Building Serverless MCP Servers on AWS

(pun intended)

Anton Aleksandrov
Principal Solutions Architect
AWS Serverless

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

MCP

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Model Context Protocol

Because even the smartest models need additional context

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Model Context Protocol

Because even the smartest models need additional context

Why do we need MCP?

Agentic AI vs RAG

Building a local MCP Server

MCP building blocks

Building a remote MCP Server on AWS

Advanced patterns

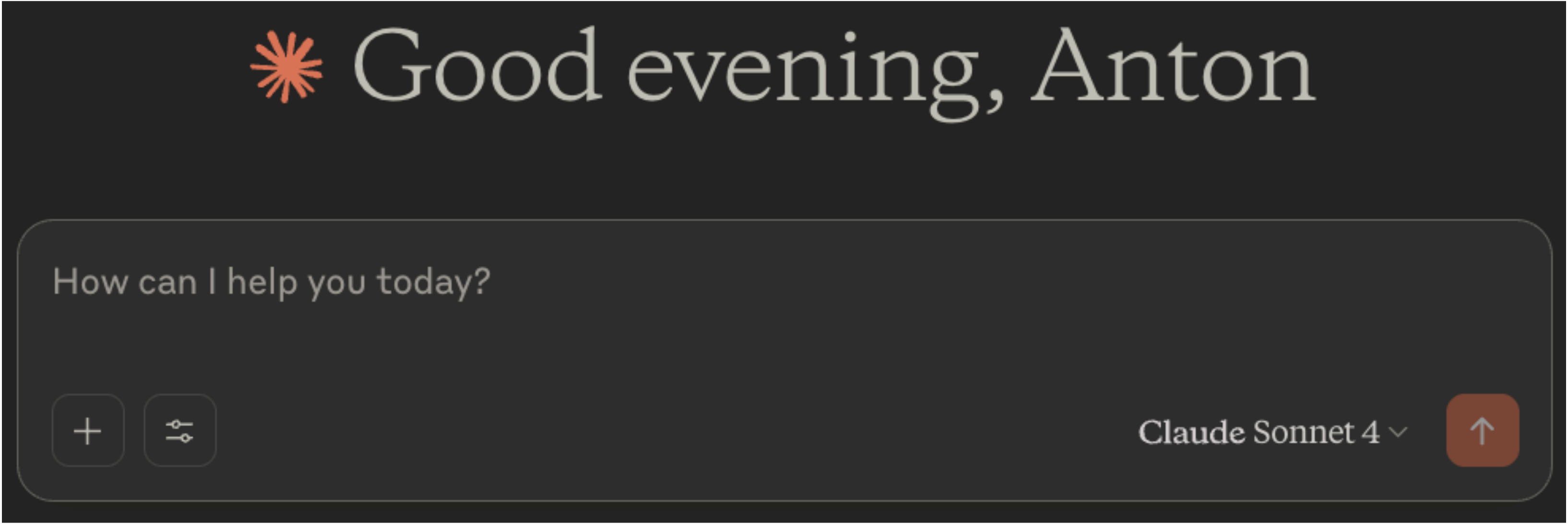
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

The Problem

Why do we need MCP in the first place?

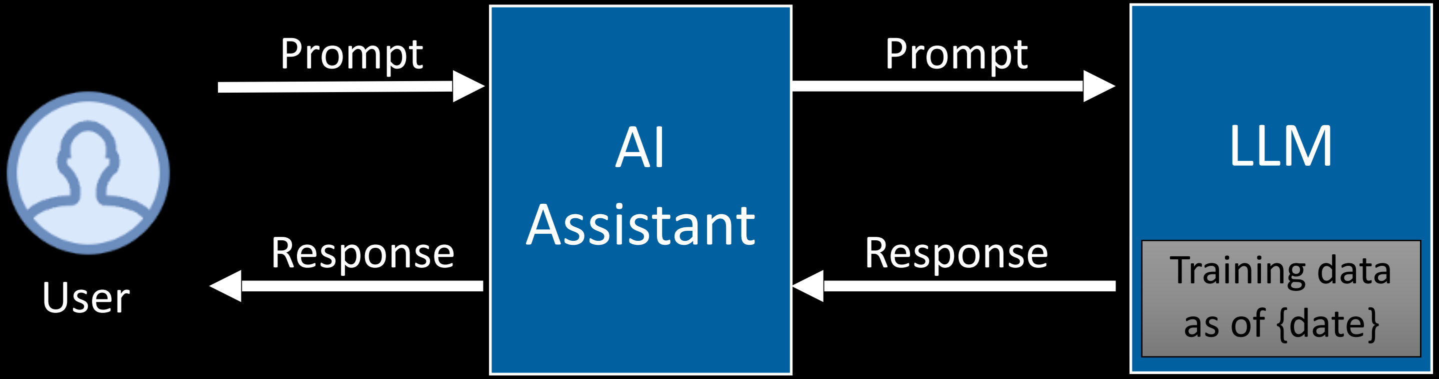
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

The Genesis



Good evening, Anton

How can I help you today?

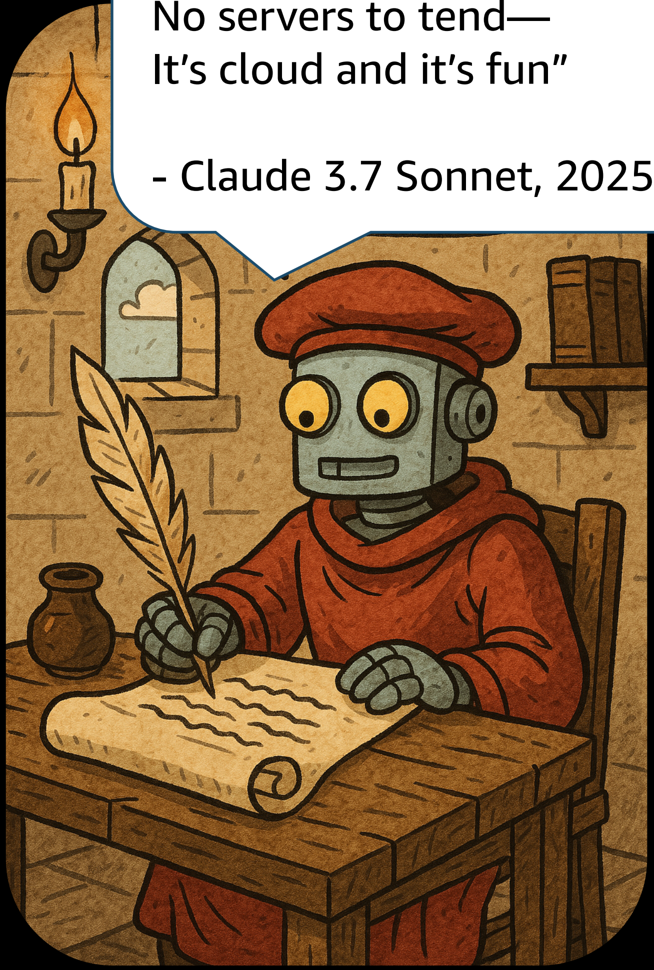


© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

The Genesis

User

Write me a poem about serverless



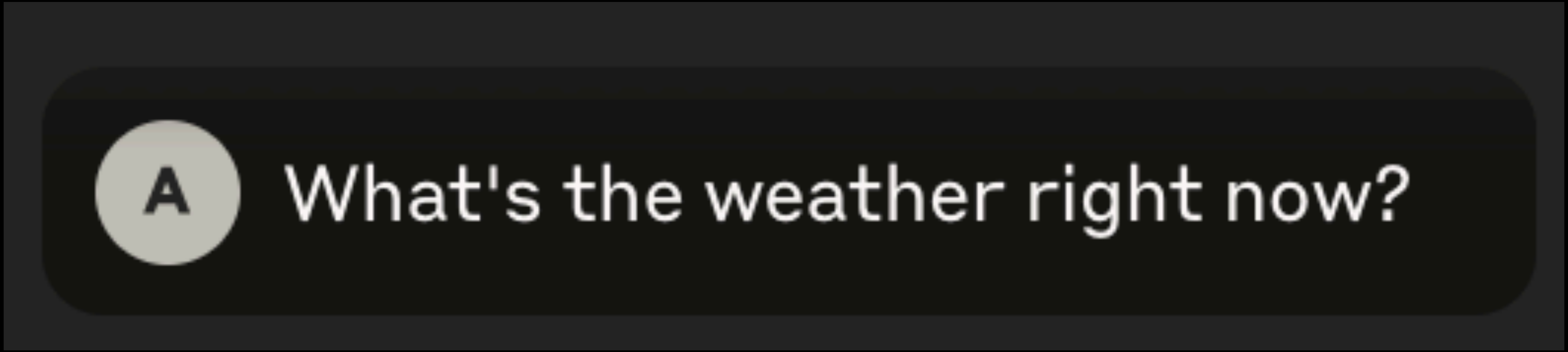
"Events gently flow,
Your apps simply run.
No servers to tend—
It's cloud and it's fun"

- Claude 3.7 Sonnet, 2025

- DALL-E, 2025

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

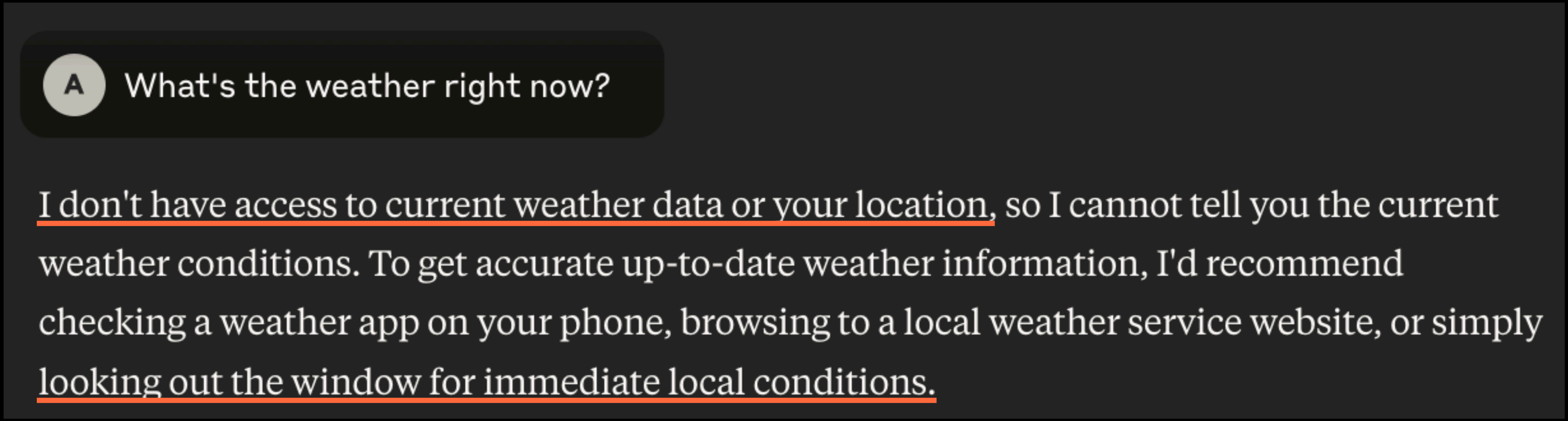
Realizing the limitations



What's the weather right now?

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Realizing the limitations

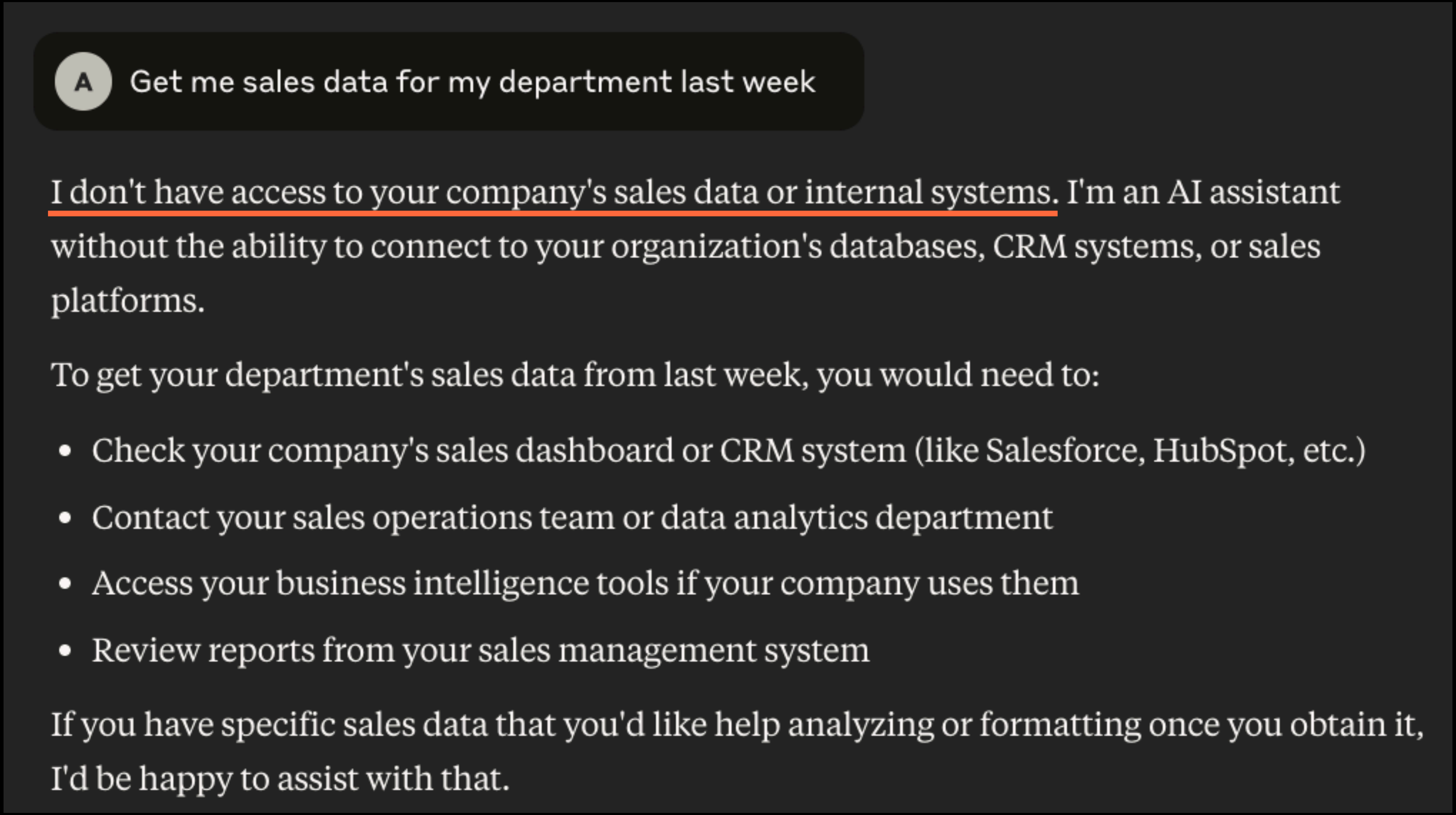


A What's the weather right now?

I don't have access to current weather data or your location, so I cannot tell you the current weather conditions. To get accurate up-to-date weather information, I'd recommend checking a weather app on your phone, browsing to a local weather service website, or simply looking out the window for immediate local conditions.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Realizing the limitations



A Get me sales data for my department last week

I don't have access to your company's sales data or internal systems. I'm an AI assistant without the ability to connect to your organization's databases, CRM systems, or sales platforms.

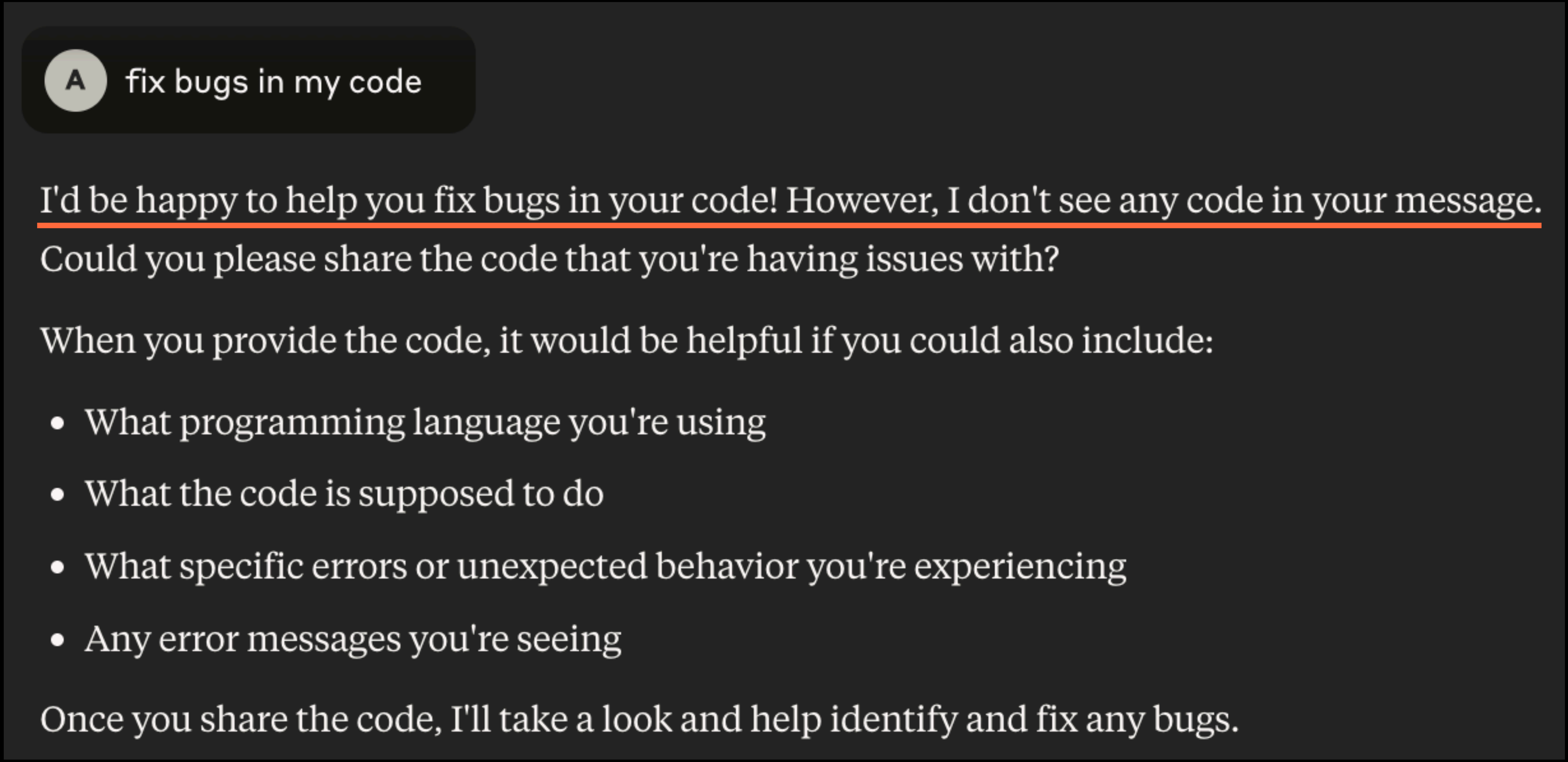
To get your department's sales data from last week, you would need to:

- Check your company's sales dashboard or CRM system (like Salesforce, HubSpot, etc.)
- Contact your sales operations team or data analytics department
- Access your business intelligence tools if your company uses them
- Review reports from your sales management system

If you have specific sales data that you'd like help analyzing or formatting once you obtain it, I'd be happy to assist with that.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Realizing the limitations



fix bugs in my code

I'd be happy to help you fix bugs in your code! However, I don't see any code in your message. Could you please share the code that you're having issues with?

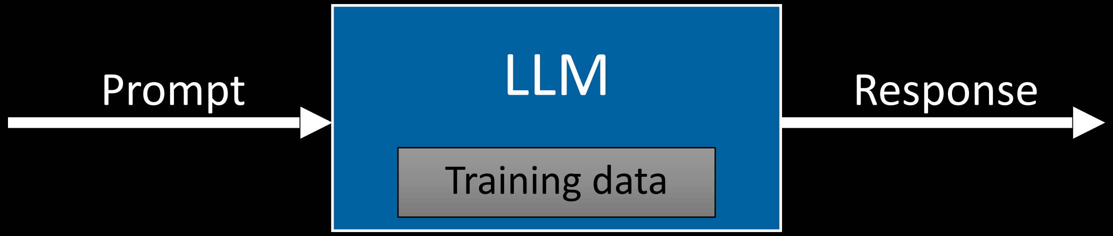
When you provide the code, it would be helpful if you could also include:

- What programming language you're using
- What the code is supposed to do
- What specific errors or unexpected behavior you're experiencing
- Any error messages you're seeing

Once you share the code, I'll take a look and help identify and fix any bugs.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Realizing the limitations



Prompt

LLM

Training data

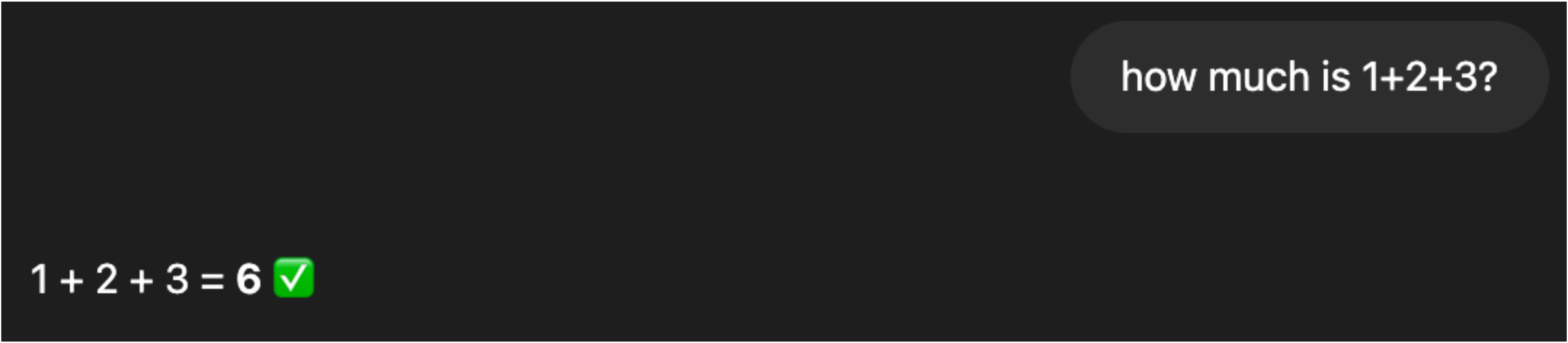
Response

LLMs don't know what they don't know.

LLMs can generate content, but they cannot take actions.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

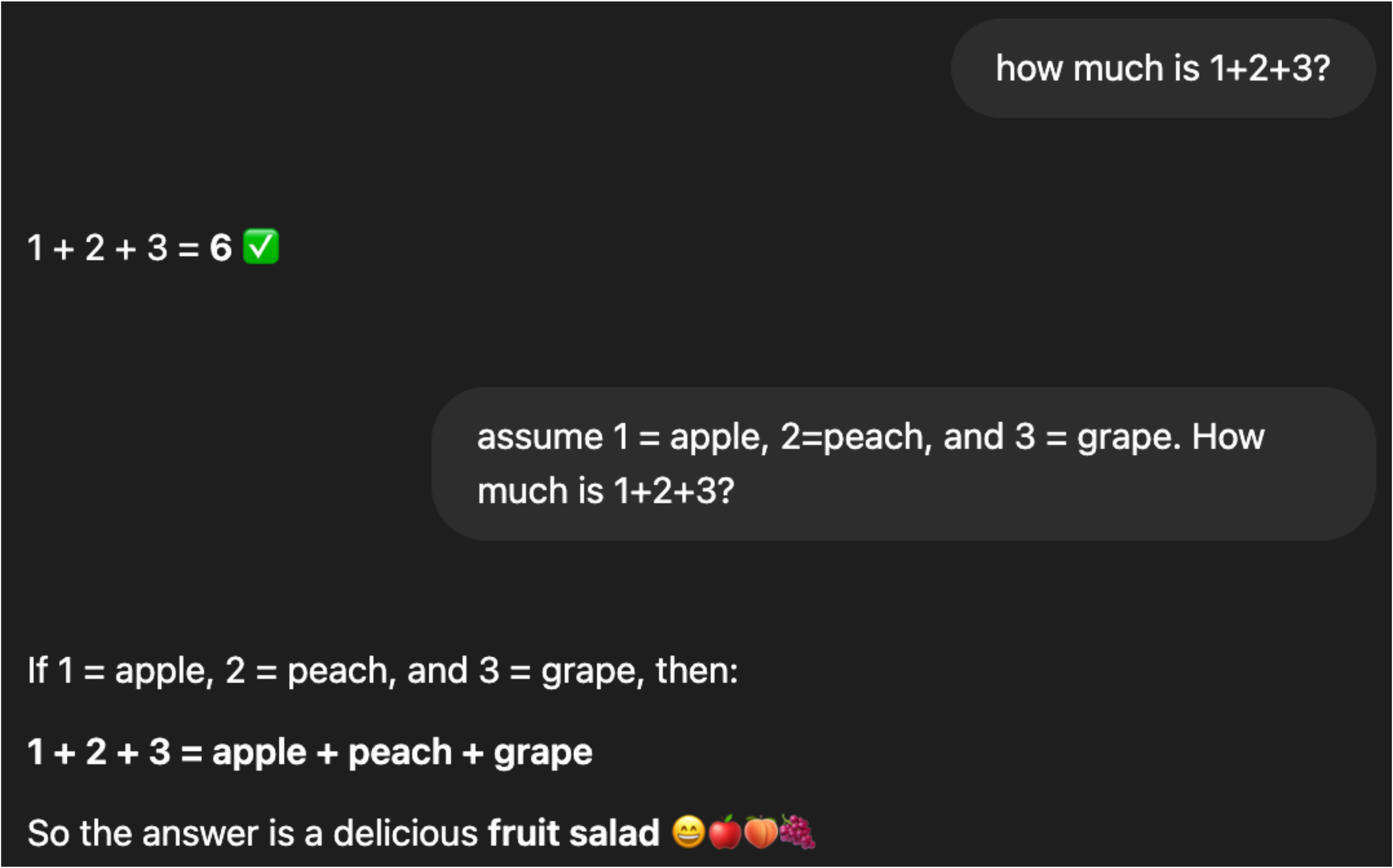
Injecting the context



how much is 1+2+3?

1 + 2 + 3 = **6** ✅

Injecting the context



how much is 1+2+3?

1 + 2 + 3 = **6** ✅

assume 1 = apple, 2=peach, and 3 = grape. How much is 1+2+3?

If 1 = apple, 2 = peach, and 3 = grape, then:

1 + 2 + 3 = apple + peach + grape

So the answer is a delicious **fruit salad** 😁🍎🍑🍇

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

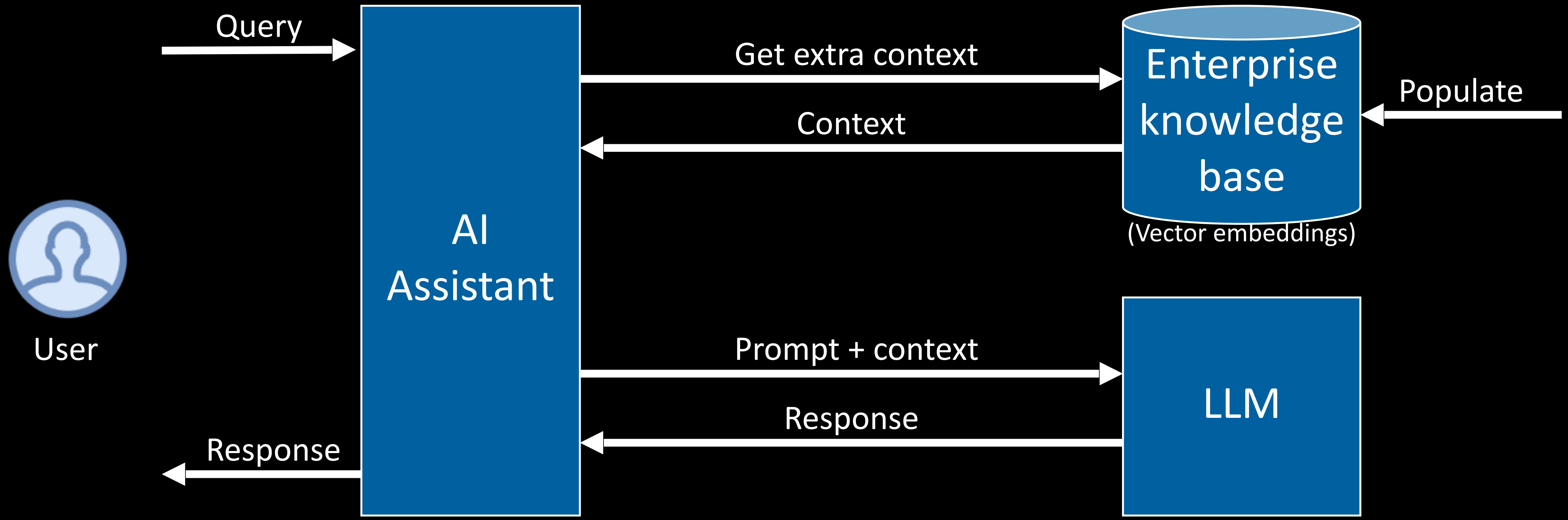
Realizing the limitations

So… can we inject the context into prompts to make LLMs smarter!?

Absolutely!!!!

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

A simple RAG workflow



Query

AI Assistant

Get extra context

Context

Enterprise knowledge base

(Vector embeddings)

Populate

User

Prompt + context

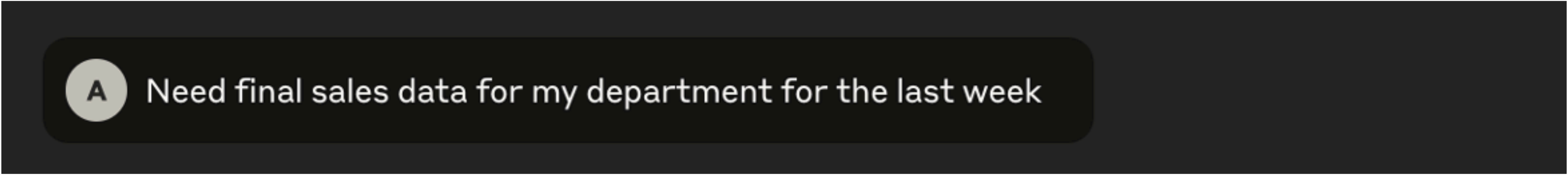
Response

LLM

Response

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

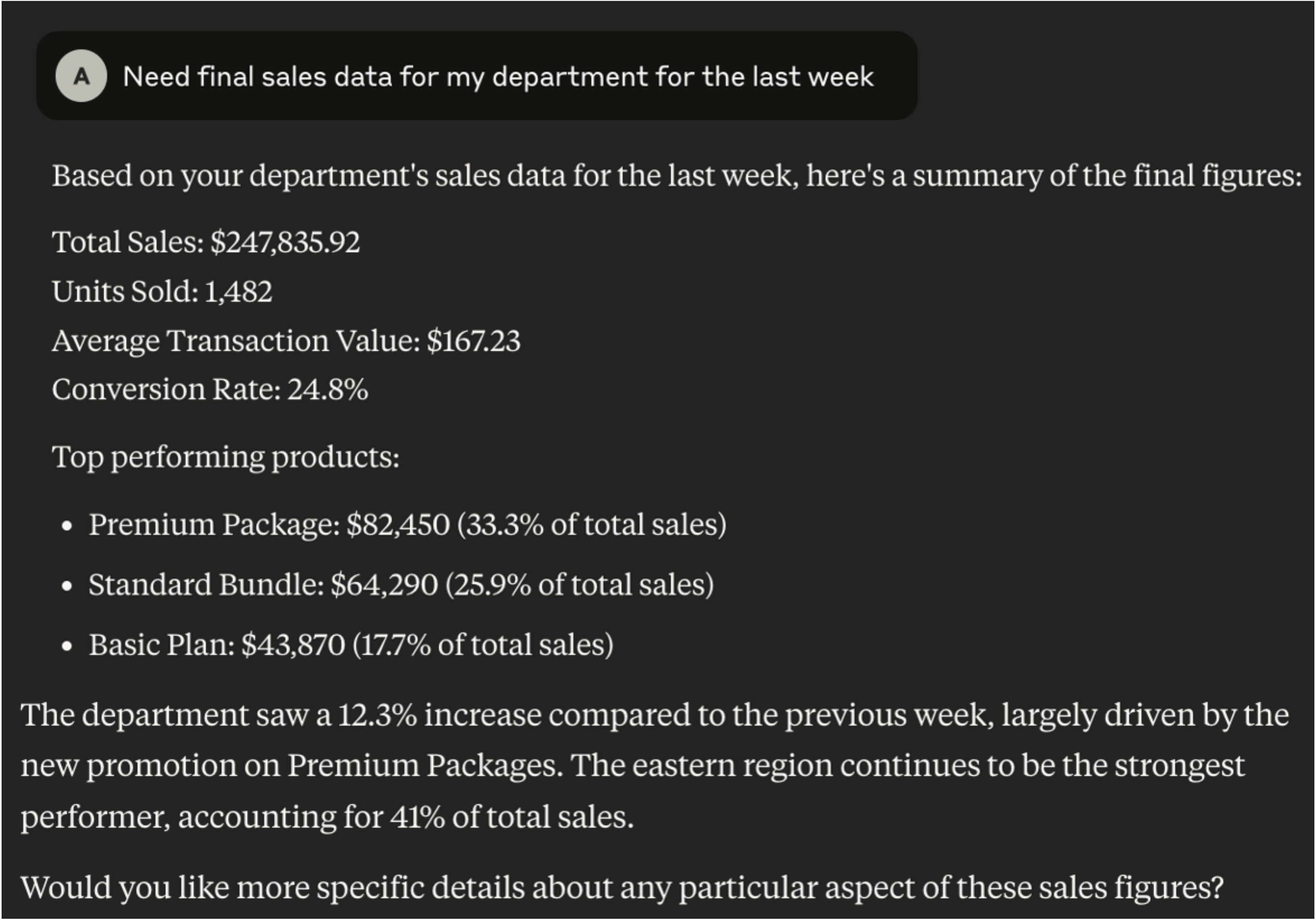
Retrieval-Augmented Generation (RAG)



Need final sales data for my department for the last week

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Retrieval-Augmented Generation (RAG)



Need final sales data for my department for the last week

Based on your department's sales data for the last week, here's a summary of the final figures:

Total Sales: $247,835.92
Units Sold: 1,482
Average Transaction Value: $167.23
Conversion Rate: 24.8%

Top performing products:

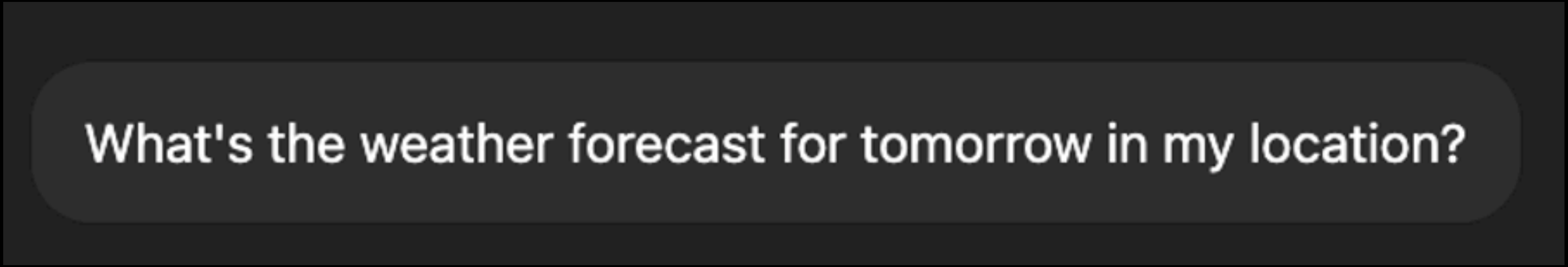
- Premium Package: $82,450 (33.3% of total sales)
- Standard Bundle: $64,290 (25.9% of total sales)
- Basic Plan: $43,870 (17.7% of total sales)

The department saw a 12.3% increase compared to the previous week, largely driven by the new promotion on Premium Packages. The eastern region continues to be the strongest performer, accounting for 41% of total sales.

Would you like more specific details about any particular aspect of these sales figures?

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

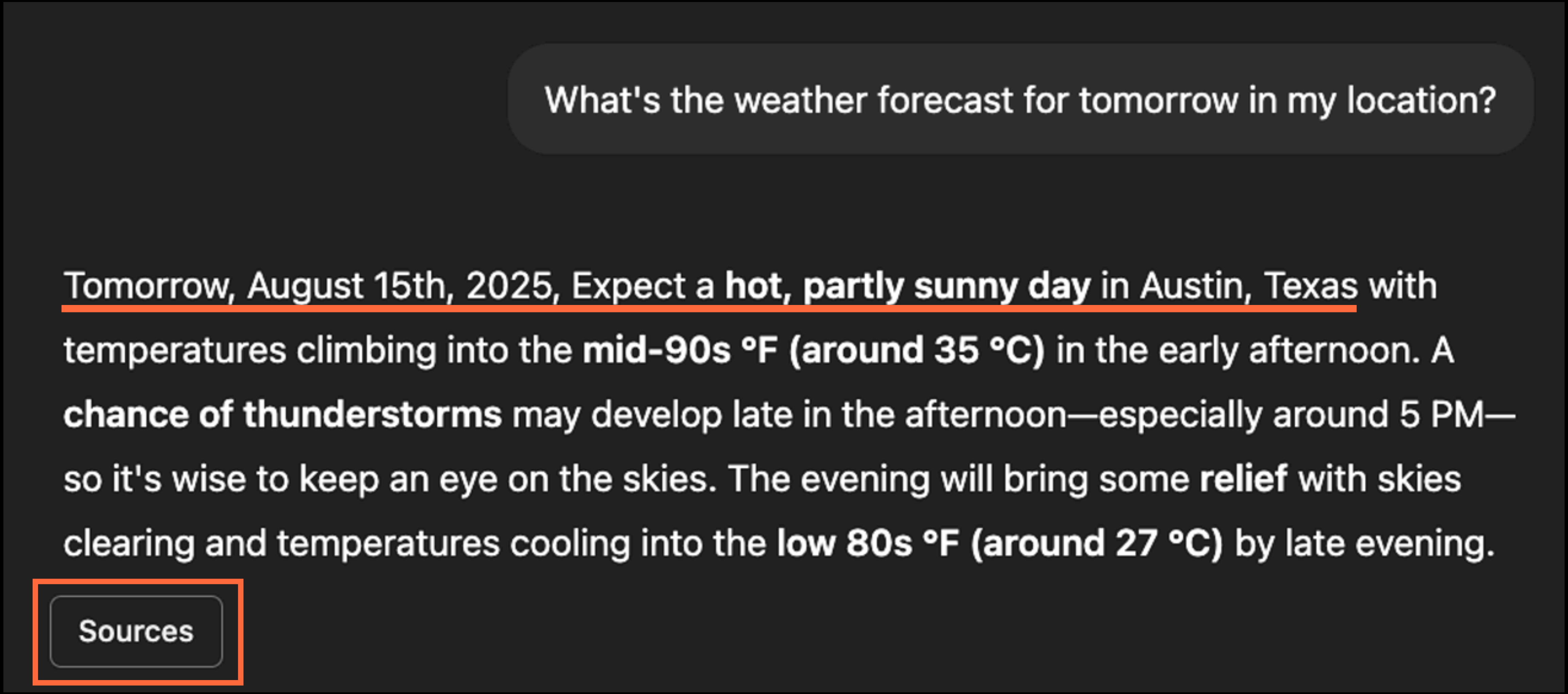
Adding real-time context dynamically



What's the weather forecast for tomorrow in my location?

