GOSIM

GOSIM AI Paris 2025

Streamlining AI App Development with Docker

Models and AI Tools That Just Work

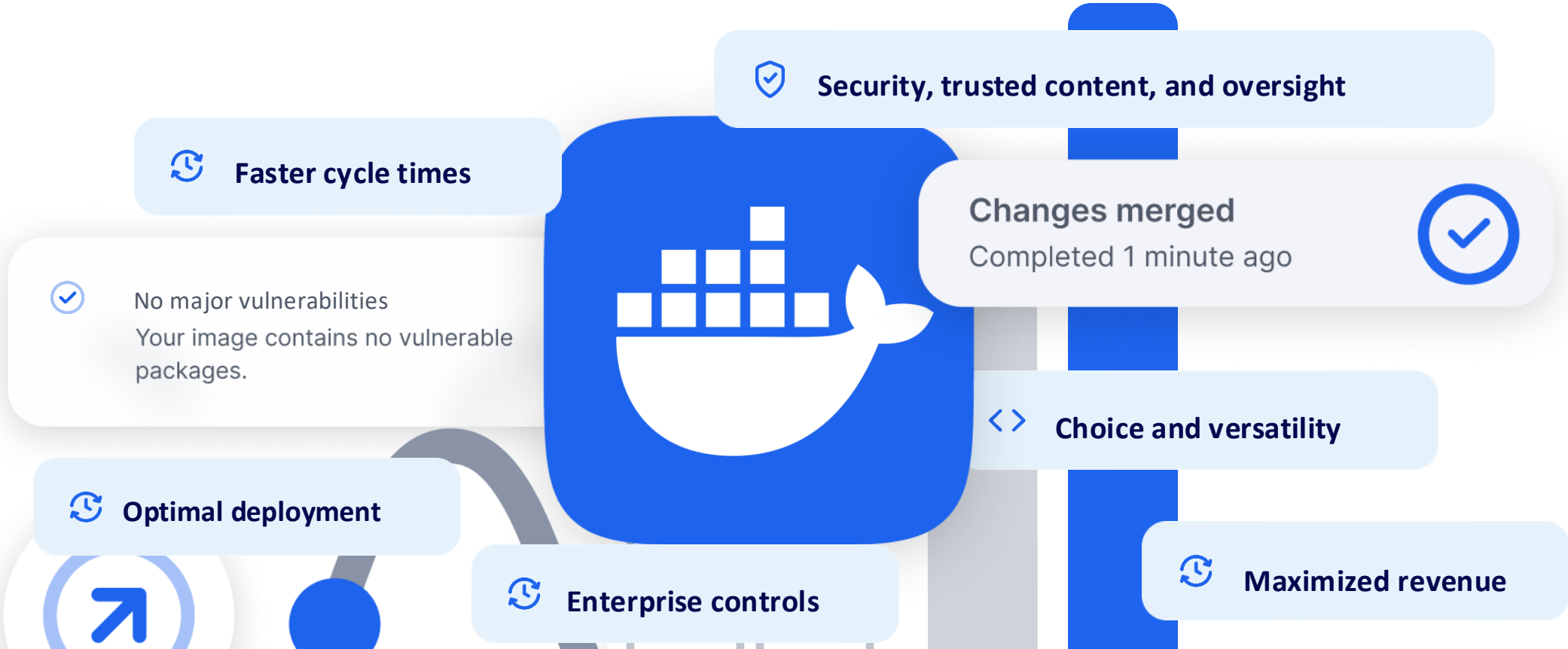
Jean Laurent de Morlhon

Sr Vice President, GenAI Acceleration

Docker

Docker is the cornerstone of software development

#1 Developer-Loved Tool



Docker maintains open-source projects

Open Source Developer Tools Developed by Docker

Docker Compose

Test Containers

Docker CLI

Docker BuildX

Docker Github Action

Essential Projects

Moby

BuildKit

Runc

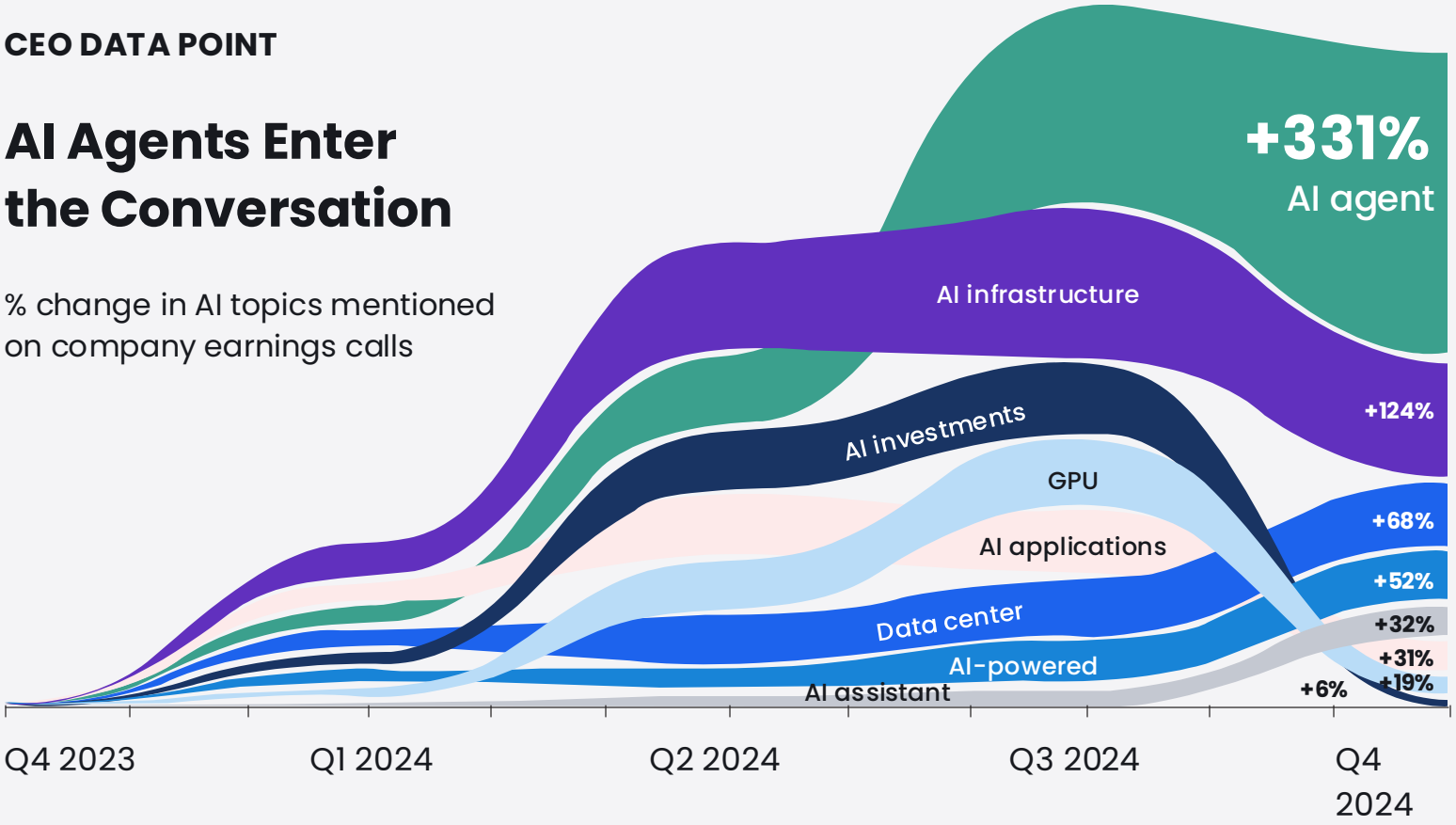
Containerd

Lots of AI Agents are being built

CEO DATA POINT

AI Agents Enter the Conversation

% change in AI topics mentioned on company earnings calls



Source: IoT Analytics, BCG, PRNewswire

44.8%

CAGR Growth

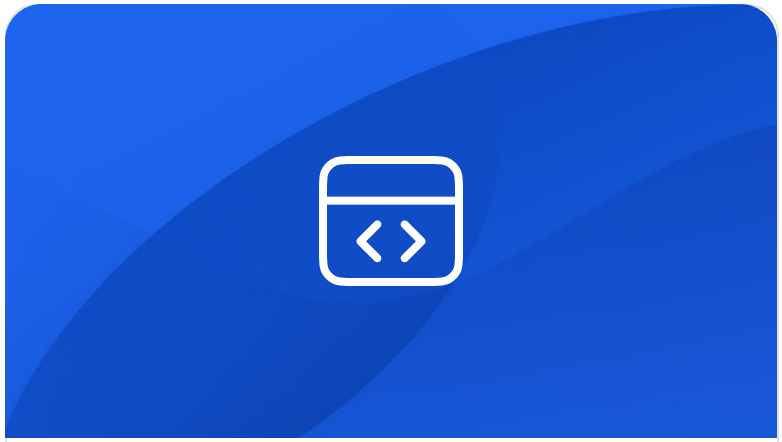
30-50%

Reduction in time to decision for Bloomberg

30%+

