for PDF query.
2025-10-20 15:30:40 - rag_app - INFO - Gemini API call succeeded. Response received.
2025-10-20 15:30:40 - rag_app - INFO - PDF RAG response successfully generated via Gemini.
INFO: 127.0.0.1:62943 - "POST /chat/query/pdf HTTP/1.1" 200 OK
```



Q4) Can you share the MSSQL server versions supported by eG Enterprise?

Answer) eG Enterprise supports Microsoft SQL Server versions 7.0, 2000, 2005, 2008, 2012, 2014, 2016, 2017, 2019, and 2022.



```
2025-10-20 15:32:14 - rag_app - INFO - Controller: Received PDF query: Can you share the MSSQL server versions supported by eG Enterprise?
2025-10-20 15:32:14 - rag_app - INFO - PDF RAG: Executing Retrieve phase for query: Can you share the MSSQL server versions supported by eG Enterprise?
2025-10-20 15:32:15 - rag_app - INFO - Retrieved 5 relevant chunks from index.
2025-10-20 15:32:15 - rag_app - INFO - GeminiRAG: Preparing prompt for PDF query.
2025-10-20 15:32:17 - rag_app - INFO - Gemini API call succeeded. Response received.
2025-10-20 15:32:17 - rag_app - INFO - PDF RAG response successfully generated via Gemini.
INFO: 127.0.0.1:63535 - "POST /chat/query/pdf HTTP/1.1" 200 OK
```



Q5) Does eG Enterprise supports 'Open Telemetry'?

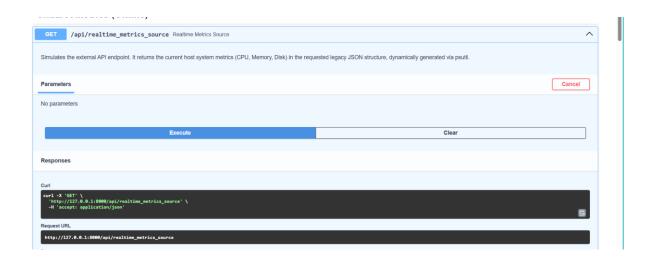
Answer) No



```
2025-10-20 15:34:03 - rag_app - INFO - Controller: Received PDF query: Does eG Enterprise supports 'Open Telemetry' ?
2025-10-20 15:34:03 - rag_app - INFO - PDF RAG: Executing Retrieve phase for query: Does eG Enterprise supports 'Open Telemetry' ?
2025-10-20 15:34:03 - rag_app - INFO - Retrieved 5 relevant chunks from index.
2025-10-20 15:34:06 - rag_app - INFO - GeminiRAG: Preparing prompt for PDF query.
2025-10-20 15:34:06 - rag_app - INFO - Gemini API call succeeded. Response received.
2025-10-20 15:34:06 - rag_app - INFO - PDF RAG response successfully generated via Gemini.
INFO: 127.0.0.1:54765 - "POST /chat/query/pdf HTTP/1.1" 200 OK
```



LIVE API: Measures the CPU Utilization and Free Space (API: /api/realtime_metrics_source)



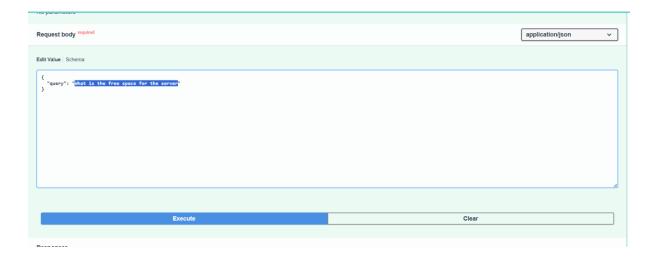
```
llama_perf_context_print: eval time = 15211.46 ms / 300 runs ( 50.70 ms per token, 19.72 tokens per second)
llama_perf_context_print: total time = 18069.31 ms / 548 tokens
llama_perf_context_print: graphs reused = 290
2025-10-23 10:29:29 - rag_app - INFO - Offline LLM RAG generation succeeded.
INFO: 127.00.0.1:56717 - "POST /api/query/api HTTP/1.1" 200 OK
2025-10-23 10:30:05 - rag_app - INFO - Controller: Successfully fetched real-time system metrics.
INFO: 127.00.0.1:6670 - "GET /api/realtime_metrics_source HTTP/1.1" 200 OK
2025-10-23 10:35:07 - rag_app - INFO - Controller: Successfully fetched real-time system metrics.
INFO: 127.0.0.1:49743 - "GET /api/realtime_metrics_source HTTP/1.1" 200 OK
```



Query Chat Bot: (API: /api/query/api)

Q1) What is the free space for the server

Answer) The free space for the server is 335855 MB, which is available. The context JSON provides the last measurement time for the server, which is Oct 23, 2025 10:37:25. The value for CPU Utilization is 15.2%, which is within the normal range of 0% to 100%. Memory Usage is also within the normal range of 0% to 100%. Therefore, the system metrics monitor should state that the server has free space of 335855 MB.



```
2025-10-23 10:37:25 - rag_app - INFO - Controller: Successfully retrieved context via HTTP from http://127.0.0.1:8000/api/realtime_metrics_source.
2025-10-23 10:37:25 - rag_app - INFO - Offline RAG: Preparing prompt for metrics query.

