

GOOGLE CLOUD NEXT '24

OVERVIEW DEI PRINCIPALI ANNUNCI

23 Aprile 2024

Giorgia Fortuna, Giuseppe Di Carlo



MACHINE LEARNING REPLY

Machine Learning Reply dal 2018 sviluppa soluzioni di Machine Learning e AI con un offering completa che comprende: **data strategy consulting, innovation management consulting e AI solutions development.**



QUALCHE NUMERO

- **15** Certified Data/ML Engineers and Cloud Developers
- **10+** soluzioni di GenAI sviluppate nell'ultimo anno su GCP
- **8+** clienti con cui abbiamo lavorato su progetti Google nell'ultimo anno

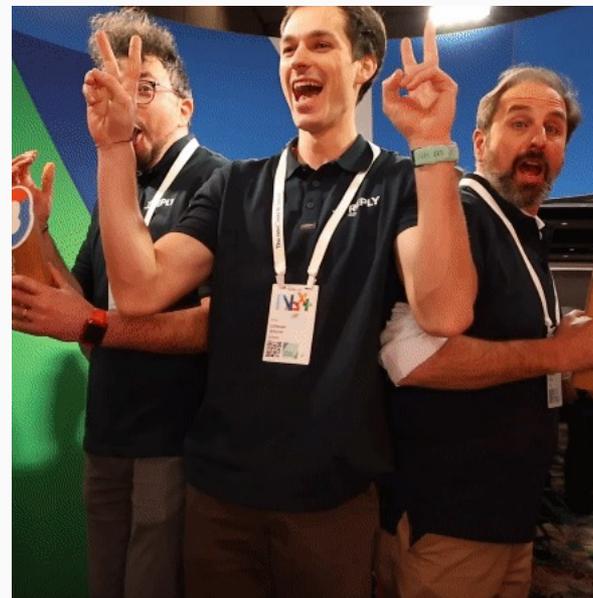
Machine Learning Reply dispone di Certified Data Engineer sulla Google Cloud Platform che hanno permesso all'azienda di acquisire la **Machine Learning Specialization**. Possiamo aiutarti a utilizzare i servizi AI e ML di Google Cloud per applicazioni di analisi dei dati, riconoscimento vocale e riconoscimento delle immagini



GO REPLY

Il partner di fiducia per abbracciare la trasformazione digitale con la potenza di **Google Cloud**. La nostra specializzazione si traduce in una vasta gamma di servizi all'avanguardia:

- **TRASFORMAZIONE CLOUD** - guidiamo le aziende nell'era digitale, migrando con successo i workload in cloud;
- **MODERNIZZAZIONE DELLE APPLICAZIONI** - ottimizziamo le applicazioni esistenti per prestazioni migliori e maggiore agilità;
- **CLOUD DATA & AI** - sfruttiamo il potere dei dati e dell'intelligenza artificiale per generare insights strategici;
- **MACHINE LEARNING E GENERATIVE AI** - portiamo l'AI al cuore delle processi aziendali per migliorare l'efficienza e stimolare l'innovazione.



195 GOOGLE CLOUD CERTIFICATIONS

57 Professional Cloud Architects	16 Professional Cloud Security Engineers
48 Professional Data Engineers	13 Professional ML Engineers
26 Professional Cloud Developers	13 Professional Network Engineer
19 Professional DevOps Engineer	3 Professional Cloud Database Engineers



GOOGLE CLOUD NEXT

Las Vegas, 9 -11 Aprile 2024



Generative AI is built on a unified technology stack

Gemini for Google Cloud
Gemini for Workspace

Vertex AI

Gemini Models

AI-optimized infrastructure



GEMINI



GEMINI MODELS

Google Next & Gemini



Multimodalità



Modelli open source

La gran parte delle news che si sono discusse riguardano l'apertura in GA di modelli che permettono l'utilizzo di un **context più esteso** e abilitano lo sviluppo di **nuove applicazioni**, rilasciate esse stesse o come prodotto embeddato nella GCP o come soluzione utilizzabile stand-alone

Gemini

Rilasciato il modello **Gemini 1.5 Pro**, un modello che prende in input fino a 1 milione di token, equiparabile 1 ora di video e 11 ore di audio.

