

Planning for Conversational Agents



Planning for Conversational Agents



Objectives

- 01 List the prerequisite platform considerations needed to design Conversational Agents.
- 02 Explain the key considerations to account for when designing Conversational Agents



Objectives

- 01 List the prerequisite platform considerations needed to design Conversational Agents.
- 02 Explain the key considerations to account for when designing Conversational Agents



Objectives

- 01 List the prerequisite platform considerations needed to design Conversational Agents.
- 02 Explain the key considerations to account for when designing Conversational Agents



Architecture building Conversational Agents (1 of 2)

During the discovery and planning phase, ask the following key questions.

- What's the scope? One agent or multiple agents?
- What are the data residency, availability, and disaster recovery requirements?
- What languages will your agent support? Which time zones?
- What Networking and Security (Authentication & authorization) architecture will be implemented?
- What are considerations for Data Privacy (collection and storage)?



Architecture building Conversational Agents (1 of 2)

During the discovery and planning phase, ask the following key questions.

- What's the scope? One agent or multiple agents?
- What are the data residency, availability, and disaster recovery requirements?
- What languages will your agent support? Which time zones?
- What Networking and Security (Authentication & authorization) architecture will be implemented?
- What are considerations for Data Privacy (collection and storage)?



Architecture building Conversational Agents (1 of 2)

During the discovery and planning phase, ask the following key questions.

- What's the scope? One agent or multiple agents?
- What are the data residency, availability, and disaster recovery requirements?
- What languages will your agent support? Which time zones?
- What Networking and Security (Authentication & authorization) architecture will be implemented?
- What are considerations for Data Privacy (collection and storage)?



Architecture building Conversational Agents (1 of 2)

During the discovery and planning phase, ask the following key questions.

- What's the scope? One agent or multiple agents?
- What are the data residency, availability, and disaster recovery requirements?
- What languages will your agent support? Which time zones?
- What Networking and Security (Authentication & authorization) architecture will be implemented?
- What are considerations for Data Privacy (collection and storage)?



Architecture building Conversational Agents (2 of 2)

During the discovery and planning phase, ask the following key questions.

- What channels will be supported?
- How will each channel be designed and supported?
- Why are session and context management for Conversational Agents and CES needed?
- How do you design interactions? How do you mix predefining flows with Generative AI such as data store agents and generative agents?

Note that these are advanced topics will be covered in detail in later modules of the CES curriculum.



Architecture building Conversational Agents (2 of 2)

During the discovery and planning phase, ask the following key questions.

- What channels will be supported?
- How will each channel be designed and supported?
- Why are session and context management for Conversational Agents and CES needed?
- How do you design interactions? How do you mix predictive flows with Generative AI such as data store agents and generative agents?

Note that these are advanced topics will be covered in detail in later modules of the CES curriculum.



Architecture building Conversational Agents (2 of 2)

During the discovery and planning phase, ask the following key questions.

- What channels will be supported?
- How will each channel be designed and supported?
- Why are session and context management for Conversational Agents and CES needed?
- How do you design interactions? How do you mix predefining flows with Generative AI such as data store agents and generative agents?

Note that these are advanced topics will be covered in detail in later modules of the CES curriculum.



Architecture building Conversational Agents (2 of 2)

During the discovery and planning phase, ask the following key questions.

- What channels will be supported?
- How will each channel be designed and supported?
- Why are session and context management for Conversational Agents and CES needed?
- How do you design interactions? How do you mix predictive flows with Generative AI such as data store agents and generative agents?

Note that these are advanced topics will be covered in detail in later modules of the CES curriculum.



Architecture building Conversational Agents (2 of 2)

During the discovery and planning phase, ask the following key questions.

- What channels will be supported?
- How will each channel be designed and supported?
- Why are session and context management for Conversational Agents and CES needed?
- How do you design interactions? How do you mix predefining flows with Generative AI such as data store agents and generative agents?

Note that these are advanced topics will be covered in detail in later modules of the CES curriculum.



Conversational Agents availability

There are two methods for achieving high service availability.

Consider these two approaches for highly available Conversational Agents:

- [1] Multiple load-balanced regional agents.
- [2] A single global agent.



Conversational Agents availability

There are two methods for achieving high service availability.