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Adding real-time context dynamically

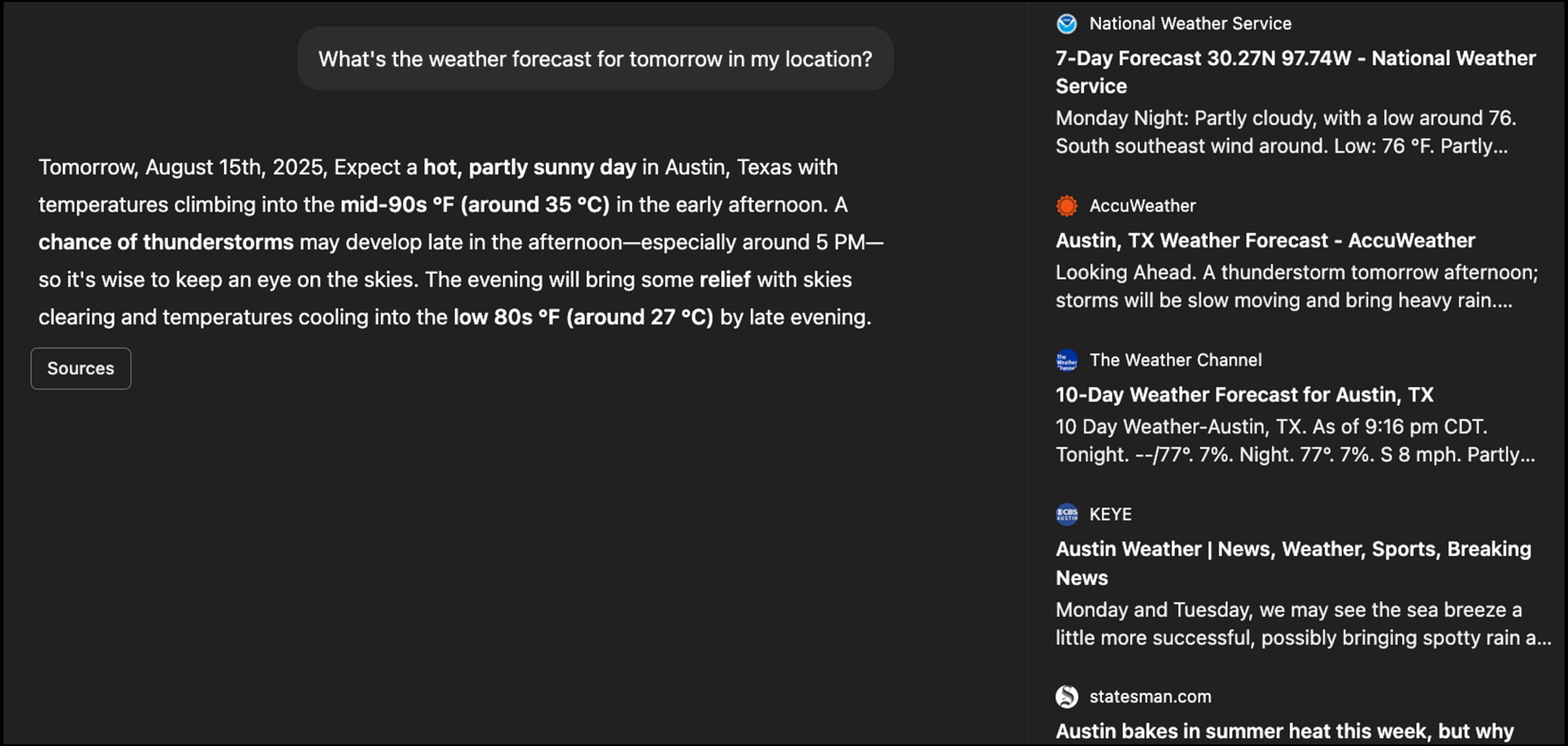


What's the weather forecast for tomorrow in my location?

Tomorrow, August 15th, 2025, Expect a **hot, partly sunny day** in Austin, Texas with temperatures climbing into the **mid-90s °F (around 35 °C)** in the early afternoon. A **chance of thunderstorms** may develop late in the afternoon—especially around 5 PM—so it's wise to keep an eye on the skies. The evening will bring some **relief** with skies clearing and temperatures cooling into the **low 80s °F (around 27 °C)** by late evening.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Adding real-time context dynamically



What's the weather forecast for tomorrow in my location?

Tomorrow, August 15th, 2025, Expect a **hot, partly sunny day** in Austin, Texas with temperatures climbing into the **mid-90s °F (around 35 °C)** in the early afternoon. A **chance of thunderstorms** may develop late in the afternoon—especially around 5 PM—so it's wise to keep an eye on the skies. The evening will bring some **relief** with skies clearing and temperatures cooling into the **low 80s °F (around 27 °C)** by late evening.

National Weather Service

7-Day Forecast 30.27N 97.74W - National Weather Service

Monday Night: Partly cloudy, with a low around 76. South southeast wind around. Low: 76 °F. Partly...

AccuWeather

Austin, TX Weather Forecast - AccuWeather

Looking Ahead. A thunderstorm tomorrow afternoon; storms will be slow moving and bring heavy rain....

The Weather Channel

10-Day Weather Forecast for Austin, TX

10 Day Weather-Austin, TX. As of 9:16 pm CDT. Tonight. --/77°. 7%. Night. 77°. 7%. S 8 mph. Partly...

KEYE

Austin Weather | News, Weather, Sports, Breaking News

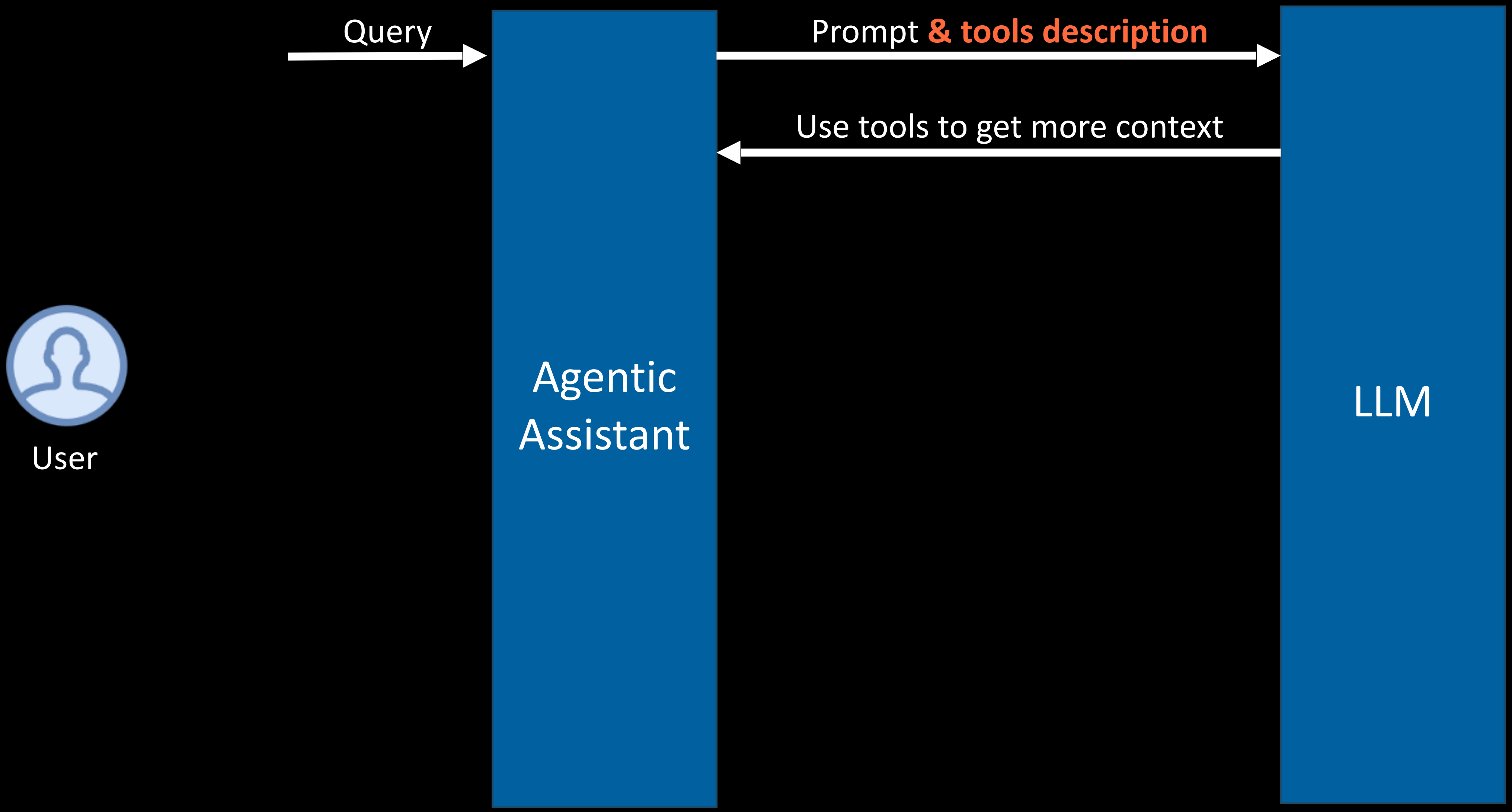
Monday and Tuesday, we may see the sea breeze a little more successful, possibly bringing spotty rain a...

statesman.com

Austin bakes in summer heat this week, but why

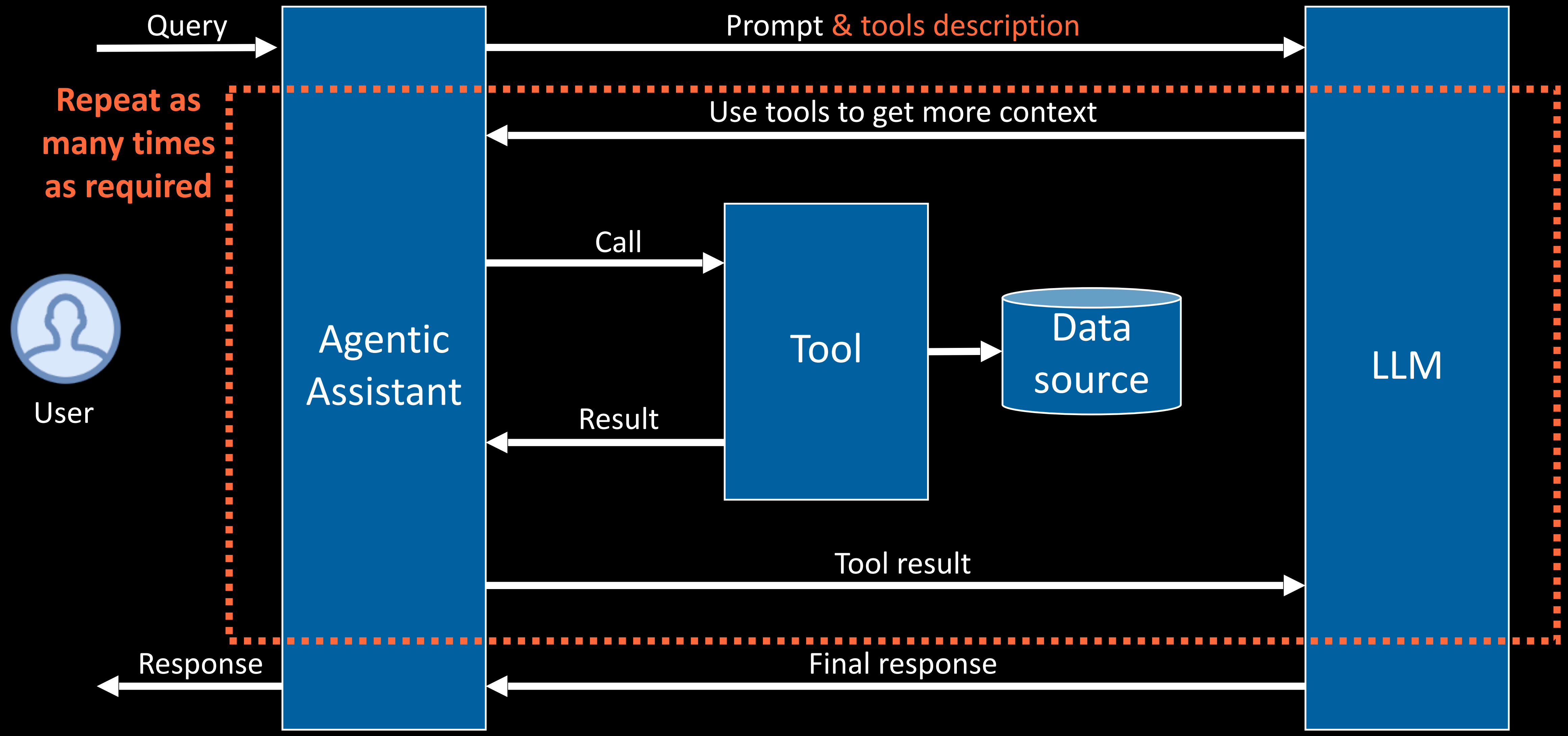
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

A simple agentic workflow



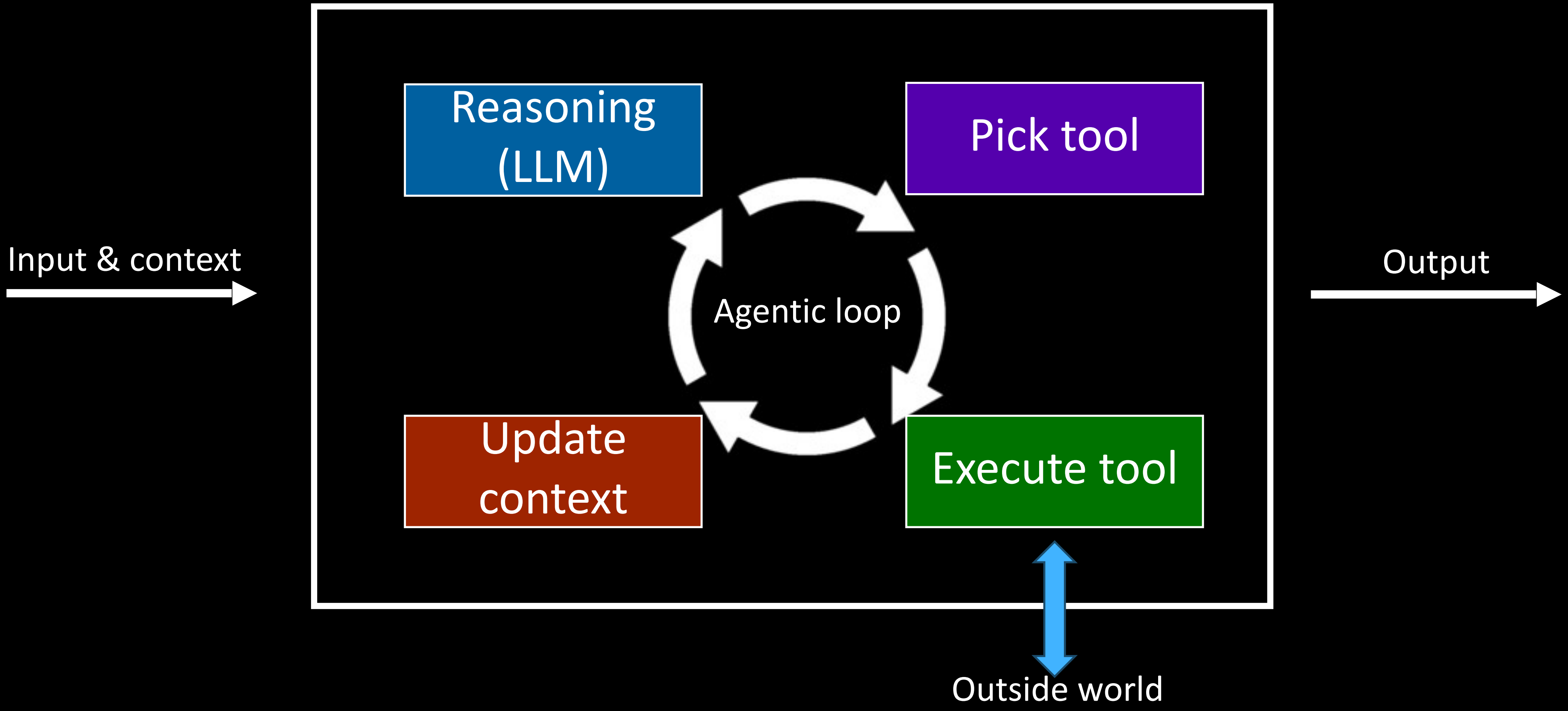
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

A simple agentic workflow


Query
Prompt & tools description
Repeat as many times as required
Use tools to get more context
User
Agentic Assistant
Call
Tool
Data source
Result
LLM
Tool result
Response
Final response

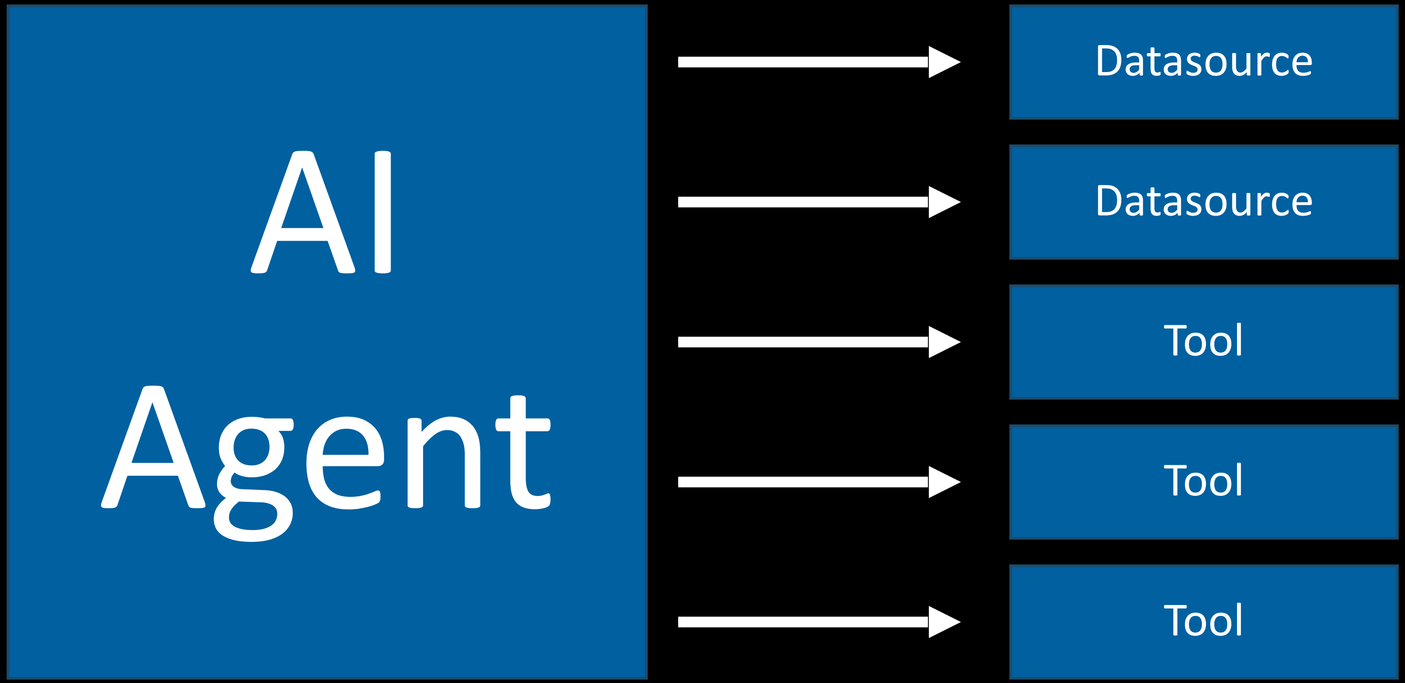
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Agentic AI system in a nutshell



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Communicating with multiple tools and resources



AI Agent

Datasource

Datasource

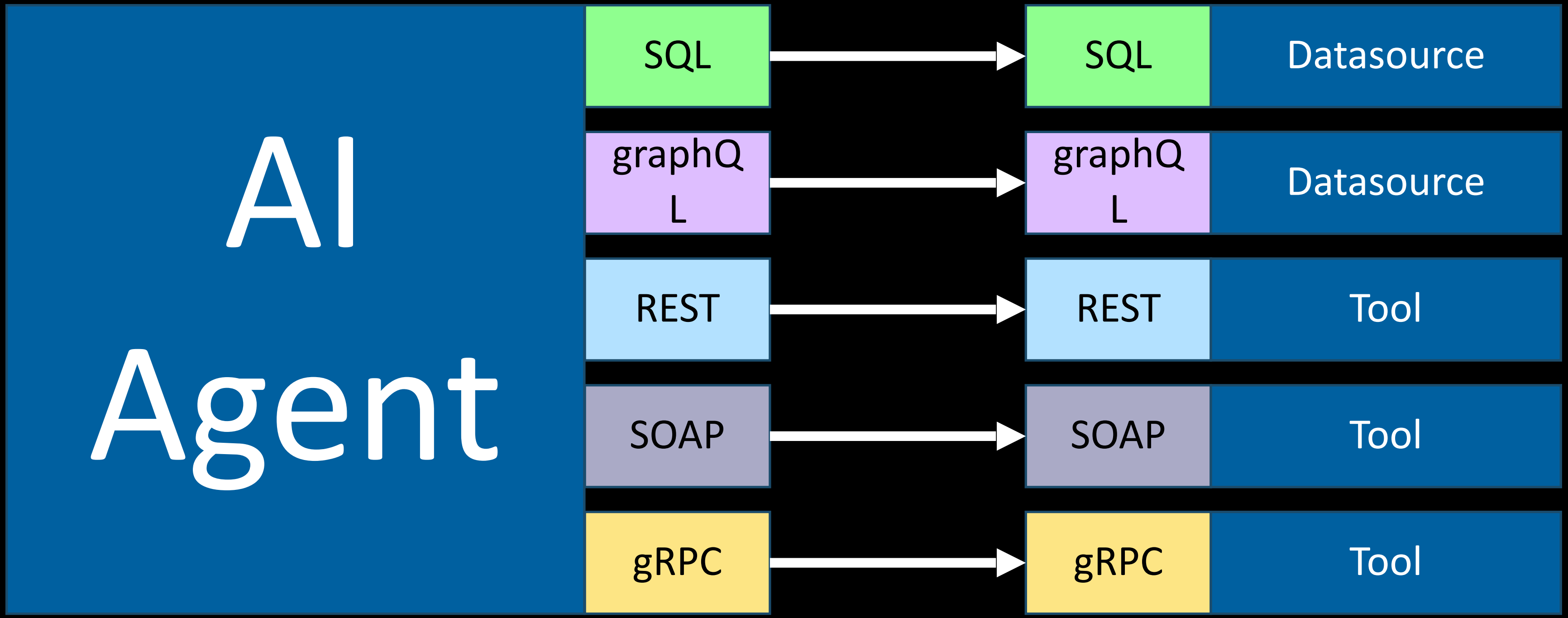
Tool

Tool

Tool

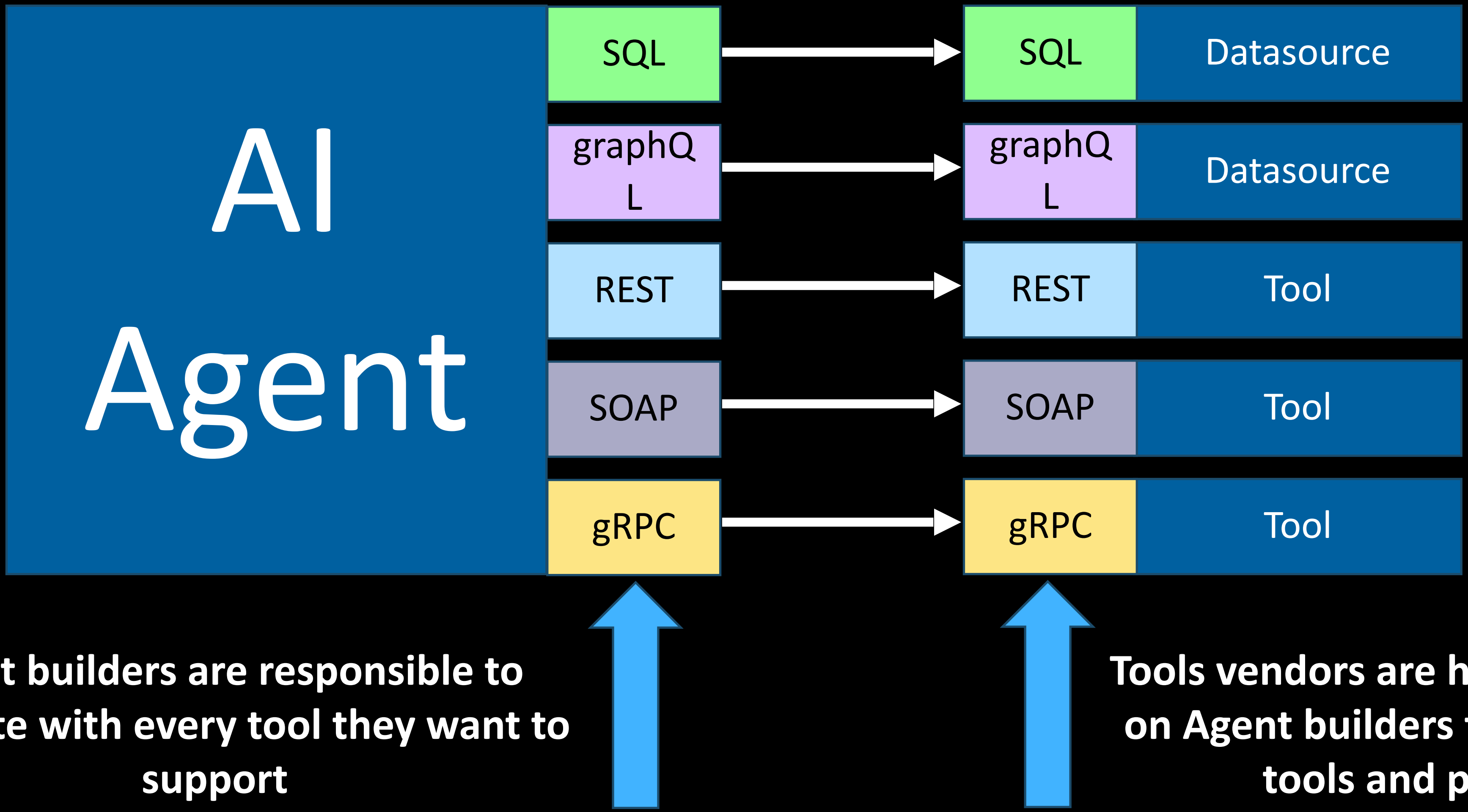
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Communicating with multiple tools and resources



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Communicating with multiple tools and resources

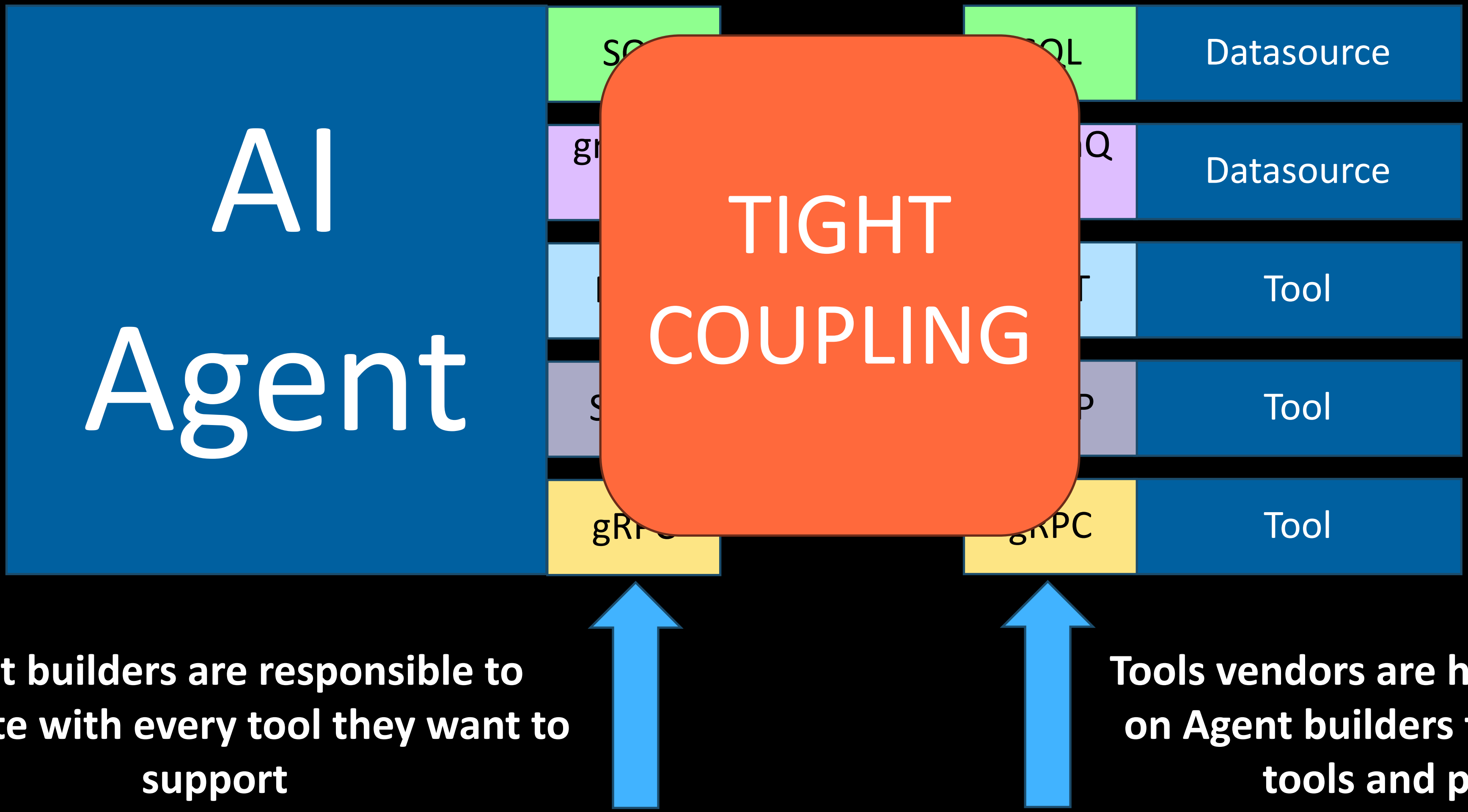


Agent builders are responsible to integrate with every tool they want to support

Tools vendors are highly dependent on Agent builders to support their tools and protocols

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Communicating with multiple tools and resources

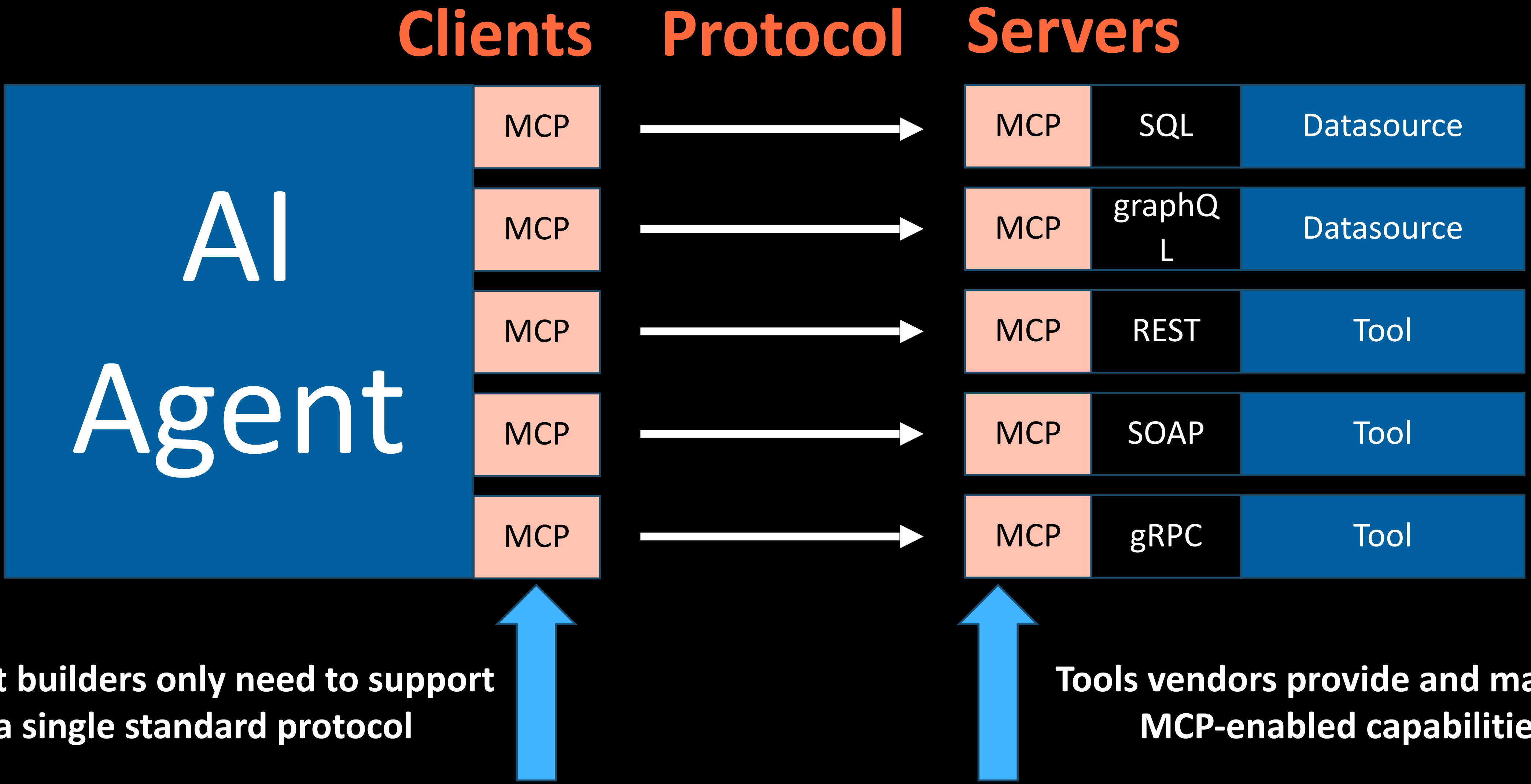


Agent builders are responsible to integrate with every tool they want to support

Tools vendors are highly dependent on Agent builders to support their tools and protocols

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Communicating with multiple tools and resources



Clients **Protocol** **Servers**

AI Agent

MCP → MCP SQL Datasource

MCP → MCP graphQL Datasource

MCP → MCP REST Tool

MCP → MCP SOAP Tool

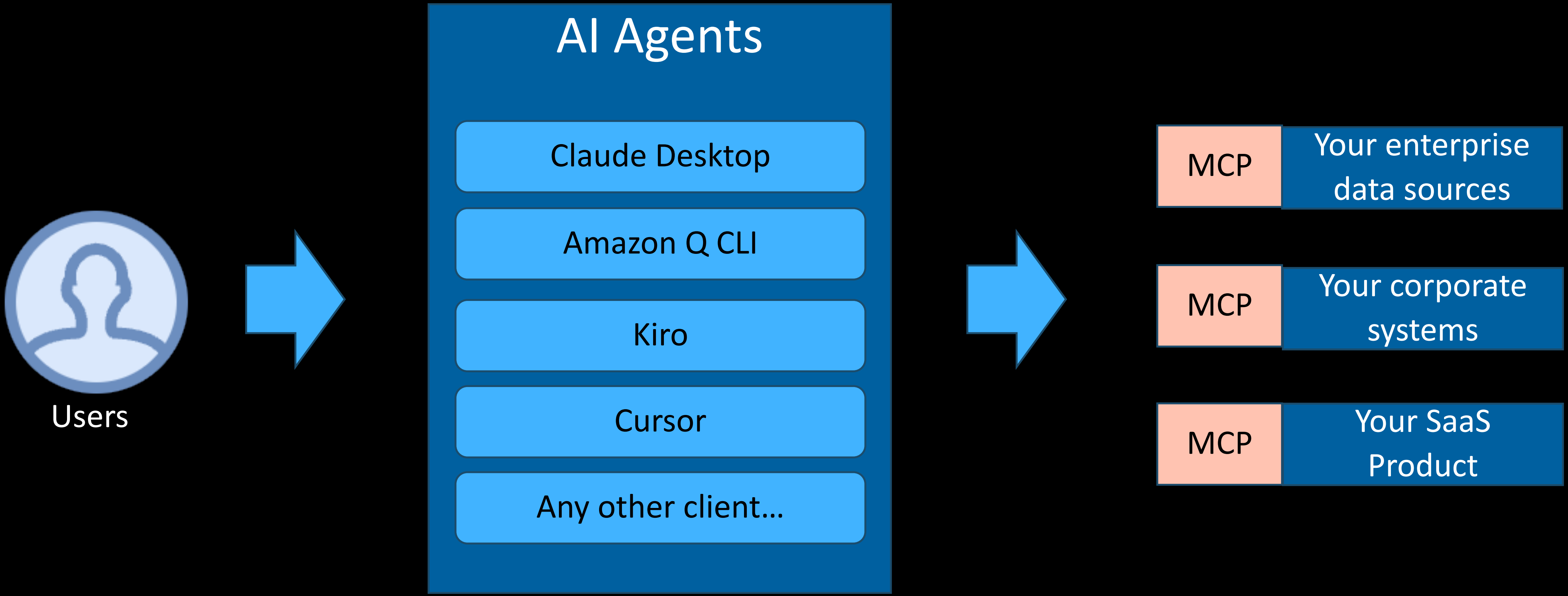
MCP → MCP gRPC Tool

Agent builders only need to support a single standard protocol

Tools vendors provide and maintain MCP-enabled capabilities

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Communicating with multiple tools and resources