Multimodal

Large context



Rilasciato in General Availability una famiglia di modelli **open source** più leggeri e disponibili in Vertex AI Model Garden

Small

Open Source

Gemini Code Assist: Assistente da aggiungere nell'IDE che supporta nella scrittura del codice prendendo in input tutta la codebase

CodeGemma: versione di Gemma fine-tuned per code generation e assistenza


Gemini 1.0 Pro

The best performing Gemini model with features for a wide range of tasks


Gemini 1.5 Pro

Created from the ground up to be multimodal (text, images, videos) and to scale across a wide range of tasks


Gemini 1.0 Pro Vision

Created from the ground up to be multimodal (text, images, videos) and to scale across a wide range of tasks


Gemma

Open checkpoint variants of Google Deepmind's Gemini model suited for a variety of text generation tasks


CodeGemma

Open code model variants of Gemma model suited for text-to-text and text-to-code tasks.

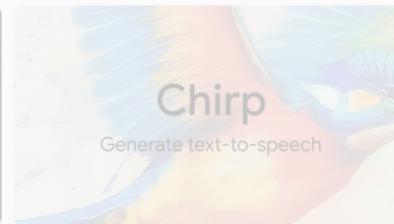


VERTEX AI





Più genericamente, Vertex AI è la piattaforma di Google per lo sviluppo di soluzioni di AI. All'interno del Model Garden viene data la possibilità di scegliere fra una gamma molto ampia di modelli



01. Rilasciato in General Availability il modello Imagen 2. Rilasciata la feature di text-to-live video capability e nuove features di photo editing



Funzionalità di editing fotografico mediante Imagen 2 (generation, editing)

02. Reso disponibile nel Model Garden anche **Claude 3** (disponibile nella sezione relativa ai «modelli dei partner»), che può essere ora facilmente integrato

Resi inoltre disponibili su model garden i modelli open-source di Hugging Face che sono ora facilmente deployabili su Vertex AI



Hugging Face

 Claude 3 Opus (Preview) Anthropic's most powerful vision and text model, navigating complex tasks and open-ended scenarios with remarkable fluency	 Claude 3 Sonnet A balanced vision and text model with an ideal mix of intelligence and speed for enterprise workloads. Maximum utility at a	 Claude 3 Haiku Anthropic's fastest, most compact vision and text model for near-instant responses to simple queries, meant for seamless AI
--	---	--

Imagen 2.0 Text to Live Image

Closeups with dramatic lighting,
emphasizing sculpted cheekbones
and highlighting techniques



VERTEX AI SEARCH

Vertex AI Search è una piattaforma completamente gestita che consente di creare esperienze di ricerca basate su AI Generativa su dati strutturati e non strutturati

Sono state rilasciate alcune nuove funzionalità che vanno ad ampliare le possibilità di personalizzazione dello strumento core per quanto riguarda la gestione di fonti documentali

NEW

Document Processing API

Smart Processing per estrarre informazioni utili ad un LLM dai documenti

UPDATED

Vertex text embedding

Vector Search

NEW

Ranking API

Re-ranking dei documenti usando la rilevanza degli stessi rispetto alla domanda usando LLM

NEW

Grounded Generation API

Servizio per generare "well grounded" answers. L'API ritorna una risposta con le fonti/citazioni esplicitate

New: Si può anche accedere al grounding mediante l'impiego della Google Search



Esempi di applicazione Vertex AI Search e approccio a sistemi RAG



VERTEX AI SEARCH

Vertex AI Search è una piattaforma completamente gestita che consente di creare esperienze di ricerca basate su AI Generativa su dati strutturati e non strutturati

Sono state rilasciate alcune nuove funzionalità che vanno ad ampliare le possibilità di personalizzazione dello strumento core per quanto riguarda la gestione di fonti documentali

NEW

Document Processing API

UPDATED

Vertex text embedding

Vector Search

NEW

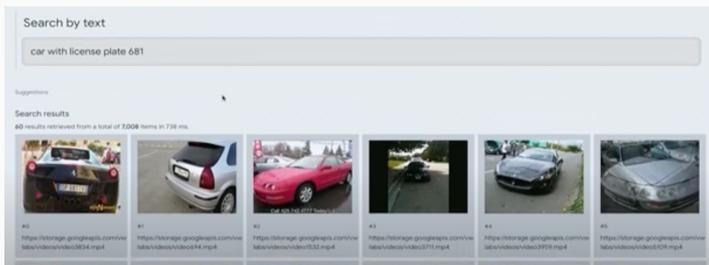
Ranking API

NEW

Grounded Generation API

Updated **multimodal embeddings** per confrontare facilmente testi e immagini

Nuove querying capabilities usando vector indexing direttamente sui DB Google Cloud



AlloyDB



Google BigQuery

e altri
...