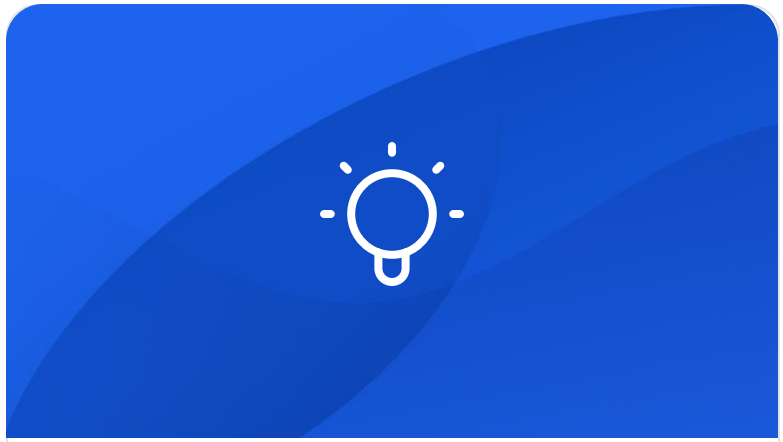
Reduction in cycle time for booking.com

… but building them are still challenging for many developers



Build AI applications easily

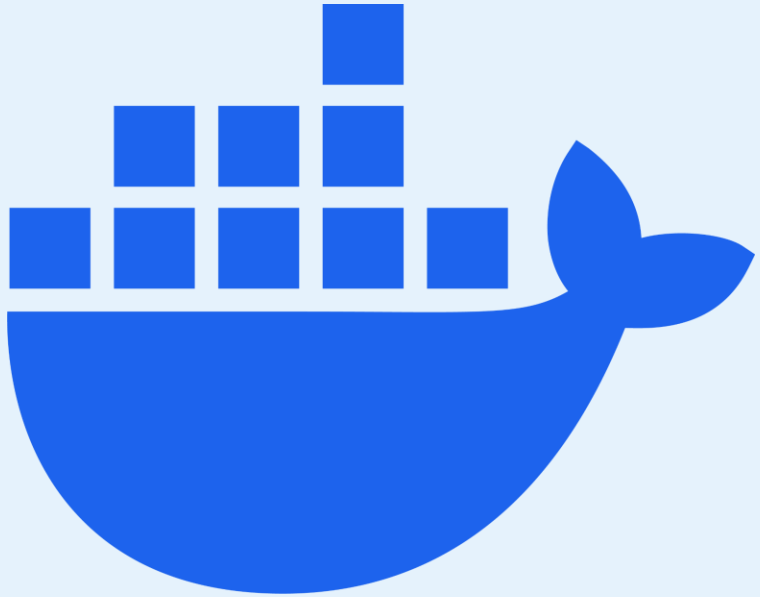
Ensure Security and Compliance



Improve productivity with AI

Containers provide the solid foundation to bring apps, including Generative and agentic AI apps to life

- Flexible
- Scalable
- Secure



Containers are the standard solution for GenAI apps

7/10

plan to containerize GenAI apps

75%

AI deployment will use containers

133%

increase in AI image pulls

Source: CIODIVE

Docker's AI efforts

Build GenAI apps and agents easily

Model Runner, MCP Catalog & Toolkit

Boost productivity with AI Agents

Gordon

Secure & Familiar workflows

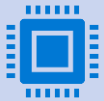
The Docker Model Runner

Same processes.
Same tools.
New stack.

