Reference Architecture
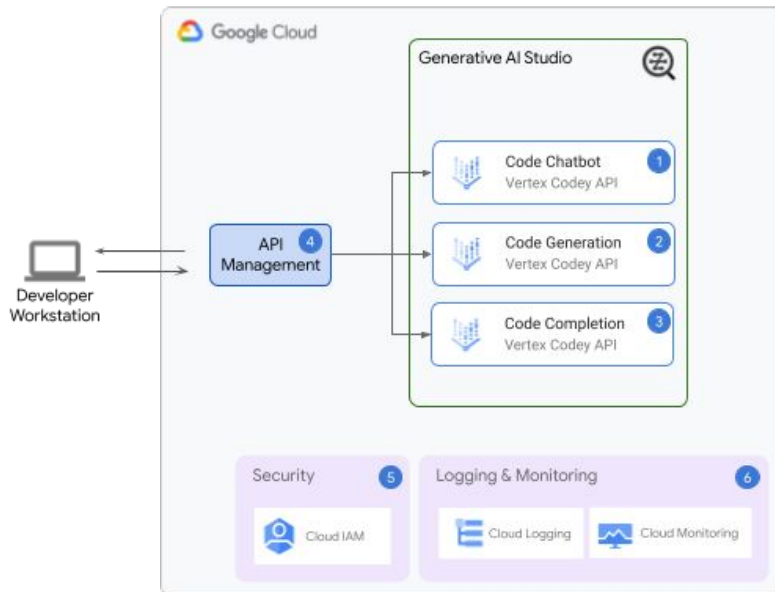# Developer Efficiency

This reference architecture shows how Generative AI Studio and Codey APIs can be used to offer assistance to developers.

Use case: Developers spend a lot of time writing code that is often repeatable. To save time they search for common solutions in various web forums and blogs. Once they find the right solution, it gets copied into their base code and adapted to fit the specific feature they are developing. Additionally, significant development time is spent on documentation and testing. Google Cloud's Codey APIs can boost developers' productivity by providing code completion, code generation including unit tests, documentation, code explainability and chat functionality for AI pair programming. Writing code faster and better can allow developers to do more and spend time on the most critical parts of the development cycle.

- Provide me with quick sort algorithm function in python
- Add Docstring to the following function
- Write unit tests for this python class
- Which type of data structure should I use for this problem and why?

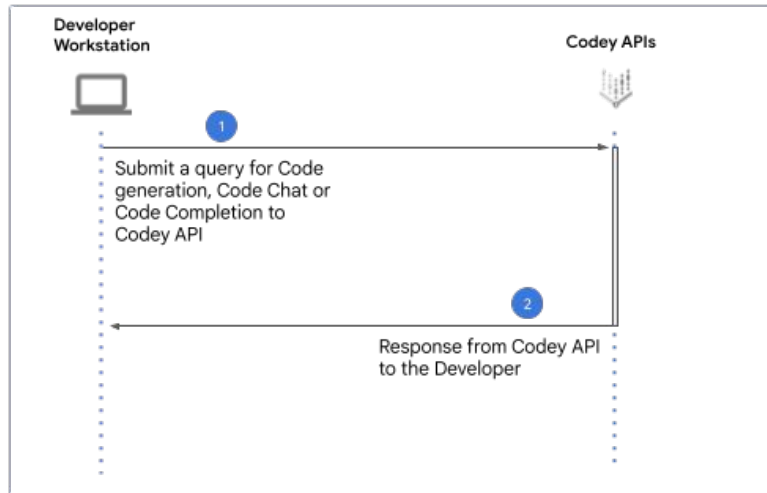### Components Description

1. **Code Chatbot** is a coding bot API that provides assistance with developer's coding questions.

2. **Code Generation** is an API that generates a block of code from natural language descriptions

3. **Code Completion** is an API that can provide developers the current and next lines of code as they develop, and in low-latency

4. **API Management** to build, manage, and secure API endpoints with tools such as Cloud Functions, Cloud Run, API Gateway, Apigee, etc.

5. **Security:** Leverage all the security and data residency features in Vertex AI

6. **Logging/Monitoring:** Use logging and monitoring features to understand data usage and resources consumption over time.

User Interaction Diagram
Developer Efficiency

This diagram shows the user interaction flow between components in the reference architecture using the Codey APIs

Use case: Developers spend a lot of time writing code that is often repeatable. To save time they search for common solutions in various web forums and blogs. Once they find the right solution, it gets copied into their base code and adapted to fit the specific feature they are developing. Additionally, significant development time is spent on documentation and testing. Google Cloud's Codey APIs can boost developers' productivity by providing code completion, code generation including unit tests, documentation, code explainability and chat functionality for AI pair programming. Writing code faster and better can allow developers to do more and spend time on the most critical parts of the development cycle.

Developer Workstation

Codey APIs

1   Submit a query for Code generation, Code Chat or Code Completion to Codey API

2   Response from Codey API to the Developer

User Interaction Flow

1   Developers from their workstation (using a chatbot interface or ide plugin) send requests to Codey API

2   Codey API replies back to the query with the appropriate payload (Code snippet / next lines of code / Answer to a coding question etc)