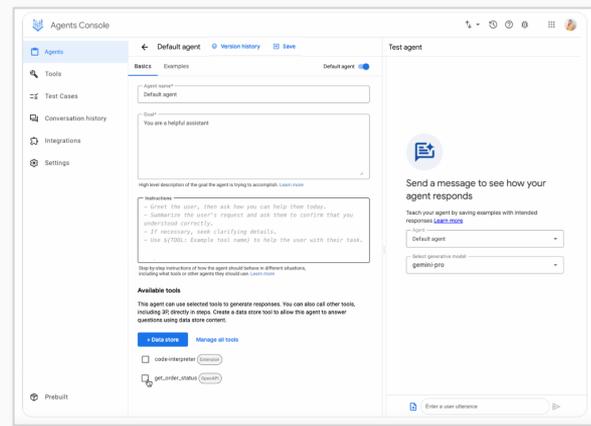


AI AGENTS & MULTI-AGENTS



Nel 2023 Google ha ampliato le capabilities di Dialogflow CX per portare le tecnologie generative a bordo di un chatbot classico tramite una serie di strumenti:

- Generative fallback
- Data Store
- Generators



Queste capabilities sono state integrate ed evolute in **Agent Builder**, in cui è possibile creare agenti **puramente generativi** caricando un datastore di riferimento e usando custom **function-calling**. Si ha un cambio di paradigma, verso un **insieme di agenti generativi** con capacità specifiche (**multi-agent**) rispetto ad un classico chatbot omnicomprensivo.



Estensione funzionalità generative su piattaforma Dialogflow CX



GEMINI FOR GOOGLE CLOUD

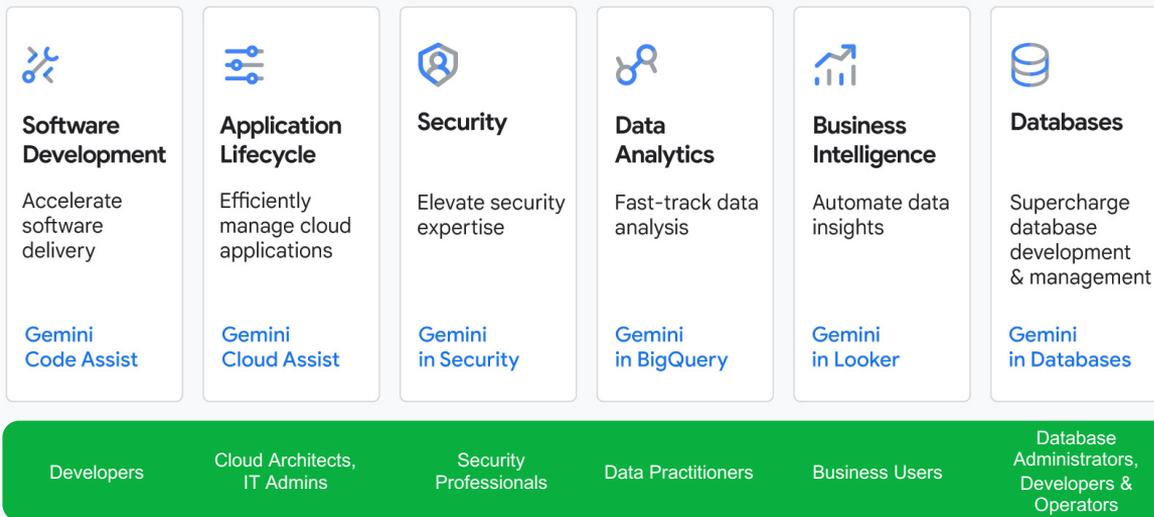


GEMINI FOR GOOGLE CLOUD

Introduzione

Gemini per Google Cloud è una gamma di assistenti AI che aiutano a semplificare le attività e a rendere i team più produttivi.