The Docker Model Runner

Pull models from Docker Hub or any other container registry

Run LLMs completely locally



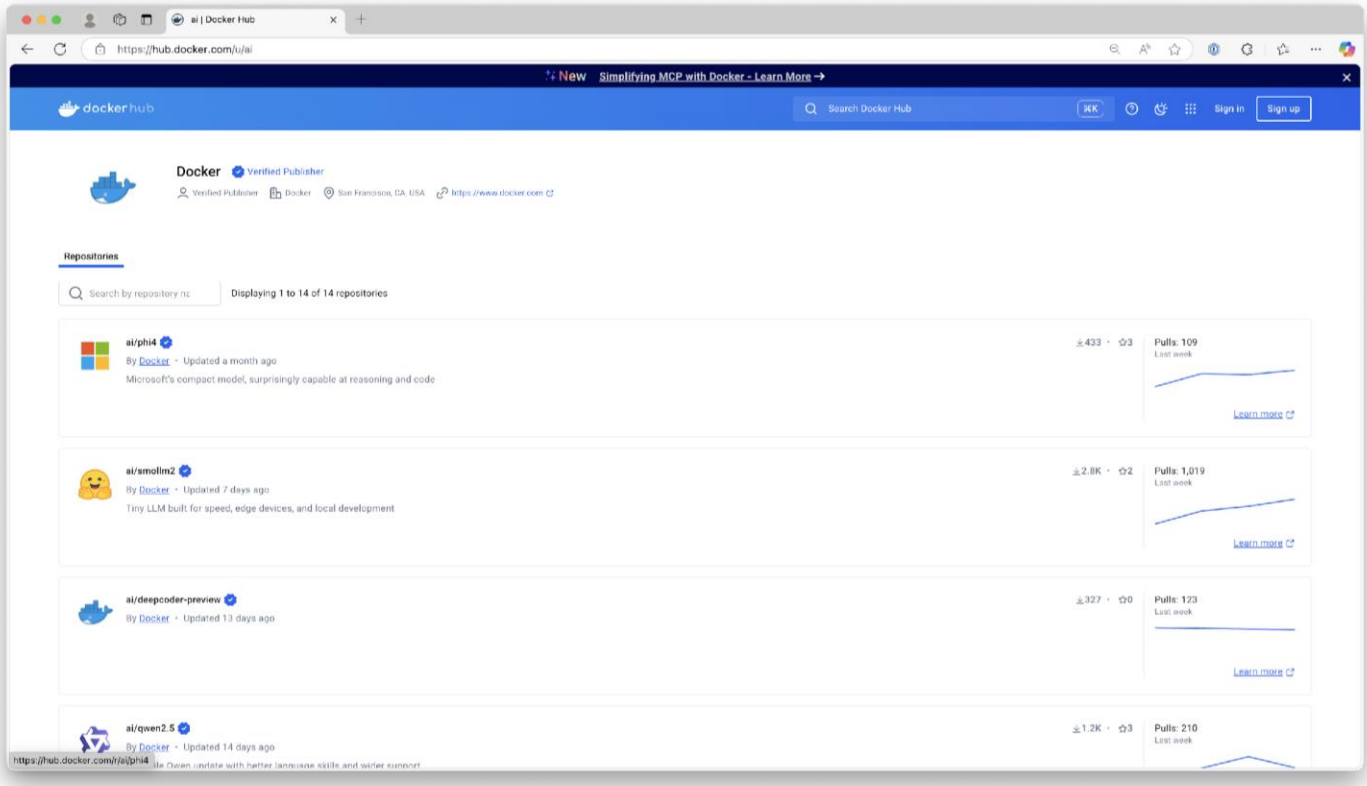
Full GPU support

OpenAI API-compatible endpoints

Interacting with models via the CLI

```
# Pull a model from Docker Hub
docker model pull ai/phi4

# List all downloaded models
docker model list

# Send a prompt to the model
docker model run ai/phi4 "What should I do in Paris after GOSIM?"
```

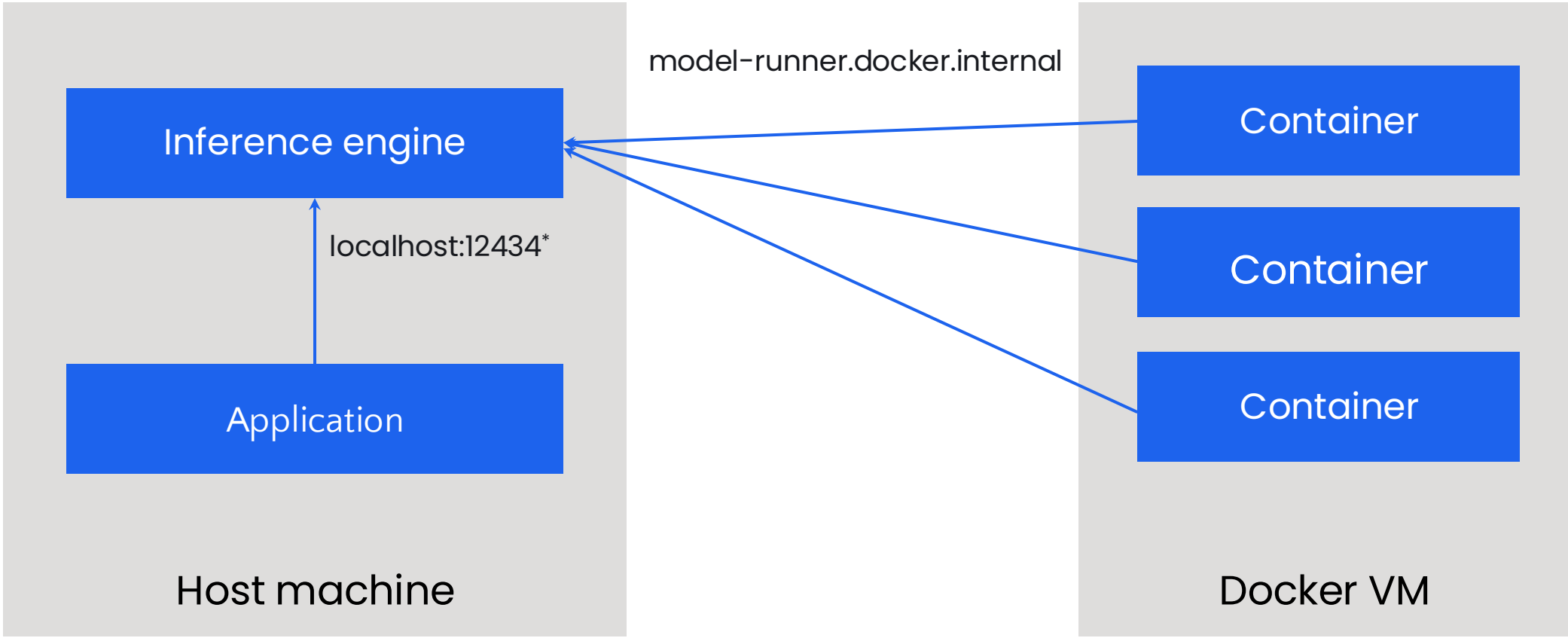
An entire catalog of models available on Docker Hub



