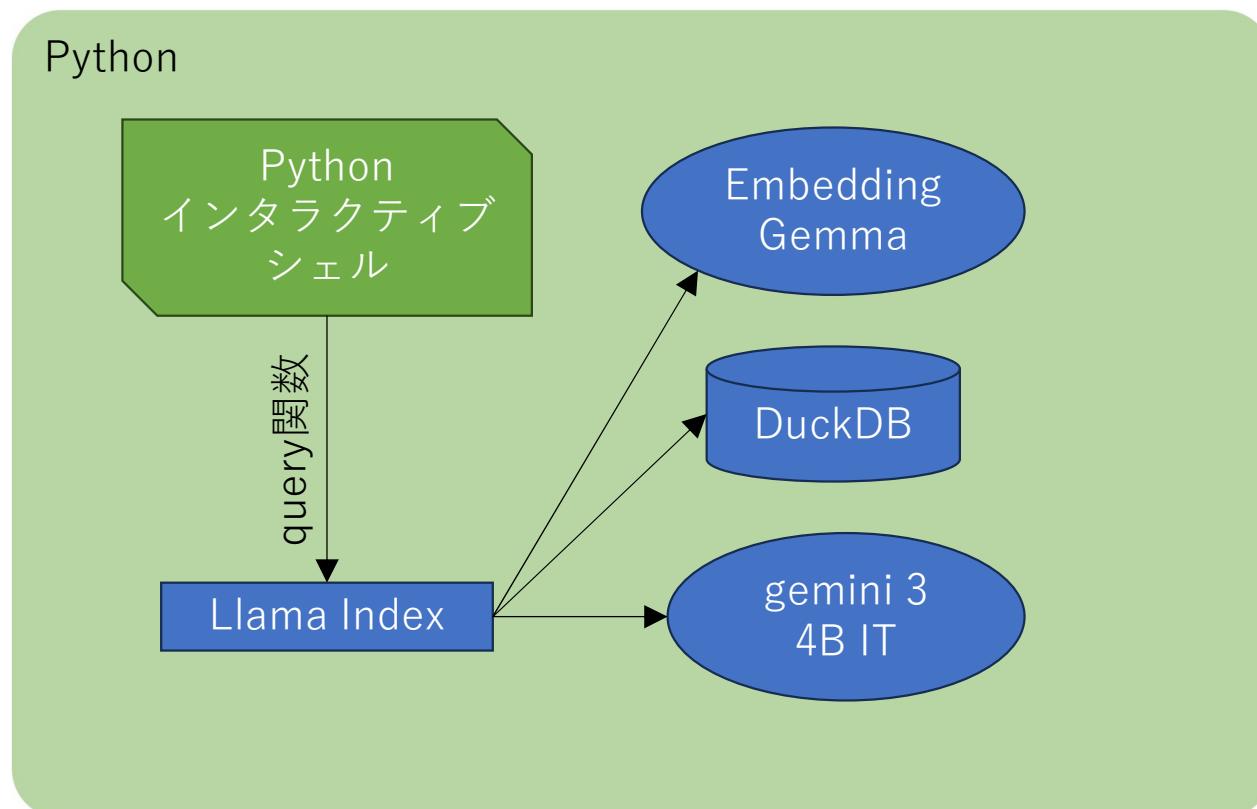