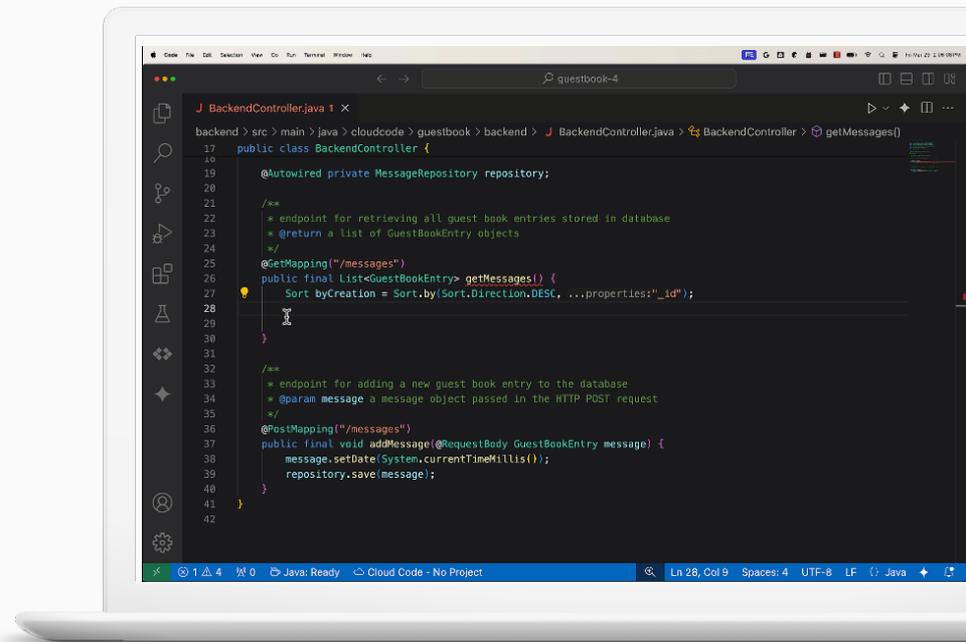
Questi agenti aiutano gli utenti a gestire i servizi di Google Cloud in modo più efficiente, a scrivere codice migliore e più sicuro, a comprendere più facilmente i dati e ad identificare e risolvere le minacce alla sicurezza informatica.



GEMINI CODE ASSIST

Accelerate software delivery with AI assistance

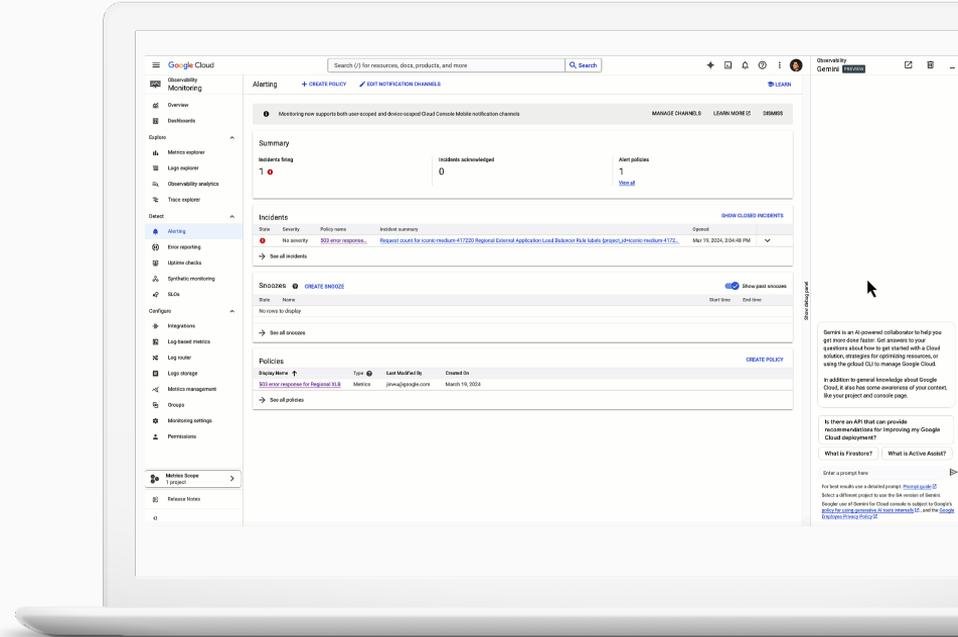
- Sviluppo e personalizzazione del codice con l'utilizzo dell' AI
- Comprensione del contest
- Enterprise security, privacy & IP protection
- Ecosistema partner in continua crescita