Publishing models

```
# Package and push a GGUF model with licenses
docker model package \
    --licenses license1.txt \
    --licenses license2.txt \
    ./model.gguf \
    registry.example.com/ai/custom-model

# Publish a model
docker model push registry.example.com/ai/custom-model
```

Accessing the models



* TCP host socket not enabled by default

Connecting apps

```
// JavaScript/Node.js with OpenAI client
const client = new OpenAI({
    baseURL: "http://localhost:12434/engines/v1",
    apiKey: "not-required",
});


// .NET with OpenAI ChatClient
ChatClient client = new(
  model: "ai/phi4",
  credential: new ApiKeyCredential("not-required"),
  new OpenAIClientOptions()
  {
    Endpoint = new Uri("http://localhost:12434/engines/v1")
  }
);
```

Going beyond the CLI - Compose

compose.yaml

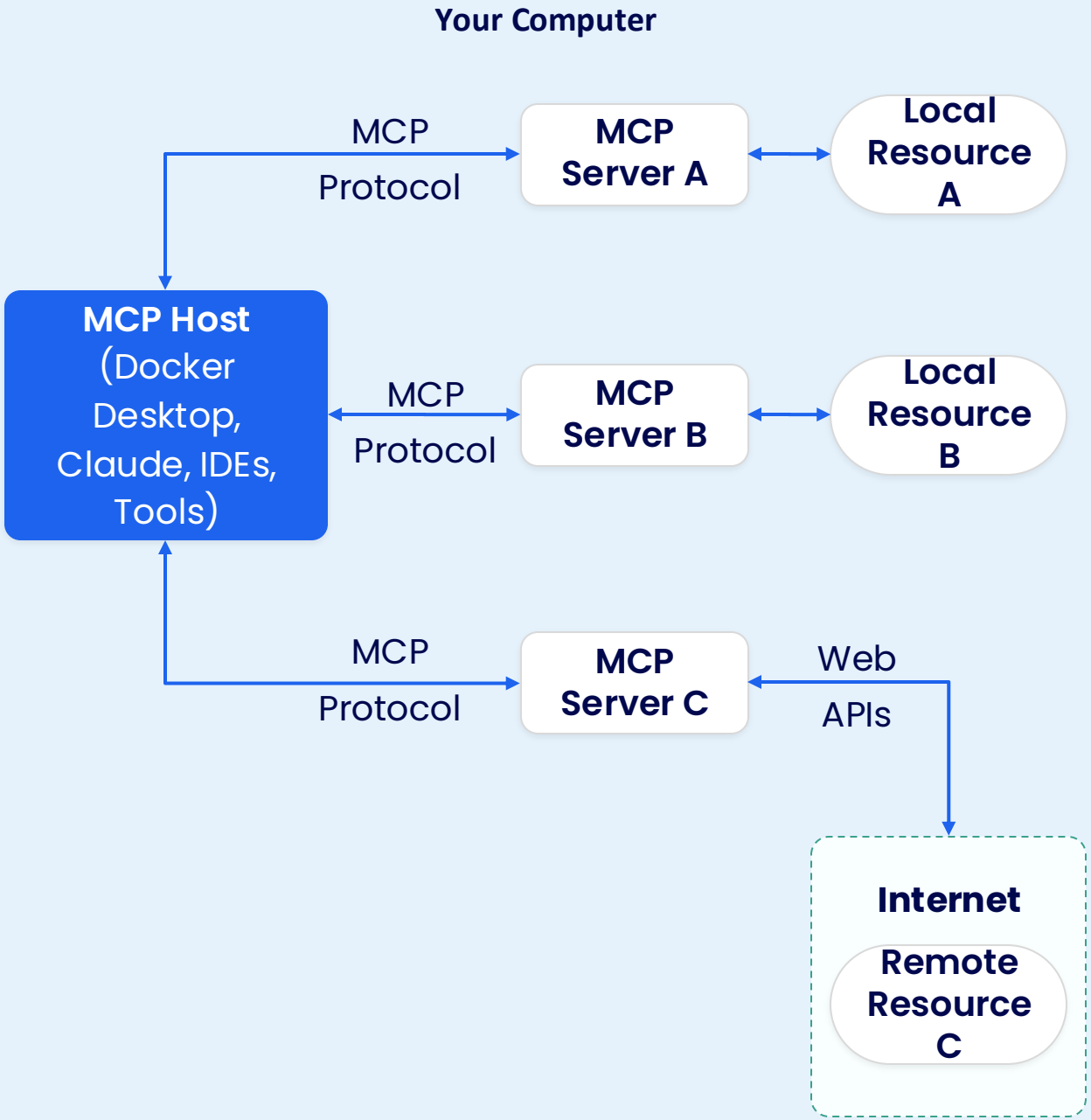
```yaml
services:
  phi:
    provider:
      type: model
      options:
        model: ai/phi4:14B-Q4_K_M

  app:
    image: node:lts-alpine
    depends_on:
      - phi
    ...

  db:
    image: postgres
    ...
```