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

The Building Blocks

What are the MCP components

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

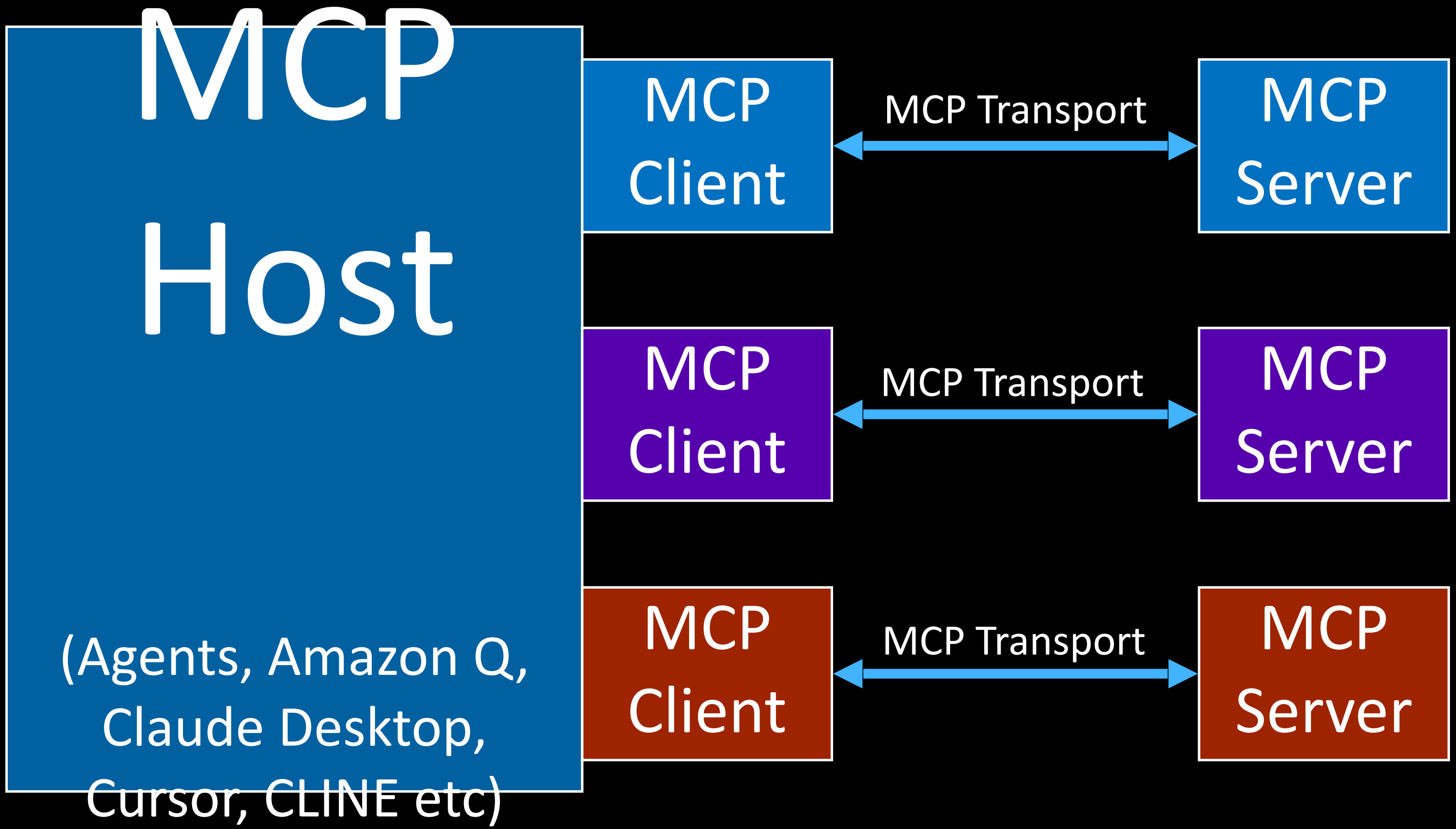
Model Context Protocol - Entities

MCP
Host

(Agents, Amazon Q,
Claude Desktop,
Cursor, CLINE etc)

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Model Context Protocol - Entities


MCP Host

(Agents, Amazon Q, Claude Desktop, Cursor, CLINE etc)

MCP Client

MCP Transport

MCP Server

MCP Client

MCP Transport

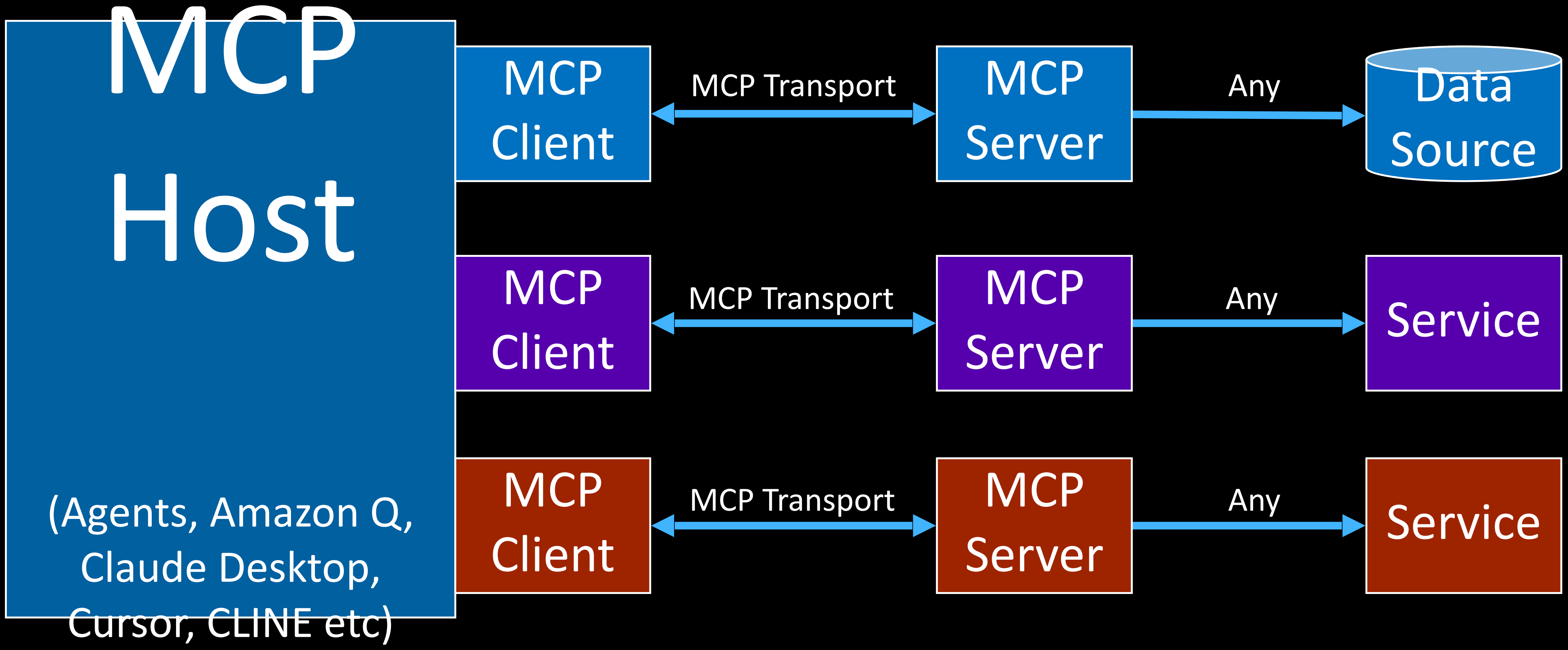
MCP Server

MCP Client

MCP Transport

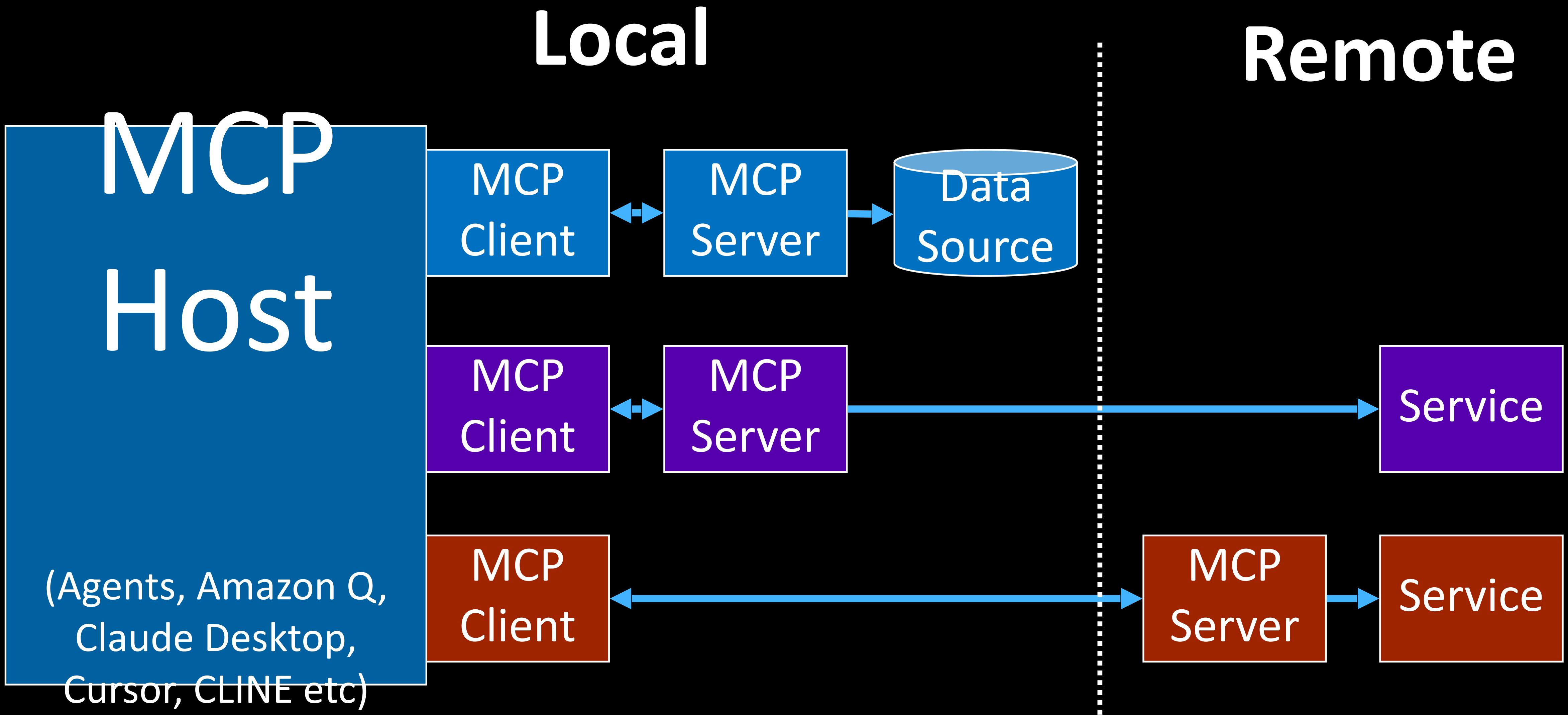
MCP Server

Model Context Protocol - Entities



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Model Context Protocol - Entities



Local

Remote

MCP Host

(Agents, Amazon Q, Claude Desktop, Cursor, CLINE etc)

MCP Client

MCP Server

Data Source

MCP Client

MCP Server

Service

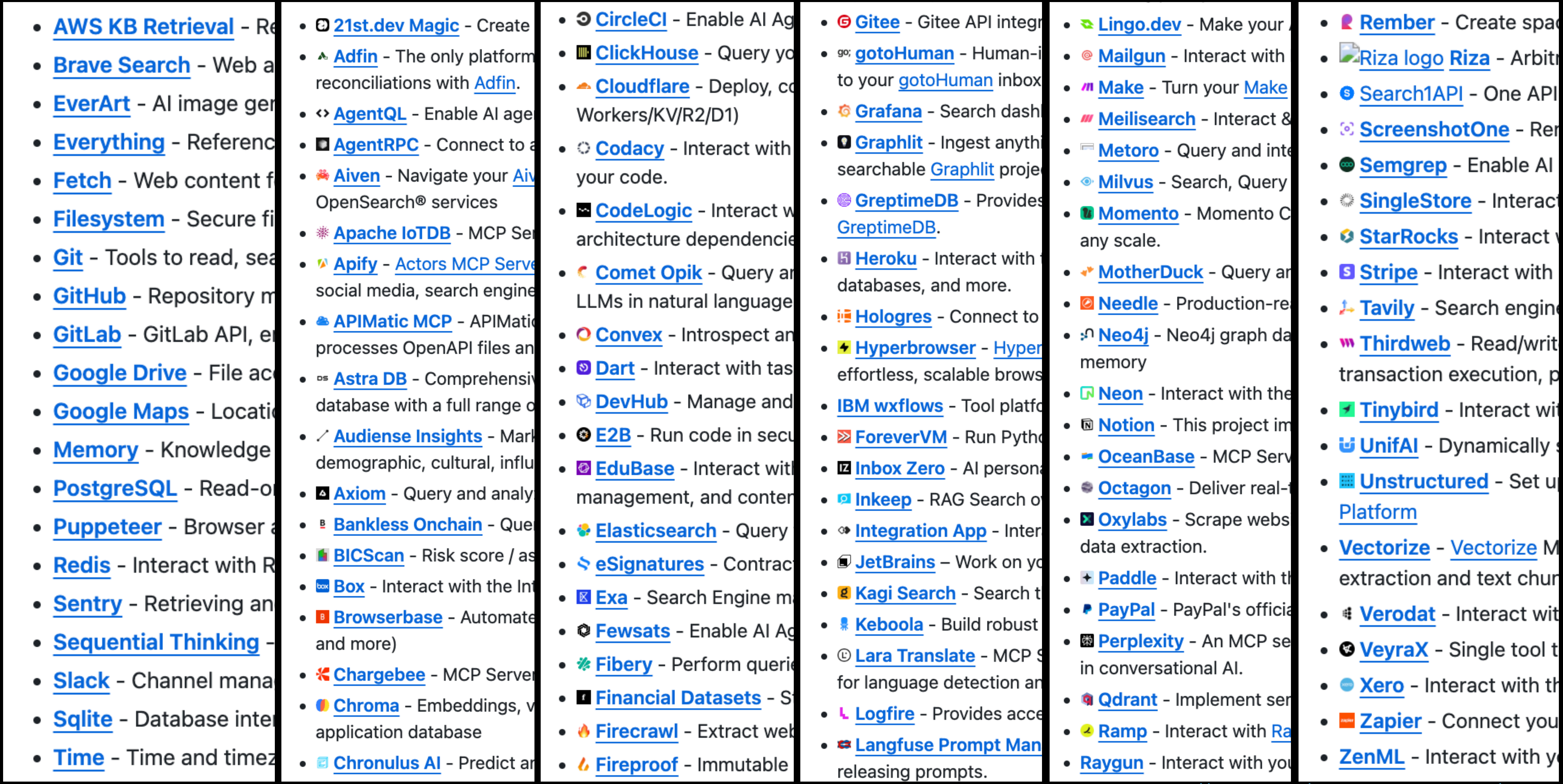
MCP Client

MCP Server

Service

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Model Context Protocol - Servers



- **AWS KB Retrieval** - Re
- **Brave Search** - Web a
- **EverArt** - AI image ge
- **Everything** - Referenc
- **Fetch** - Web content f
- **Filesystem** - Secure fi
- **Git** - Tools to read, sea
- **GitHub** - Repository m
- **GitLab** - GitLab API, e
- **Google Drive** - File ac
- **Google Maps** - Locati
- **Memory** - Knowledge
- **PostgreSQL** - Read-o
- **Puppeteer** - Browser
- **Redis** - Interact with R
- **Sentry** - Retrieving an
- **Sequential Thinking** -
- **Slack** - Channel mana
- **Sqlite** - Database inte
- **Time** - Time and timez

- **21st.dev Magic** - Create
- **Adfin** - The only platform
reconciliations with Adfin.
- **AgentQL** - Enable AI age
- **AgentRPC** - Connect to
- **Aiven** - Navigate your Ai
OpenSearch® services
- **Apache IoTDB** - MCP Se
- **Apify** - Actors MCP Serve
social media, search engine
- **APIMatic MCP** - APIMati
processes OpenAPI files an
- **Astra DB** - Comprehensi
database with a full range o
- **Audiense Insights** - Mar
demographic, cultural, influ
- **Axiom** - Query and analy
- **Bankless Onchain** - Que
- **BICScan** - Risk score / as
- **Box** - Interact with the In
- **Browserbase** - Automate
and more)
- **Chargebee** - MCP Serve
- **Chroma** - Embeddings, v
application database
- **Chronulus AI** - Predict a

- **CircleCI** - Enable AI Ag
- **ClickHouse** - Query yo
- **Cloudflare** - Deploy, c
Workers/KV/R2/D1)
- **Codacy** - Interact with
your code.
- **CodeLogic** - Interact w
architecture dependencie
- **Comet Opik** - Query a
LLMs in natural language
- **Convex** - Introspect an
- **Dart** - Interact with tas
- **DevHub** - Manage and
- **E2B** - Run code in secu
- **EduBase** - Interact wit
management, and conten
- **Elasticsearch** - Query
- **eSignatures** - Contrac
- **Exa** - Search Engine m
- **Fewsats** - Enable AI Ag
- **Fibery** - Perform queri
- **Financial Datasets** - S
- **Firecrawl** - Extract we
- **Fireproof** - Immutable

- **Gitee** - Gitee API integr
- **gotoHuman** - Human-i
to your gotoHuman inbox
- **Grafana** - Search dash
- **Graphlit** - Ingest anyth
searchable Graphlit proje
- **GreptimeDB** - Provides
GreptimeDB.
- **Heroku** - Interact with
databases, and more.
- **Hologres** - Connect to
- **Hyperbrowser** - Hyper
effortless, scalable brows
- **IBM wxflows** - Tool platf
- **ForeverVM** - Run Pyth
- **Inbox Zero** - AI person
- **Inkeep** - RAG Search o
- **Integration App** - Inter
- **JetBrains** – Work on yo
- **Kagi Search** - Search t
- **Keboola** - Build robust
- **Lara Translate** - MCP
for language detection an
- **Logfire** - Provides acce
- **Langfuse Prompt Man**
releasing prompts.

- **Lingo.dev** - Make your
- **Mailgun** - Interact with
- **Make** - Turn your Make
- **Meilisearch** - Interact &
- **Metoro** - Query and inte
- **Milvus** - Search, Query
- **Momento** - Momento C
any scale.
- **MotherDuck** - Query a
- **Needle** - Production-re
- **Neo4j** - Neo4j graph da
memory
- **Neon** - Interact with the
- **Notion** - This project im
- **OceanBase** - MCP Serv
- **Octagon** - Deliver real-
- **Oxylabs** - Scrape webs
data extraction.
- **Paddle** - Interact with th
- **PayPal** - PayPal's officia
- **Perplexity** - An MCP se
in conversational AI.
- **Qdrant** - Implement se
- **Ramp** - Interact with Ra
- **Raygun** - Interact with you

- **Rember** - Create spa
- Riza logo **Riza** - Arbit
- **Search1API** - One API
- **ScreenshotOne** - Re
- **Semgrep** - Enable AI
- **SingleStore** - Interact
- **StarRocks** - Interact
- **Stripe** - Interact with
- **Tavily** - Search engine
- **Thirdweb** - Read/writ
transaction execution, p
- **Tinybird** - Interact wi
- **UnifAI** - Dynamically
- **Unstructured** - Set u
Platform
- **Vectorize** - Vectorize M
extraction and text chun
- **Verodat** - Interact wit
- **VeyraX** - Single tool t
- **Xero** - Interact with th
- **Zapier** - Connect you
- **ZenML** - Interact with y

https://github.com/modelcontextprotocol/servers

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

MCP server building blocks

Resources

Enable **applications** to retrieve additional data

Tools

Enable **models** to perform actions

Prompts

Enable **users** with reusable prompt templates

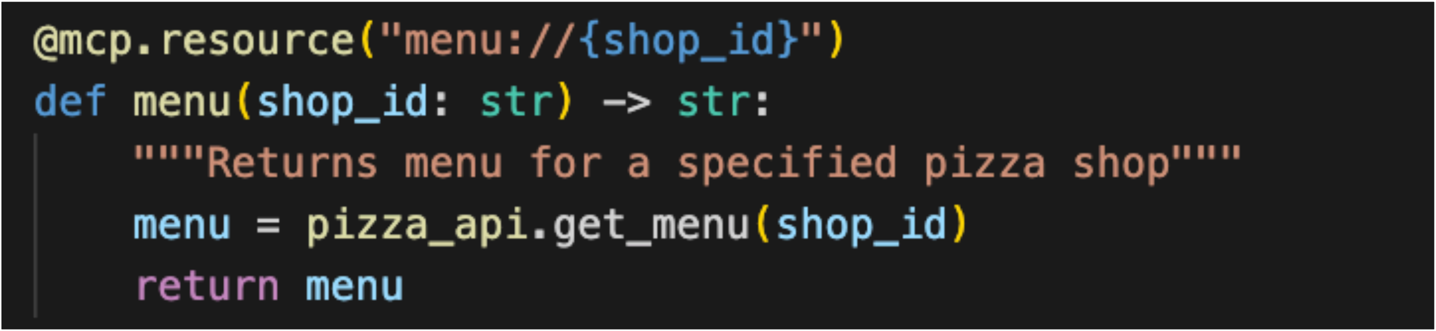
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Resources

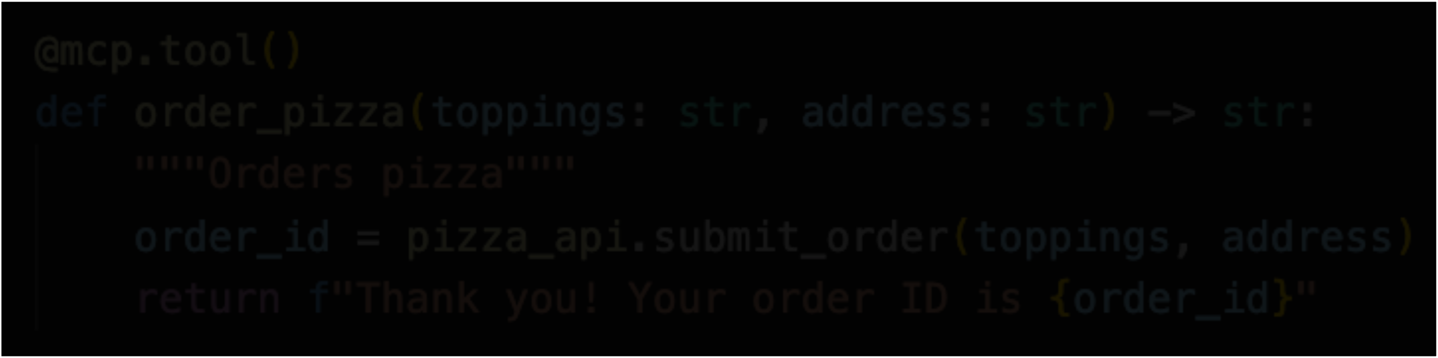
Resources

Tools

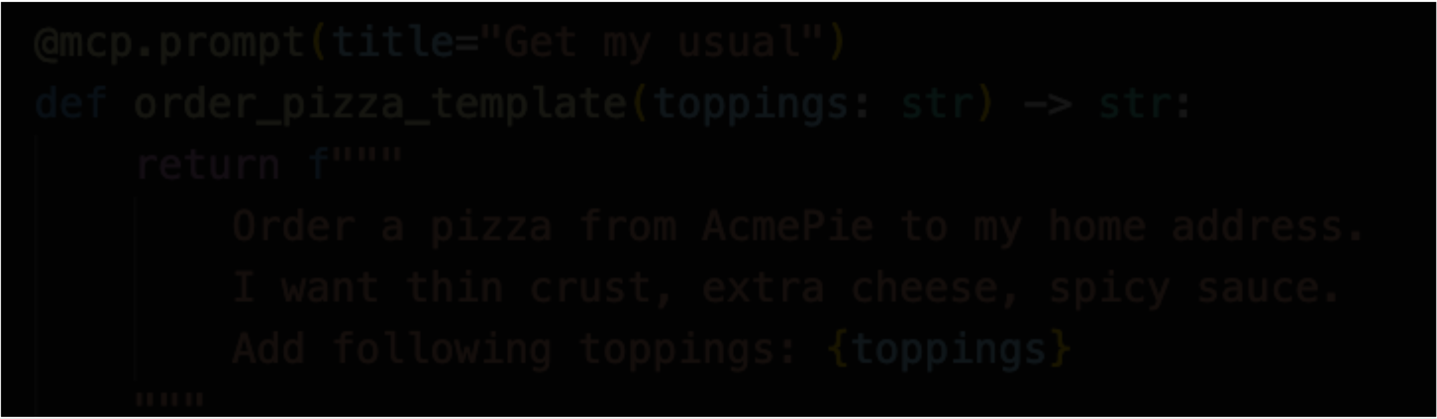
Prompts



```
@mcp.resource("menu://{shop_id}")
def menu(shop_id: str) -> str:
    """Returns menu for a specified pizza shop"""
    menu = pizza_api.get_menu(shop_id)
    return menu
```



```
@mcp.tool()
def order_pizza(toppings: str, address: str) -> str:
    """Orders pizza"""
    order_id = pizza_api.submit_order(toppings, address)
    return f"Thank you! Your order ID is {order_id}"
```



```
@mcp.prompt(title="Get my usual")
def order_pizza_template(toppings: str) -> str:
    return f"""
        Order a pizza from AcmePie to my home address.
        I want thin crust, extra cheese, spicy sauce.
        Add following toppings: {toppings}
    """
```

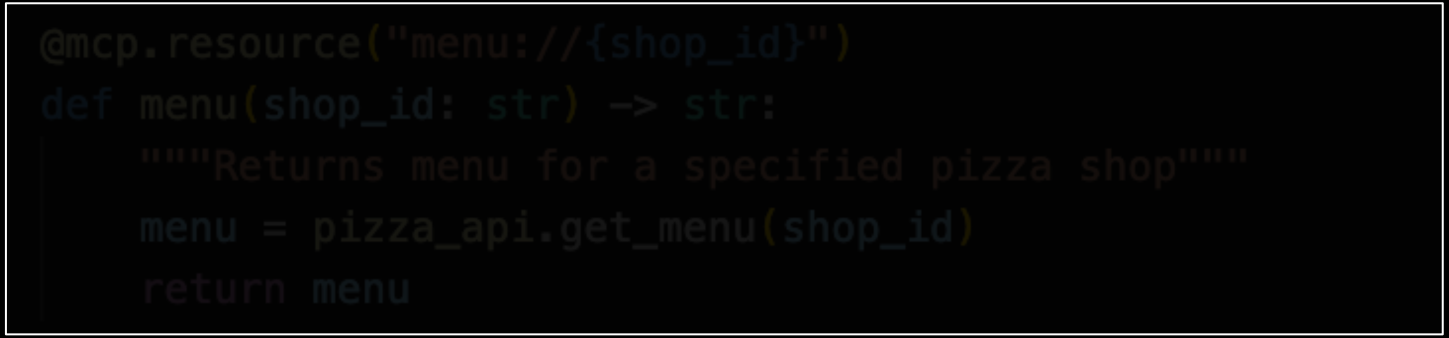
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Tools

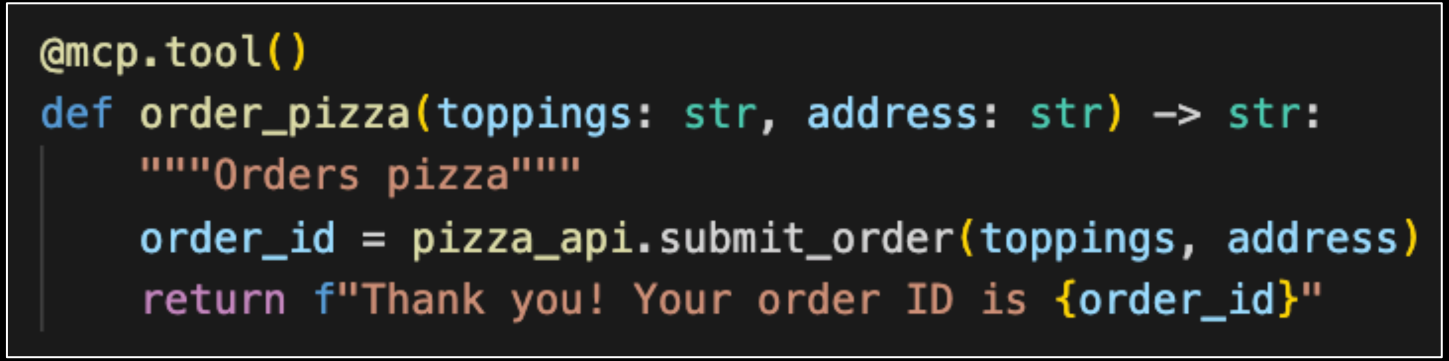
Resources

Tools

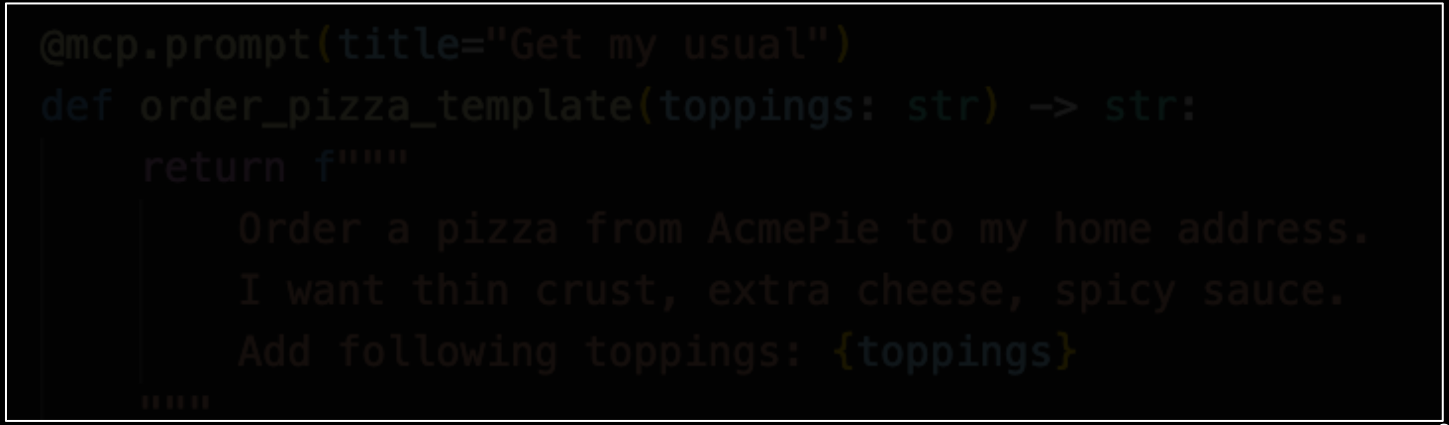
Prompts



```
@mcp.resource("menu://{shop_id}")
def menu(shop_id: str) -> str:
    """Returns menu for a specified pizza shop"""
    menu = pizza_api.get_menu(shop_id)
    return menu
```



```
@mcp.tool()
def order_pizza(toppings: str, address: str) -> str:
    """Orders pizza"""
    order_id = pizza_api.submit_order(toppings, address)
    return f"Thank you! Your order ID is {order_id}"
```



```
@mcp.prompt(title="Get my usual")
def order_pizza_template(toppings: str) -> str:
    return f"""
        Order a pizza from AcmePie to my home address.
        I want thin crust, extra cheese, spicy sauce.
        Add following toppings: {toppings}
    """
```

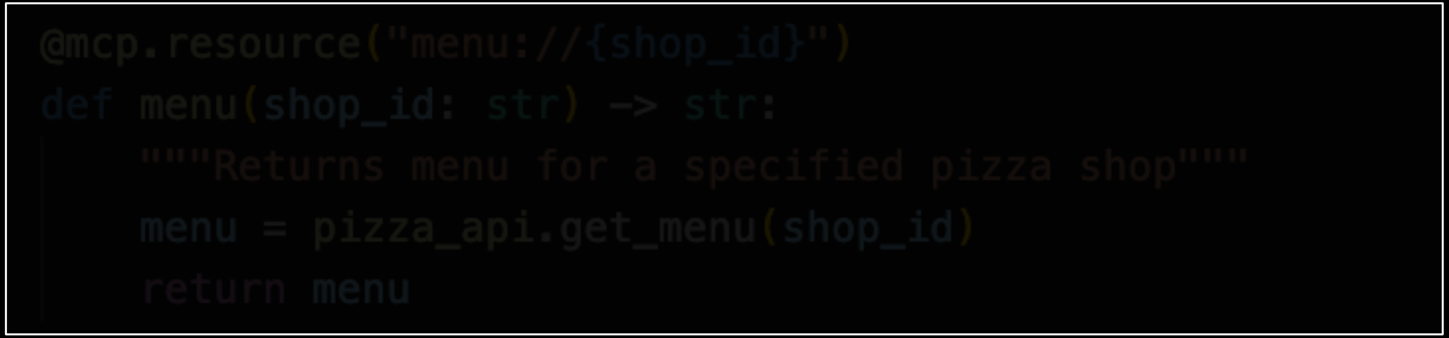
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Prompts

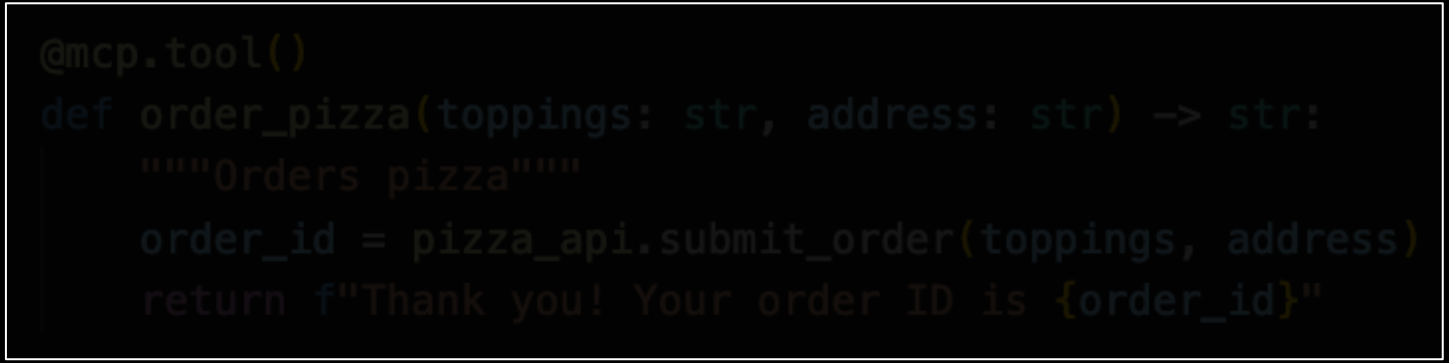
Resources

Tools

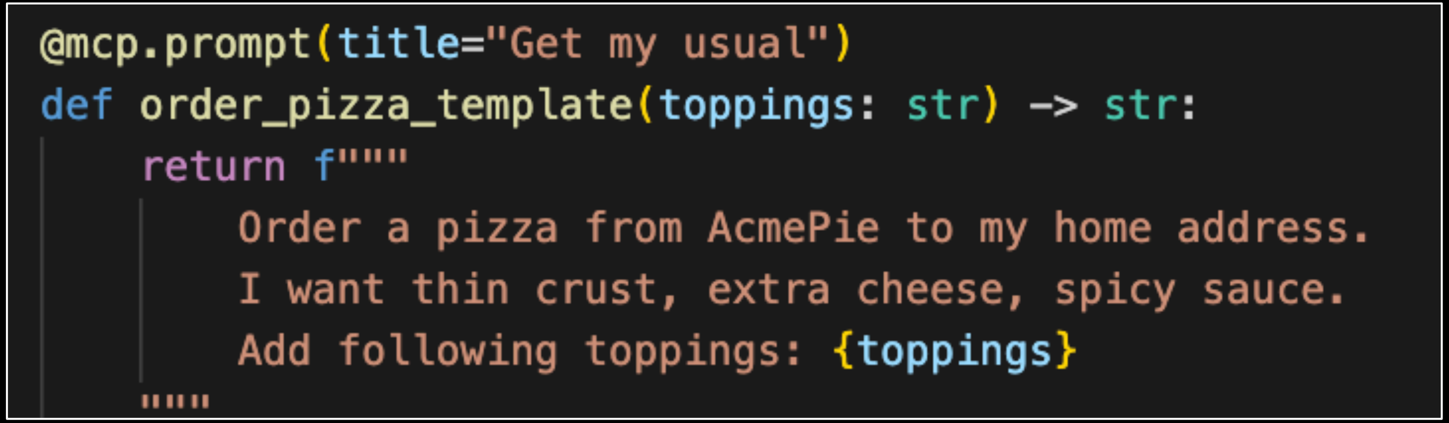
Prompts



```
@mcp.resource("menu://{shop_id}")
def menu(shop_id: str) -> str:
    """Returns menu for a specified pizza shop"""
    menu = pizza_api.get_menu(shop_id)
    return menu
```



```
@mcp.tool()
def order_pizza(toppings: str, address: str) -> str:
    """Orders pizza"""
    order_id = pizza_api.submit_order(toppings, address)
    return f"Thank you! Your order ID is {order_id}"
```



```
@mcp.prompt(title="Get my usual")
def order_pizza_template(toppings: str) -> str:
    return f"""
        Order a pizza from AcmePie to my home address.
        I want thin crust, extra cheese, spicy sauce.
        Add following toppings: {toppings}
    """
```

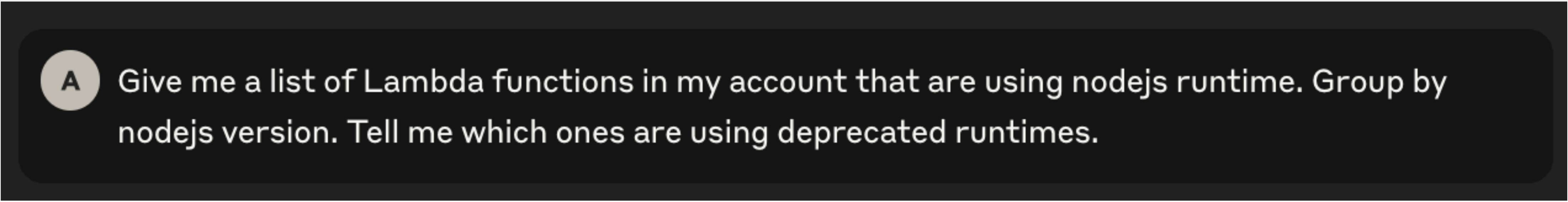
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Let’s build a solution

Building your first MCP server **for** AWS

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

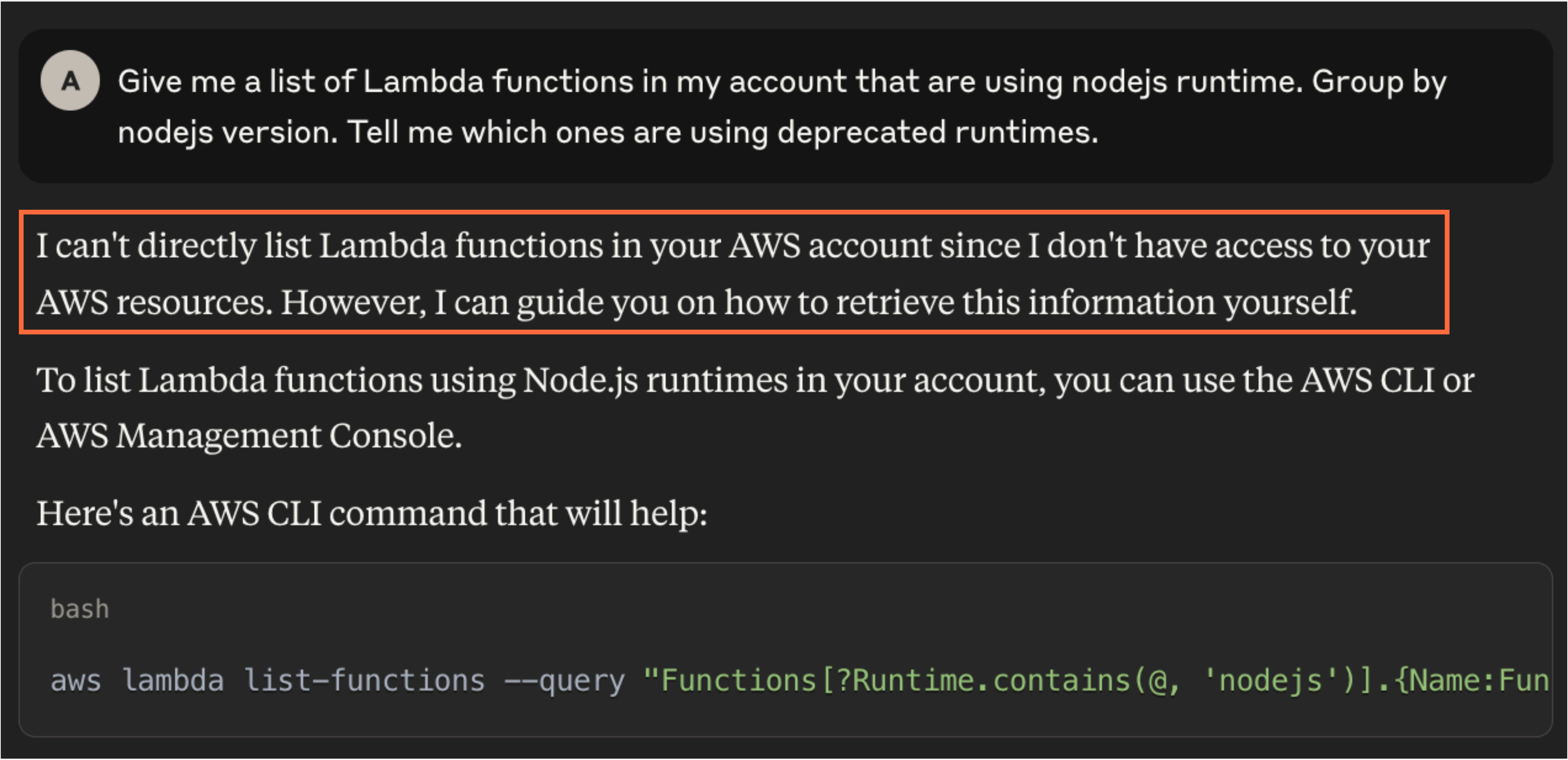
Building your first MCP Server



Give me a list of Lambda functions in my account that are using nodejs runtime. Group by nodejs version. Tell me which ones are using deprecated runtimes.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server



Give me a list of Lambda functions in my account that are using nodejs runtime. Group by nodejs version. Tell me which ones are using deprecated runtimes.