GEMINI CLOUD ASSIST

Efficiently manage cloud applications with AI assistance

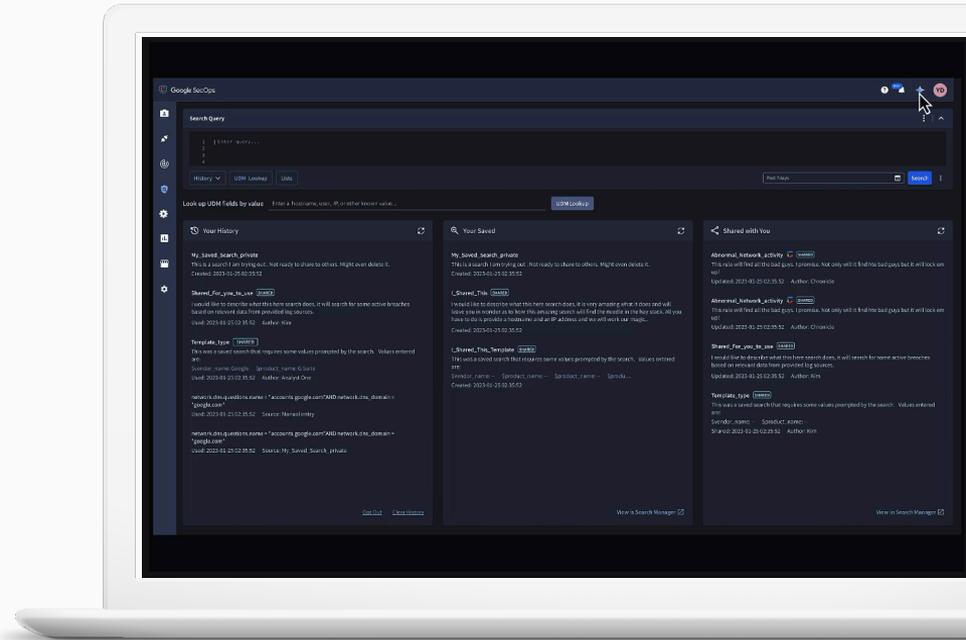
- Goal-driven Design
- Diagnostica e risoluzione problemi guidata
- Ottimizzazioni personalizzate



GEMINI IN SECURITY

Elevate security expertise

- Gemini in Security Operations
- Gemini in Security Command Center
- Gemini in Threat Intelligence