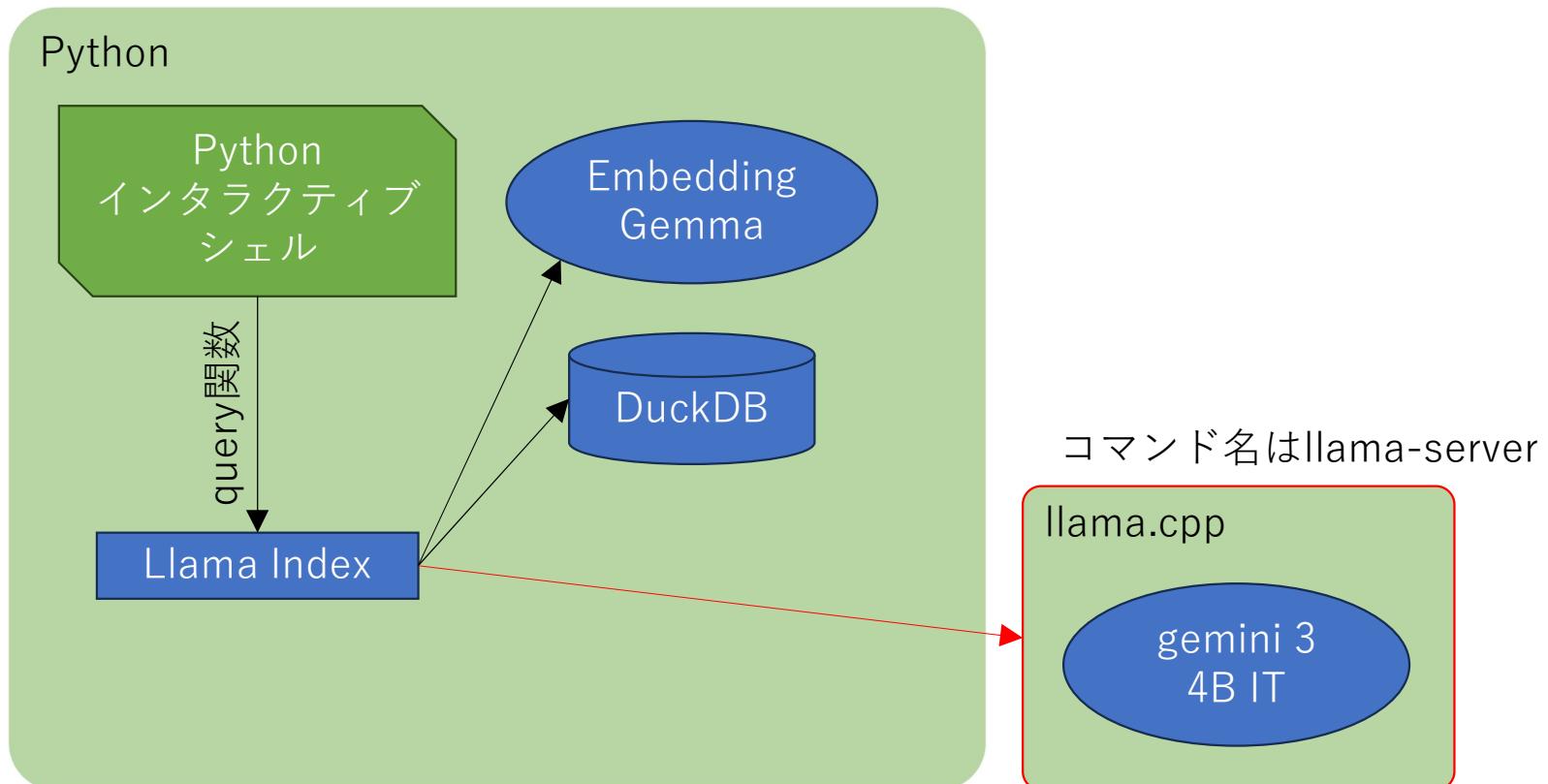


