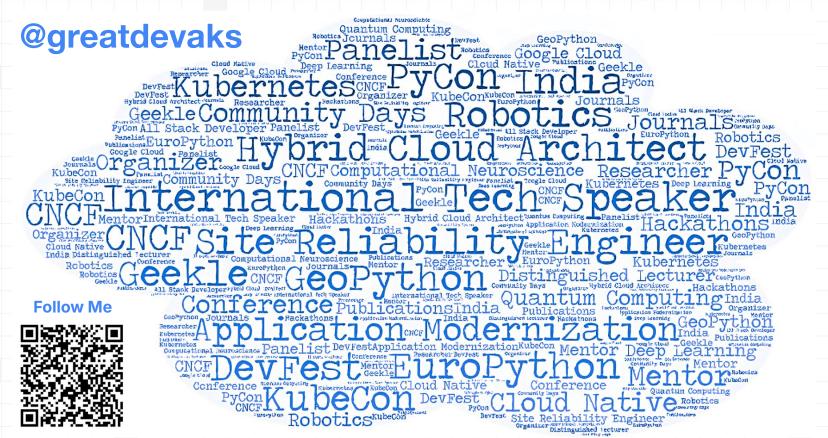


Serverless Magic: Bringing Al Agents to Life with MCP and Cloud Run

Anmol Krishan Sachdeva Sr. Hybrid Cloud Architect, Google



cloud community days 0 2025 **



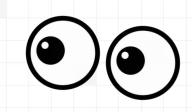
Serverless Magic: Bringing Al Agents to Life with MCP and Cloud Run | Google Cloud Community Day Jaipur 2025

Anmol Krishan Sachdeva (@greatdevaks) | Sr. Hybrid Cloud Architect, Google



THE CONTENT AND VIEWS PRESENTED DURING THE SESSION ARE AUTHOR'S OWN AND NOT OF ANY ORGANIZATIONS THEY ARE ASSOCIATED WITH.

SOME IMAGES IN THIS PRESENTATION WERE GENERATED WITH THE ASSISTANCE OF ARTIFICIAL INTELLIGENCE. SUCH ILLUSTRATIVE REPRESENTATIONS MAY NOT CONVEY ACCURATE OR FACTUALLY CORRECT INFORMATION.





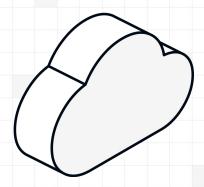
Agenda

- 1 Drawing Parallels
- (2) (MCP Architecture)
- (3) (Key Terminologies and Processes)
- (4) (Hands-On
- (5) Conclusion and Key Takeaways



Chapter One

Drawing Parallels



Hi XYZ, I got to know you run a startup...what all resources does one need to run a startup?







Oh, great...
How do you connect with all these different service providers?

Actually, each service provider has a different on-boarding process...it is a very tedious job.



Hmm...
Is just on-boarding a service provider enough?

No! On-boarding a service provider is of no use unless one is able to continuously coordinate with them.



I understand...
How do you get visibility into status of each service provider?

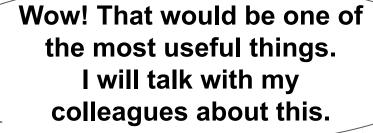
Well, there is no real visibility as of now.

My team has other priority tasks to fulfill.



I wonder if you were to get a centralized dashboard for on-boarding and operating the service providers...







Since you are taking interest, just having a dashboard is not enough...
No compromise on security.

I almost totally forgot about this important aspect.
Access controls and data handling are important.





Imagine you got this central platform, will that be all?

We will aim to design the platform as a unified facade with integrations and flexibility in mind:)





Startup Founder Scenario	LLM Scenario	Terminology / Area
Services needed to run a startup: HR, Tech., Legal,	Resources needed to run effectively: Data Schemas,	Resources and Tools
Finance, Marketing, Operations, etc.	Data Records, Files, APIs, additional context, etc.	[Provided by MCP Servers]
Contracts with Service	Integrations with data	Custom Integrations
Providers	sources and tools	[Pre-MCP Scenario]
Service Provider Integration	API Contract and Data	Developer Overhead
Standardization	Schema Standardization	[Pre-MCP Scenario]
Service Provider	Message Exchange	Protocols and Transports
Communication	between components	[MCP Architecture]
Centralized Dashboard	Universal Adapter	Protocols and Architecture

Serverless Magic: Bringing Al Agents to Life with MCP and Cloud Run | Google Cloud Community Day Jaipur 2025

Anmol Krishan Sachdeva (@greatdevaks) | Sr. Hybrid Cloud Architect, Google

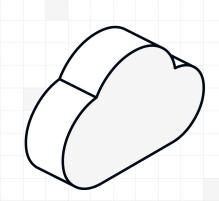


Startup Founder Scenario	LLM Scenario	Terminology / Area
Top-Down / Bottom-Up Communication	Unidirectional / Bidirectional Message Streams	Messages (Bidirectional) Notifications (Unidirectional)
Requirements for services by Startup	Need for accessing resources, tools, context, etc.	MCP Server capabilities
Requirements from Service Providers	Need for getting user inputs or non-sensitive privileged data	MCP Client capabilities
Need for Strategic Guidance from Founder	LLM requiring additional context without burdening the MCP Server	Sampling