Demo 1 : Docker Model Runner

Docker MCP Catalog and Toolkit

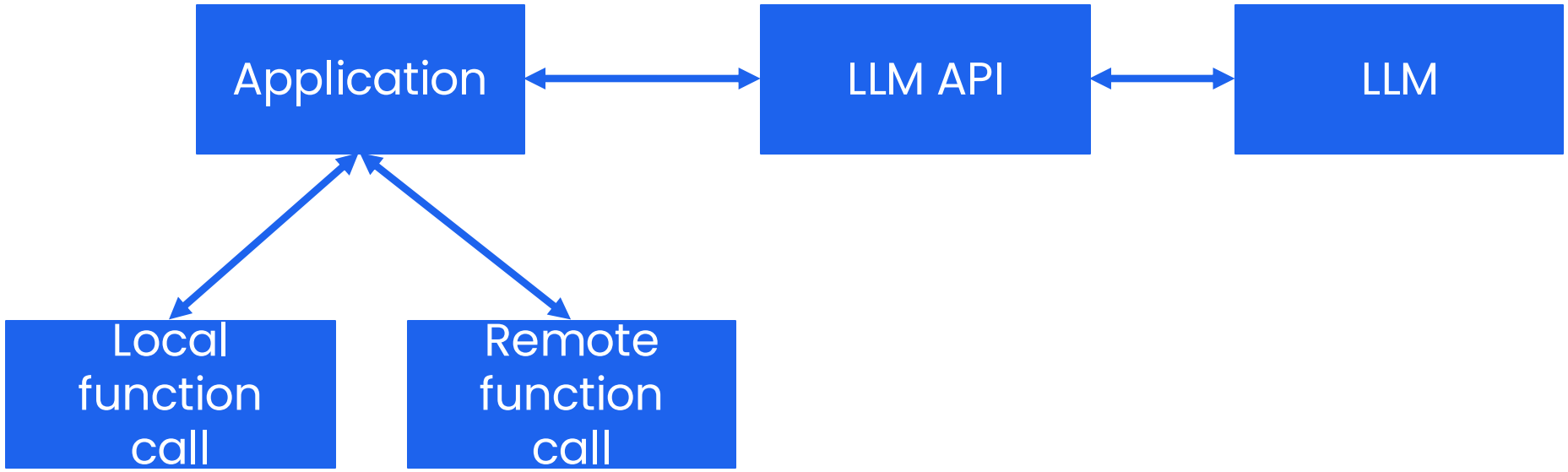
What is MCP

- A new protocol open-sourced by Anthropic
- Standardized interface for AI Agents, apps and LLMs
- Integrate with external data sources and tools
- Maybe it's an API for enhancing LLM capabilities
- It's the next evolution after OpenAI introduced Tool Call



Source: Model Context Protocol Architecture

Tool/function calling



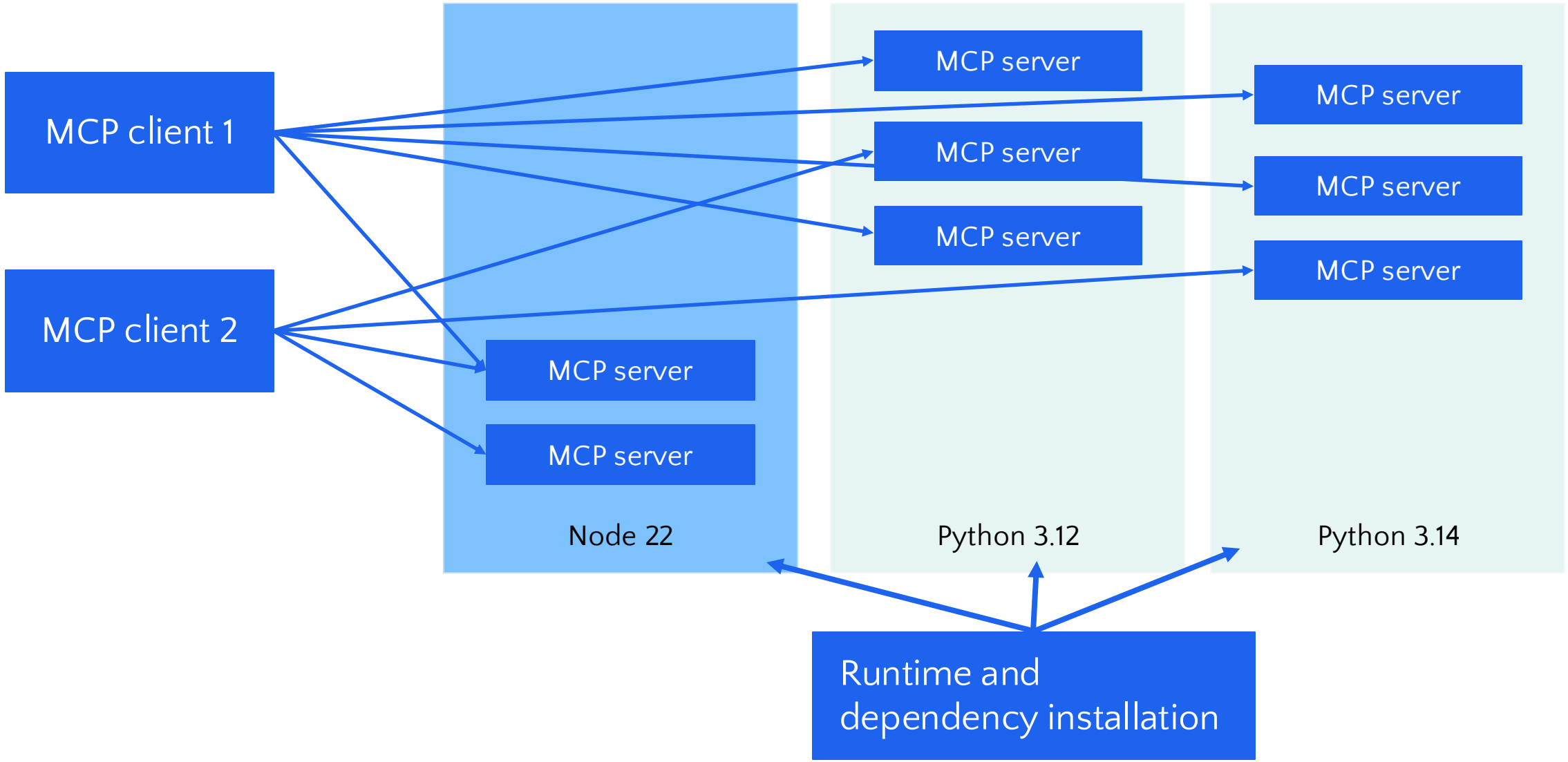
Configuring a MCP server - npx

```
{
  "servers": {
    "github": {
      "command": "npx",
      "args": [
        "-y",
        "@modelcontextprotocol/server-github"
      ],
      "env": {
        "GITHUB_PERSONAL_ACCESS_TOKEN": "<YOUR_TOKEN>”
      }
    }
  }
}
```