Consider these two approaches for highly available Conversational Agents:

- [1] Multiple load-balanced regional agents.
- [2] A single global agent.



Conversational Agents availability

There are two methods for achieving high service availability.

Consider these two approaches for highly available Conversational Agents:

- [1] Multiple load-balanced regional agents.
- [2] A single global agent.



Conversational Agents availability

There are two methods for achieving high service availability.

Consider these two approaches for highly available Conversational Agents:

- [1] Multiple load-balanced regional agents.
- [2] A single global agent.



Multiple regional agents (1 of 2)

Use a load balancer for a single entry point to multiple regional agents.

Scenario:

You use multiple regional agents to provide availability.

A regional failure occurs causing the underlying platform to failover to a second region. Conversational Agents API endpoint.



Multiple regional agents (1 of 2)

Use a load balancer for a single entry point to multiple regional agents.

Scenario:

You use multiple regional agents to provide availability.

A regional failure occurs causing the underlying platform to failover to a second region. Conversational Agents API endpoint.



Multiple regional agents (2 of 2)

Use a load balancer for a single entry point to multiple regional agents.

Infrastructure complexity

Availability is achieved via an L7 proxy that manages health checks and routes requests to the nearest available region.

Operational complexity

Multiple regional agents are deployed independently and have unique agent IDs. You need to keep your agent configurations in sync.



Multiple regional agents (2 of 2)

Use a load balancer for a single entry point to multiple regional agents.

Infrastructure complexity

Availability is achieved via an L7 proxy that manages health checks and routes requests to the nearest available region.

Operational complexity

Multiple regional agents are deployed independently and have unique agent IDs. You need to keep your agent configurations in sync.



Multiple regional agents (2 of 2)

Use a load balancer for a single entry point to multiple regional agents.

Infrastructure complexity

Availability is achieved via an L7 proxy that manages health checks and routes requests to the nearest available region.

Operational complexity

Multiple regional agents are deployed independently and have unique agent IDs. You need to keep your agent configurations in sync.



Multiple regional agents (2 of 2)

Use a load balancer for a single entry point to multiple regional agents.

Infrastructure complexity

Availability is achieved via an L7 proxy that manages health checks and routes requests to the nearest available region.

Operational complexity

Multiple regional agents are deployed independently and have unique agent IDs. You need to keep your agent configurations in sync.



Multiple regional agents (2 of 2)

Use a load balancer for a single entry point to multiple regional agents.

Infrastructure complexity

Availability is achieved via an L7 proxy that manages health checks and routes requests to the nearest available region.

Operational complexity

Multiple regional agents are deployed independently and have unique agent IDs. You need to keep your agent configurations in sync.



Multiple regional agents (2 of 2)

Use a load balancer for a single entry point to multiple regional agents.

Infrastructure complexity

Availability is achieved via an L7 proxy that manages health checks and routes requests to the nearest available region.

Operational complexity

Multiple regional agents are deployed independently and have unique agent IDs. You need to keep your agent configurations in sync.



One global Conversational Agent

User agents to provide availability for your Conversational Agent.

Our recommended option

Regardless of which Google Cloud AI endpoint is used, the request will be routed to the nearest available region.

Your Conversational Agent will respond using the global Agent ID.



One global Conversational Agent

User agents to provide availability for your Conversational Agent.

Our recommended option

Regardless of which Google Cloud AI endpoint is used, the request will be routed to the nearest available region.

Your Conversational Agent will respond using the global Agent ID.



One global Conversational Agent

User agents to provide availability for your Conversational Agent.

Our recommended option

Regardless of which Google Cloud AI endpoint is used, the request will be routed to the nearest available region.

Your Conversational Agent will respond using the global Agent ID.



Agent languages

Which languages do your fulfillments need to support?

Understand which languages your customers speak and how many languages your agent needs to support.



Agent languages

Which languages do your fulfillment need to support?

Understand which languages your customers speak and how many languages your agent needs to support.



Agent languages

Which languages do your fulfillments need to support?

Understand which languages your customers speak and how many languages your agent needs to support.



Agent channels

Which presentation-layer capabilities do your Fulfillments need to support?

There are many different channels customers will use to access your CTS solution:

- Chat from web
- Chat from mobile app
- Chat from WhatsApp
- Voice from Skype
- Voice from Twilio



Agent channels

Which presentation-layer capabilities do your Fulfillments need to support?