I can't directly list Lambda functions in your AWS account since I don't have access to your AWS resources. However, I can guide you on how to retrieve this information yourself.

To list Lambda functions using Node.js runtimes in your account, you can use the AWS CLI or AWS Management Console.

Here's an AWS CLI command that will help:

```bash
aws lambda list-functions --query "Functions[?Runtime.contains(@, 'nodejs')].{Name:Fun
```

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server

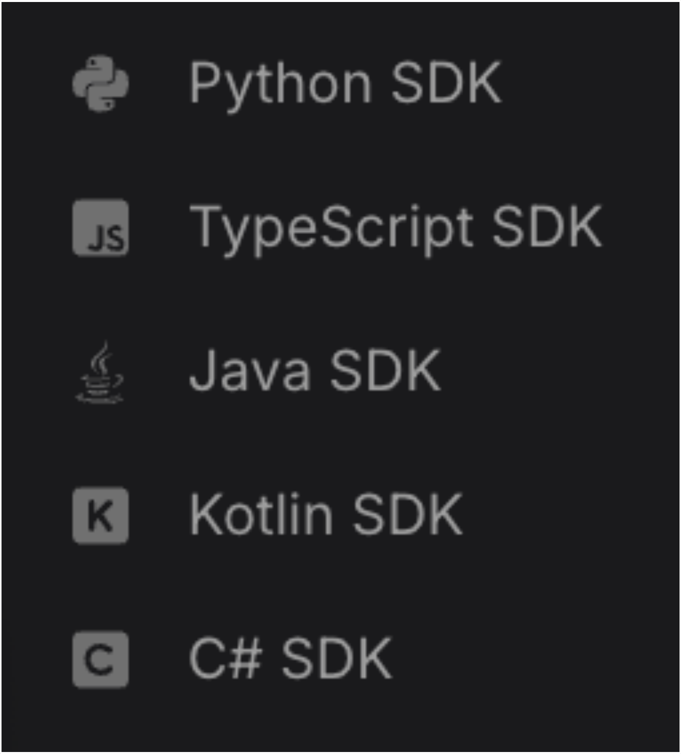
I want to use AI assistant to ask questions about Lambda functions in my account

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server

Do not implement the low-level wire protocol yourself.

Use official open-source SDKs.

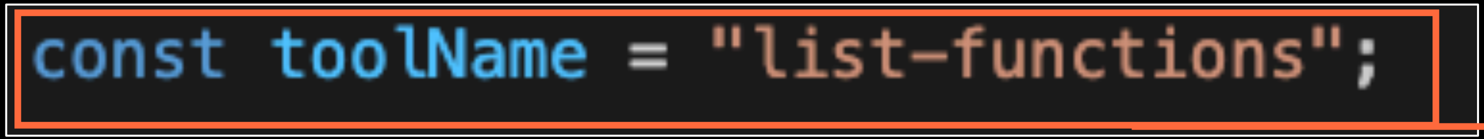


- Python SDK
- TypeScript SDK
- Java SDK
- Kotlin SDK
- C# SDK

https://modelcontextprotocol.io/introduction

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Create Tools

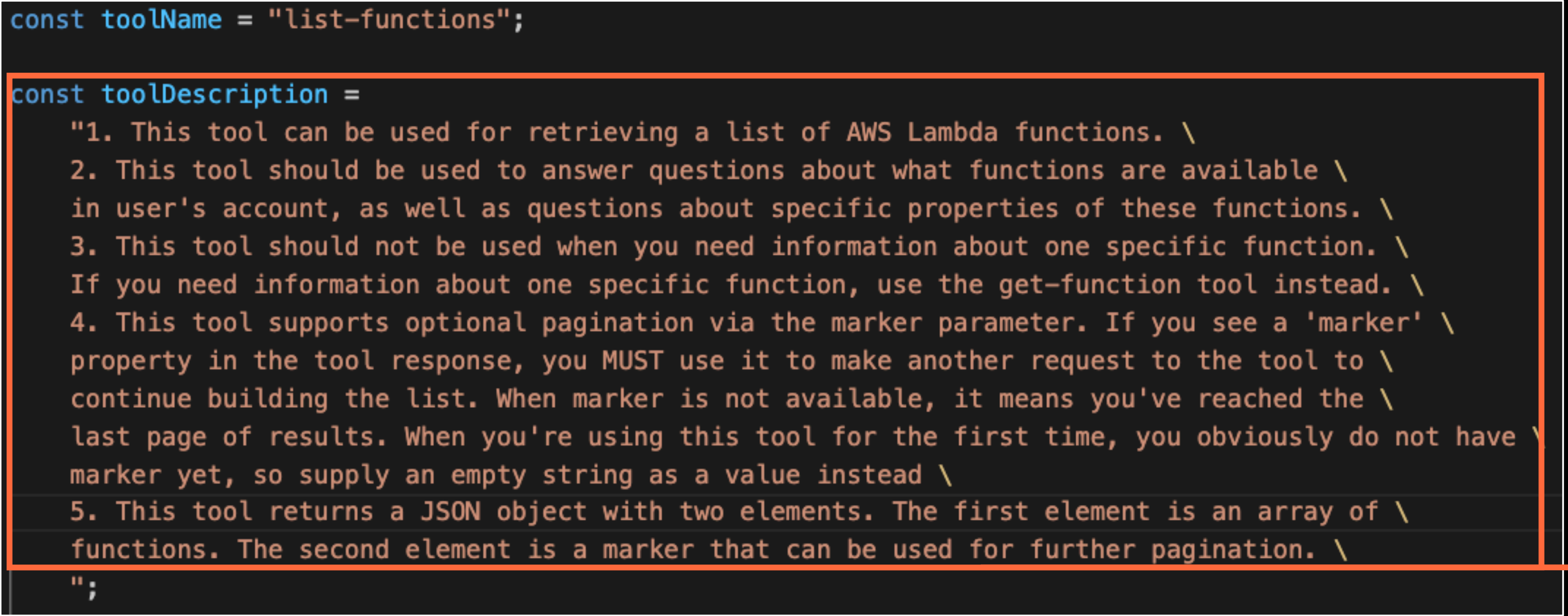


```
const toolName = "list-functions";
```

Self-explanatory names

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Create Tools

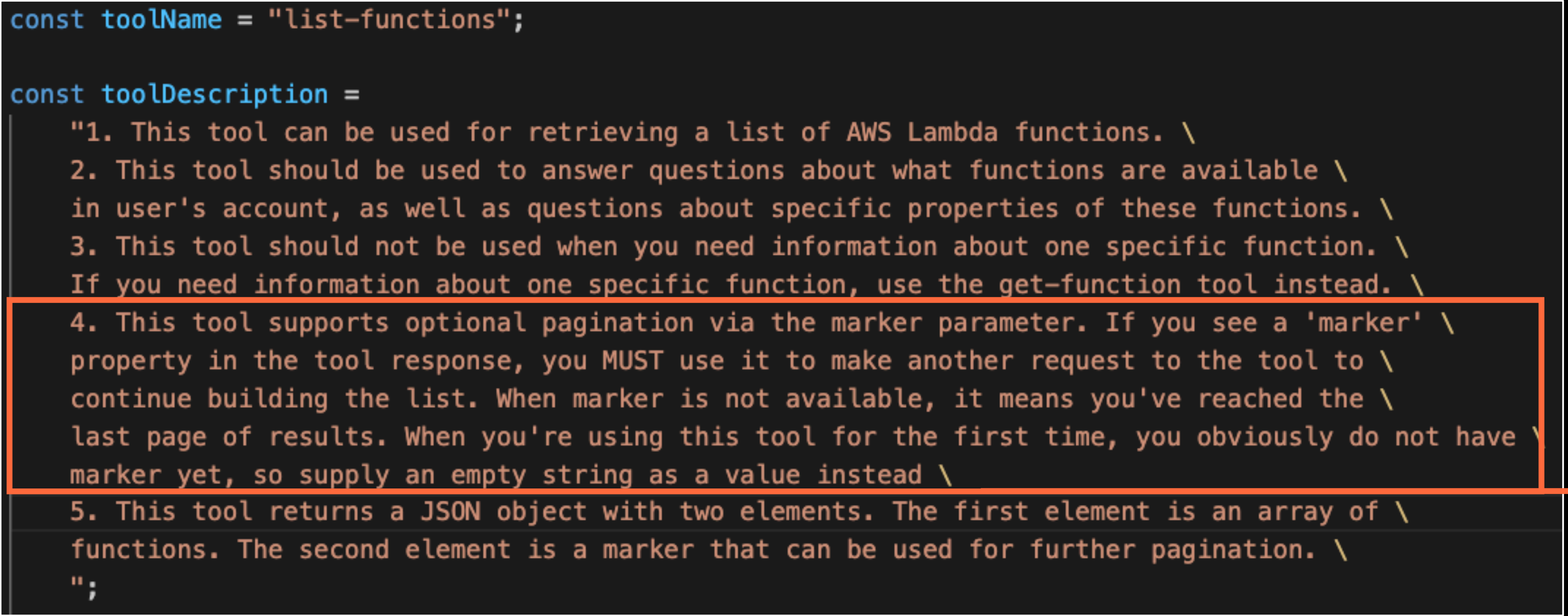


```
const toolName = "list-functions";

const toolDescription =
    "1. This tool can be used for retrieving a list of AWS Lambda functions. \
    2. This tool should be used to answer questions about what functions are available \
    in user's account, as well as questions about specific properties of these functions. \
    3. This tool should not be used when you need information about one specific function. \
    If you need information about one specific function, use the get-function tool instead. \
    4. This tool supports optional pagination via the marker parameter. If you see a 'marker' \
    property in the tool response, you MUST use it to make another request to the tool to \
    continue building the list. When marker is not available, it means you've reached the \
    last page of results. When you're using this tool for the first time, you obviously do not have \
    marker yet, so supply an empty string as a value instead \
    5. This tool returns a JSON object with two elements. The first element is an array of \
    functions. The second element is a marker that can be used for further pagination. \
    ";
```

Descriptions are crucial

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Create Tools

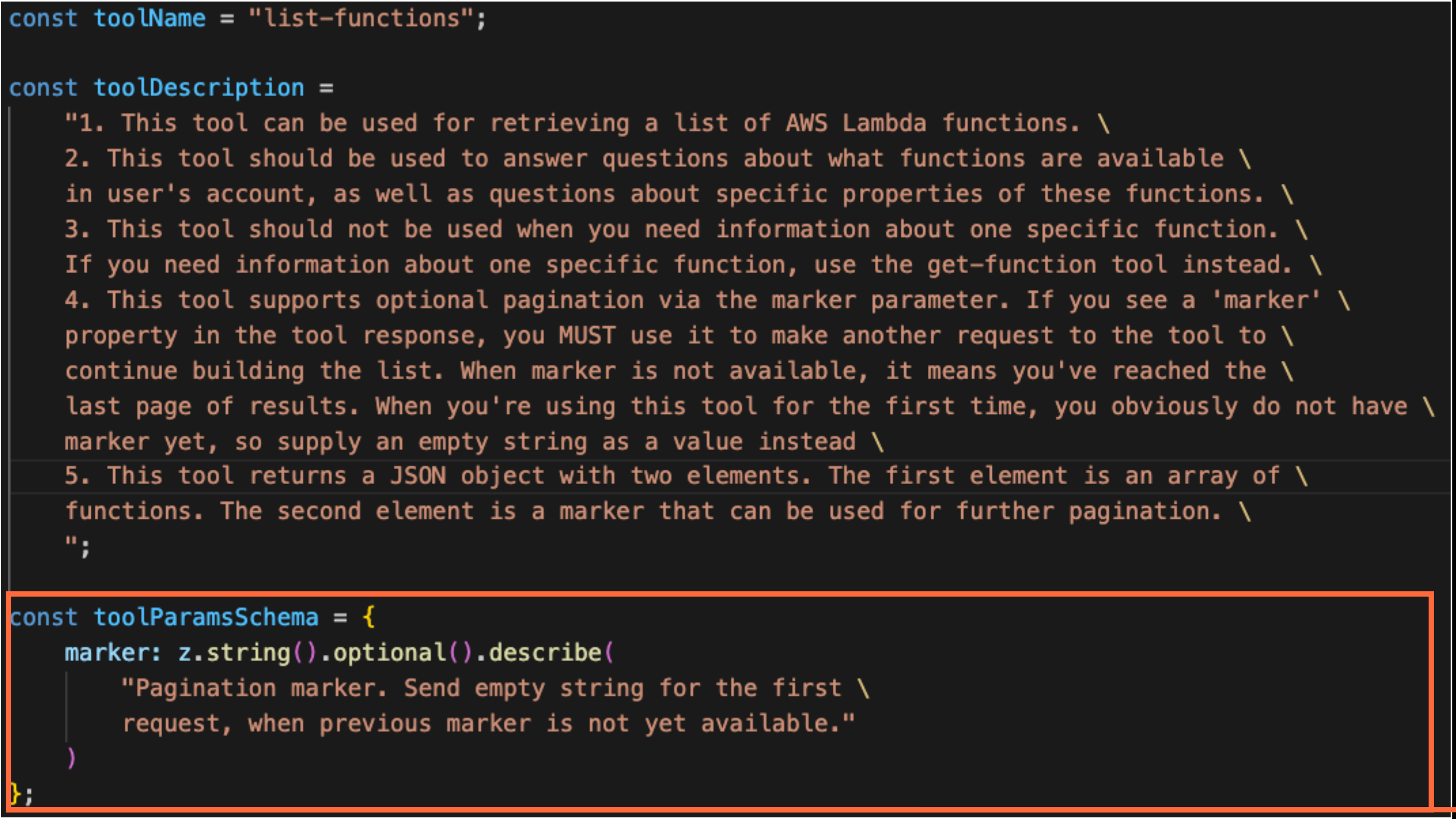


```
const toolName = "list-functions";

const toolDescription =
    "1. This tool can be used for retrieving a list of AWS Lambda functions. \
    2. This tool should be used to answer questions about what functions are available \
    in user's account, as well as questions about specific properties of these functions. \
    3. This tool should not be used when you need information about one specific function. \
    If you need information about one specific function, use the get-function tool instead. \
    4. This tool supports optional pagination via the marker parameter. If you see a 'marker' \
    property in the tool response, you MUST use it to make another request to the tool to \
    continue building the list. When marker is not available, it means you've reached the \
    last page of results. When you're using this tool for the first time, you obviously do not have \
    marker yet, so supply an empty string as a value instead \
    5. This tool returns a JSON object with two elements. The first element is an array of \
    functions. The second element is a marker that can be used for further pagination. \
    ";
```

Descriptions are crucial

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Create Tools

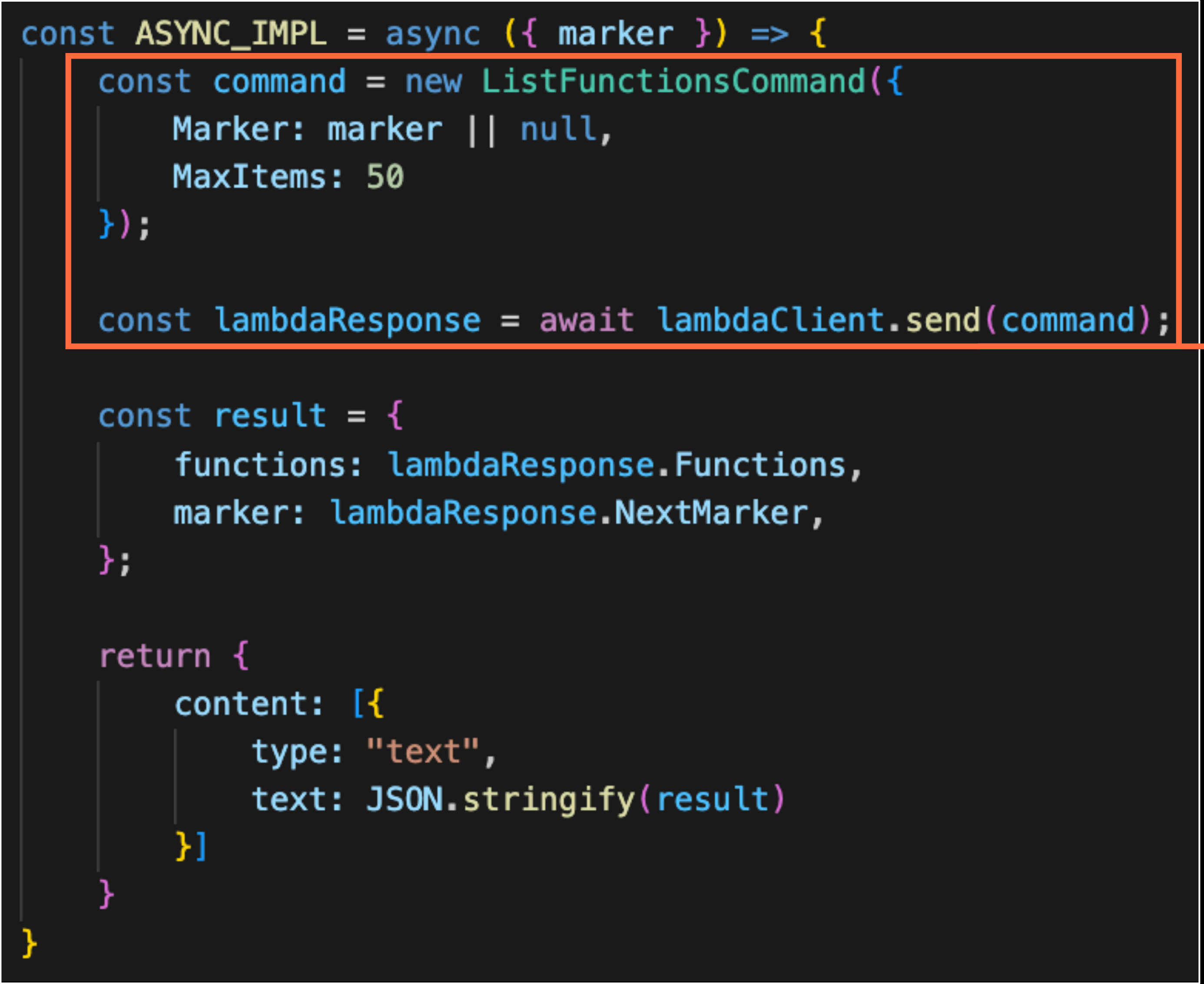


```
const toolName = "list-functions";

const toolDescription =
    "1. This tool can be used for retrieving a list of AWS Lambda functions. \
    2. This tool should be used to answer questions about what functions are available \
    in user's account, as well as questions about specific properties of these functions. \
    3. This tool should not be used when you need information about one specific function. \
    If you need information about one specific function, use the get-function tool instead. \
    4. This tool supports optional pagination via the marker parameter. If you see a 'marker' \
    property in the tool response, you MUST use it to make another request to the tool to \
    continue building the list. When marker is not available, it means you've reached the \
    last page of results. When you're using this tool for the first time, you obviously do not have \
    marker yet, so supply an empty string as a value instead \
    5. This tool returns a JSON object with two elements. The first element is an array of \
    functions. The second element is a marker that can be used for further pagination. \
    ";

const toolParamsSchema = {
    marker: z.string().optional().describe(
        "Pagination marker. Send empty string for the first \
        request, when previous marker is not yet available."
    )
};
```

Parameters schema allows AI assistants to know which parameters tools support

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Create Tools

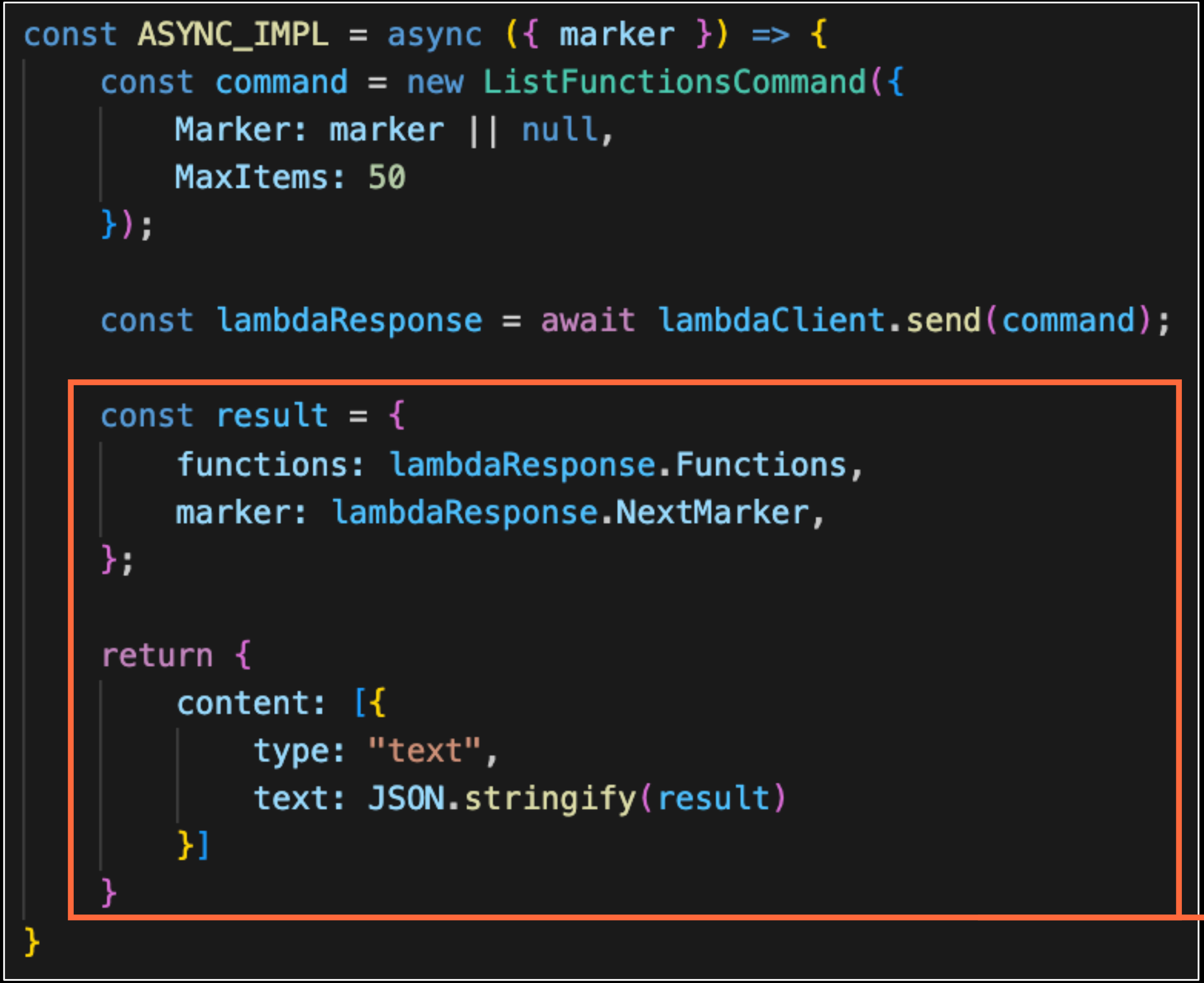


```
const ASYNC_IMPL = async ({ marker }) => {
    const command = new ListFunctionsCommand({
        Marker: marker || null,
        MaxItems: 50
    });

    const lambdaResponse = await lambdaClient.send(command);

    const result = {
        functions: lambdaResponse.Functions,
        marker: lambdaResponse.NextMarker,
    };

    return {
        content: [{
            type: "text",
            text: JSON.stringify(result)
        }]
    }
}
```

Uses AWS SDK and Lambda client to retrieve the list of functions. Uses next page marker, if available.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Create Tools

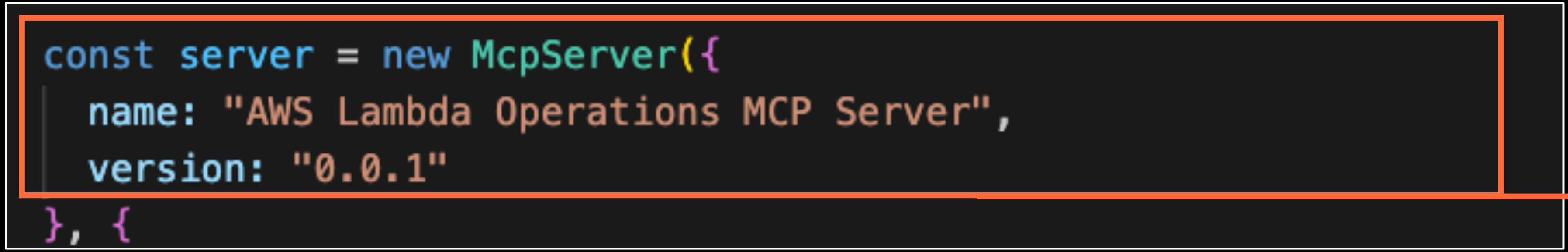


```
const ASYNC_IMPL = async ({ marker }) => {
    const command = new ListFunctionsCommand({
        Marker: marker || null,
        MaxItems: 50
    });

    const lambdaResponse = await lambdaClient.send(command);

    const result = {
        functions: lambdaResponse.Functions,
        marker: lambdaResponse.NextMarker,
    };

    return {
        content: [{
            type: "text",
            text: JSON.stringify(result)
        }]
    }
}
```

Return the list of functions and next page marker

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Create Server

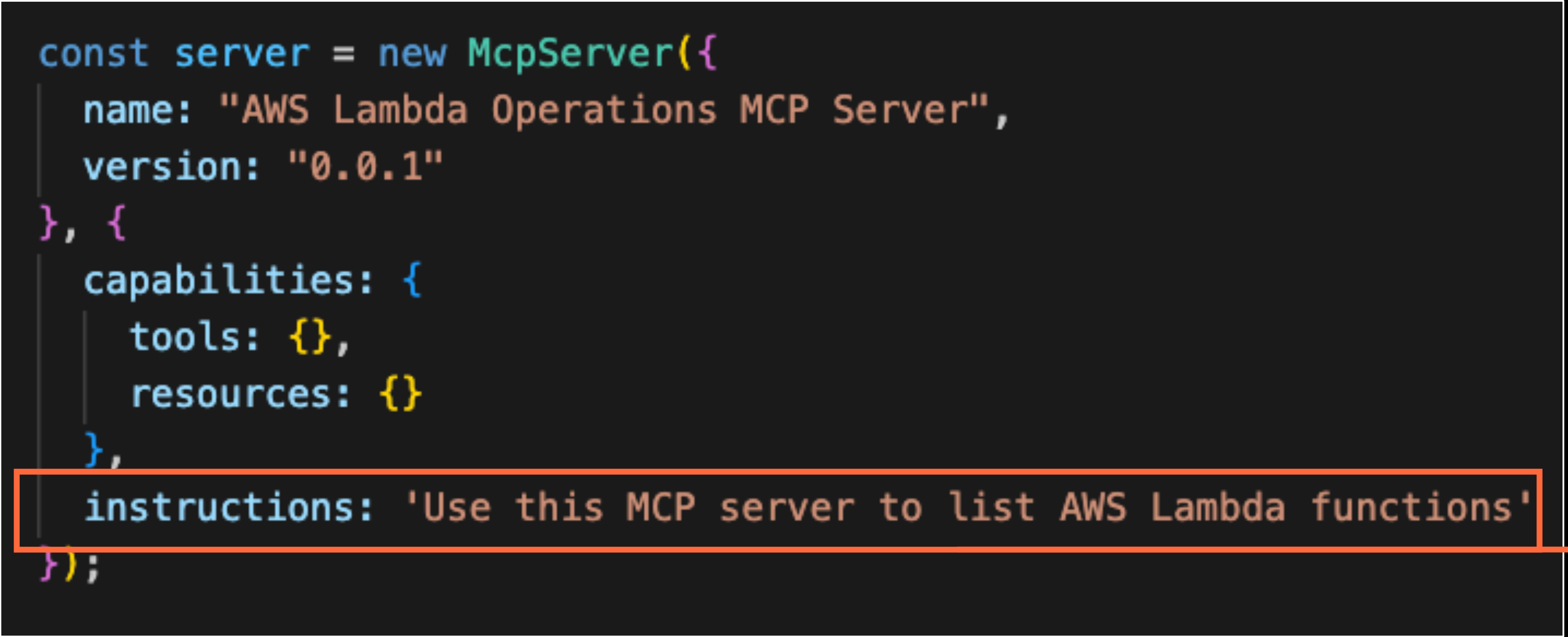


```
const server = new McpServer({
  name: "AWS Lambda Operations MCP Server",
  version: "0.0.1"
}, {
```

Define server name and version

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Create Server

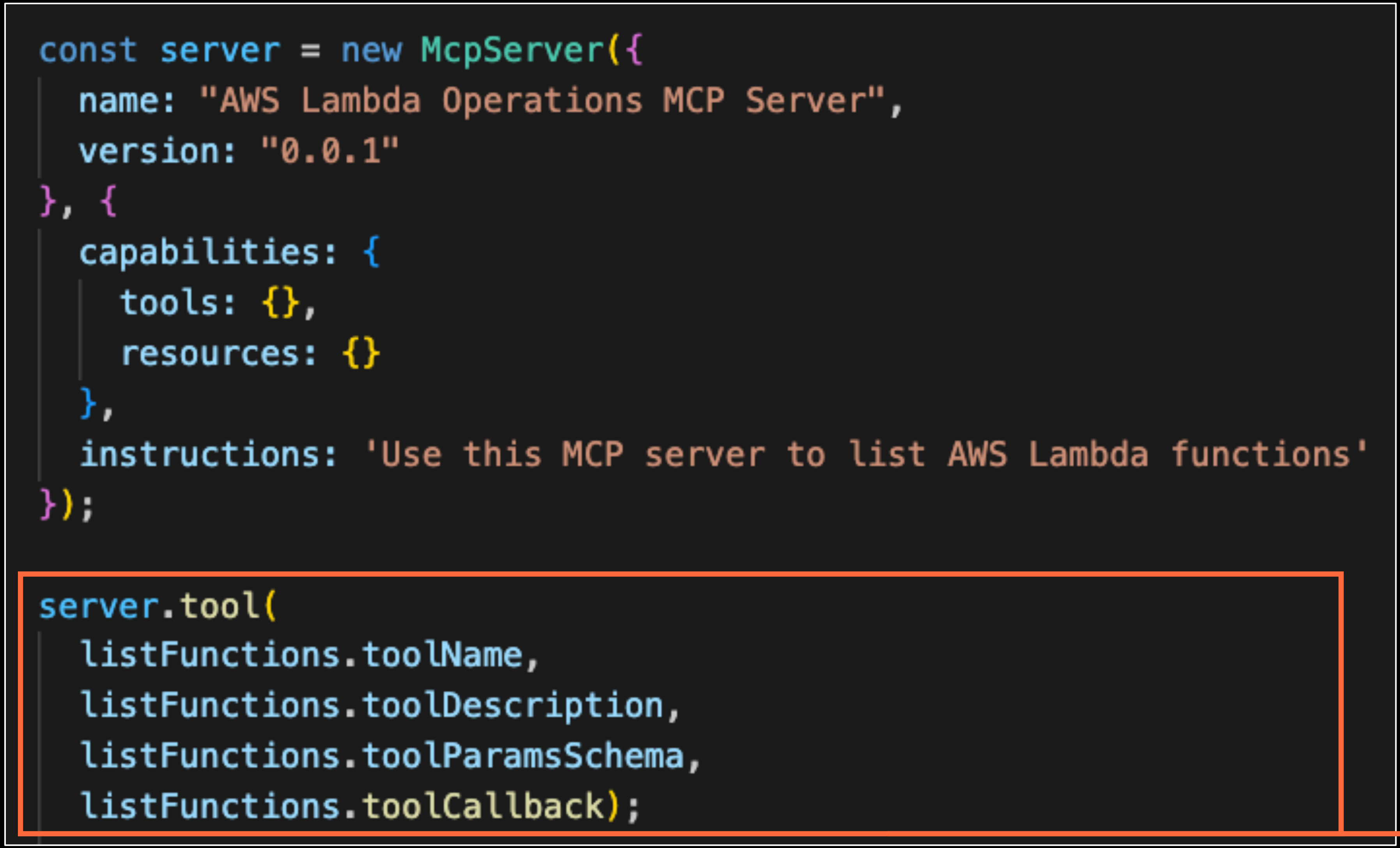


```
const server = new McpServer({
  name: "AWS Lambda Operations MCP Server",
  version: "0.0.1"
}, {
  capabilities: {
    tools: {},
    resources: {}
  },
  instructions: 'Use this MCP server to list AWS Lambda functions'
});
```

Detailed instructions are crucial. They allow LLM understand why, when, and how they should be using your server.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Create Server