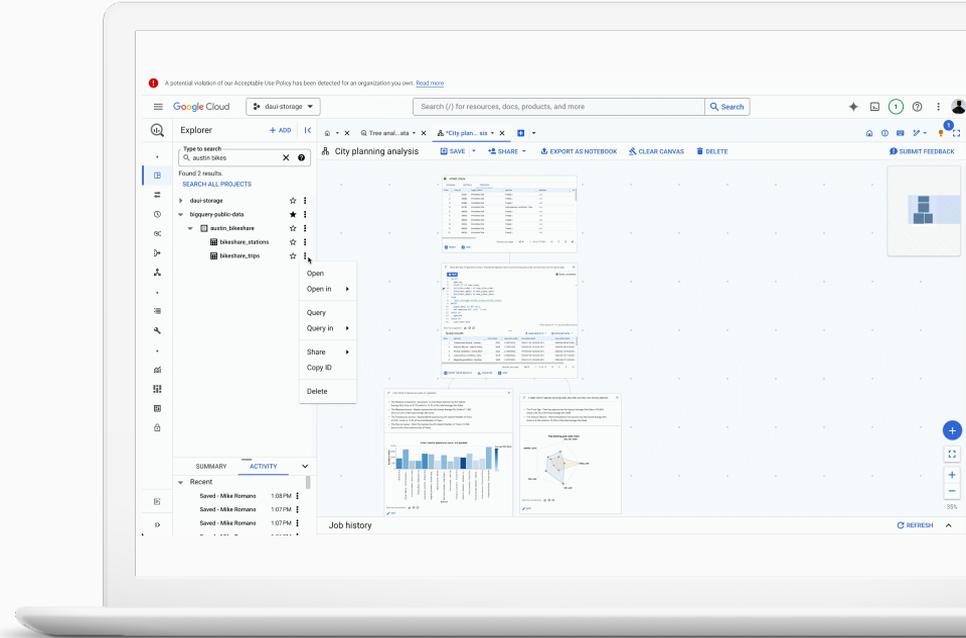
GEMINI IN BIGQUERY

Accelerate end to end data & analytics lifecycle with AI assistance

- Sviluppo pipeline
- Discovery e analisi
- Scrittura query

BigQuery Data Canvas

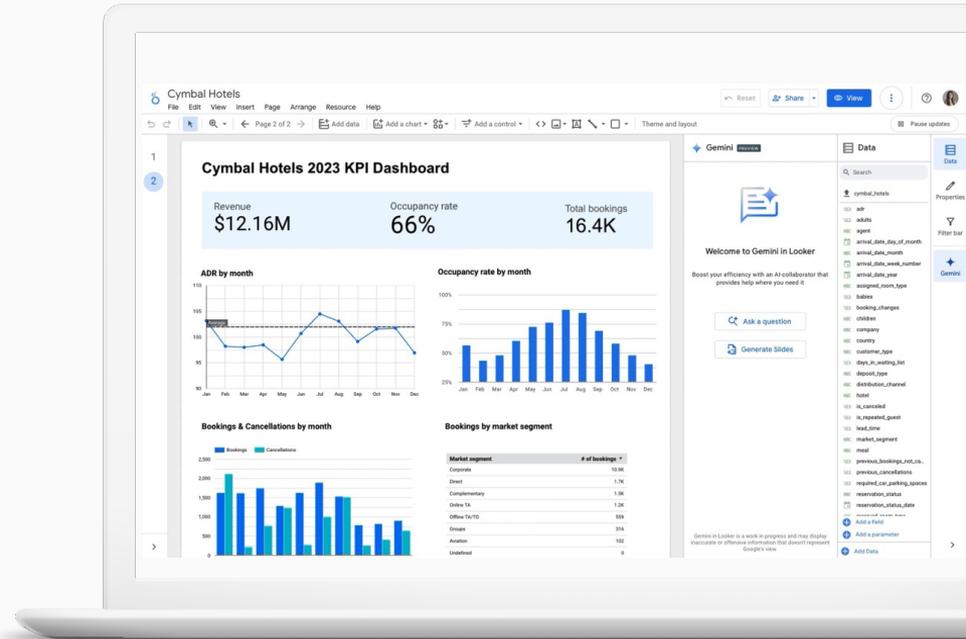
- Esplorazione e visualizzazione dei dati con experience GenAI-centric
- Incluso in BigQuery Studio
- Generazione automatica di notebook python