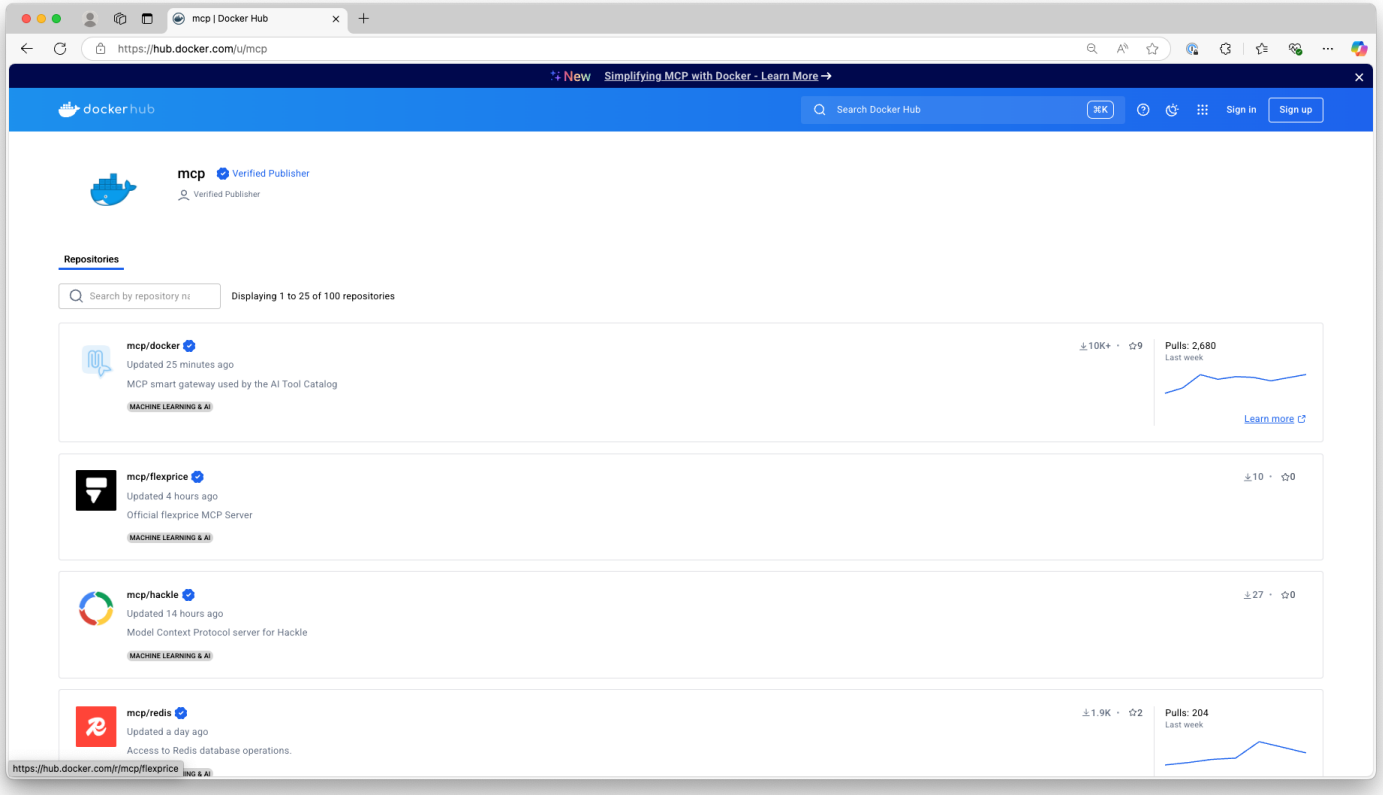
Configuring a MCP server - uvx

```
{
  "servers": {
    "time": {
      "command": "uvx",
      "args": [
        "mcp-server-time"
      ]
    }
  }
}
```

MCP struggles



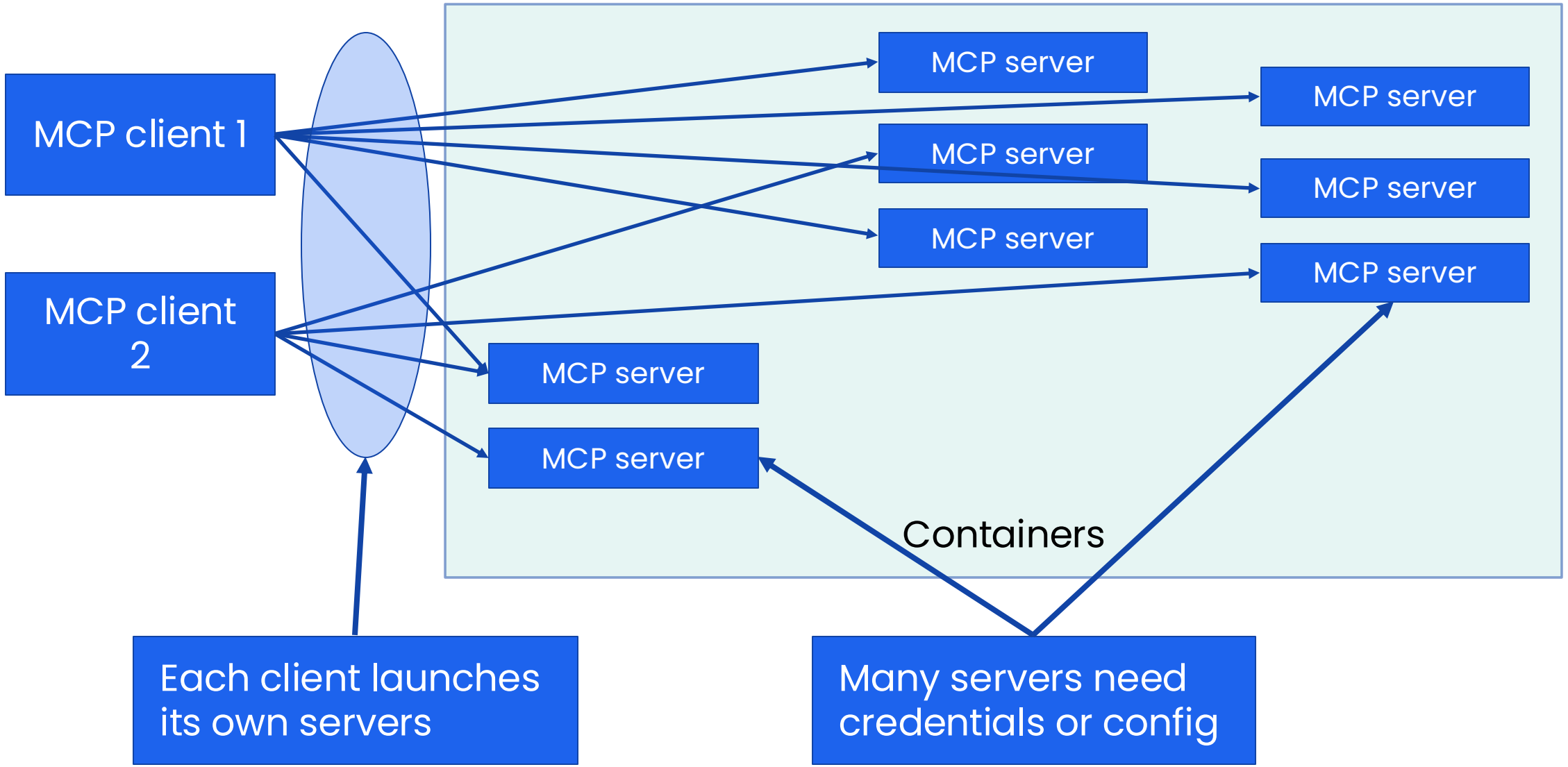
The MCP Catalog provides a collection of containerized MCP servers