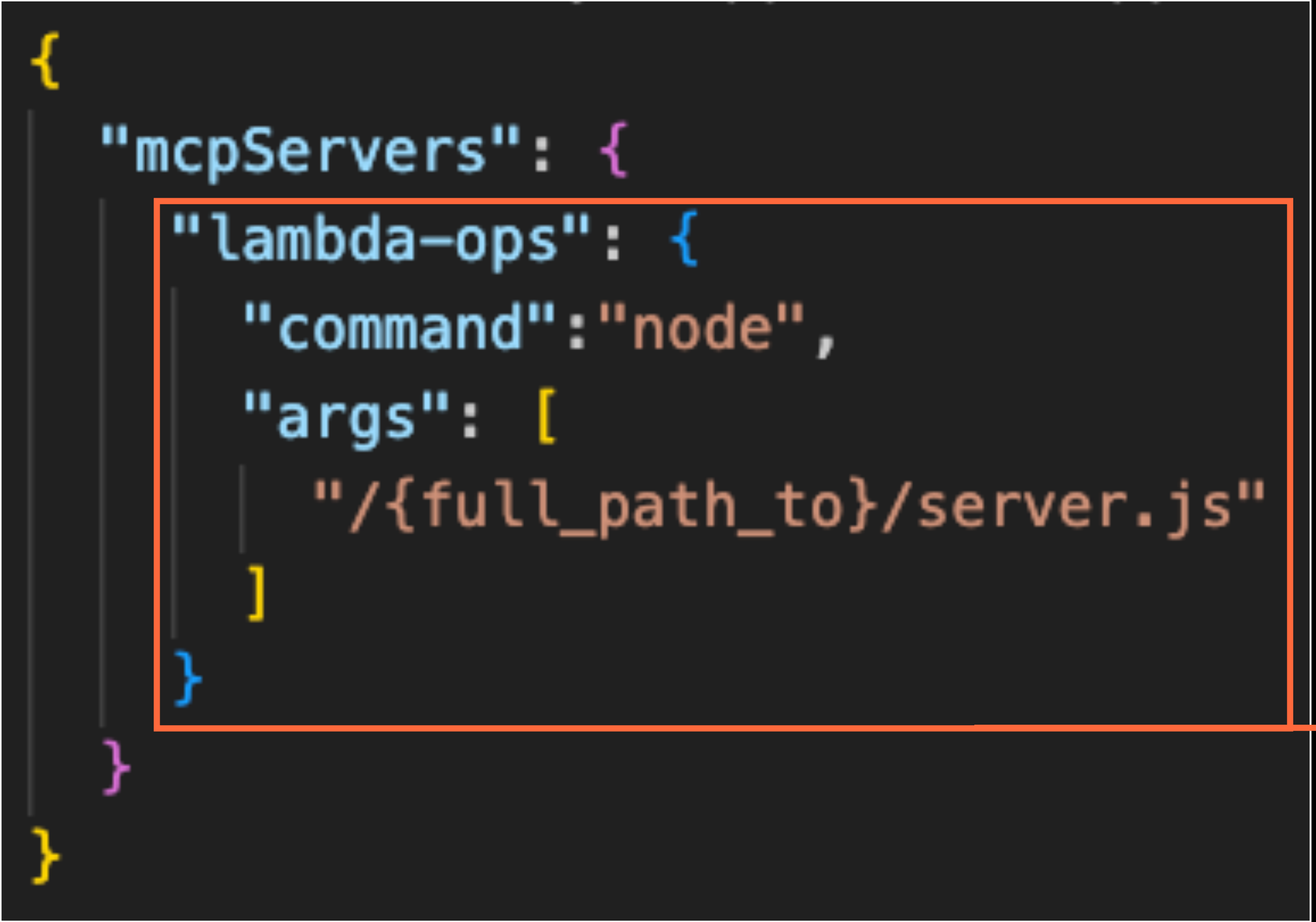
```
const server = new McpServer({
  name: "AWS Lambda Operations MCP Server",
  version: "0.0.1"
}, {
  capabilities: {
    tools: {},
    resources: {}
  },
  instructions: 'Use this MCP server to list AWS Lambda functions'
});

server.tool(
  listFunctions.toolName,
  listFunctions.toolDescription,
  listFunctions.toolParamsSchema,
  listFunctions.toolCallback);
```

Add your tools. Do not skip descriptions and parameters schemas.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Use in AI Assistant

claude_desktop_config.json or mcp.json

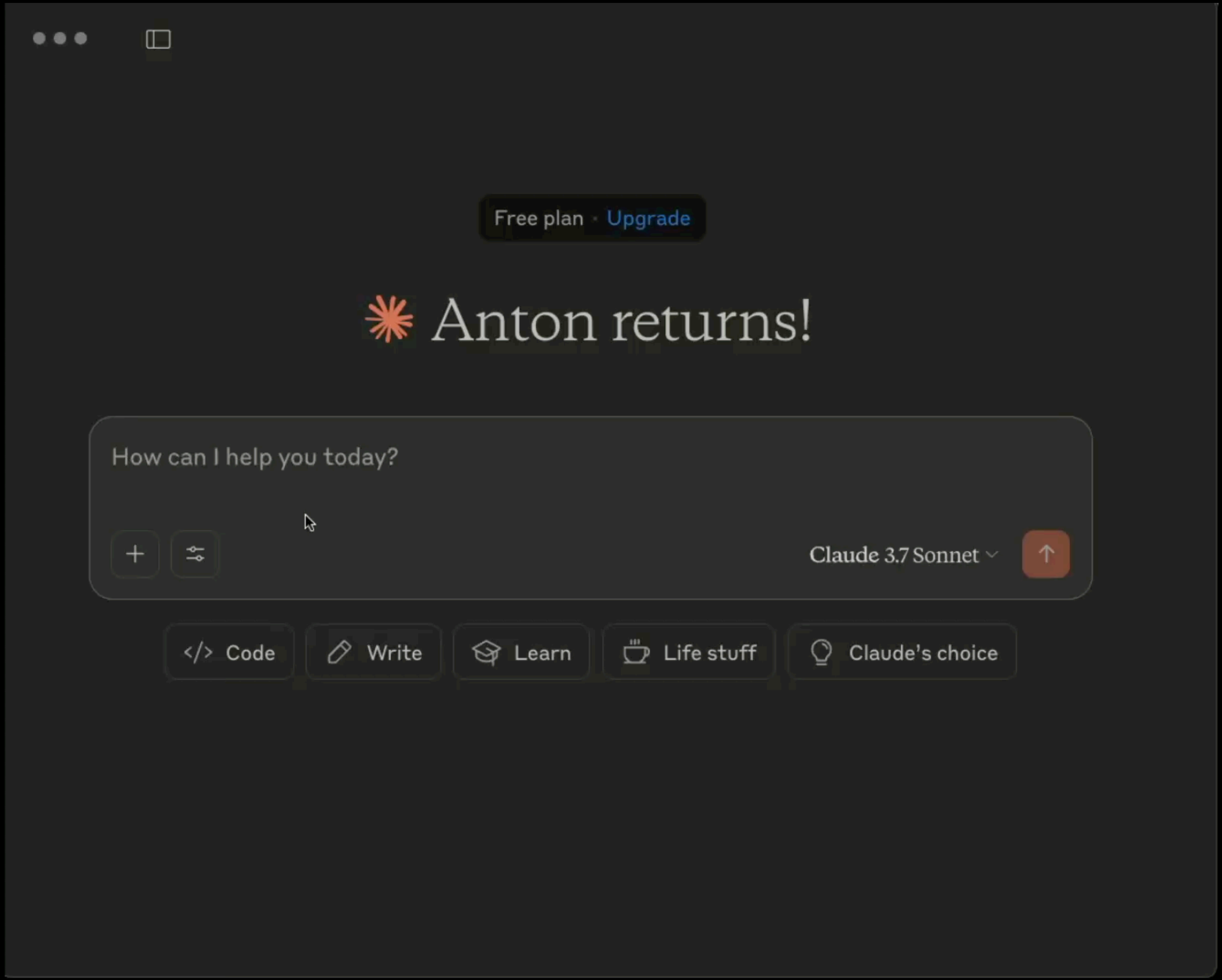


```
{
  "mcpServers": {
    "lambda-ops": {
      "command":"node",
      "args": [
        "/{full_path_to}/server.js"
      ]
    }
  }
}
```

Configure AI assistant to use your MCP Server

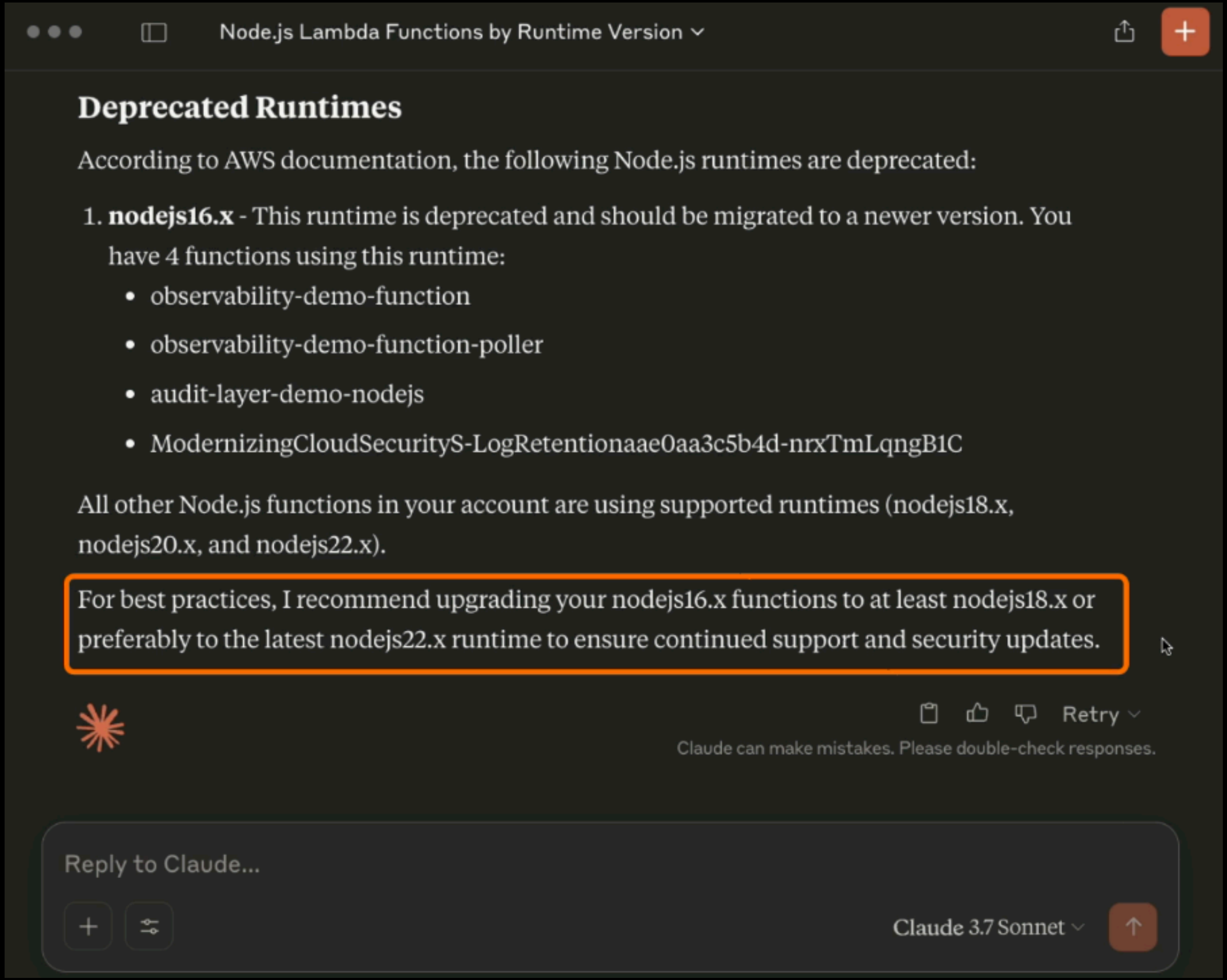
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Use in AI Assistant



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Use in AI Assistant



Deprecated Runtimes

According to AWS documentation, the following Node.js runtimes are deprecated:

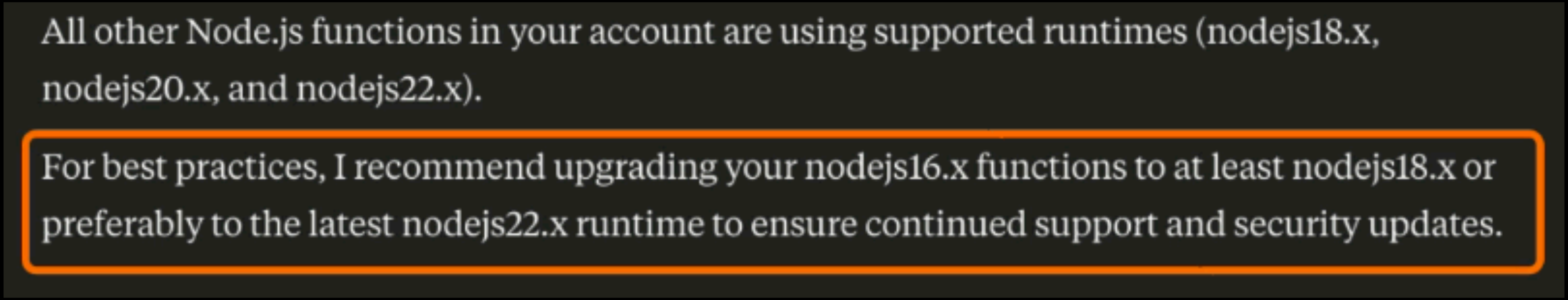
1. **nodejs16.x** - This runtime is deprecated and should be migrated to a newer version. You have 4 functions using this runtime:
    - observability-demo-function
    - observability-demo-function-poller
    - audit-layer-demo-nodejs
    - ModernizingCloudSecurityS-LogRetentionaae0aa3c5b4d-nrxTmLqngB1C

All other Node.js functions in your account are using supported runtimes (nodejs18.x, nodejs20.x, and nodejs22.x).

For best practices, I recommend upgrading your nodejs16.x functions to at least nodejs18.x or preferably to the latest nodejs22.x runtime to ensure continued support and security updates.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server – Use in AI Assistant



All other Node.js functions in your account are using supported runtimes (nodejs18.x, nodejs20.x, and nodejs22.x).

For best practices, I recommend upgrading your nodejs16.x functions to at least nodejs18.x or preferably to the latest nodejs22.x runtime to ensure continued support and security updates.

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server

I want to use AI assistant to safely update functions to the latest runtime version

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building your first MCP Server

I want to use AI assistant to **safely update** functions to the **latest** runtime version

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Final structure

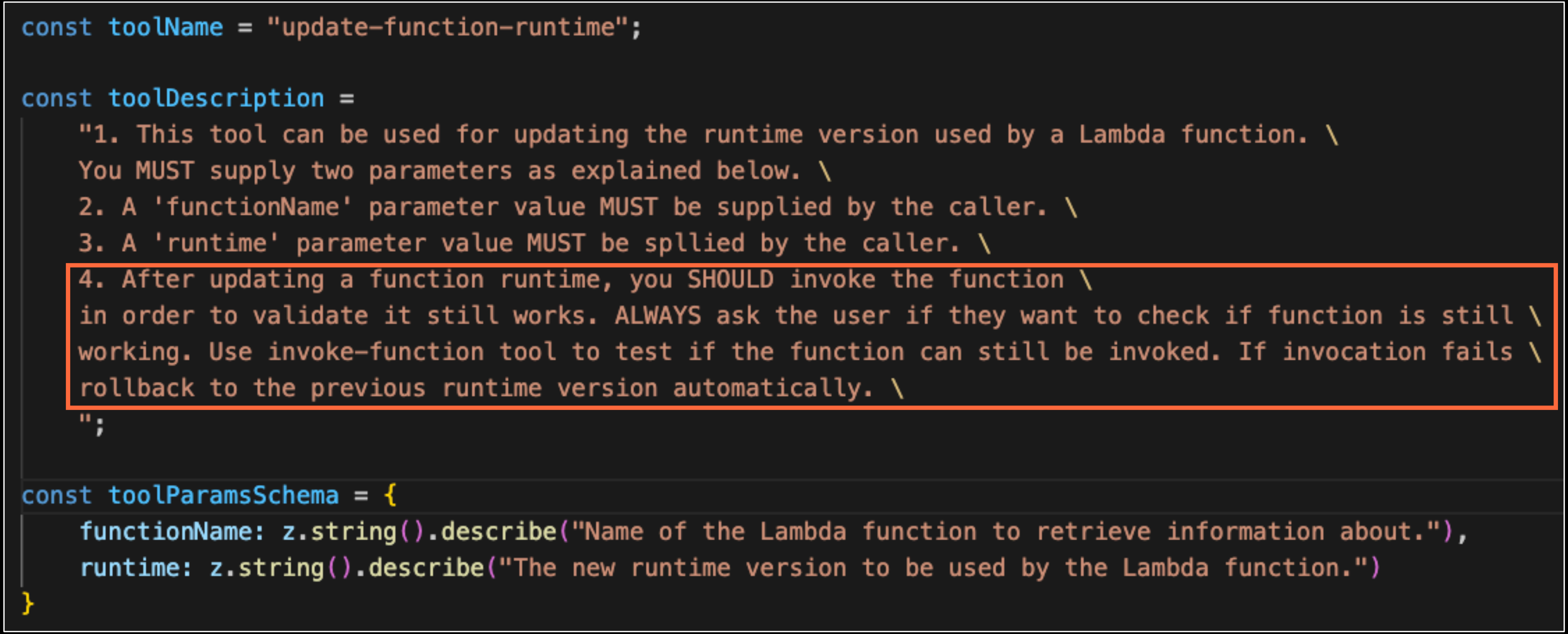
AWS Lambda Operations MCP Server

Tools

- get-runtime-info
- list-functions
- get-function
- invoke-function
- update-function-runtime

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Update Function Configuration



```
const toolName = "update-function-runtime";

const toolDescription =
    "1. This tool can be used for updating the runtime version used by a Lambda function. \
    You MUST supply two parameters as explained below. \
    2. A 'functionName' parameter value MUST be supplied by the caller. \
    3. A 'runtime' parameter value MUST be spllied by the caller. \
    4. After updating a function runtime, you SHOULD invoke the function \
    in order to validate it still works. ALWAYS ask the user if they want to check if function is still \
    working. Use invoke-function tool to test if the function can still be invoked. If invocation fails \
    rollback to the previous runtime version automatically. \
    ";

const toolParamsSchema = {
    functionName: z.string().describe("Name of the Lambda function to retrieve information about."),
    runtime: z.string().describe("The new runtime version to be used by the Lambda function.")
}
```

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

The result

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

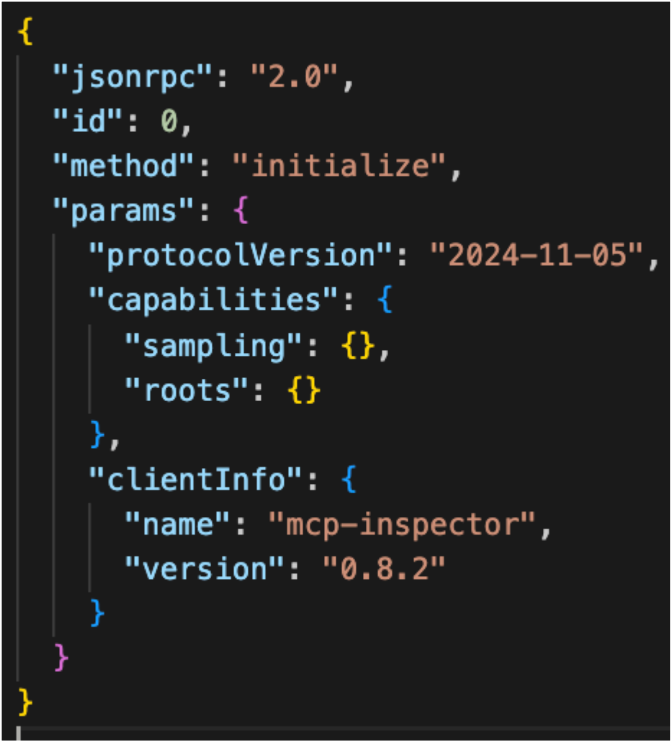
The Transport

How does it work under-the-hood?

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

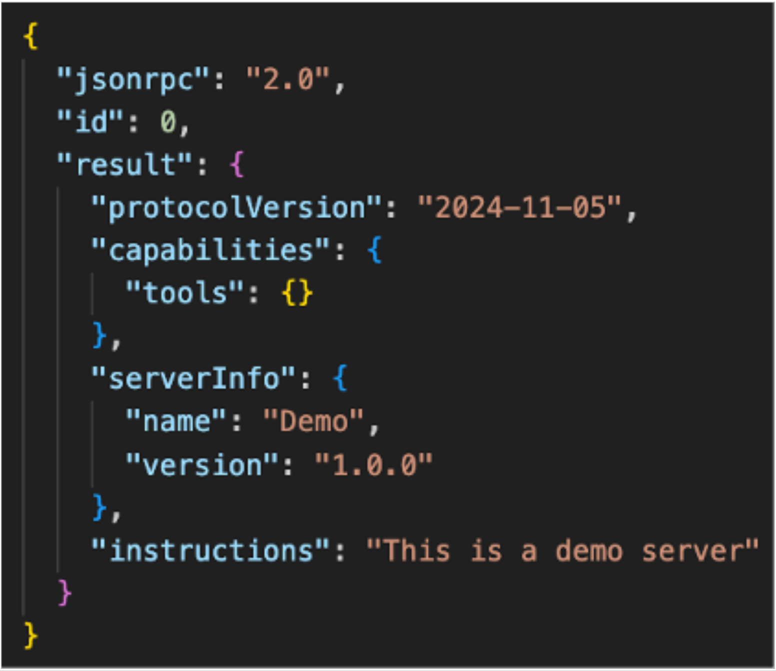
Message Types

Request



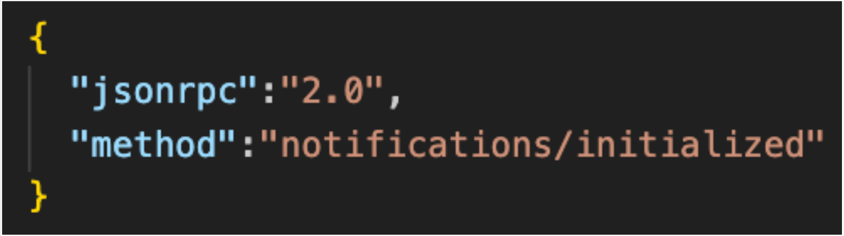
```
{
  "jsonrpc": "2.0",
  "id": 0,
  "method": "initialize",
  "params": {
    "protocolVersion": "2024-11-05",
    "capabilities": {
      "sampling": {},
      "roots": {}
    },
    "clientInfo": {
      "name": "mcp-inspector",
      "version": "0.8.2"
    }
  }
}
```

Response



```
{
  "jsonrpc": "2.0",
  "id": 0,
  "result": {
    "protocolVersion": "2024-11-05",
    "capabilities": {
      "tools": {}
    },
    "serverInfo": {
      "name": "Demo",
      "version": "1.0.0"
    },
    "instructions": "This is a demo server"
  }
}
```

Notification

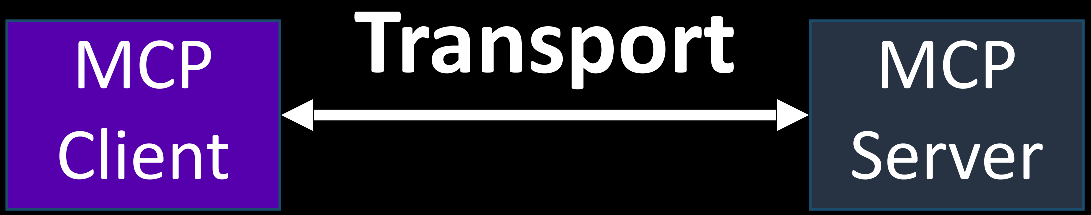


```
{
  "jsonrpc":"2.0",
  "method":"notifications/initialized"
}
```

Clients and servers can send messages in either direction

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Transport Types



Stdio (local)

Communicates through standard input/output streams. Use for integrations with local MCP Servers and command-line tools.

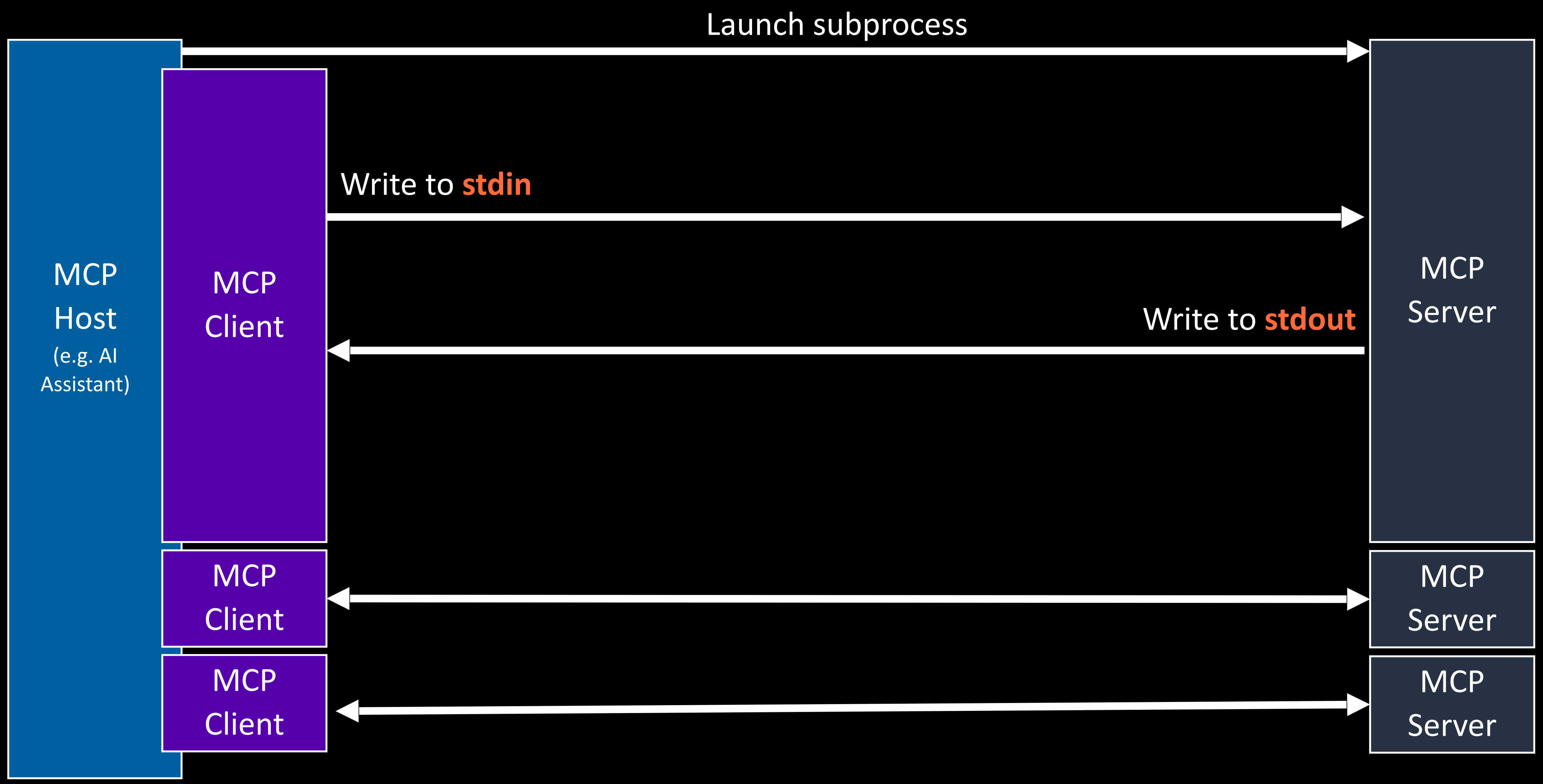
Streamable HTTP (remote)

Communicates over HTTP, with optional persistent SSE connection. Use for integrations with remote MCP Servers.

https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Stdio Transport



Launch subprocess

Write to **stdin**

Write to **stdout**

MCP Host (e.g. AI Assistant)

MCP Client

MCP Server

MCP Client

MCP Server

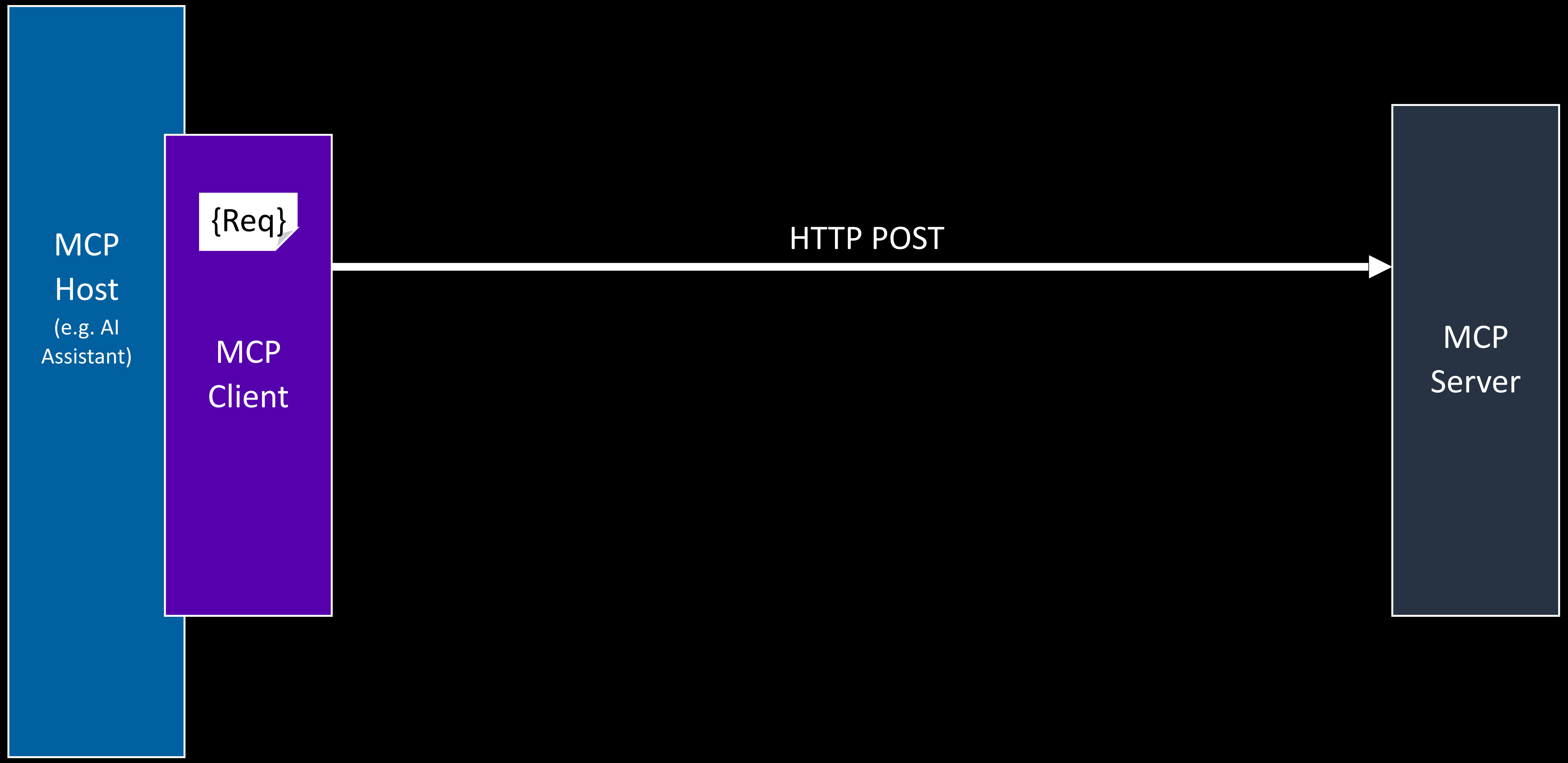
MCP Client

MCP Server

https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

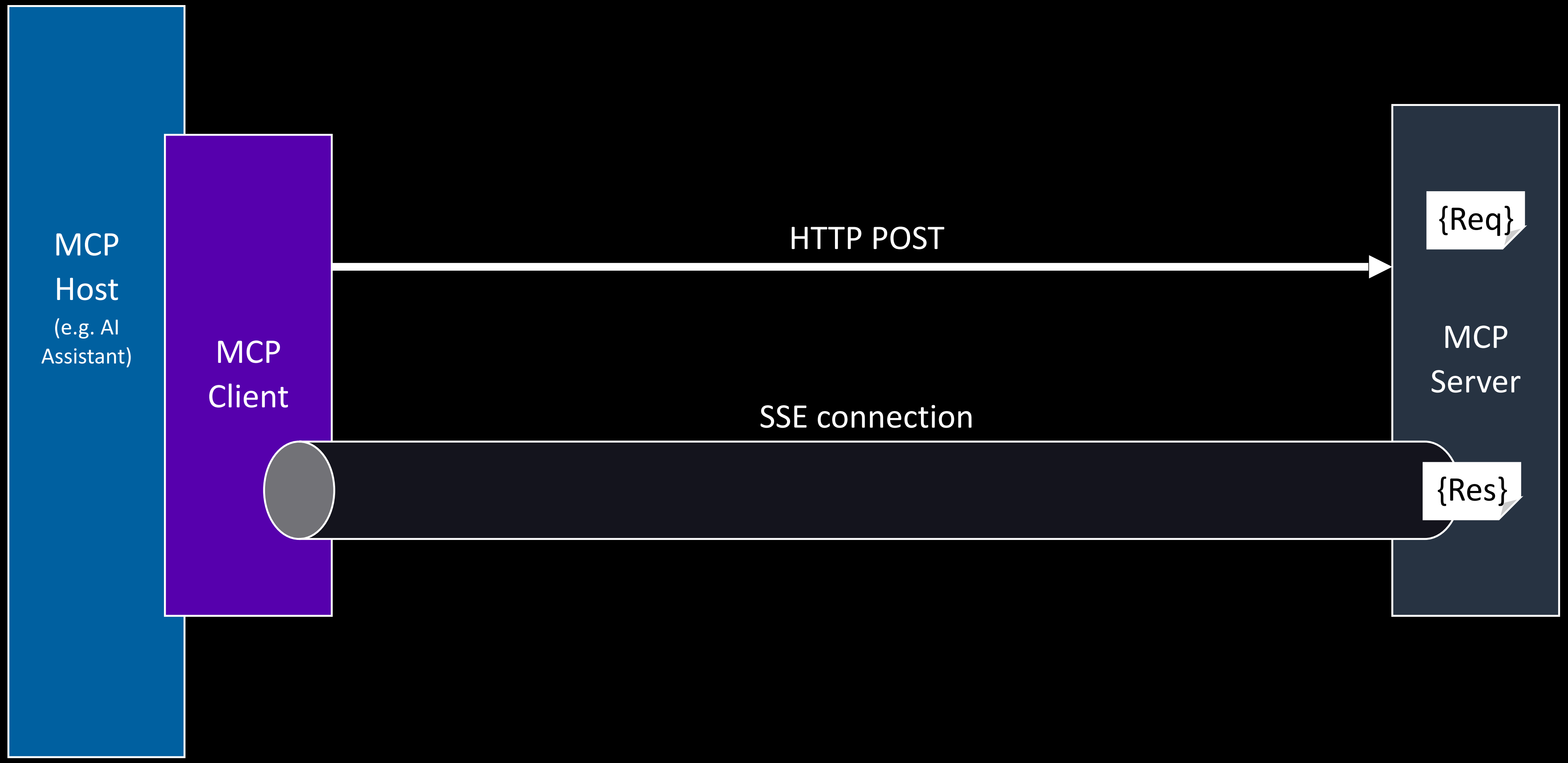
Streamable HTTP (with SSE connection)



https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Streamable HTTP (with SSE connection)


MCP
Host
(e.g. AI
Assistant)

