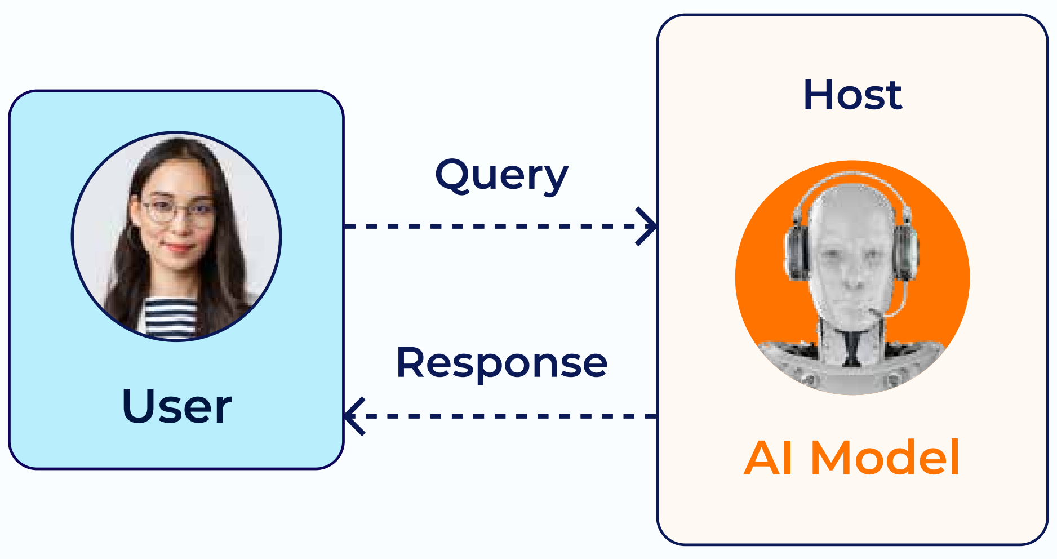
SearchUnify®

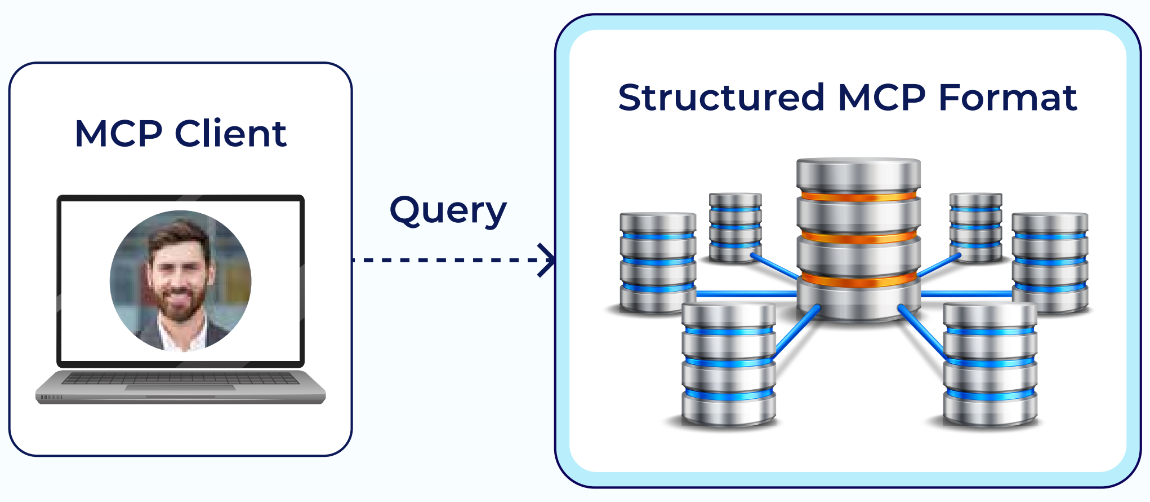
Demystifying MCP
How Model Context Protocol Works?

01 User Query Initiation

The user query is initiated through an AI model, a web application, a chat interface, a virtual assistant, or anything that is your “HOST”.