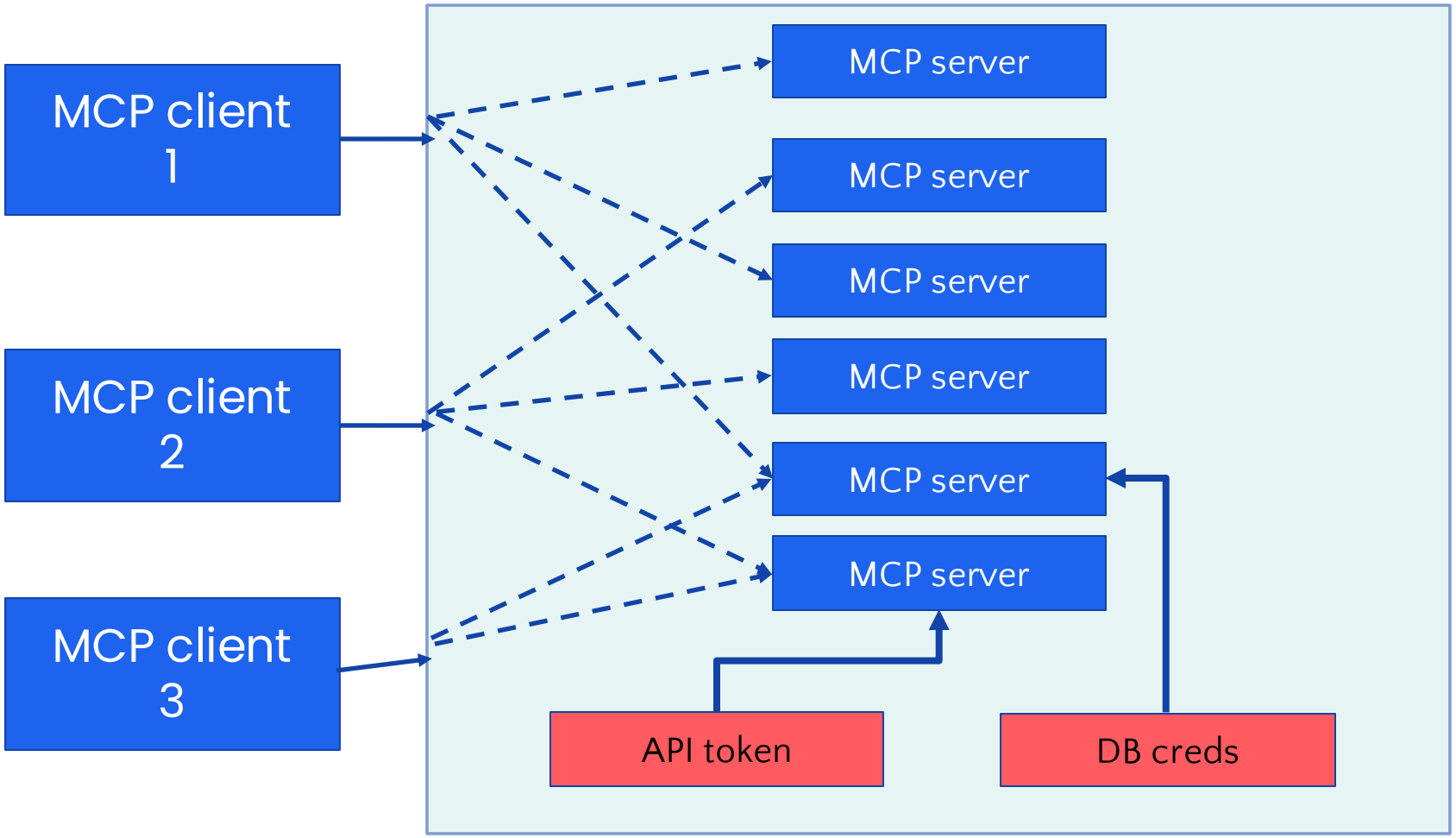
Adding a MCP server to Claude Desktop

```
{
  "mcp": {
    "servers": {
      "time-server": {
        "command": "docker",
        "args": ["run", "--rm", "-i", "mcp/time"]
      }
    }
  }
}
```