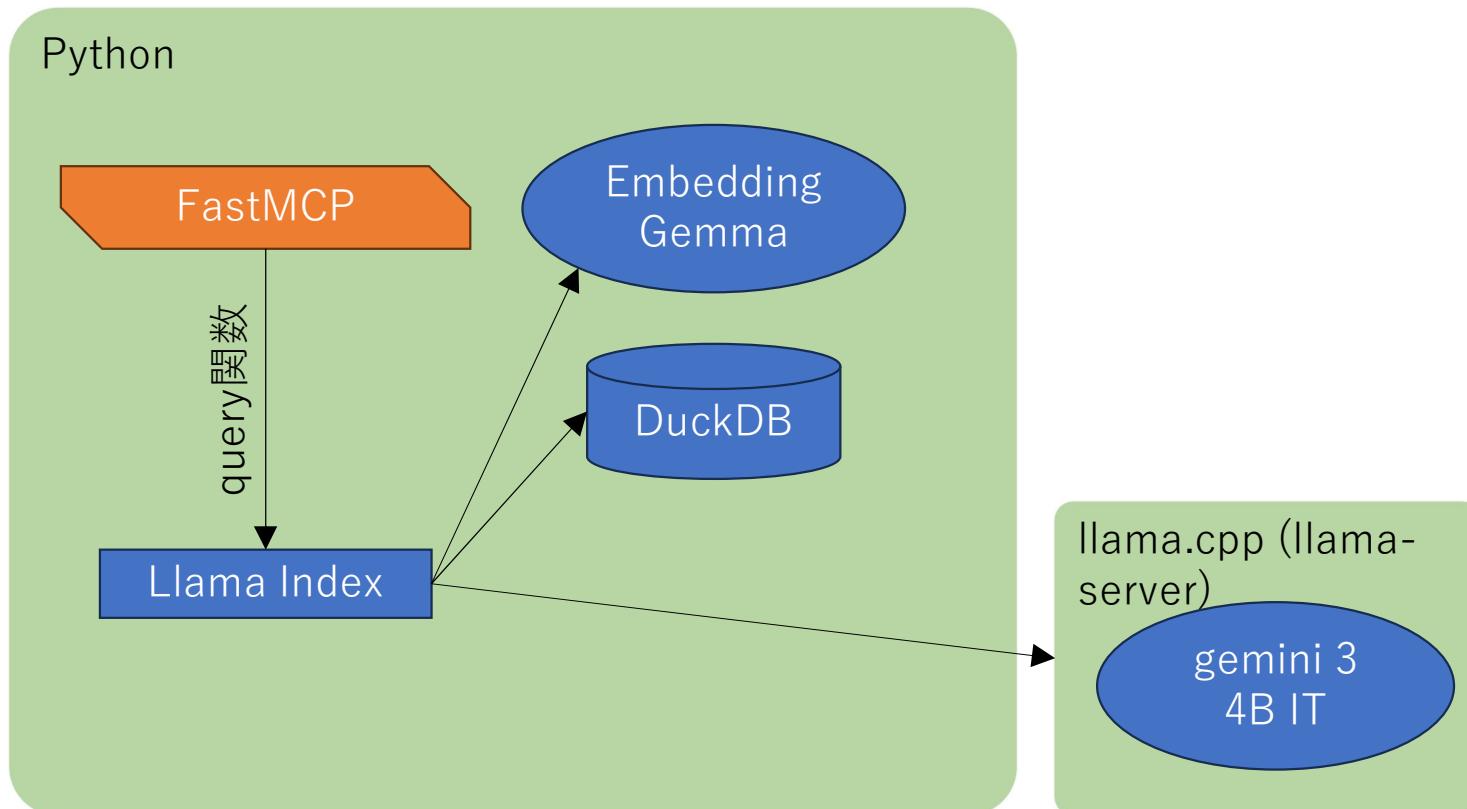
Llama Indexを用いた原始的なRAG構成



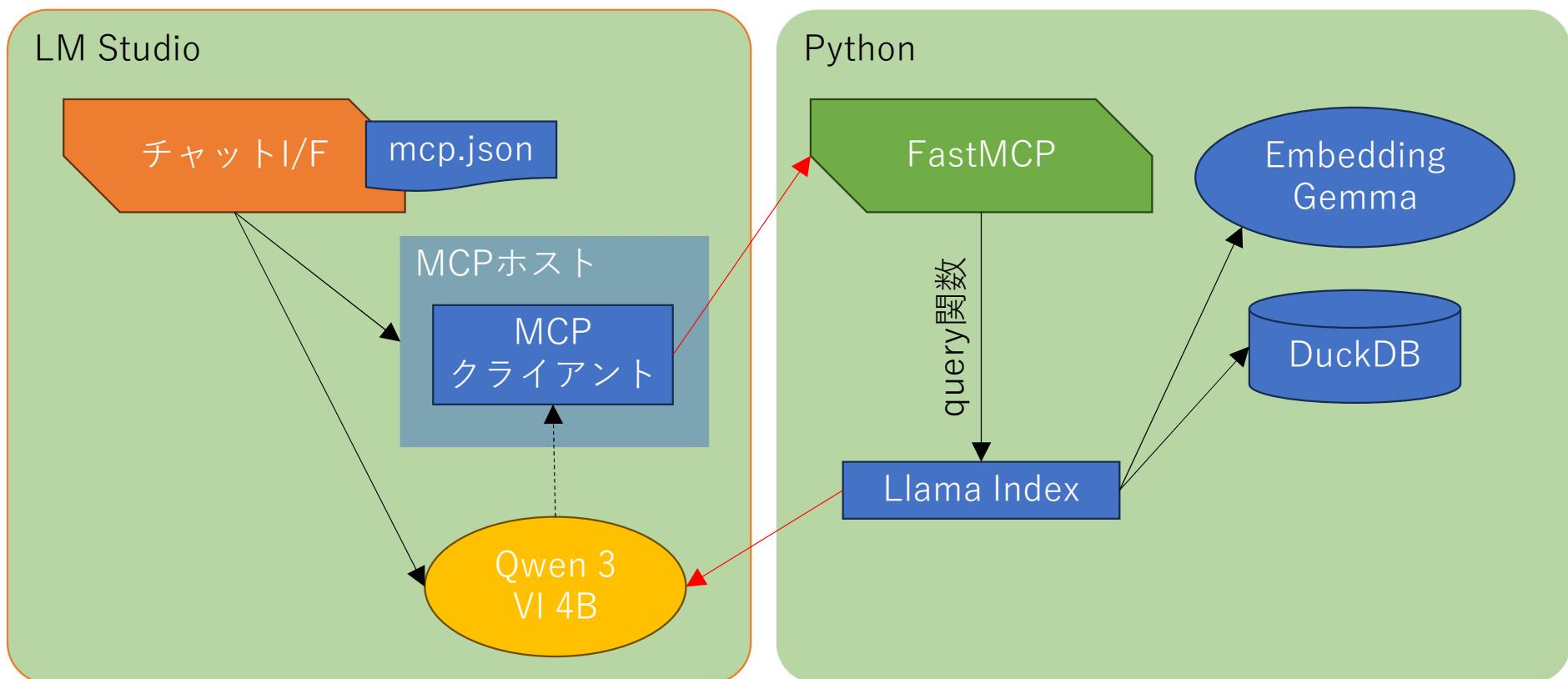
LLM高速化のためにllama.cppを導入



FastMCPを用いてMCPサーバー化



LM StudioでチャットからMCPを利用



mcp.json と MCP ホスト / クライアント / サーバー -

LLM オーケーストレイター

mcp.json

```
{  
  "mcpServers": {  
    "foo": {"url": "..."},  
    "bar": {"url": "..."},  
    "baz": {"command": "..."}  
  }  
}
```

注意: MCP サーバーを子プロセスとして起動することも可。
その際の通信は標準入出力

MCP ホスト

MCP
クライアント
foo

MCP
クライアント
bar

MCP
クライアント
bar

MCP サーバー (外部プロセス)

MCP: foo

MCP サーバー (外部プロセス)

MCP: bar

MCP サーバー (子プロセス)

MCP: baz

Diagram illustrating the MCP architecture:

- MCP Host (MCP.json configuration):** Contains three MCP clients: "foo", "bar", and another "bar".
- MCP Servers (External Processes):** Three separate green boxes representing external processes:
 - MCP: foo** (top)
 - MCP: bar** (middle)
 - MCP: baz** (bottom)
- Connections:** Arrows point from each MCP client in the host to its corresponding MCP server. The top "bar" client connects to MCP: bar, the middle "bar" client connects to MCP: baz, and the "foo" client connects to MCP: foo.