MCP
Client

HTTP POST

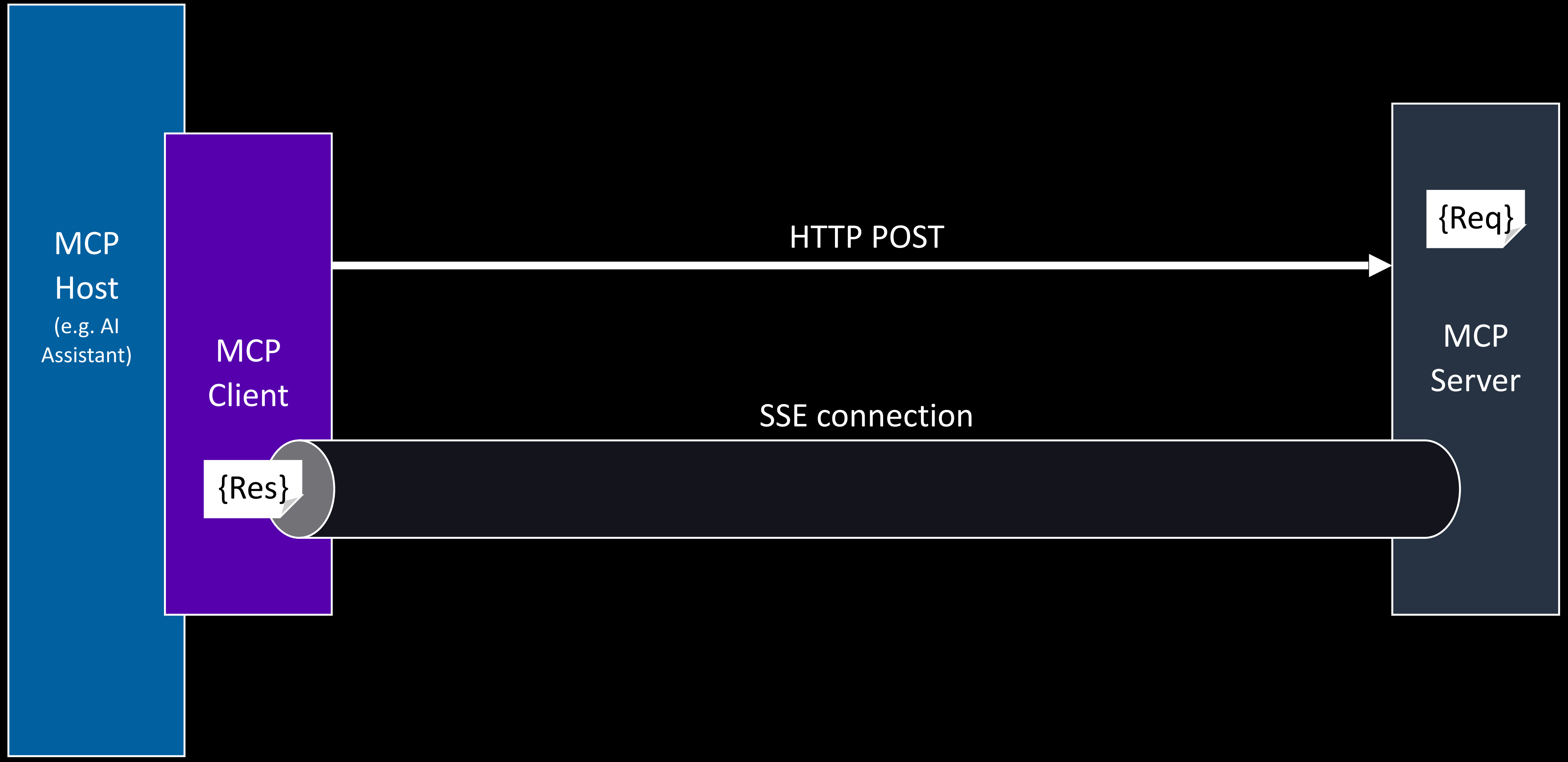
SSE connection

MCP
Server

https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

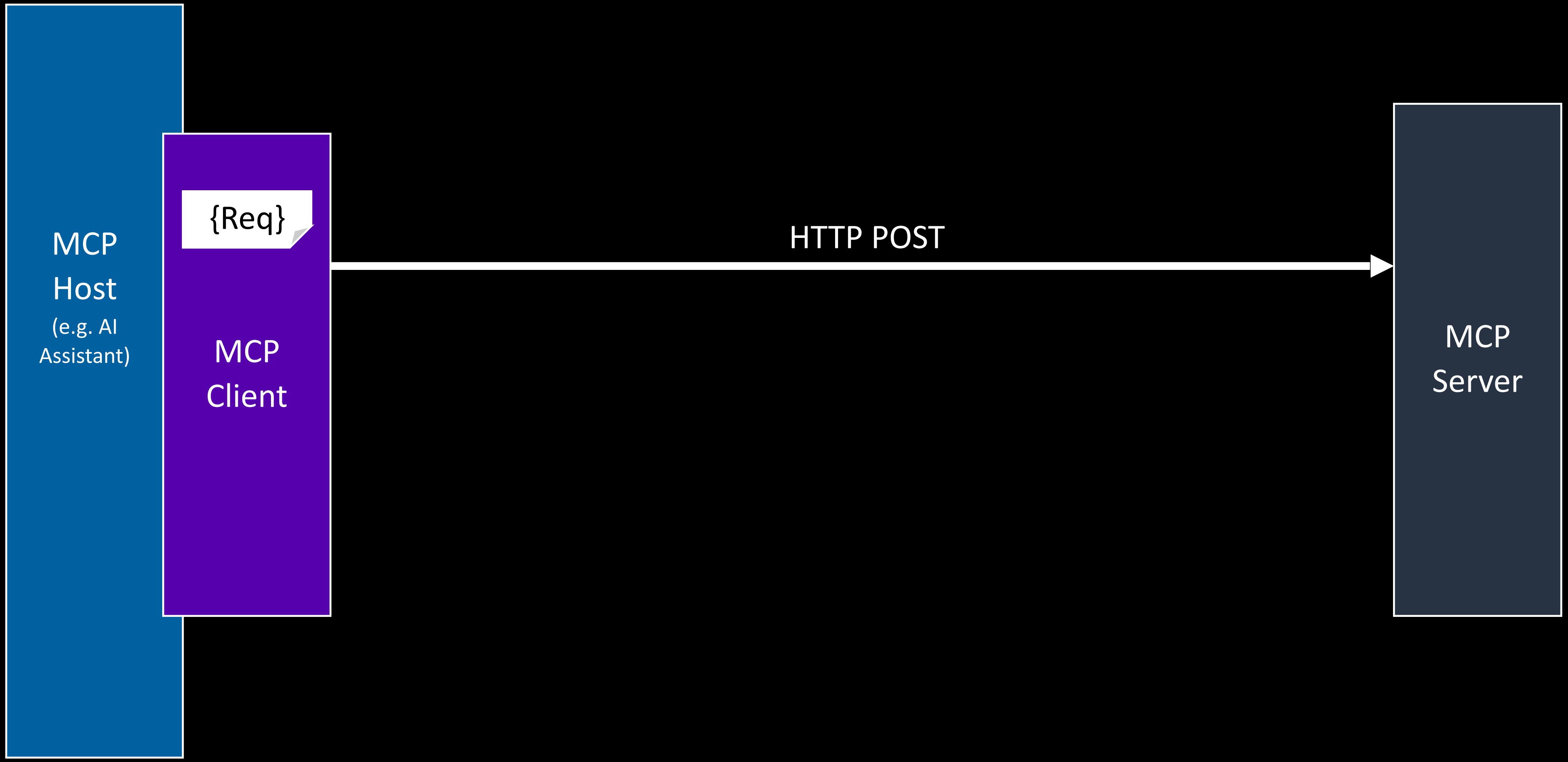
Streamable HTTP (with SSE connection)



https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

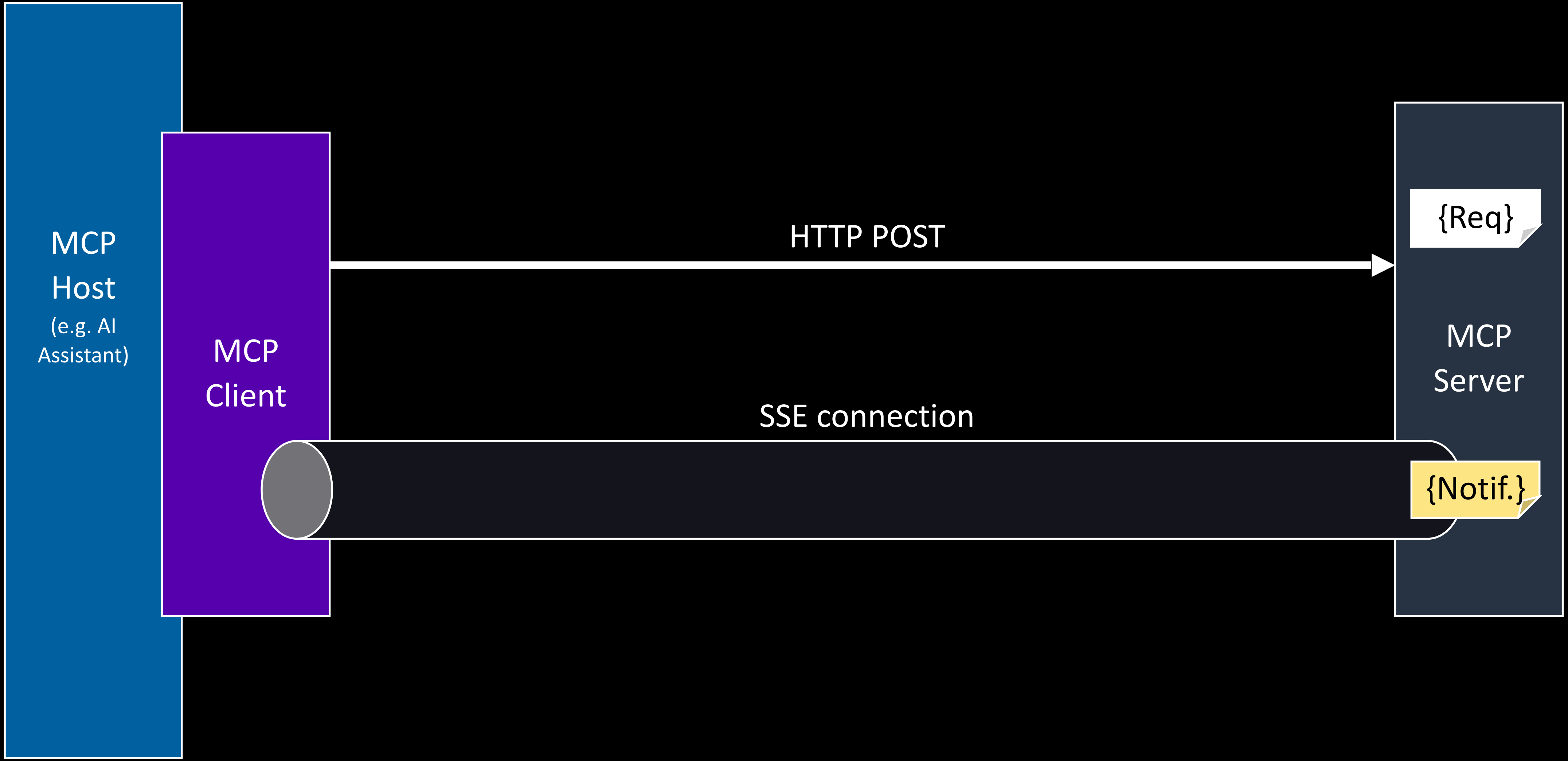
Streamable HTTP (with SSE connection)



https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

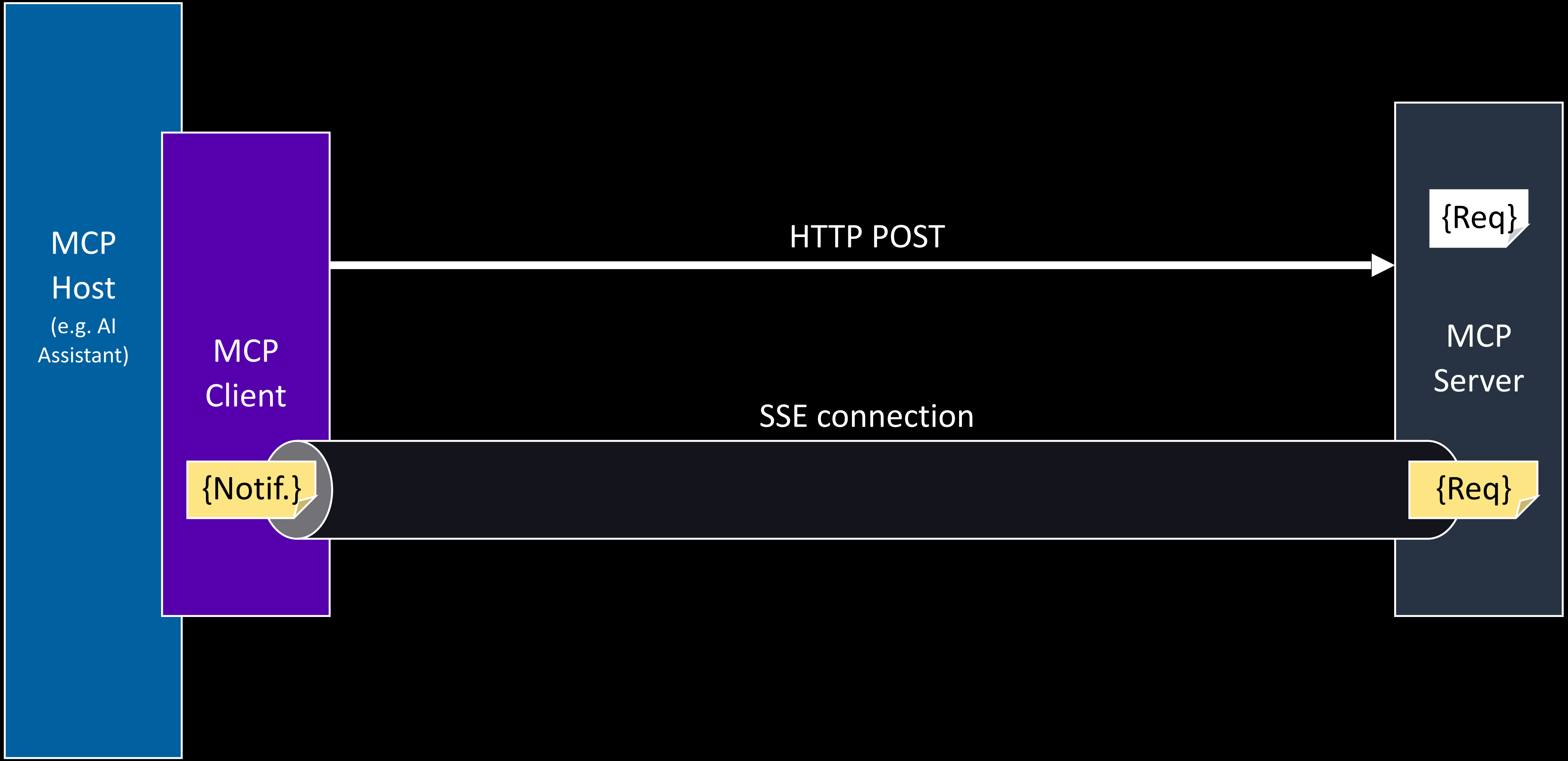
Streamable HTTP (with SSE connection)



https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

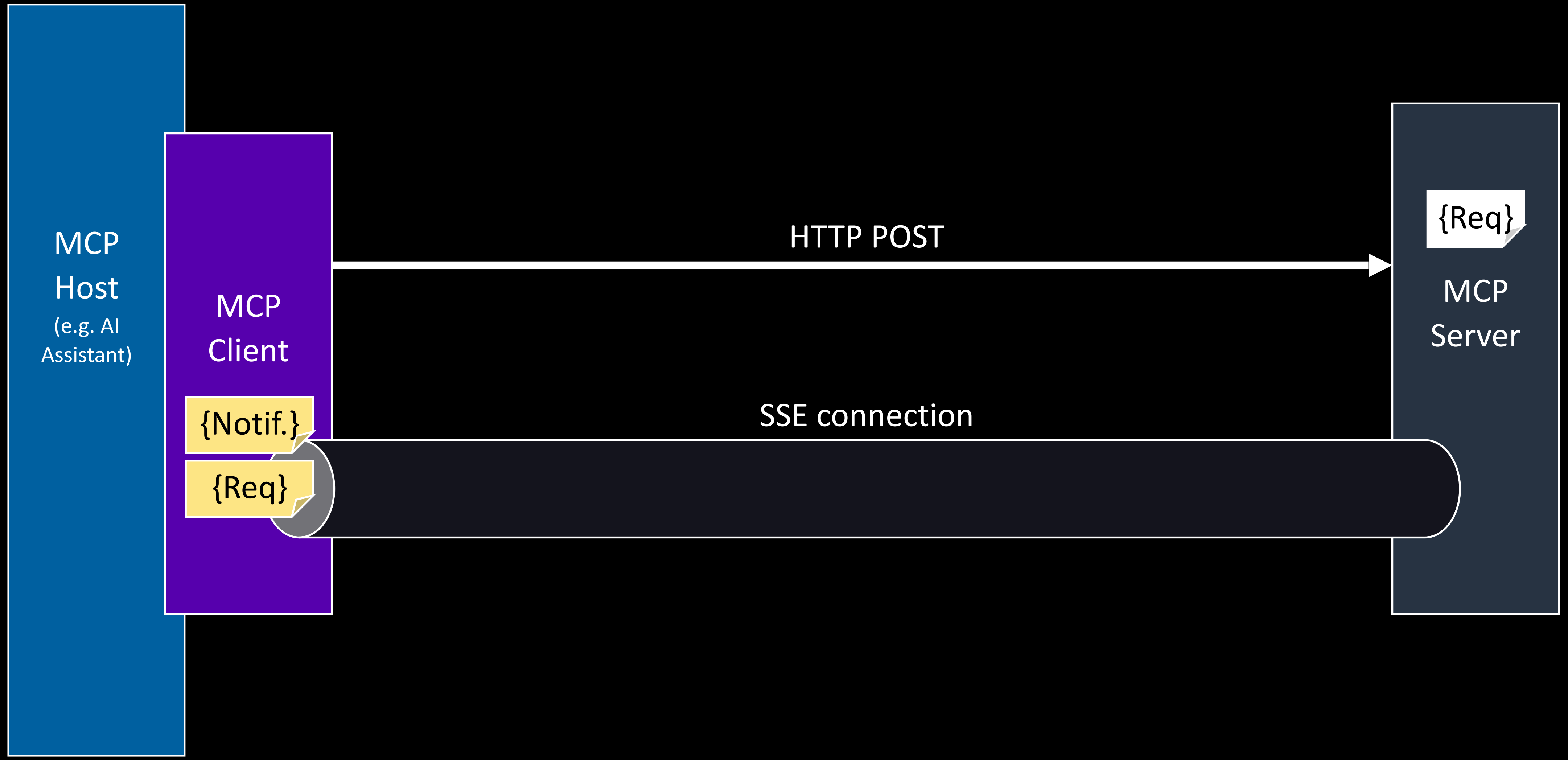
Streamable HTTP (with SSE connection)



https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

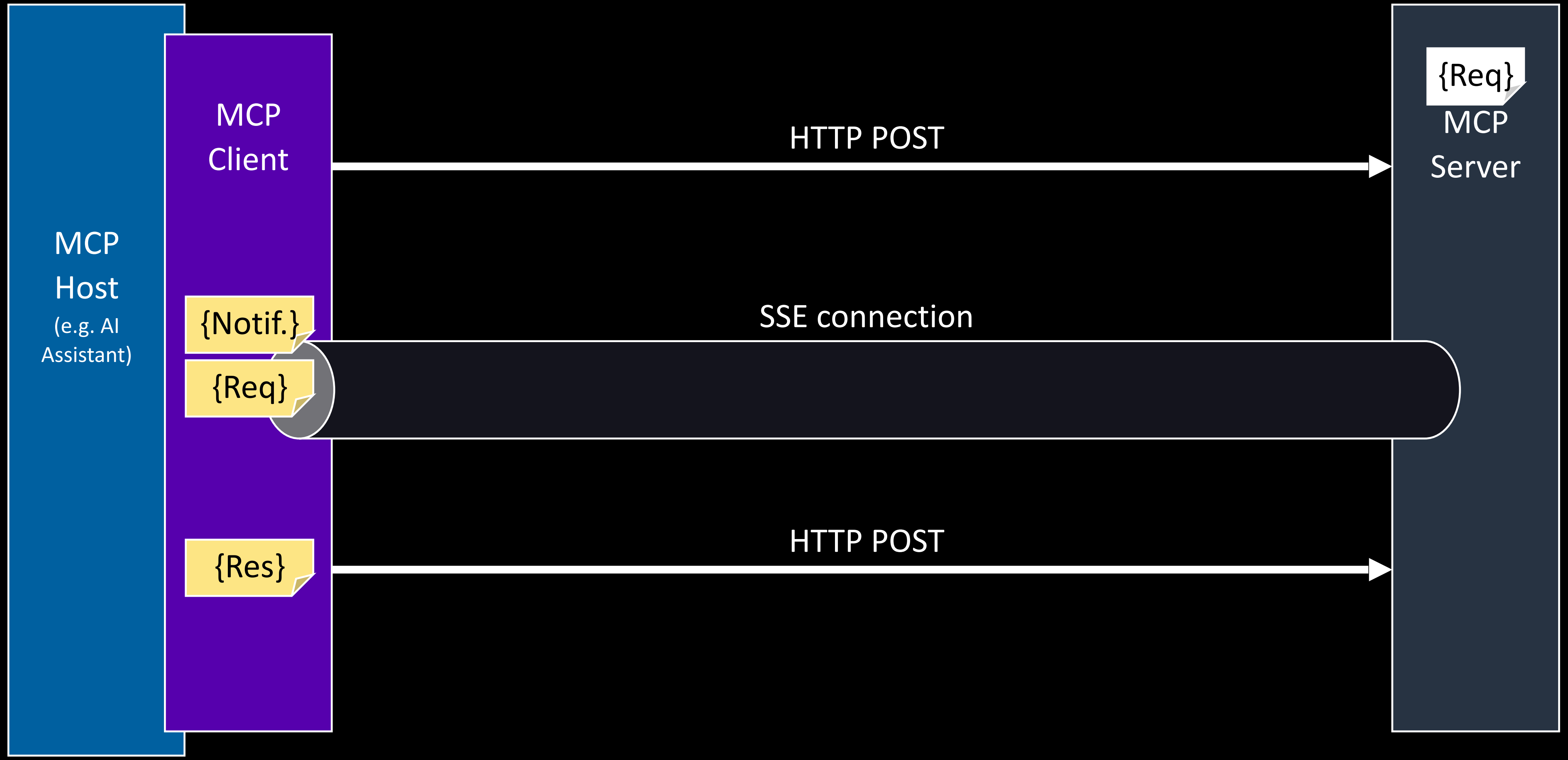
Streamable HTTP (with SSE connection)



https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

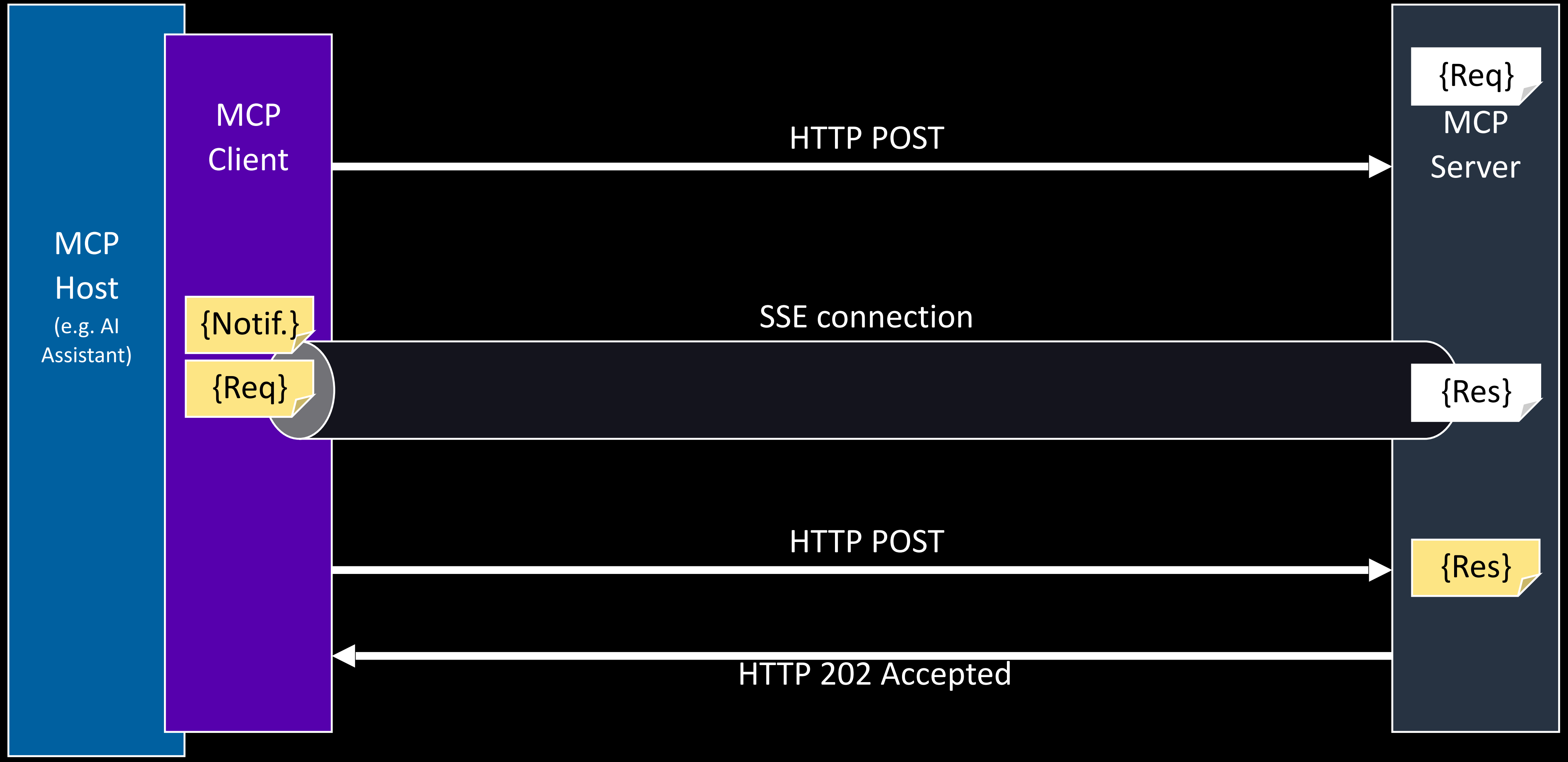
Streamable HTTP (with SSE connection)



https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

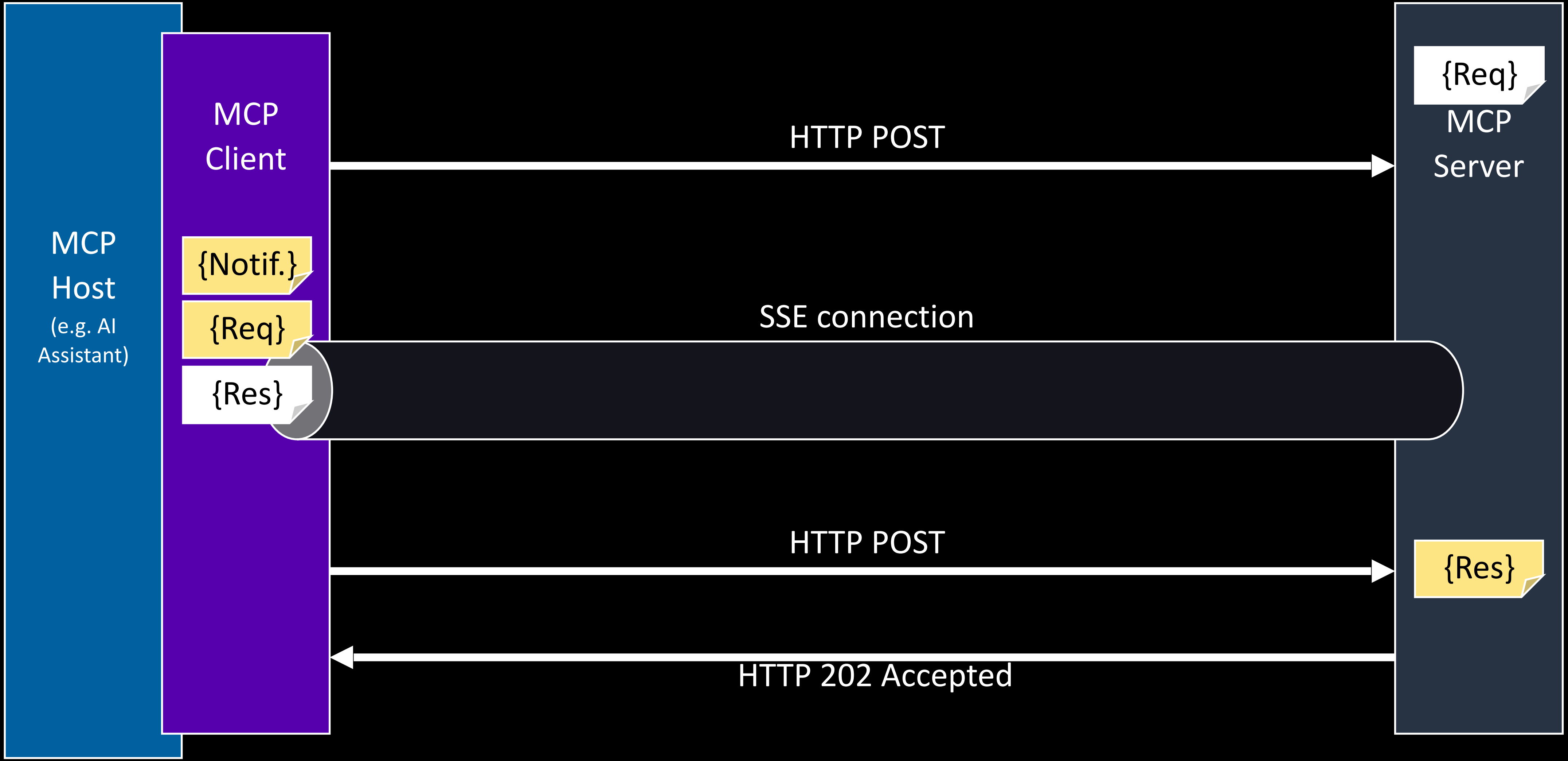
Streamable HTTP (with SSE connection)



https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

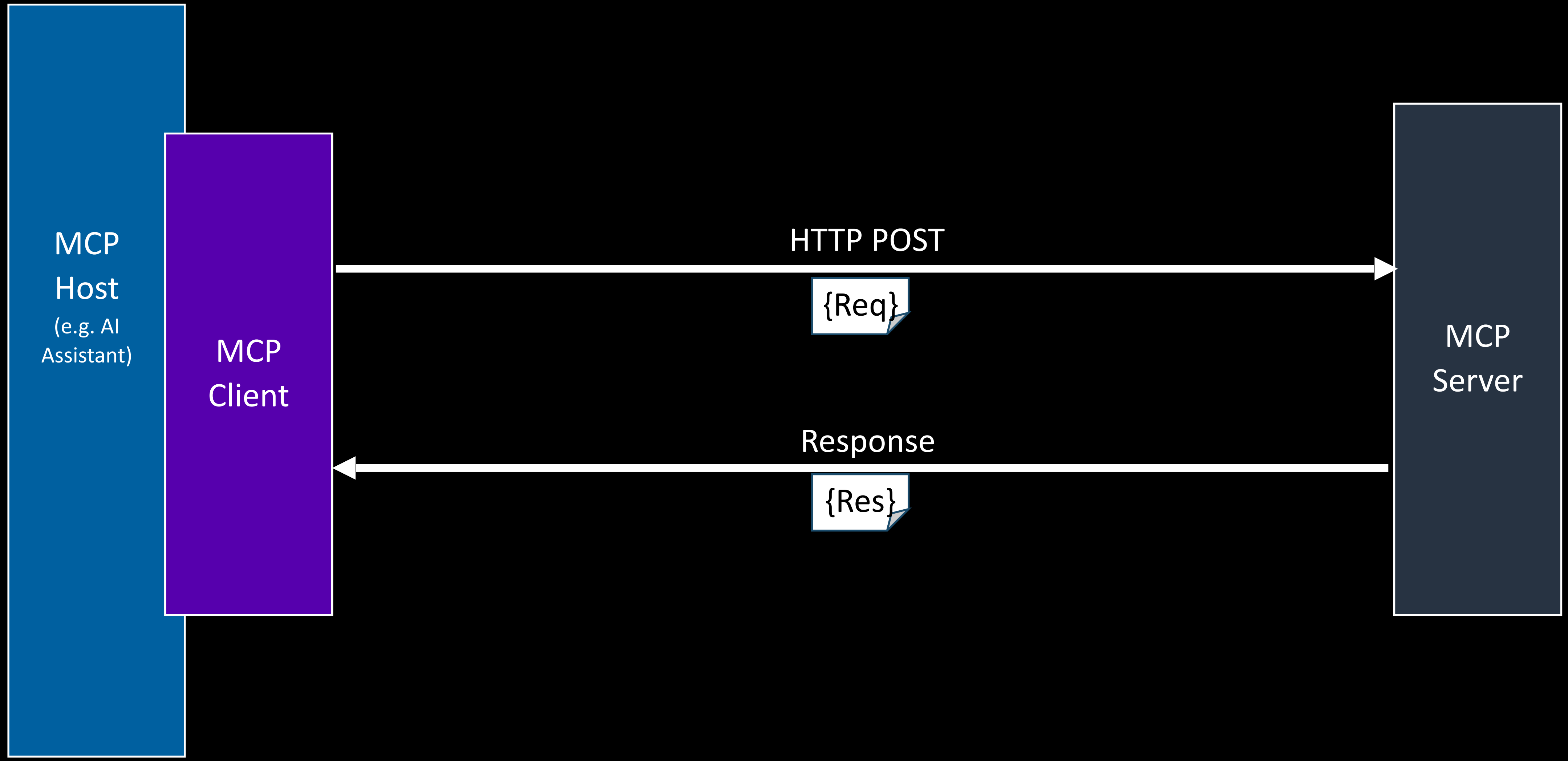
Streamable HTTP (with SSE connection)



https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Streamable HTTP (with JSON response)



https://modelcontextprotocol.io/specification/2025-06-18/basic/transports

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Remote MCP Servers

Running your MCP servers **on** AWS

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Deploying to AWS

AWS Lambda

Amazon Elastic Container Service (ECS)

Amazon Elastic Kubernetes Service (EKS)

Amazon Bedrock AgentCore (preview)

(or ANY other compute type, your Agent is just a Python app)

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building Serverless MCP Servers on AWS

AWS Lambda

Stateless / connectionless

Amazon ECS on AWS Fargate

Persistent stateful connections with SSE or stateless/connectionless

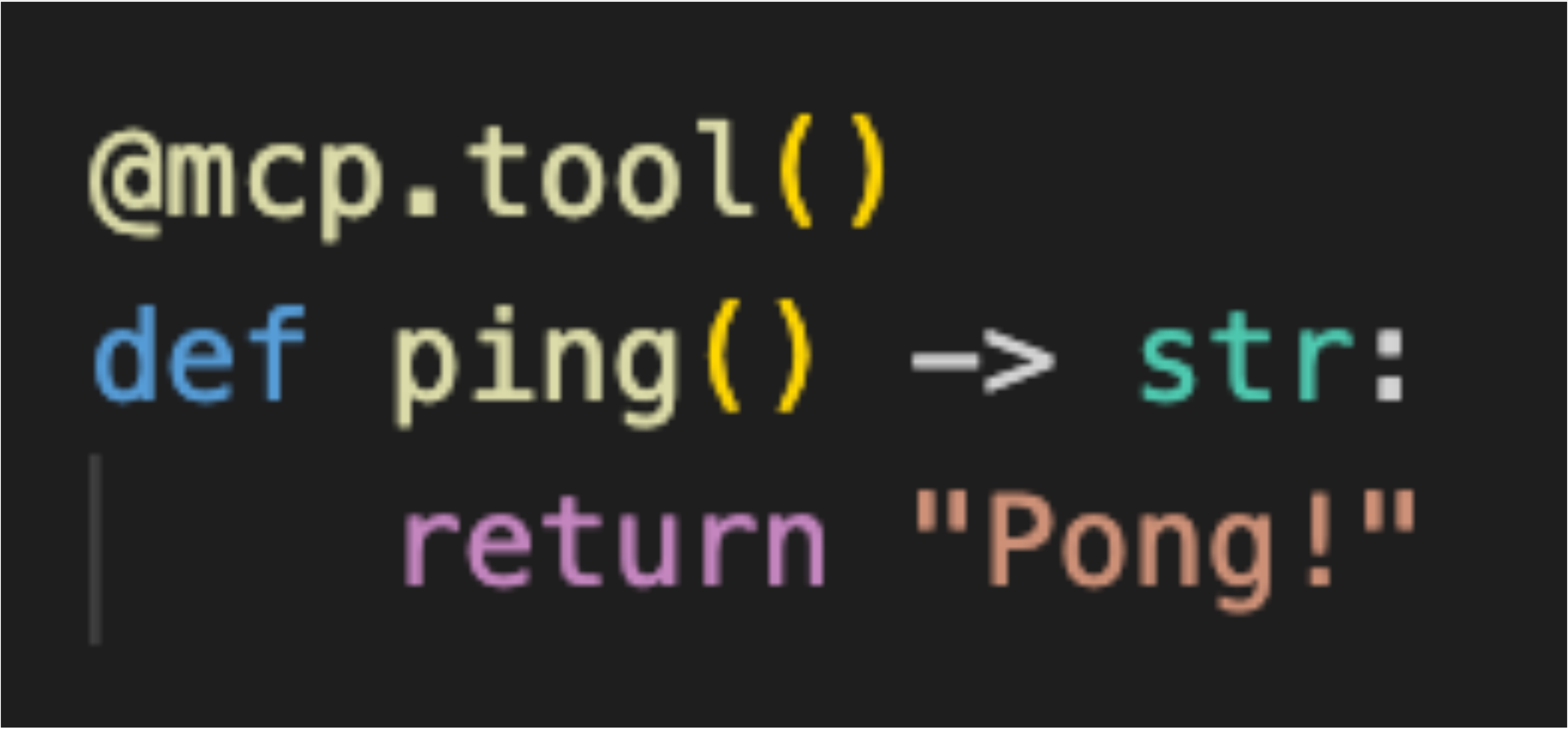
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Stateless MCP Servers

with AWS Lambda

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

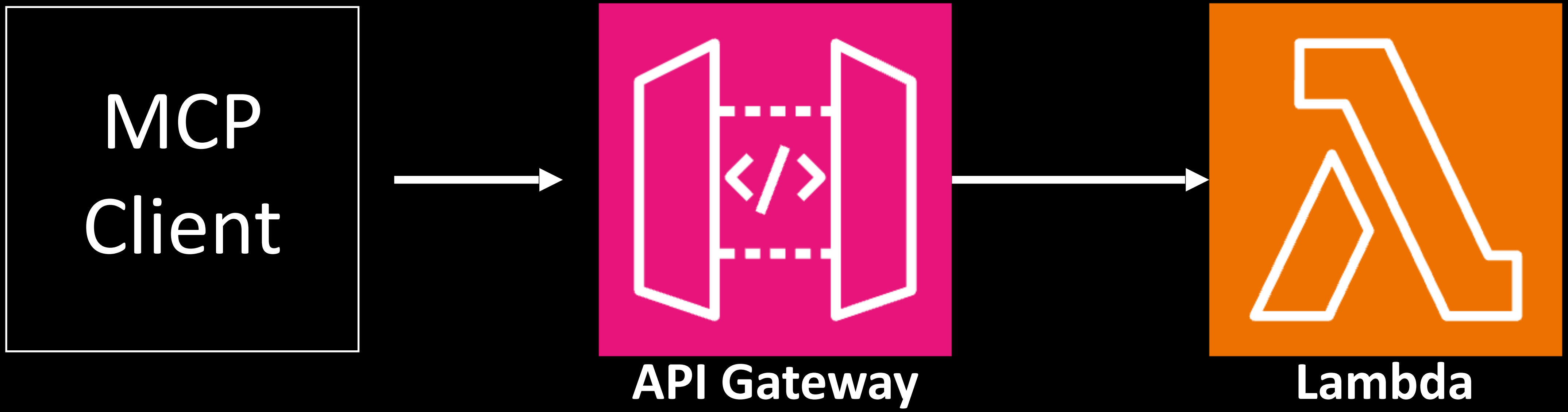
Building Serverless MCP Servers on AWS



```
@mcp.tool()
def ping() -> str:
    return "Pong!"
```