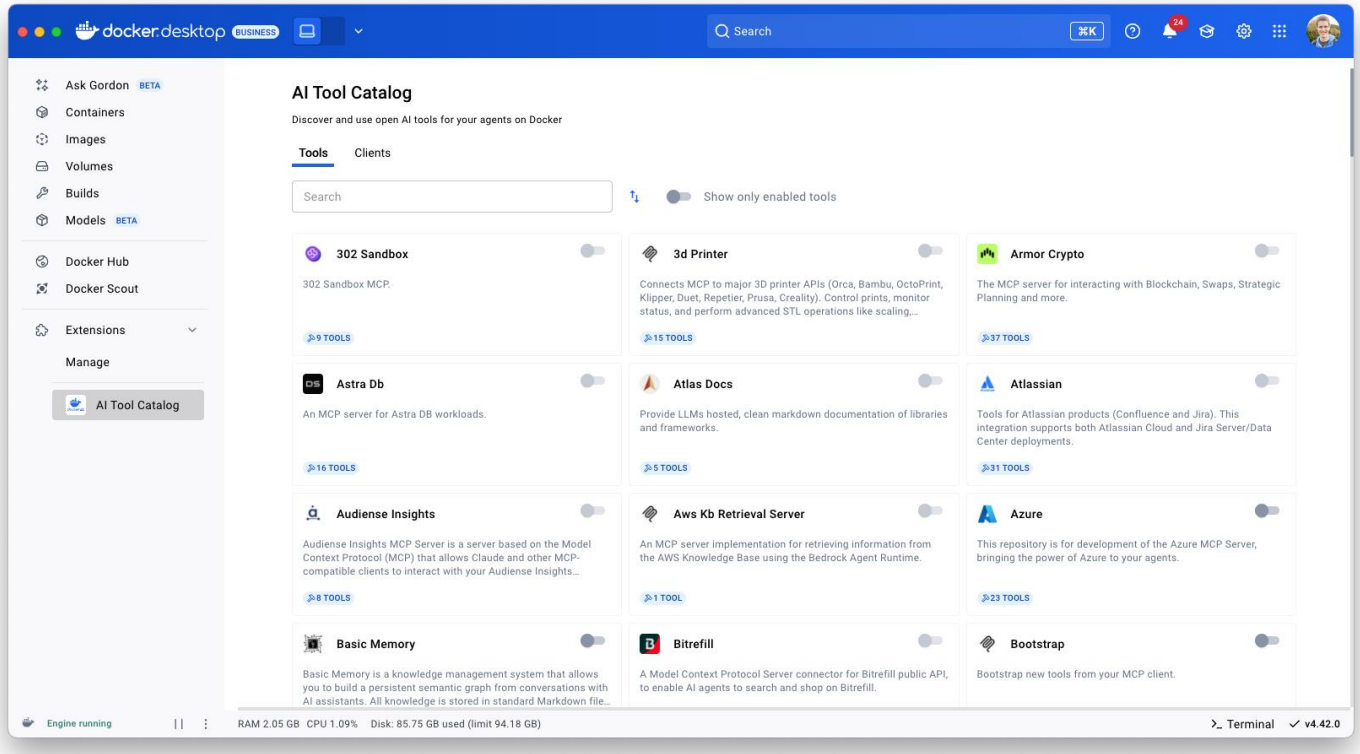
Gateway needed on top of containerized MCP tools



The MCP Toolkit



The MCP Toolkit provides easy discovery, configuration, and launching of containerized MCP servers



Demo 2 : Docker MCP Toolkit

The Docker AI Agent : Gordon

Best Practice for Building AI Agents

Specialized

Depth

Meeting devs where they are

Breadth

Customization

Flexibility

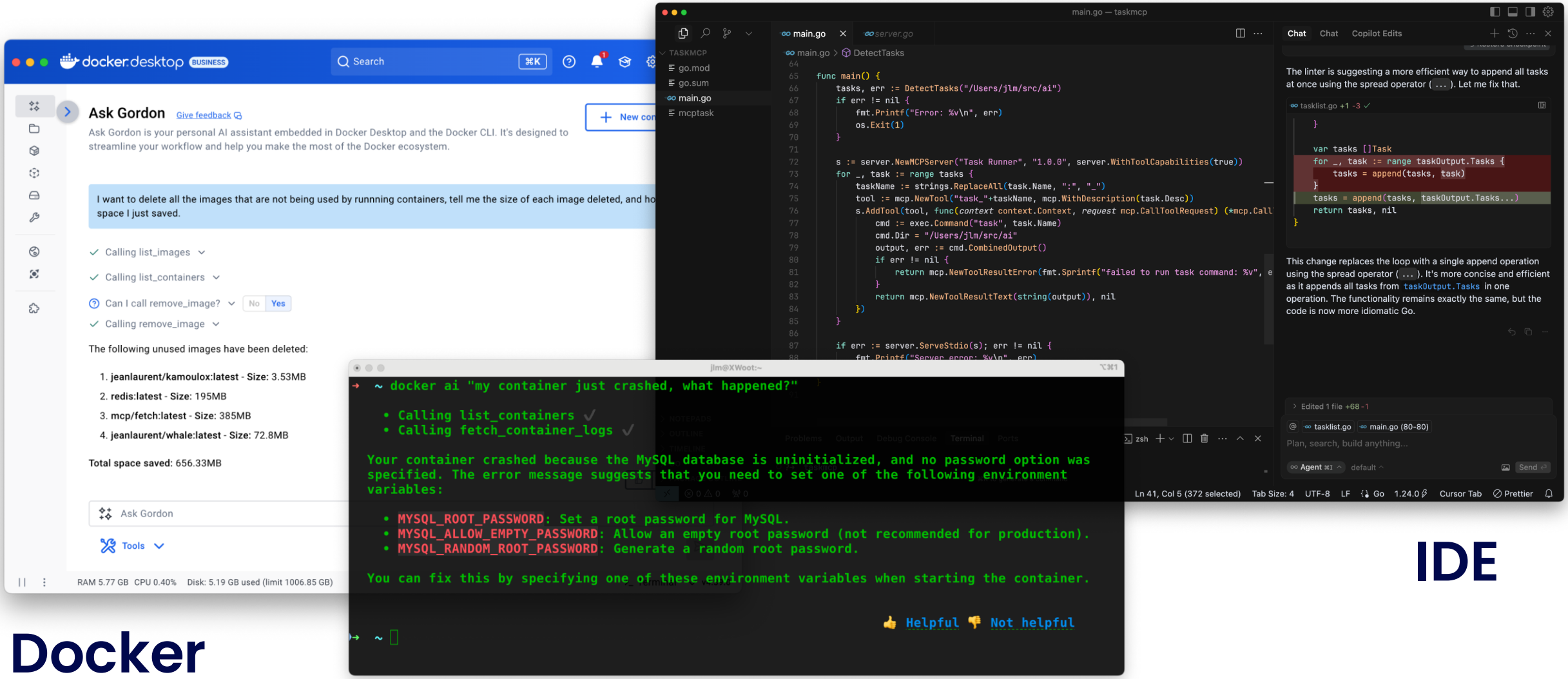
Specialized AI Agents are more useful

Access to usage data

Expert knowledge

Continuous innovation

Drive AI Agent's adoption by meeting devs where they are



IDE

Docker

Terminal

Demo 3 : Docker AI Agent (Gordon)

Get started with Docker GenAI Tooling today

Explore Docker Model Runner

Learn more about Docker MCP Catalog and Toolkit

Try the Docker AI Agent (Gordon)