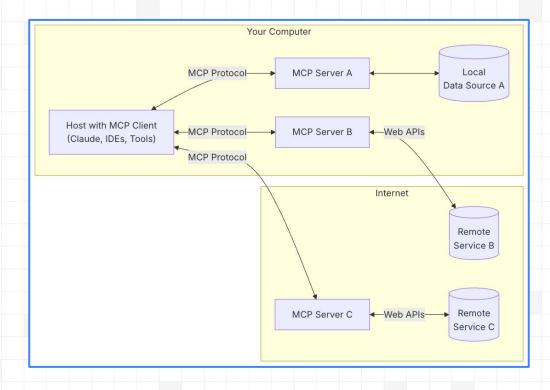
Chapter Two

MCP Architecture





The Client-Server Architecture



Source: Official Documentation for MCP



Chapter Three

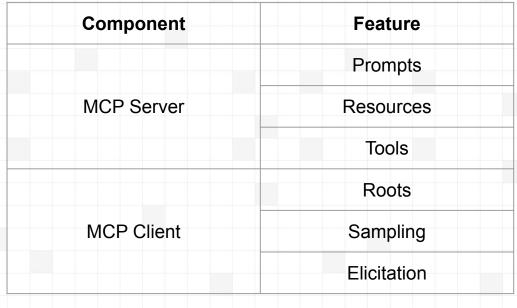
Key Terminologies and Processes

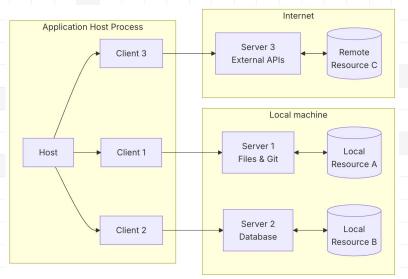




Key Terminologies and Processes

JSON-RPC 2.0 (Protocol)





Source: Official Documentation for MCP



Resources, Prompts, and Tools

Concept	Definition
	Read-only information; data, context, API response, files, etc.
Resources	Application-Controlled
Prompts	Instructions given to an LLM; better to templatize and leverage
	User-Controlled
Tools	An executable function or service that performs some action - like calculator, code generator, etc.
	Model-Controlled



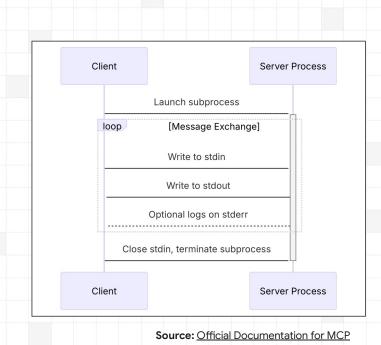
Roots, Sampling, and Elicitation

Concept	Definition	
Roots	Helps define boundaries or scope that MCP Servers should respect; hard enforcement is not guaranteed	
Sampling	Allows MCP Servers to request LLM-assisted completions from MCP Clients	
	MCP Servers shouldn't store LLM Keys/Credentials; MCP Clients can	
Elicitation	Enables interactive, user-in-the-loop flows	
	MCP Servers can request additional inputs from users	



Transports

Transport	Use-Case	
Stdio (Standard Input / Output)	Local CLI Tools Testing Embedding	
Server-Sent Events (SSE) [Deprecated - Standalone]	Real-Time Streaming over HTTP (legacy MCP)	
Streamable HTTP [With Optional SSE; JSON responses by default]	Modern default for remote servers, large payloads, or streaming use-cases	
Custom	Special requirements - WebSockets, gRPC, etc.	



Serverless Magic: Bringing Al Agents to Life with MCP and Cloud Run | Google Cloud Community Day Jaipur 2025



Chapter Four

Hands-On

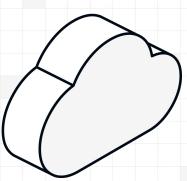
tinyurl.com/ccdj-2025





Chapter Five

Conclusion and Key Takeaways



Noteworthy Points

- 1. MCP Unlocks Module LLM Agent Workflows
- 2. Hybrid Orchestration (Local + Remote) is powerful and easy
- 3. Use Sampling for LLM-in-the-loop kind of tools; track cost, and use token limits and stop seq.
 - 4. Use Elicitation for exercising user-in-the-loop pattern
 - 5. Opt for Serverless for scale; both MCP Clients and MCP Servers can run on Serverless
 - 6. Serverless pattern can turn out to be cost efficient
 - 7. Leverage MCP Inspector for quick debugging
 - 8. Security, sanitization, and governance are critical; don't forget enforcing them
 - 9. Access Controls, Pagination, Auth*, and Rate Limiting patterns should be exercised
- 10. MCP on Embedded / IOT Devices and Shared Web Hosting (cPanel) is a BIG NO (as of today)







Anmol Krishan Sachdeva Sr. Hybrid Cloud Architect, Google

#Agents #LLM #MCP #CCDJaipur #Serverless #GoogleCloud \ @greatdevaks



Follow Me



cloud community days 0 2025 **