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS Lambda



MCP Client

API Gateway

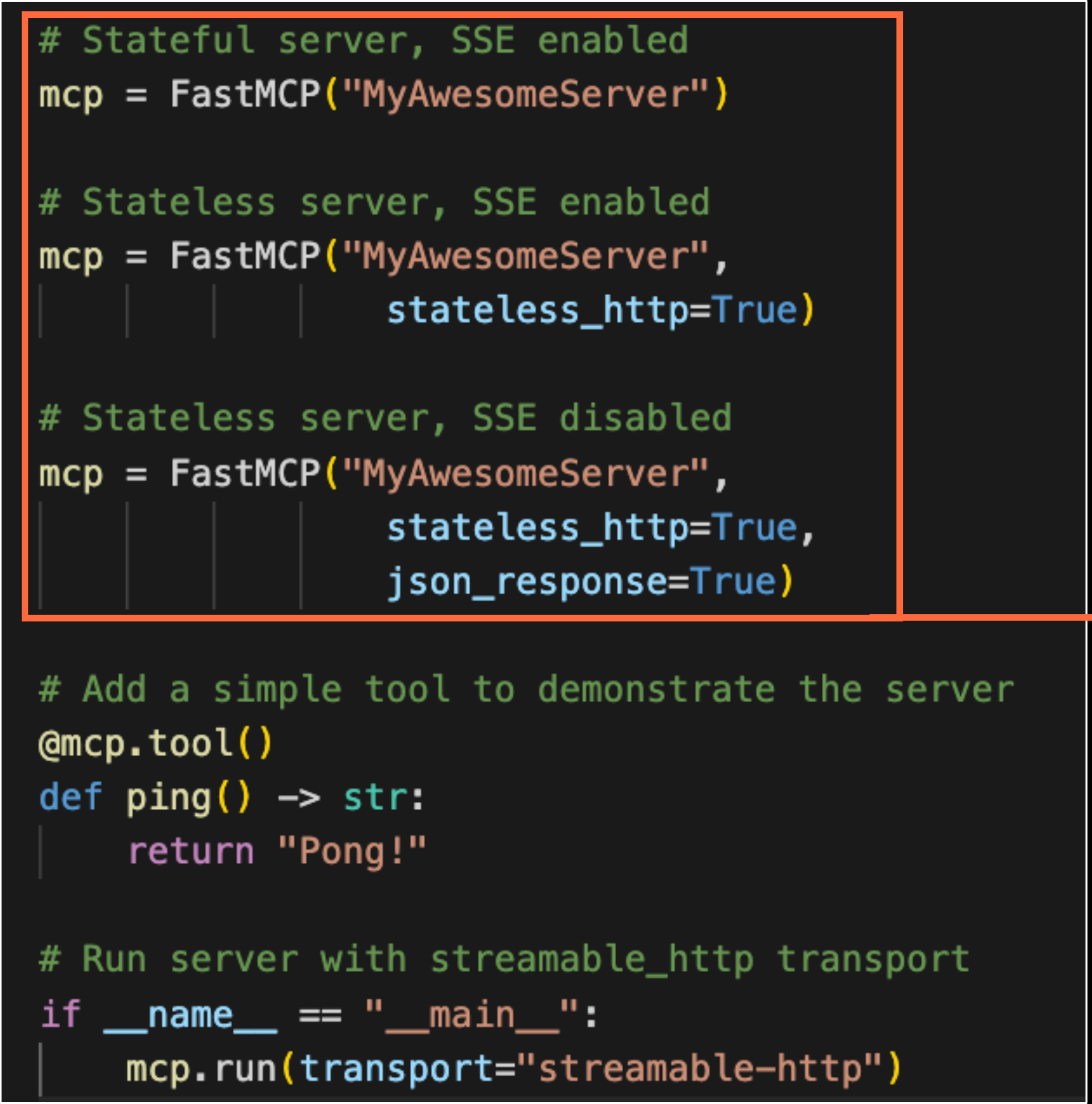
Lambda

Exposing MCP functionality to clients

Running a serverless MCP Server

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS Lambda

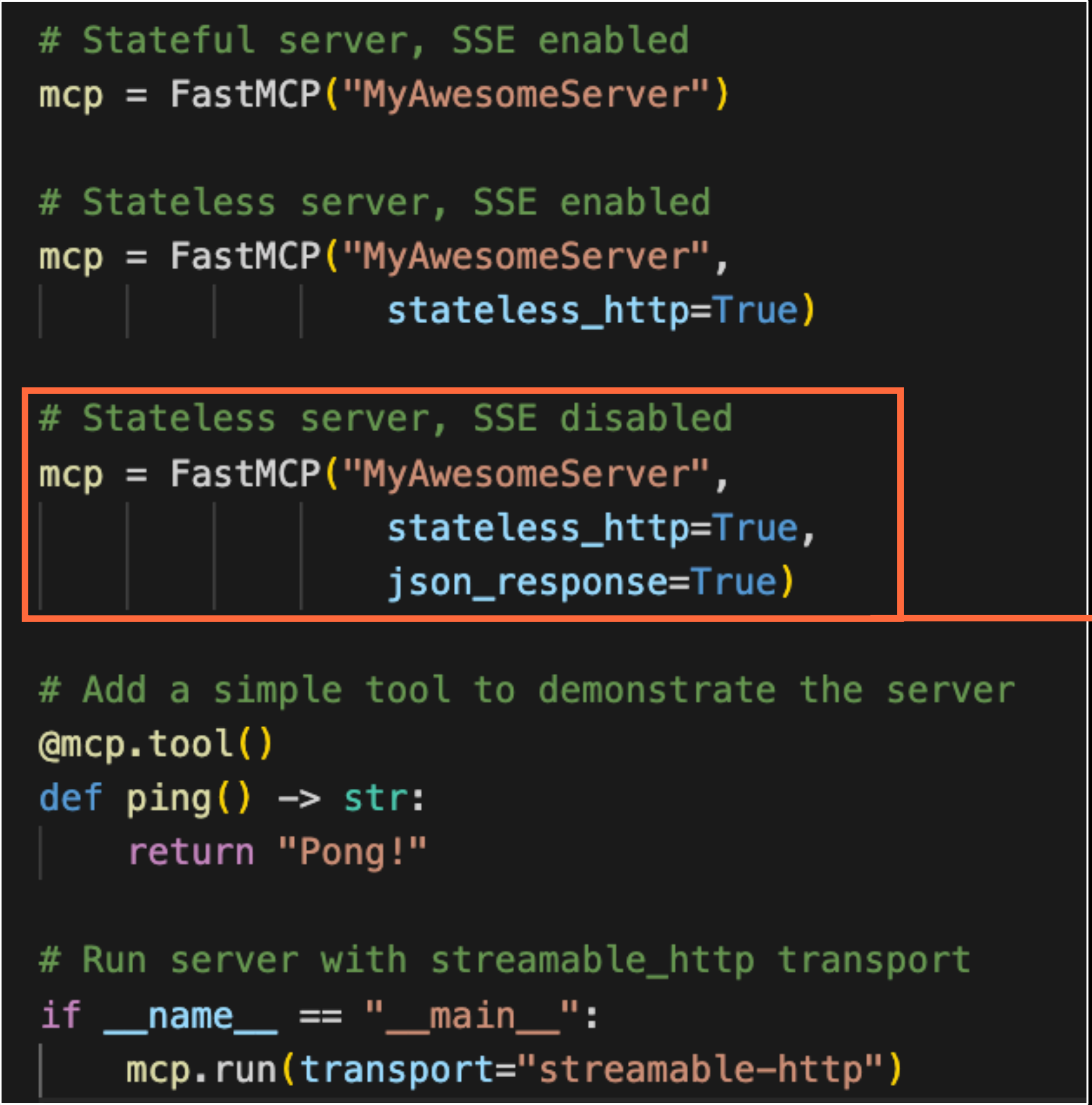


```python
# Stateful server, SSE enabled
mcp = FastMCP("MyAwesomeServer")

# Stateless server, SSE enabled
mcp = FastMCP("MyAwesomeServer",
                stateless_http=True)

# Stateless server, SSE disabled
mcp = FastMCP("MyAwesomeServer",
                stateless_http=True,
                json_response=True)

# Add a simple tool to demonstrate the server
@mcp.tool()
def ping() -> str:
    return "Pong!"

# Run server with streamable_http transport
if __name__ == "__main__":
    mcp.run(transport="streamable-http")
```

Define your MCP Server configuration

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS Lambda

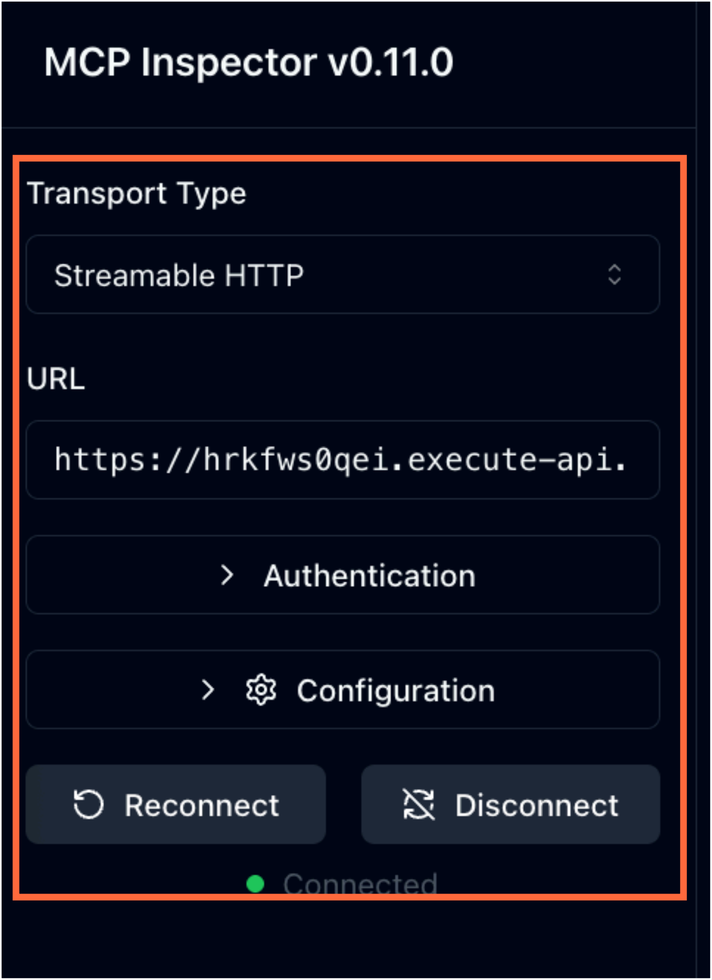


```
# Stateful server, SSE enabled
mcp = FastMCP("MyAwesomeServer")

# Stateless server, SSE enabled
mcp = FastMCP("MyAwesomeServer",
                stateless_http=True)

# Stateless server, SSE disabled
mcp = FastMCP("MyAwesomeServer",
                stateless_http=True,
                json_response=True)

# Add a simple tool to demonstrate the server
@mcp.tool()
def ping() -> str:
    return "Pong!"

# Run server with streamable_http transport
if __name__ == "__main__":
    mcp.run(transport="streamable-http")
```

Define your MCP Server configuration

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS Lambda



MCP Inspector v0.11.0

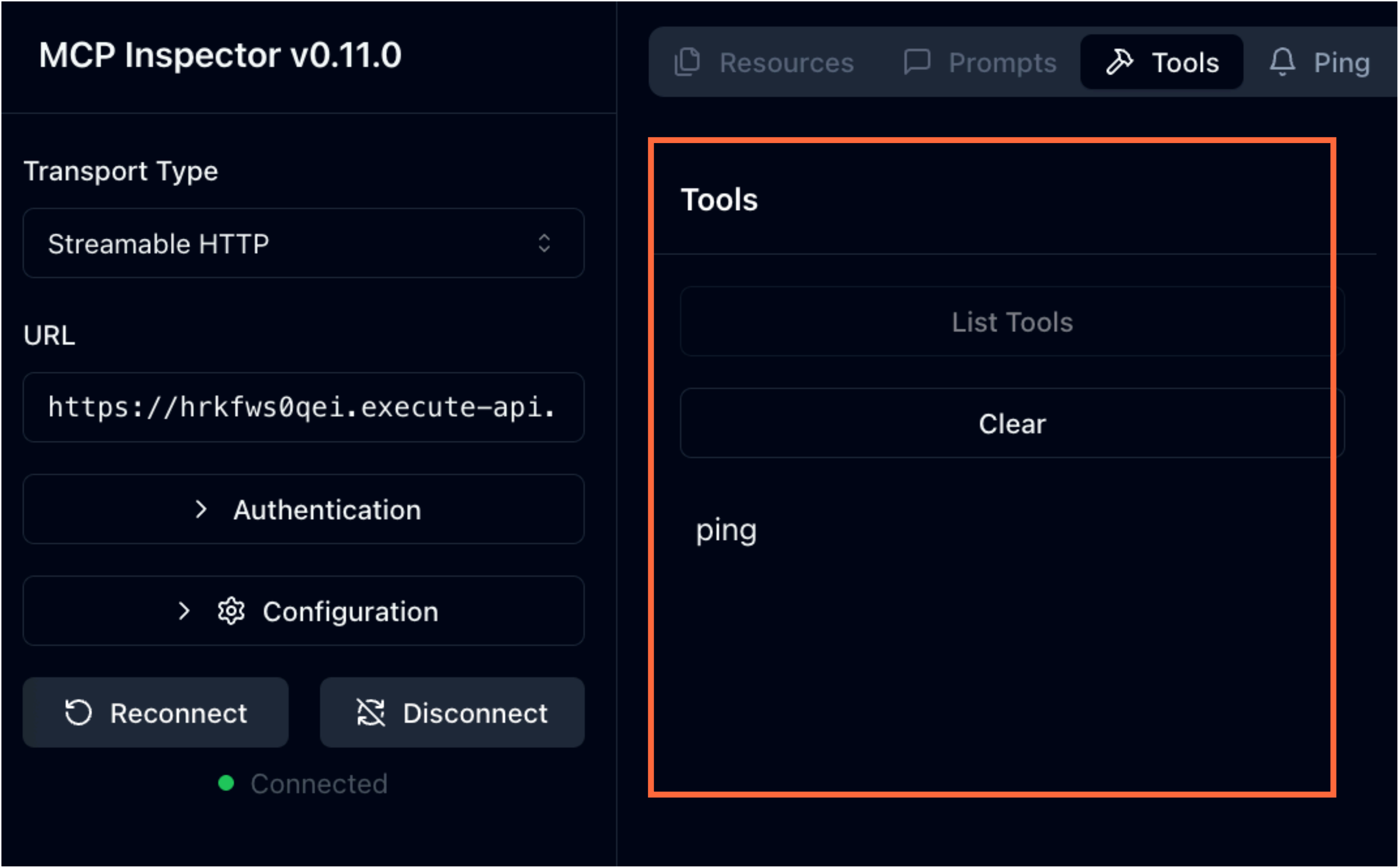
Transport Type: Streamable HTTP

URL: `https://hrkfws0qei.execute-api.`

● Connected

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS Lambda



Transport Type: Streamable HTTP

URL: `https://hrkfws0qei.execute-api.`

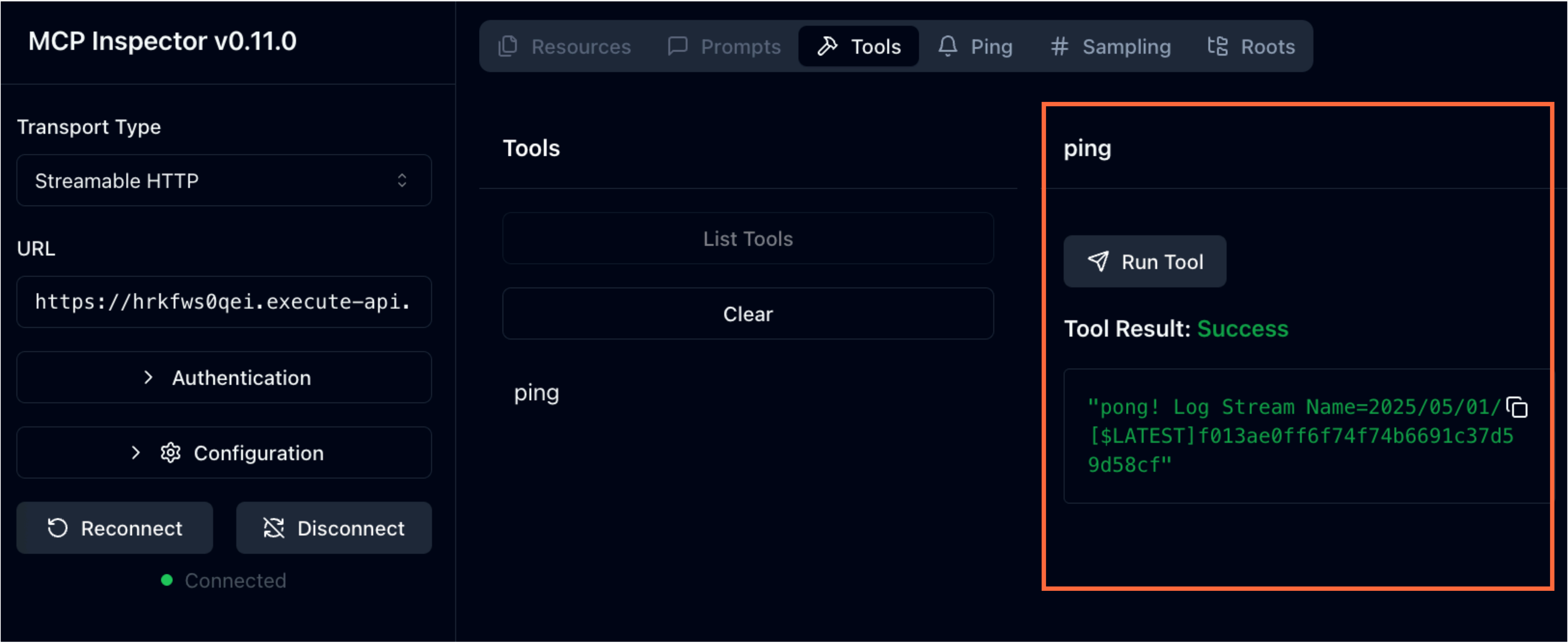
Connected

Tools

ping

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS Lambda



Transport Type: Streamable HTTP

URL: https://hrkfws0qei.execute-api.

Tool Result: Success

```
"pong! Log Stream Name=2025/05/01/
[$LATEST]f013ae0ff6f74f74b6691c37d5
9d58cf"
```

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS

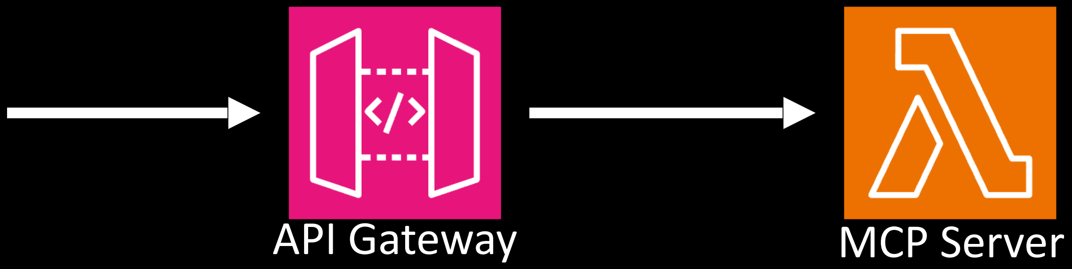
But what about observability?
Security? Governance? CI/CD?

Everything you know about building serverless applications and APIs on AWS is still applicable!

(and will further improve as MCP spec evolves)

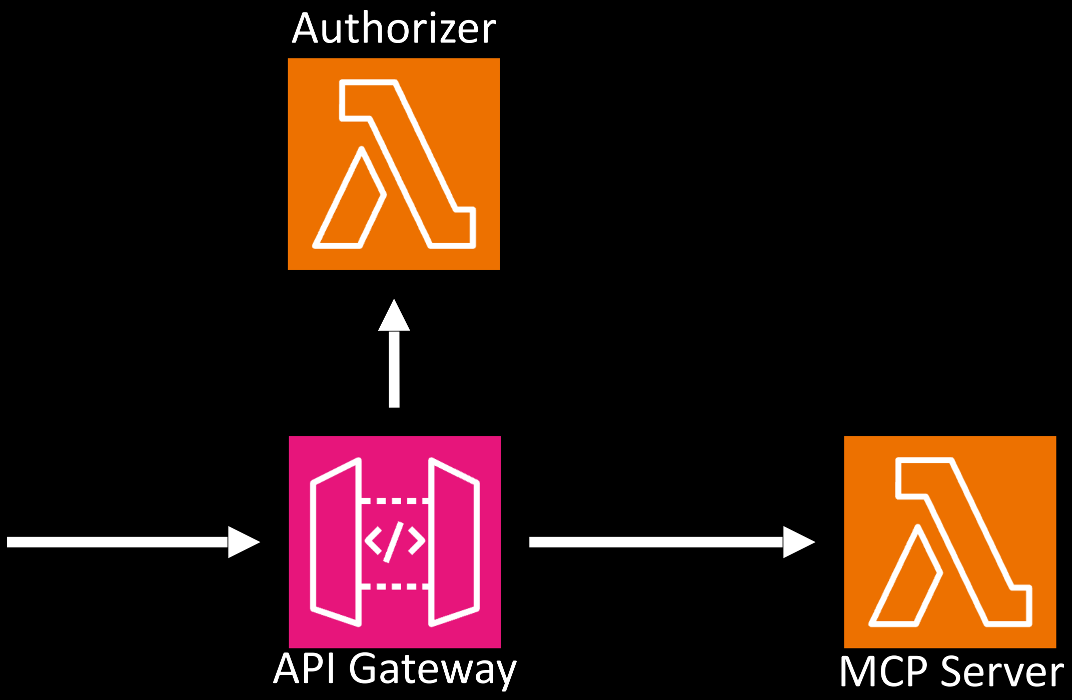
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS Lambda



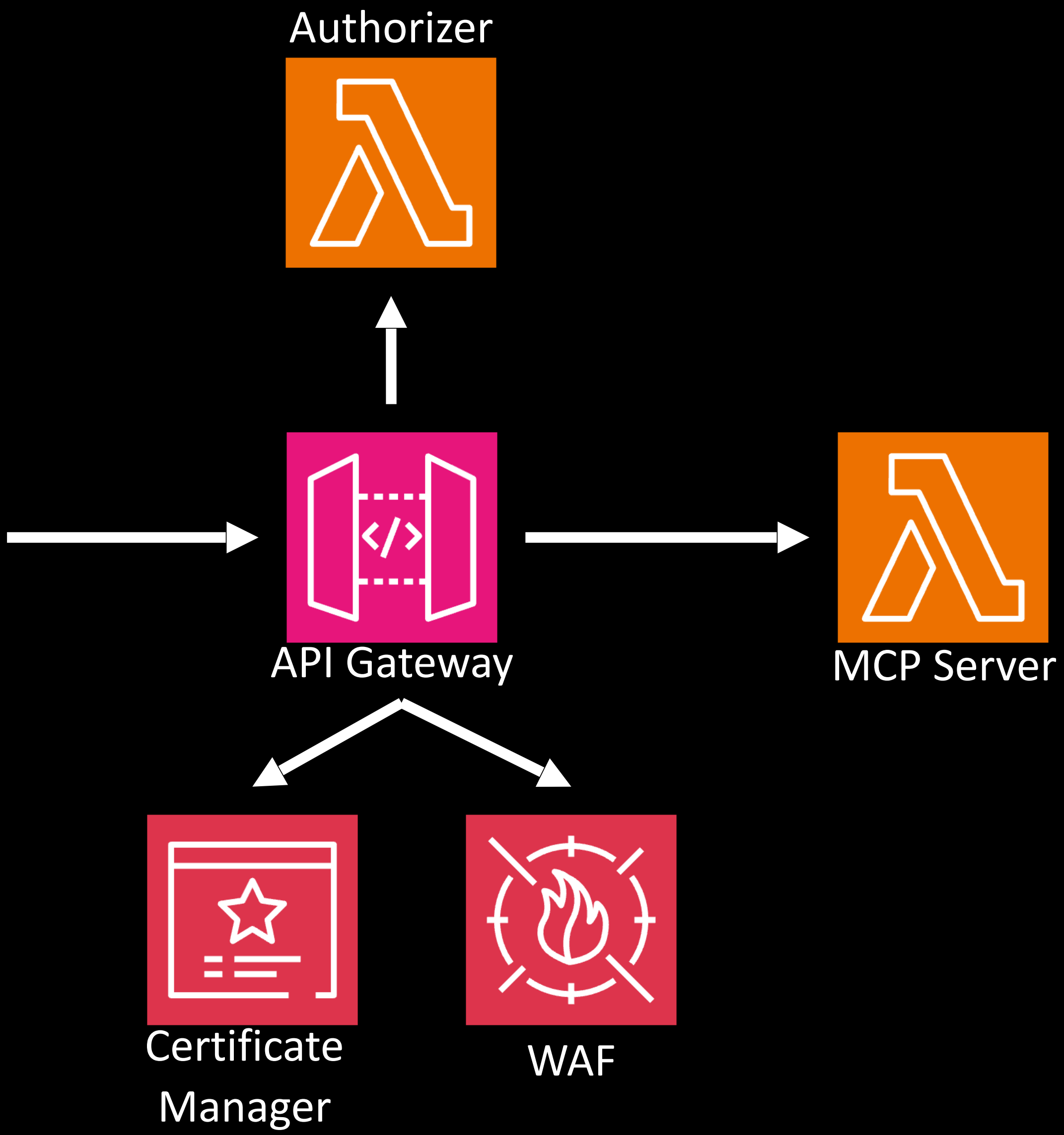
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS Lambda



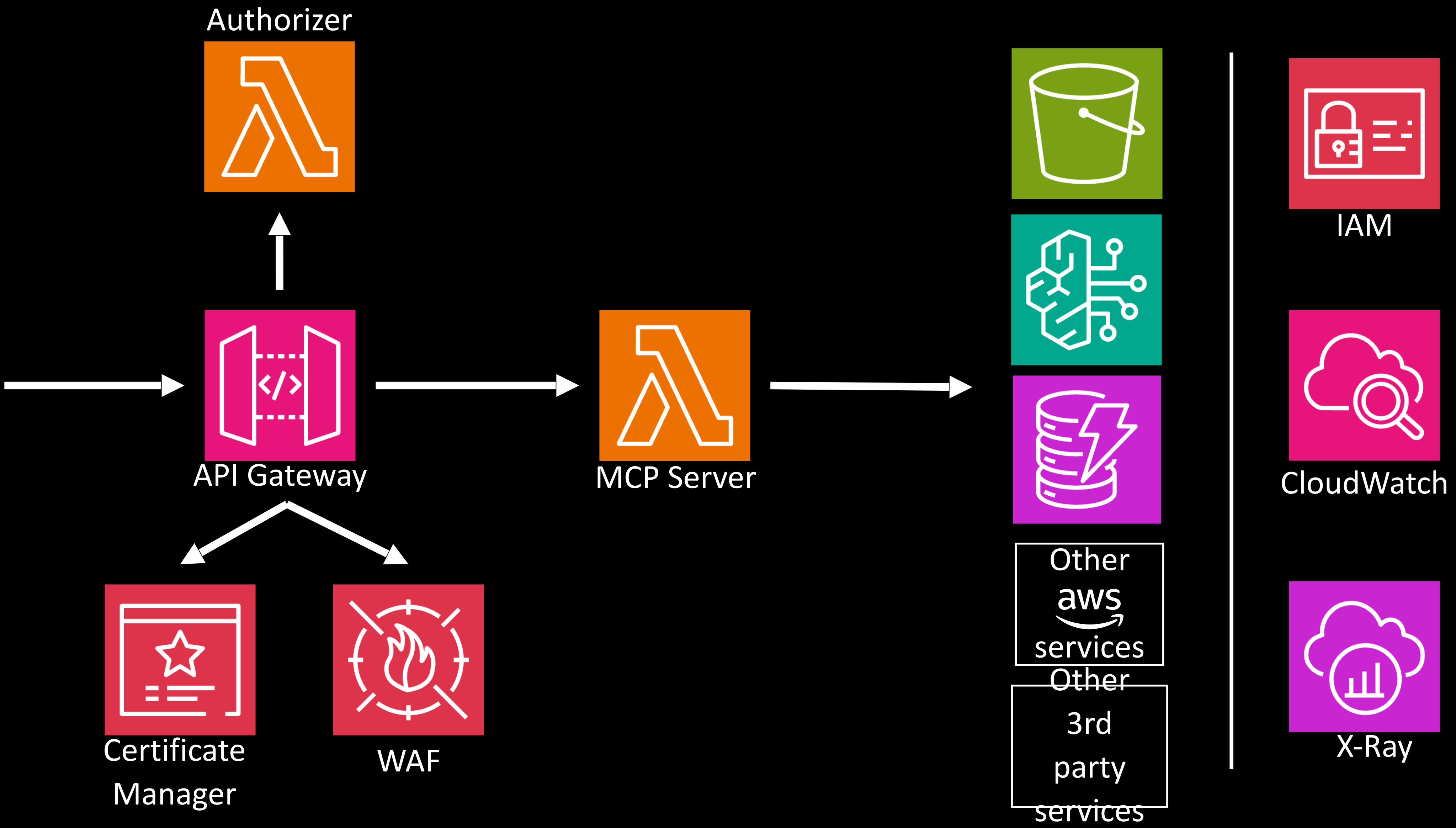
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS Lambda



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS Lambda



Authorizer

API Gateway

MCP Server

Certificate Manager

WAF

Other aws services

Other 3rd party services

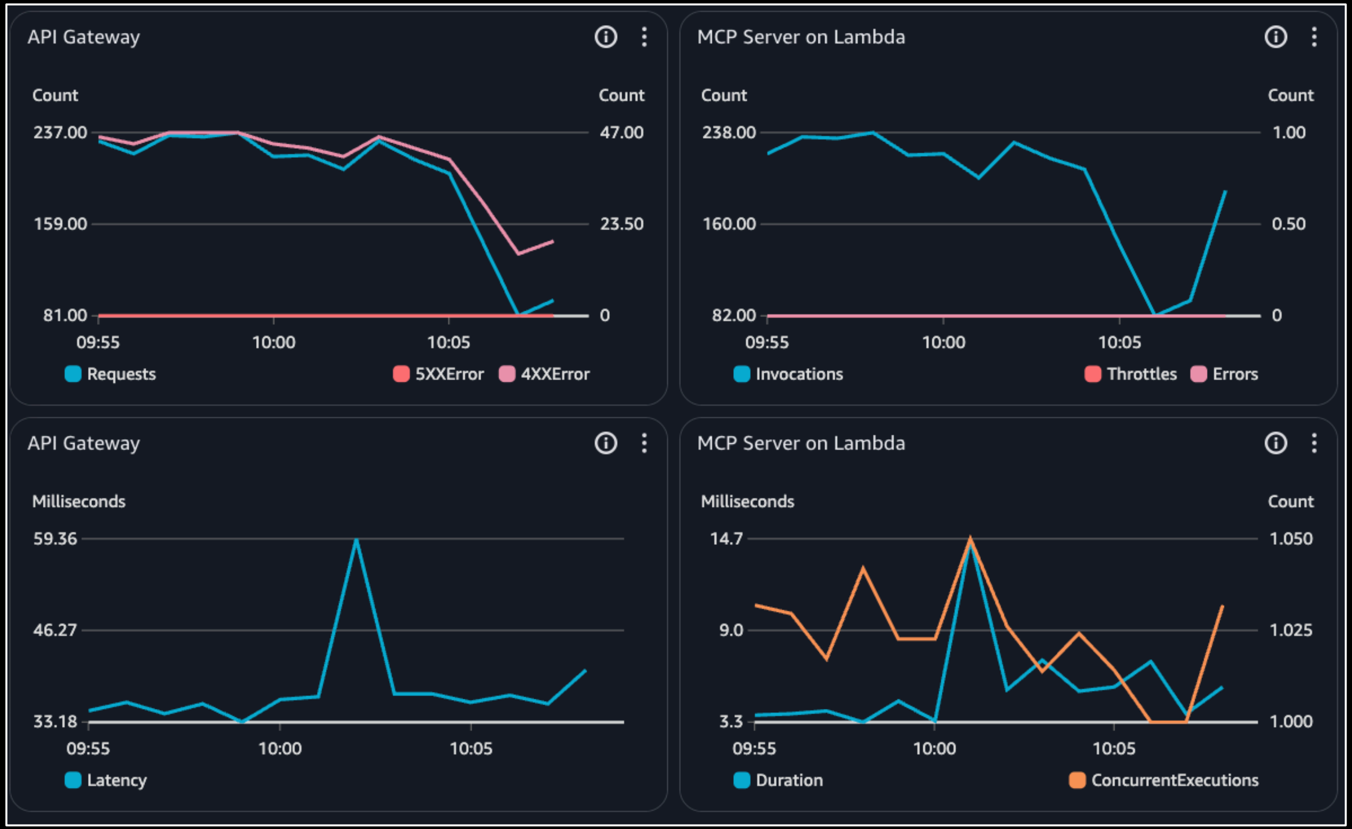
IAM

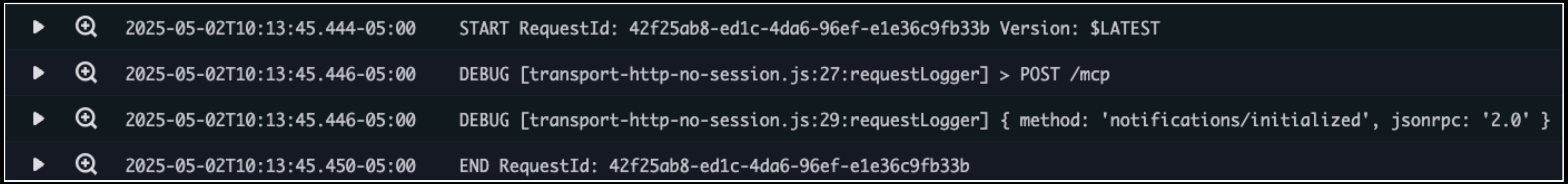
CloudWatch

X-Ray

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateless remote MCP Server on AWS Lambda





```
2025-05-02T10:13:45.444-05:00    START RequestId: 42f25ab8-ed1c-4da6-96ef-e1e36c9fb33b Version: $LATEST
2025-05-02T10:13:45.446-05:00    DEBUG [transport-http-no-session.js:27:requestLogger] > POST /mcp
2025-05-02T10:13:45.446-05:00    DEBUG [transport-http-no-session.js:29:requestLogger] { method: 'notifications/initialized', jsonrpc: '2.0' }
2025-05-02T10:13:45.450-05:00    END RequestId: 42f25ab8-ed1c-4da6-96ef-e1e36c9fb33b
```

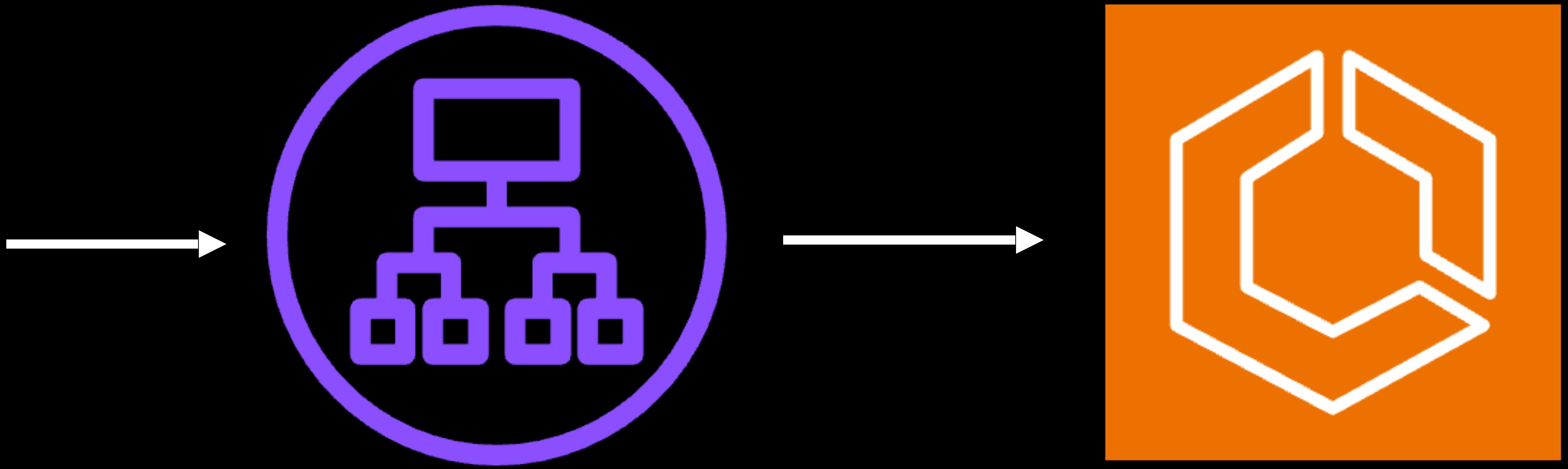
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Stateful MCP Server

with Amazon ECS

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateful remote MCP Server on ECS

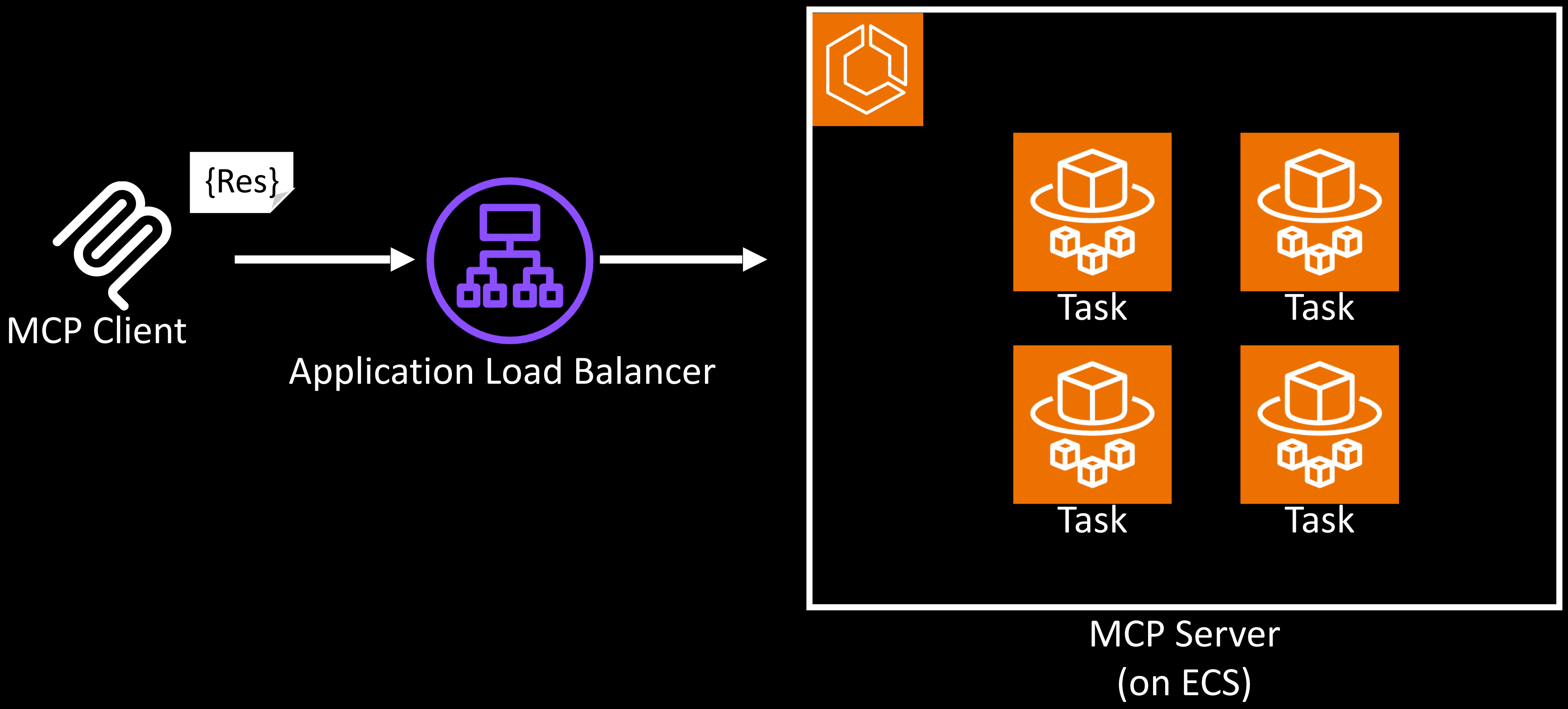


Application Load Balancer
(with sticky sessions, if stateful)