GEMINI IN LOOKER

Make analytics tasks and workflows easier with Prebuilt AI agents for BI

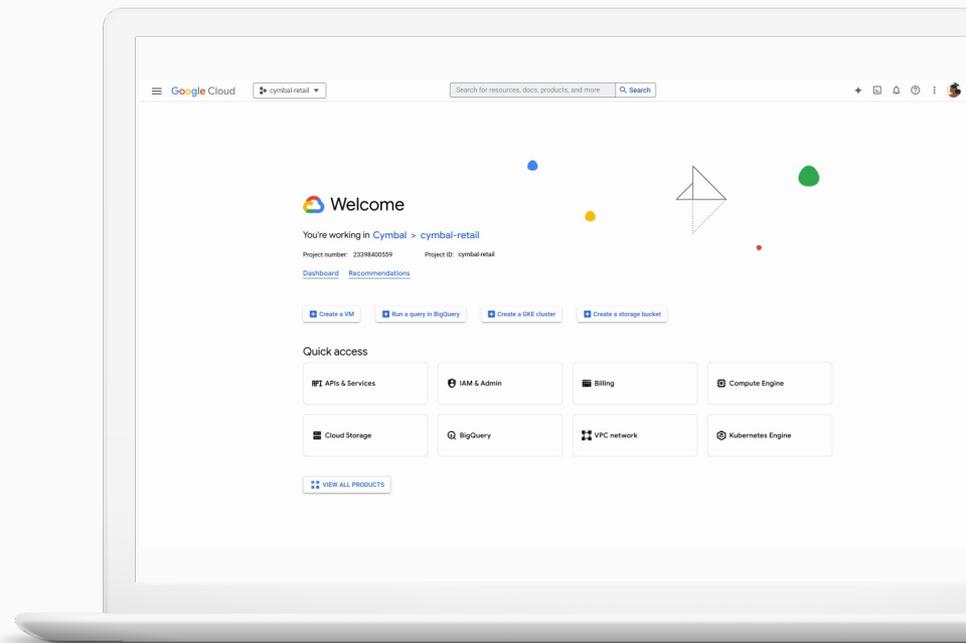
- Creazione e automazione di report e approfondimenti utilizzando il linguaggio naturale
- Assistenza con le formule e la scrittura di modelli di dati custom per la tipologia di business
- Creazione veloce con l'AI di viste, report e slides



GEMINI IN DATABASES

Supercharge database development and management with AI

- Generare, sintetizzare e correggere facilmente il codice SQL con l'ausilio dell'AI in Database Studio
- Ottimizzare il parco database da un unico pannello con Database Center
- Anticipare potenziali problemi con raccomandazioni intelligenti
- Accelerate le migrazioni di database con conversioni di codice assistite dall'intelligenza artificiale.



GEMINI FOR WORKSPACE



GOOGLE WORKSPACE AI ADD-ONS

Meeting customers where they are

Gemini Enterprise

Includes everything from Gemini Business, plus:

- Advanced meetings with translated captions in 15+ languages
- AI to automatically classify and label data to preserve confidentiality
- Full access to generative AI capabilities, no monthly usage limits

Gemini Business

AI-powered assistant for teams of all sizes:

- Gemini in Gmail, Docs, Slides, Sheets, and Meet
- Enterprise-grade data protection
- Access to Gemini with 1.0 Ultra, our most capable AI model widely available today
- Meets the needs of typical business users with monthly usage limits

[NEW] AI Security

AI-powered security and data protection:

- Automatically identify and classify sensitive files in Google Drive
- Uses privacy-preserving AI models trained uniquely for your organization
- Protect classified files with existing data loss prevention controls

[NEW] AI Meetings & Messaging

AI-enhanced video meetings and chats:

- Look and sound your best
- Generate custom backgrounds
- Advanced meetings with translated captions | 15+ languages
- Let Gemini take notes for you
- Google Chat summaries in the home view



MODERN INFRASTRUCTURE



AI HYPERCOMPUTER

Next generation AI supercomputing architecture for Cloud customers

Flexible Consumption

Dynamic Workload
Scheduler

On demand

CUD

Spot



**Dynamic resource and
cost optimization**

Open software



Multislice Training, Multihost Inference, XLA

Google Kubernetes Engine & Compute



**High flexibility, usability,
and optionality**

Performance-optimized hardware

Compute
(GPUs, TPUs)

Storage
(Block, File, Object)

Networking
(OCS, Jupiter)



**Exceptional performance
and efficiency**



RECAP

Più di 200 annunci

Google Cloud Next

Welcome to Google Cloud Next '24

April 9, 2024

Welcome to
Next '24

Google Cloud



<https://cloud.google.com/blog/topics/google-cloud-next/welcome-to-google-cloud-next24>

All 218 things we announced at Google Cloud Next '24 – a recap

April 13, 2024



<https://cloud.google.com/blog/topics/google-cloud-next/google-cloud-next-2024-wrap-up>



Scopri di più su ROSE

Puoi prenotare **meeting one-to-one** con gli esperti Reply per approfondire questo argomento e porre le tue domande.



Visita **ROSE (Reply Online Services)** per esplorare tutti i contenuti.



ASK THE EXPERT

Gemini Models: Q&A



ASK THE EXPERT

Gemini for Google Cloud and for
Workspace: Q&A

THANK YOU