Llama_generate: 178 prefix-match hit, remaining 251 prompt tokens to eval

Llama_perf_context_print: load time = 9913.97 ms

Llama_perf_context_print: prompt eval time = 2453.27 ms / 251 tokens ( 9.77 ms per token,  102.31 tokens per second)

Llama_perf_context_print: eval time = 8072.92 ms / 126 runs ( 64.07 ms per token,  15.61 tokens per second)

Llama_perf_context_print: total time = 10807.36 ms / 377 tokens

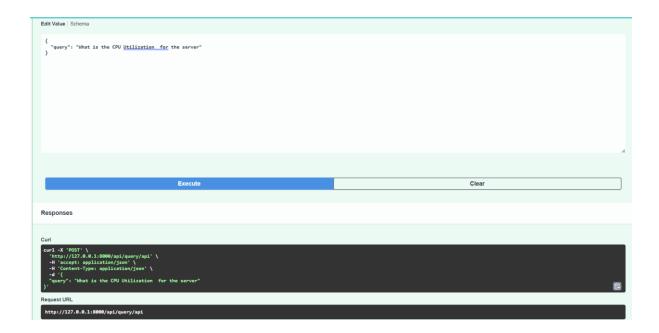
Llama_perf_context_print: graphs reused = 121

2025-10-23 10:37:36 - rag_app - INFO - Offline LLM RAG generation succeeded.

INFO: 127.0.0.1:59241 - "POST /api/query/api HTTP/1.1" 200 OK
```

Q2) What is the CPU Utilization for the server

Ans) "The CPU Utilization metric for the server is \"22.0\", which is reported as \"22.0\" with a priority of \"normal\". The unit is \"%\", indicating that the metric is expressed as a percentage. The specific requested metric is not explicitly present in the JSON, but it is present in the context.",



```
INFO: 127.0.0.1:53984 - "GET /api/realtime_metrics_source HTTP/1.1" 200 OK
2025-10-23 10:39:10 - rag app - INFO - Controller: Successfully retrieved context via HTTP from http://127.0.0.1:8000/api/realtime_metrics_source.
2025-10-23 10:39:10 - rag app - INFO - Offline RAG: Preparing prompt for metrics query.

Llama_generate: 179 prefix-match hit, remaining 250 prompt tokens to eval.

Ilama_perf_context_print: load time = 9913.97 ms

Ilama_perf_context_print: prompt eval time = 2456.58 ms / 250 tokens ( 9.83 ms per token, 101.77 tokens per second)

Ilama_perf_context_print: eval time = 4776.03 ms / 67 runs ( 71.28 ms per token, 14.03 tokens per second)

Ilama_perf_context_print: total time = 7358.54 ms / 317 tokens

Ilama_perf_context_print: graphs reused = 64

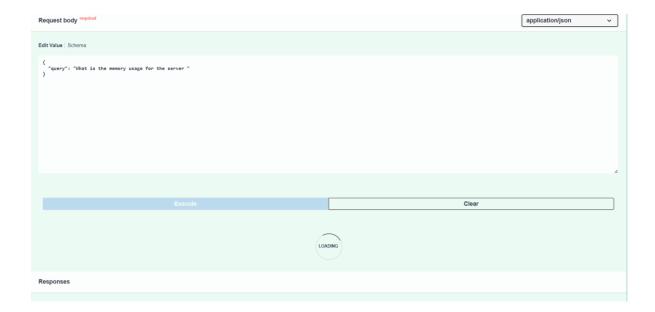
2025-10-23 10:39:18 - rag_app - INFO - Offline LLM RAG generation succeeded.

INFO: 127.0.0.1:53983 - "POST /api/query/api HTTP/1.1" 200 OK
```



Q3) What is the memory usage for the server

Ans) "The system metrics monitor for the server 10.200.2.192:7077:eG Manager is as follows:\n\n- Disk Space:\n - lastMeasurementTime: Oct 23, 2025 10:40:36\n - C:\n - Used space:\n - unit: MB\n - priority: warning\n - value: 6708\n - Drive availability:\n - priority: normal\n - unit: -\n value: Available\n - Free space:\n - unit: MB\n - priority: normal\n value: 335855\n\n- CPU Utilization:\n - lastMeasurementTime: Oct 23, 2025 10:40:36\n - value: 22.5\n - unit: %\n - priority: normal\n\n- Memory Usage:\n lastMeasurementTime: Oct 23, 2025 10:40:36\n - value: 6708\n - unit: MB\n priority: warning\n\nThe memory usage for the server 10.200.2.192:7077:eG Manager is 6708 MB."



```
INFO: 127.0.0.1:53144 - "GET /api/realtime_metrics_source HTTP/1.1" 200 OK
2025-10-23 10:40:36 - rag_app - INFO - Controller: Successfully retrieved context via HTTP from http://127.0.0.1:8000/api/realtime_metrics_source.
2025-10-23 10:40:36 - rag_app - INFO - Offline RAG: Preparing prompt for metrics query.

Llama.generate: 178 prefix-match hit, remaining 250 prompt tokens to eval

Llama.perf context_print: load time = 9913.97 ms

Llama_perf_context_print: prompt eval time = 2570.43 ms / 250 tokens ( 10.28 ms per token, 97.26 tokens per second)

Llama_perf_context_print: eval time = 19926.43 ms / 316 runs ( 63.06 ms per token, 15.86 tokens per second)

Llama_perf_context_print: total time = 22867.42 ms / 566 tokens

Llama_perf_context_print: graphs reused = 365

2025-10-23 10:40:59 - rag_app - INFO - Offline LLM RAG generation succeeded.

TNFO: 127.0.0.1:53143 - "POST /ani/nuery/ani/HTTP/1.1" 200 OK
```