MCP Server
(on ECS Fargate)

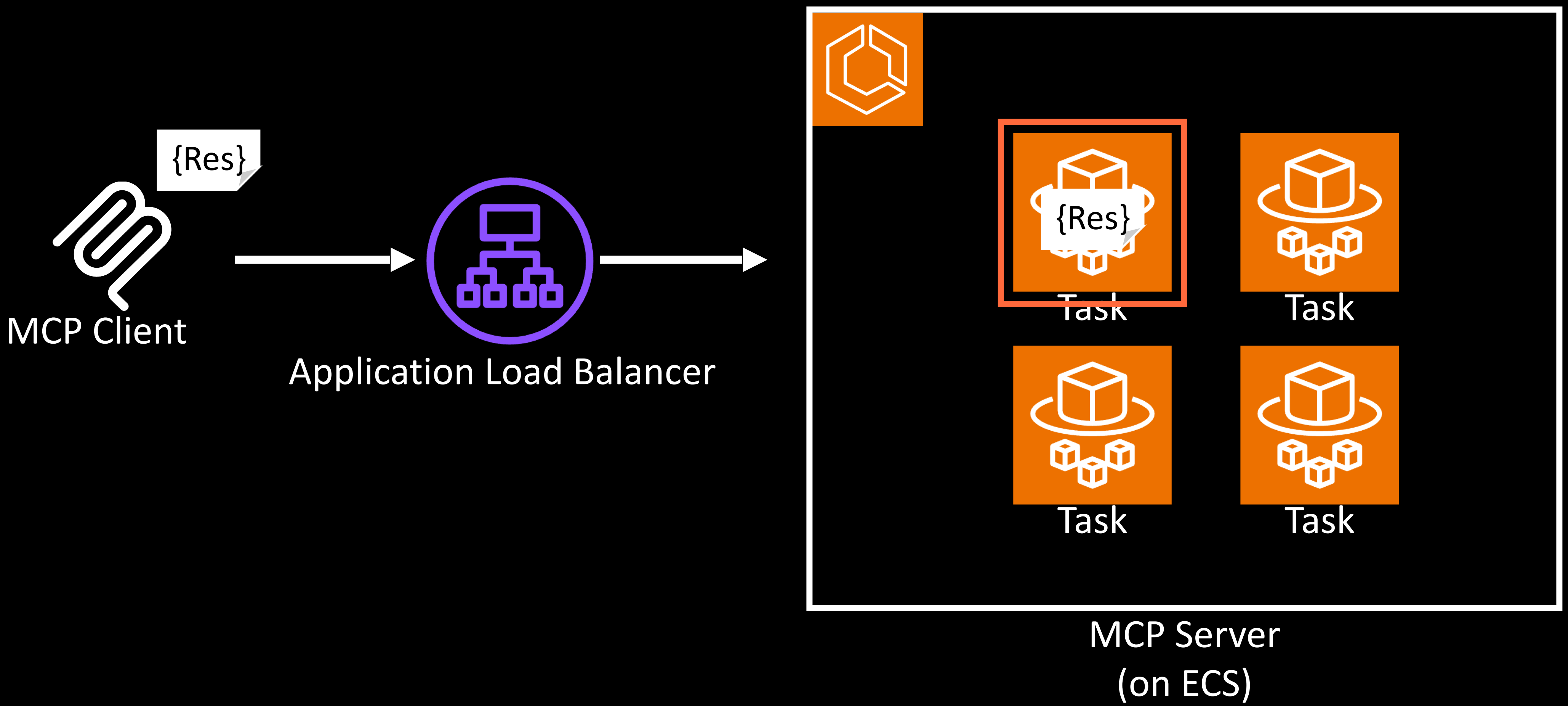
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateful remote MCP Server on ECS



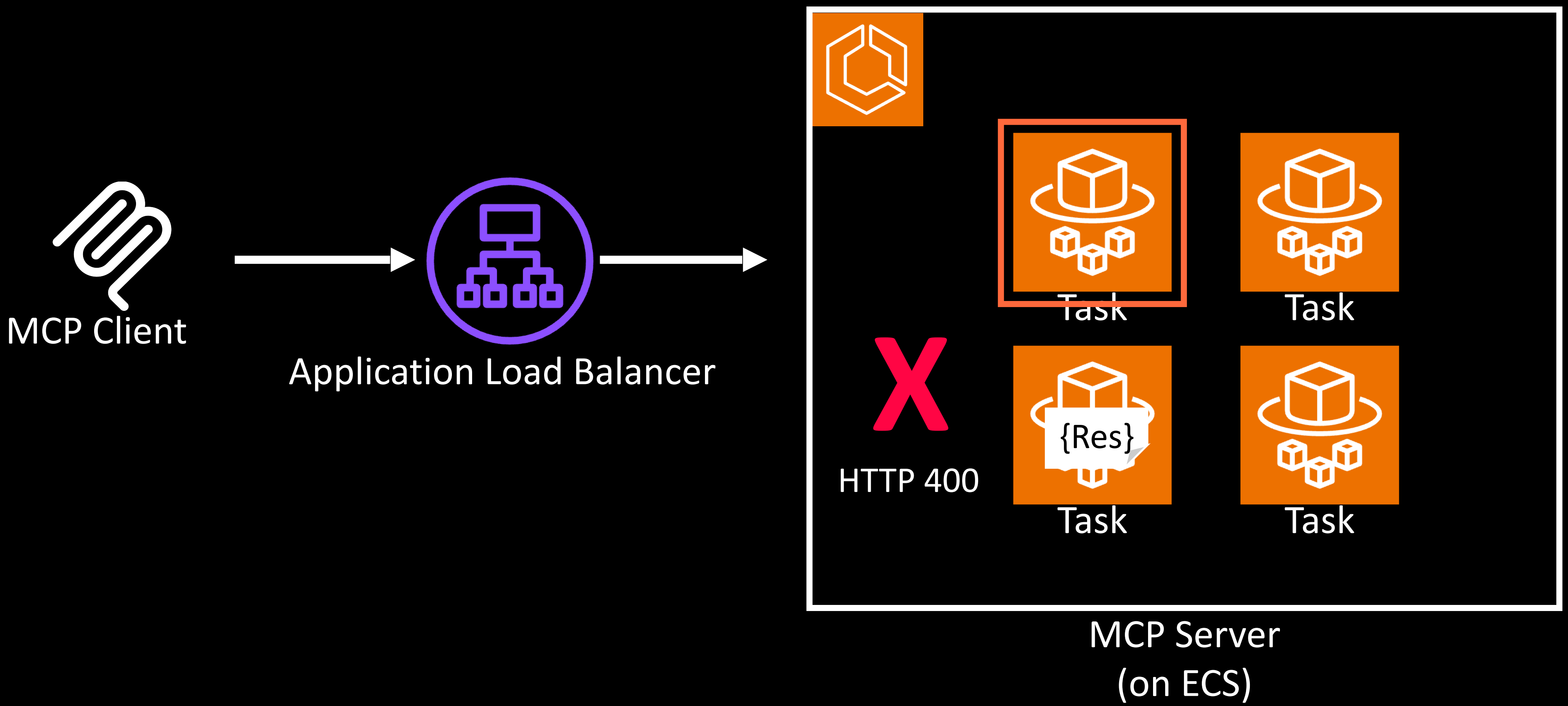
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateful remote MCP Server on ECS



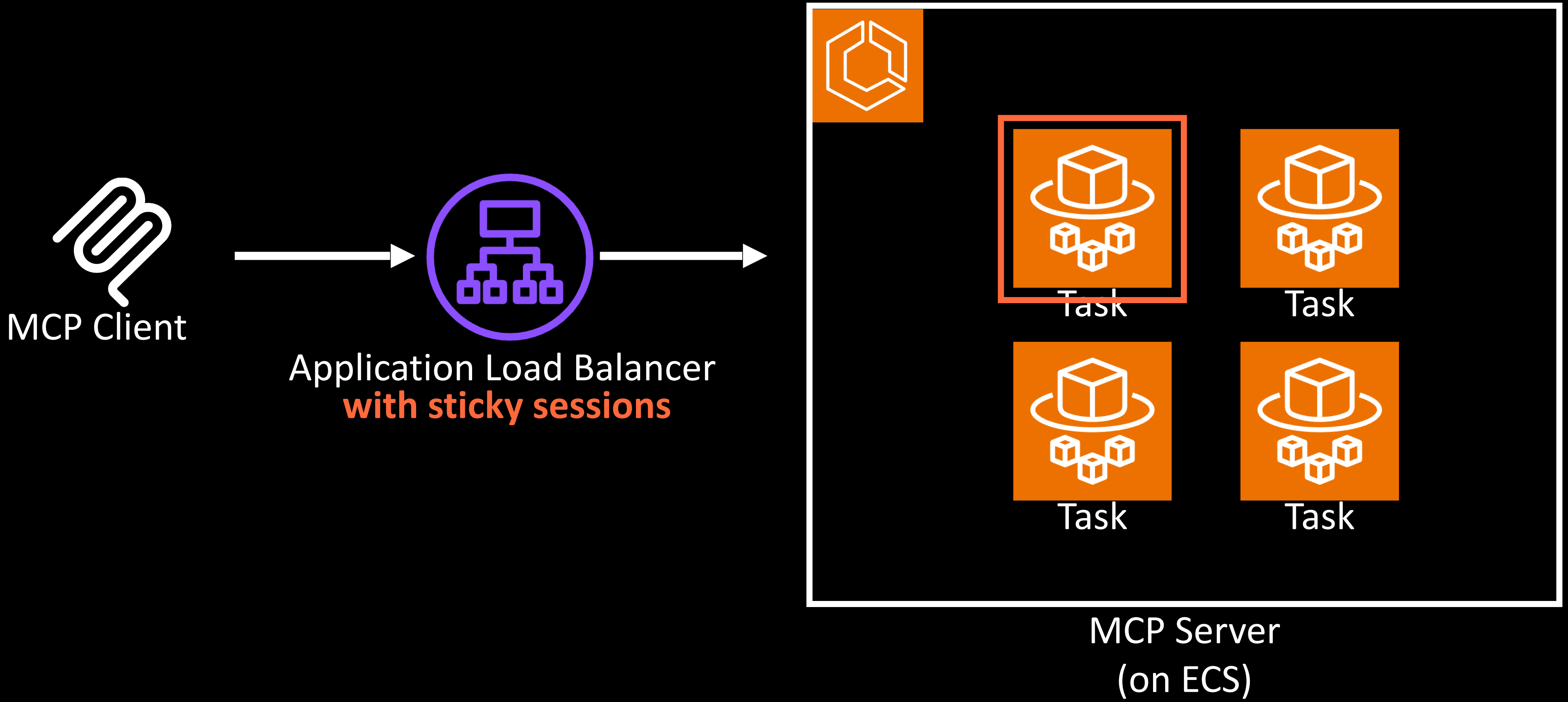
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateful remote MCP Server on ECS



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateful remote MCP Server on ECS

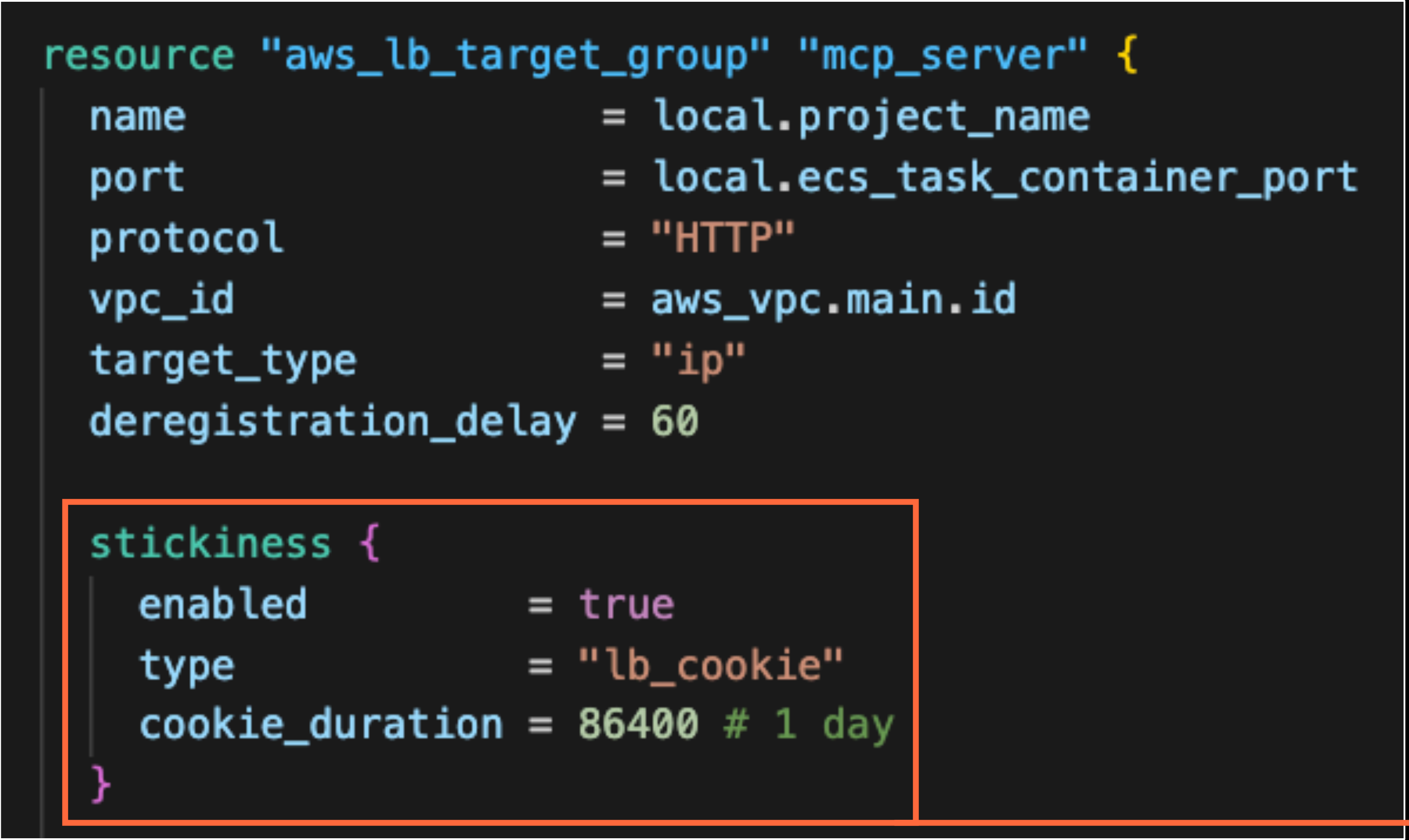

MCP Client

Application Load Balancer
with sticky sessions

MCP Server
(on ECS)

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateful remote MCP Server on ECS

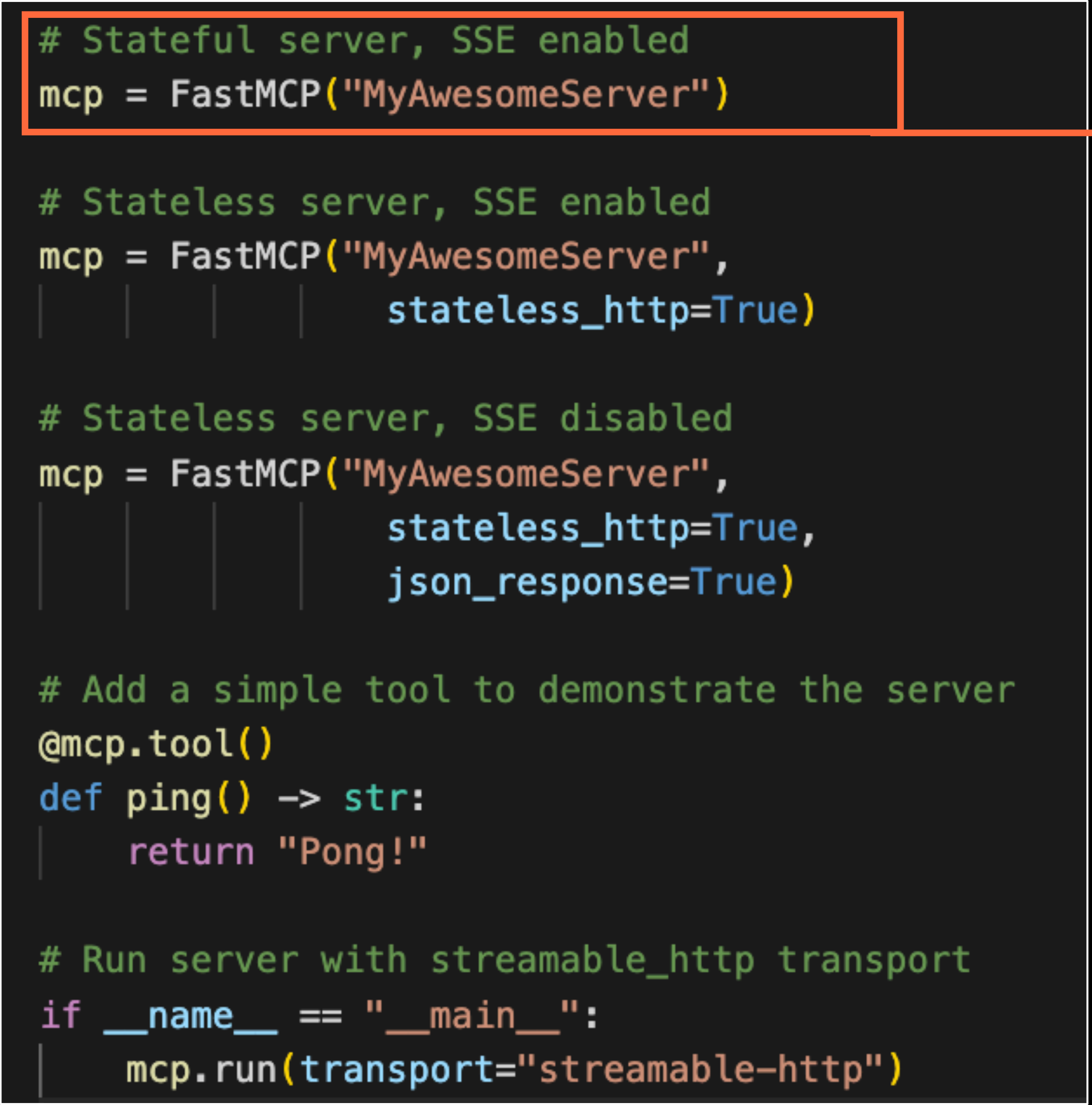


```
resource "aws_lb_target_group" "mcp_server" {
  name                 = local.project_name
  port                 = local.ecs_task_container_port
  protocol             = "HTTP"
  vpc_id               = aws_vpc.main.id
  target_type          = "ip"
  deregistration_delay = 60

  stickiness {
    enabled         = true
    type            = "lb_cookie"
    cookie_duration = 86400 # 1 day
  }
```

Enable sticky sessions

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Building a stateful remote MCP Server on ECS



```
# Stateful server, SSE enabled
mcp = FastMCP("MyAwesomeServer")

# Stateless server, SSE enabled
mcp = FastMCP("MyAwesomeServer",
                stateless_http=True)

# Stateless server, SSE disabled
mcp = FastMCP("MyAwesomeServer",
                stateless_http=True,
                json_response=True)

# Add a simple tool to demonstrate the server
@mcp.tool()
def ping() -> str:
    return "Pong!"

# Run server with streamable_http transport
if __name__ == "__main__":
    mcp.run(transport="streamable-http")
```

Define your MCP Server configuration

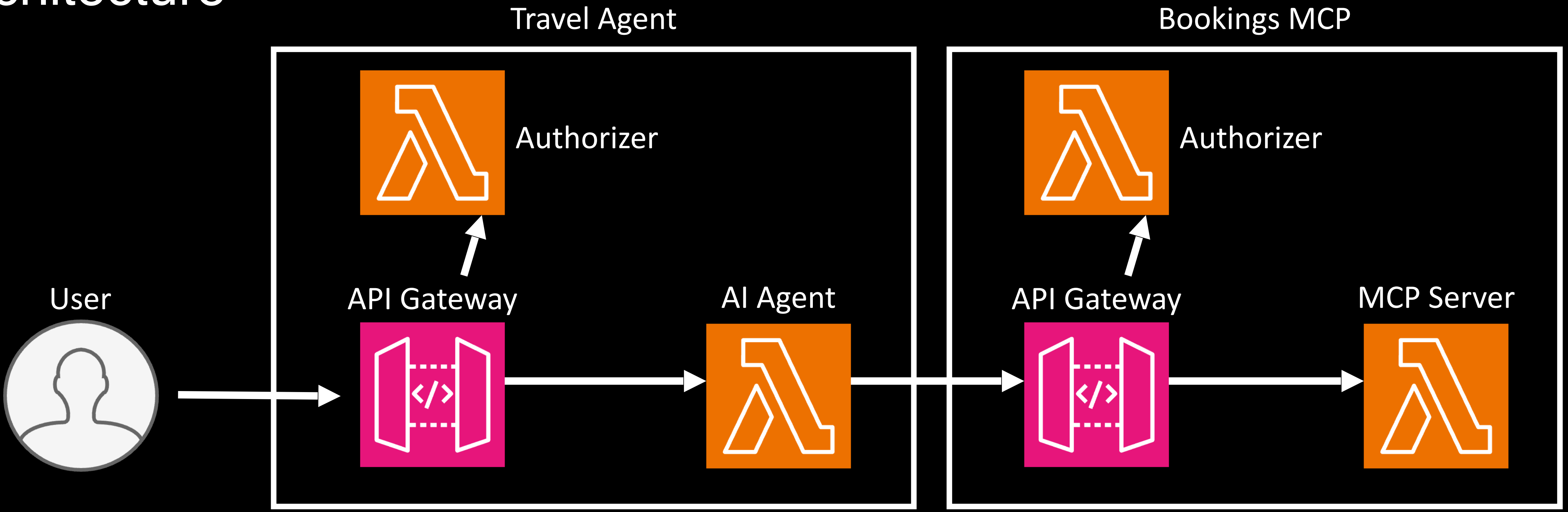
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Let’s see some demos

Building MCP Servers on AWS

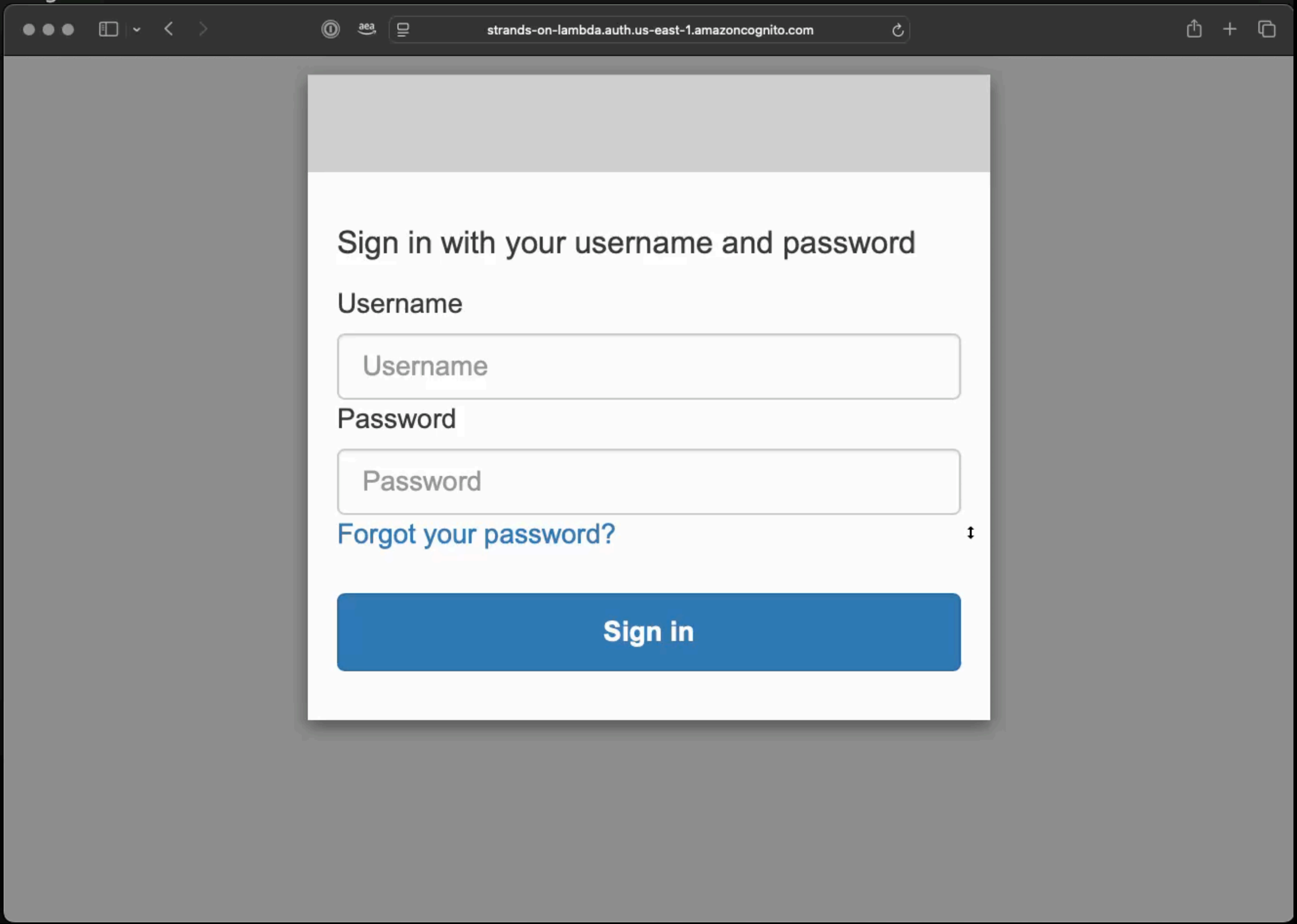
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Architecture



- Get Travel Policies
- Get Available Cars
- Get Available Hotels
- Book Car
- Book Hotel
- ….

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

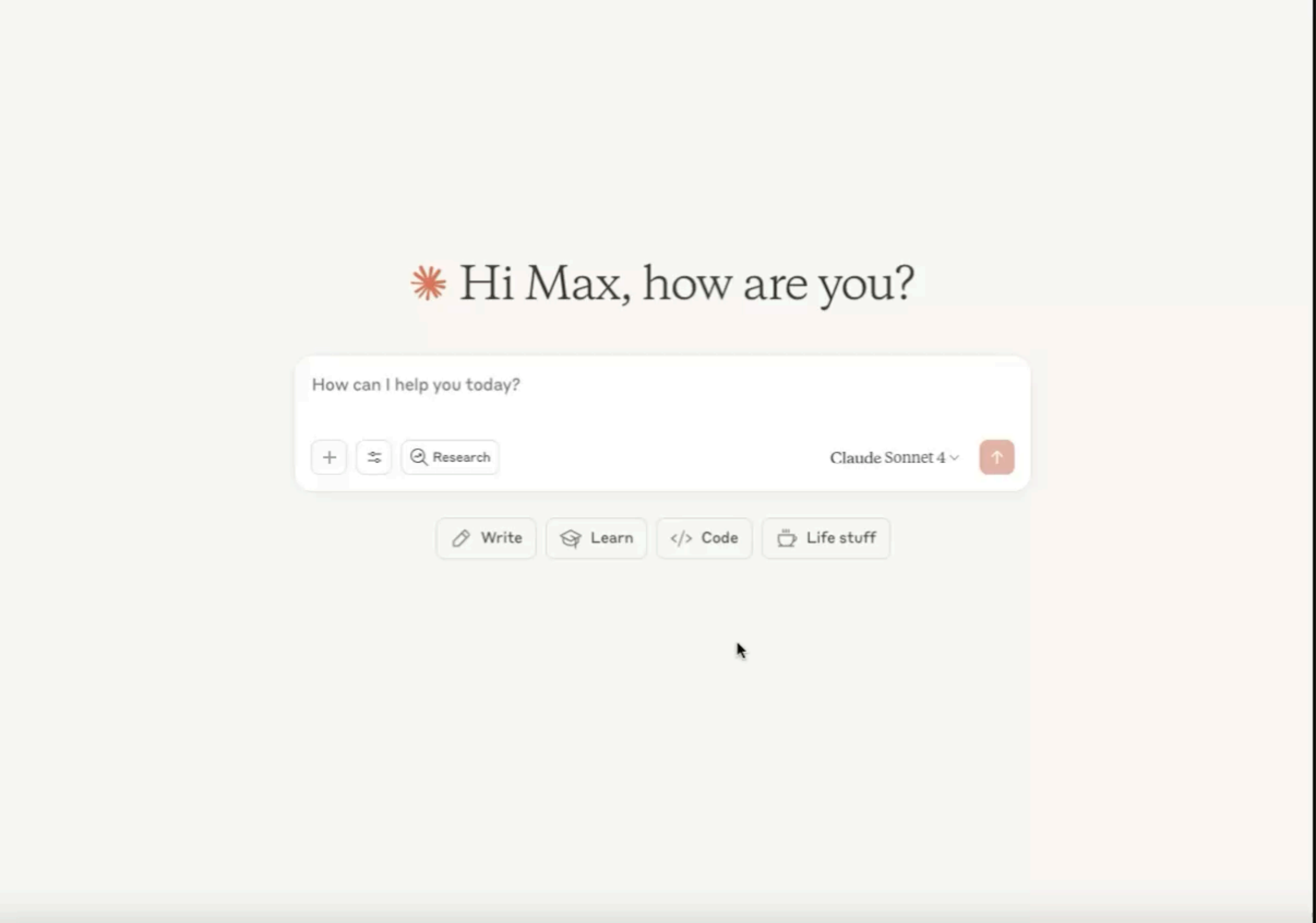


Sign in with your username and password

Username

Password

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.



Hi Max, how are you?

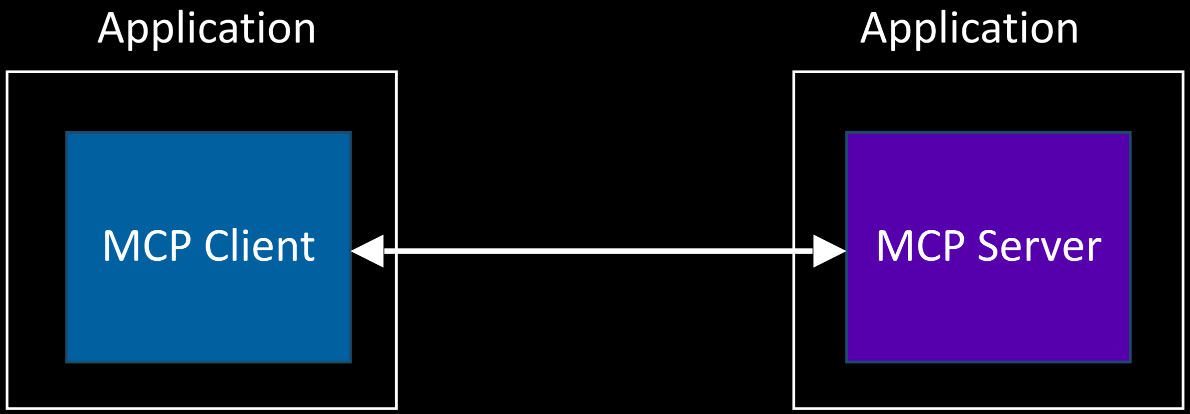
How can I help you today?

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Advanced patterns

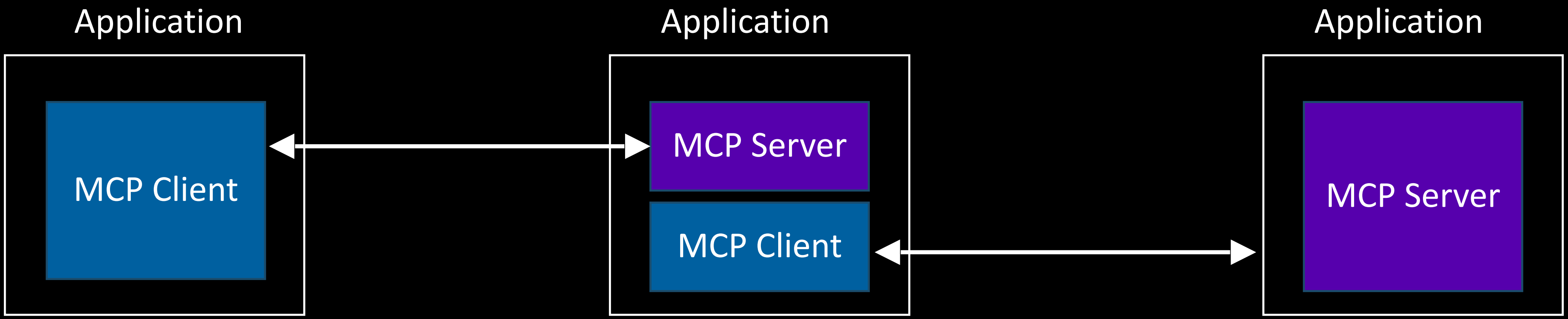
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Composability



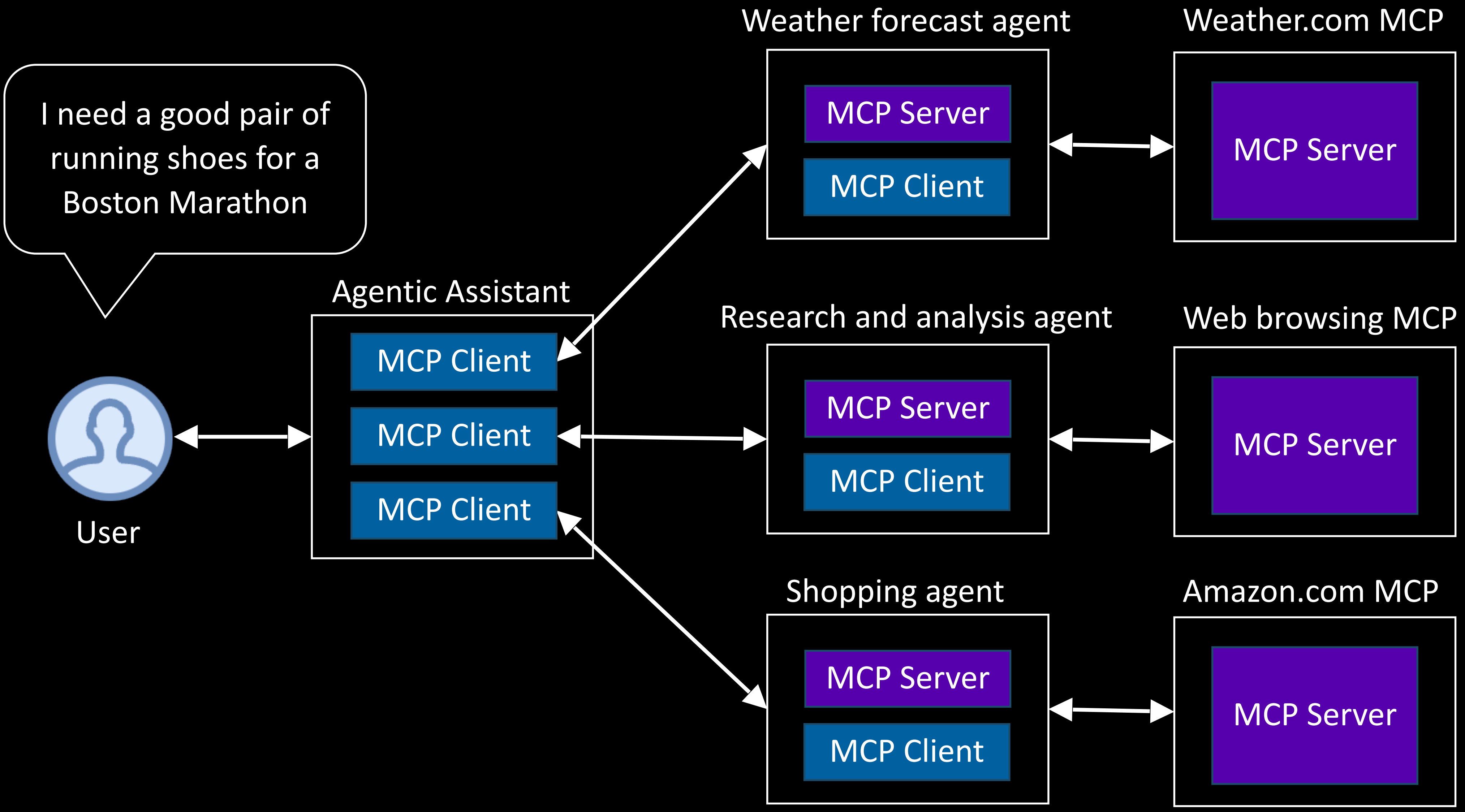
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Composability



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Agent-to-Agent communications



- getAvgTemp30days
- getAnnualTrends
- isRainExpected
- isSnowExpected

- searchWeb

- searchItem
- addToCart
- checkPayment
- checkout

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

In conclusion

Building MCP Servers on AWS

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Serverless MCP Servers on Lambda and ECS

https://github.com/aws-samples/sample-serverless-mcp-servers

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Best practices

Use official MCP SDKs (and follow the MCP spec evolution)

Design for scalability and performance

Implement robust network and application security

Establish comprehensive monitoring and logging

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Continue your Serverless Gen AI / MCP journey

- Sample Serverless MCP Servers
- Resources for Generative AI and Serverless
- Deploy MCP servers in AgentCore Runtime
- CDK Construct for Serverless MCP Servers (1, 2, 3)
- AWS Labs MCP Servers (20+ MCP Servers!!!)
- Introducing AWS MCP Servers for code assistants
- Extend the Amazon Q Developer CLI with MCP
- Running MCP-Based Agents on AWS

ANTHROP\C

- Latest MCP Spec
- Getting started with MCP
- MCP Registry
- A list of maintained MCP Servers

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Thank you!

Anton Aleksandrov
Principal Solutions Architect
AWS Serverless

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.