02 Client Encodes the Request

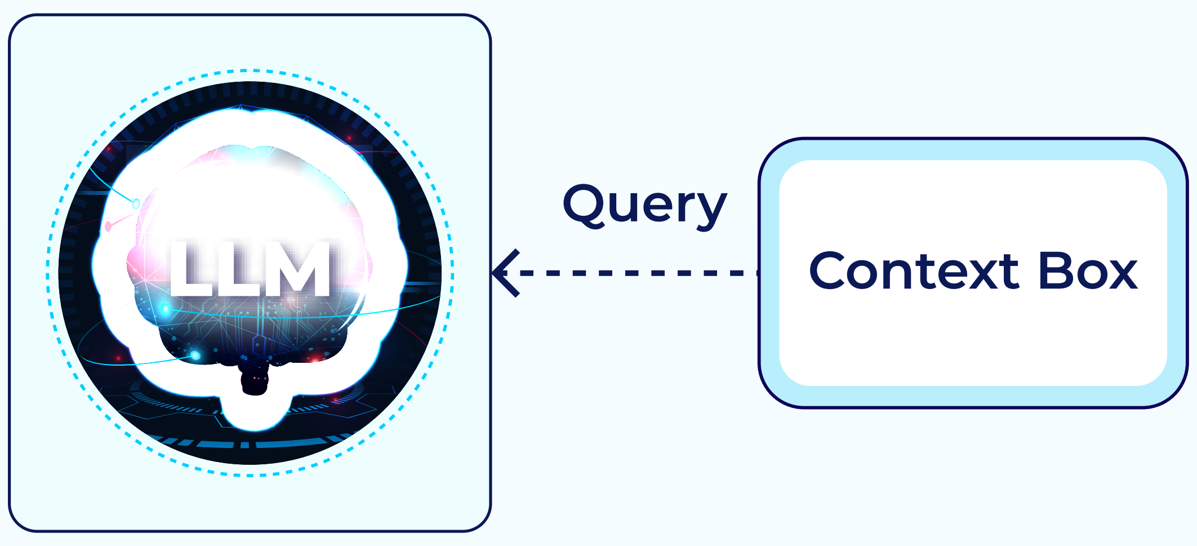


The host’s MCP client structures the query as a request object following the MCP format and sends it to the server. It also handles the security or authentication as well

03 Server Retrieves External Context

The server receives the query and determines which external tools or systems to consult—this could be an internal financial API, a SQL database, or a content management system. It pulls only the relevant data needed to answer the query.

04 Context Sent to the LLM



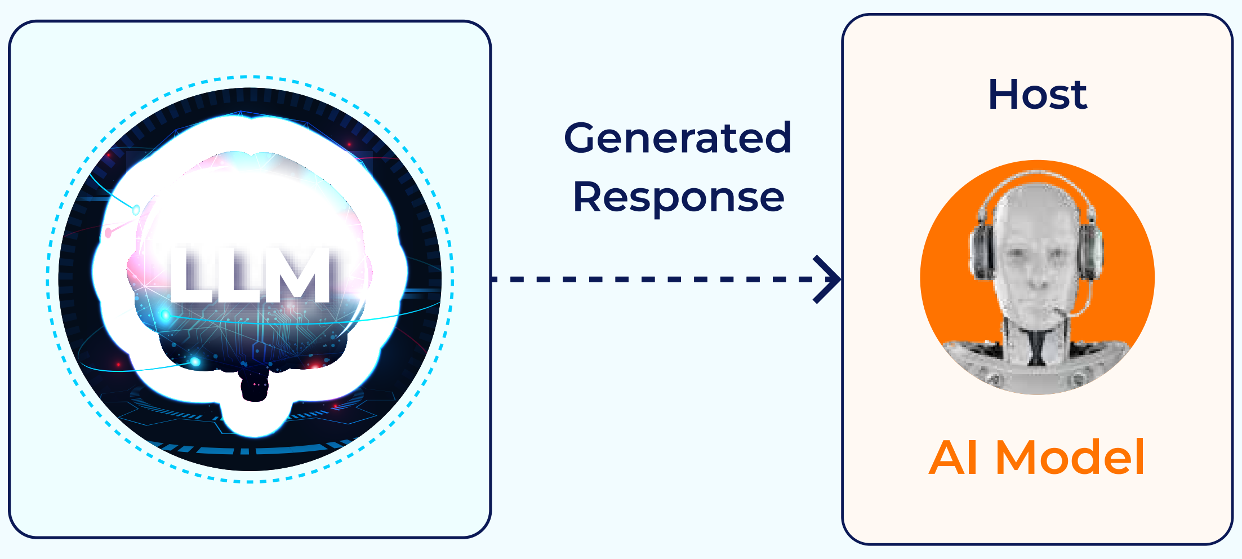
The retrieved data is formatted into a context block and sent alongside the original query to the LLM. The LLM now has both natural language and structured context to reason from.

05 LLM Generates a Response

With access to real-time context, the model generates an accurate, informed answer—grounded in actual data, not guesses.

LLM
Generate Responses

06 Response Returned to Host



The final response is returned to the Host and displayed to the user in the appropriate interface.