There are many different channels customers will use to access your CTS solution:

- Chat from web
- Chat from mobile app
- Chat from WhatsApp
- Voice from Skype
- Voice from Twilio



Agent channels

Which presentation-layer capabilities do your Fulfillments need to support?

There are many different channels customers will use to access your CTS solution:

- Chat from web
- Chat from mobile app
- Chat from SMS/WhatsApp
- Voice from Agents
- Voice from IVR



Agent channels

Which presentation-layer capabilities do your Fulfillments need to support?

There are many different channels customers will use to access your CTS solution:

- Chat from web
- Chat from mobile app
- Chat from WhatsApp
- Voice from Skype
- Voice from Twilio



Omni channel support and design (1 of 2)

Omni channel lets your contact center design use more than one channel at a time.

- Determine how to route each channel, and optimize their responses.
- Determine web-chat framework and webhooks integrations.
- Plan telephony integration, along with IVR & TTS configurations.
 - Other common attributes: Conversation profile, session ID, turn participant, call timestamp and direction of phone calls
- Integrate backend system with other integrations like CRM and ticketing.



Omni channel support and design (1 of 2)

Omni channel lets your contact center design use more than one channel at a time.

- Determine how to route each channel, and optimize their responses.
- Determine web-chat framework and webhooks integrations.
- Plan telephony integration, along with IVR & TTS configurations.
 - Other common attributes: Conversation profile, session ID, turn participant, call timestamp and direction of phone calls
- Integrate backend system with other integrations like CRM and ticketing.



Omni channel support and design (1 of 2)

Omni channel lets your contact center design use more than one channel at a time.

- Determine how to route each channel, and optimize their responses.
- Determine web-chat framework and webhooks integrations.
- Plan telephony integration, along with IVR & TTS configurations.
 - Other common attributes: Conversation profile, session ID, turn, participant, call timestamp and direction of phone calls
- Integrate backend system with other integrations like CRM and ticketing.



Omni channel support and design (1 of 2)

Omni channel lets your contact center design use more than one channel at a time.

- Determine how to route each channel, and optimize their responses.
- Determine web-chat framework and webhooks integrations.
- Plan telephony integration, along with IVR & TTS configurations.
 - Other common attributes: Conversation profile, session ID, turn participant, call timestamp and direction of phone calls
- Integrate backend system with other integrations like CRM and ticketing.



Omni channel support and design (1 of 2)

Omni channel lets your contact center design use more than one channel at a time.

- Determine how to route each channel, and optimize their responses.
- Determine web-chat framework and webhooks integrations.
- Plan telephony integration, along with IVR & TTS configurations.
 - Other common attributes: Conversation profile, session ID, turn, participant, call timestamp and direction of phone calls
- Integrate backend system with other integrations like CRM and ticketing.



Omni channel support and design (2 of 2)

Omni channel lets your contact center design use more than one channel at a time.

If supporting multiple channels, consider a mechanism to manage user data and conversational history in a central location to facilitate a consistent omnichannel experience.



Omni channel support and design (2 of 2)

Omni channel lets your contact center design use more than one channel at a time.

If supporting multiple channels, consider a mechanism to manage user data and conversational history in a central location to facilitate a consistent omni-channel experience.



User input (1 of 2)

How will user input be validated?

- Consider how you'll validate user input.
- Will you need to use user input subject to compliance, like PCI or PII data?
- There are many examples of sensitive client data, like contact, healthcare and financial information.



User input (1 of 2)

How will user input be validated?

- Consider how you'll validate user input.
- Will you need to use user input subject to compliance, like PCI or PII data?
- There are many examples of sensitive client data, like contact, healthcare and financial information.



User input (2 of 2)

How will user input be validated?

Use tools like Sensitive Data Protection and the Cloud DLP API to guard sensitive data.

Understand how this data interacts with other APIs and if it requires additional encryption.



User input (2 of 2)

How will user input be validated?

Use tools like Sensitive Data Protection and the Cloud DLP API to guard sensitive data.

Understand how this data interacts with other APIs and if it requires additional encryption.



User input (2 of 2)

How will user input be validated?

Use tools like Sensitive Data Protection and the Cloud DLP API to guard sensitive data.

Understand how this data interacts with other APIs and if it requires additional encryption.

