



XMASDEV

<oh>oh</oh>



XMAS DEV 2024



SOS LETTERINE! AZURE E L'AI SALVERANNO IL NATALE!

Nicola Paro

Platinum Sponsor



Gold Sponsor



Silver Sponsor



Technical Sponsor



Community





UNA STORIA VERA...

Era il primo giorno di Avvento, e i bambini di tutto il mondo si preparavano a scrivere le loro letterine a Babbo Natale.

Con pastelli colorati, disegnavano giocattoli, scrivevano desideri e aggiungevano piccoli cuori per mostrare il loro affetto.





UNA STORIA VERA...

Al Polo Nord, Babbo Natale e i suoi elfi si preparavano a ricevere una valanga di posta.

"Ci sono 2,2 miliardi di bambini nel mondo!" esclamò il capo degli elfi. "Dobbiamo essere super organizzati quest'anno!"

Nell'immenso magazzino delle letterine, gli elfi lavoravano senza sosta. Ogni giorno, arrivavano pacchi enormi pieni di buste colorate.





UNA STORIA VERA...

Ogni notte di Natale, infatti, Babbo Natale parte dal Polo Nord con un lotto specifico di letterine.

Nel suo sacco magico ci sono tutti i regali richiesti da quelle letterine, accuratamente impacchettati dagli elfi. Babbo Natale visita ogni casa della zona, lasciando i doni sotto gli alberi di Natale.

Una volta completato il giro per quel lotto, torna al Polo Nord per caricare il sacco con il lotto successivo: nuove letterine e nuovi regali.





UNA STORIA VERA...

Un giorno, l'elfo Pepito, il più giovane del team, era al volante di un muletto per la prima volta. "Attento a non correre troppo!" gli disse l'elfo anziano Pino.

Ma Pepito era così emozionato che premette l'acceleratore più del dovuto e... **SBAM!** Finì contro uno scaffale pieno di letterine!





UNA STORIA VERA...

Le letterine volarono in aria come fiocchi di neve e si sparpagliarono in ogni angolo del magazzino. Gli elfi si bloccarono, trattenendo il fiato. "Oh no! Non abbiamo tempo per riorganizzarle tutte!" esclamò Pino.





UNA STORIA VERA...

Babbo Natale, avisato dell'incidente, arrivò subito sul posto. Guardò la scena caotica e poi sorrise: "Non preoccupatevi, miei cari elfi. Possiamo risolvere tutto."

Dopo qualche istante di riflessione, ebbe un'idea; si accarezzò la barba e disse: "È il momento di chiamare... **XmasDev!**"





PIANO D'AZIONE

Gli elfi scattano
le foto alle
letterine

Tramite l'AI
estraiamo il
testo delle
letterine

Sempre tramite
AI, estraiamo i
giocattoli
richiesti dal
testo presente
nella letterina

Persistiamo
tutte queste
informazioni in
un database
SQL

Babbo Natale
accede a
queste info
tramite
un'applicazione
web



PROBLEMA NUMERO 1

Come ottengo la lista di giocattoli dal testo della letterina?

In maniera abbastanza facile?

Posso utilizzare un LLM!





COME DEPLOYARE UN LLM OPENAI SU AZURE?

Creo la risorsa su azure, poi posso accedere ai modelli tramite Foundry



The screenshot displays the Azure portal interface. On the left, the 'npopenaiserviceeus' resource is selected under the 'openai' resource group. The 'Overview' tab is active, showing details such as Status (Active), Location (East US), Subscription (Nicola Paro), and Subscription ID (99aa2800-5bc8-4cf7-917d-d98...). A red arrow points from the 'npopenaiserviceeus' resource to the 'Chat playground' interface in the background. The 'Chat playground' interface shows a 'Chat history' section and a 'Start with a sample prompt' section. The 'Azure OpenAI resource' sidebar on the right shows the 'npopenaiserviceeus' resource selected, with a dropdown menu for 'Resource' showing options like 'npopenaiserviceus', 'npopenaiserviceus', 'npopenaiserviceus', 'npopenaiserviceus', 'npopenaiserviceus', and 'npopenaiserviceus'. The 'Key 2' field is also visible.

Build your own secure copilot and generative AI



DOVE DEPLOYARE UN LLM OPENAI SU AZURE?

Region	o1-preview	o1-mini	gpt-4o e gpt-4o-mini	gpt-4o-realtime-preview	gpt-4 e gpt4-turbo
australiaeast	-	-	✓	-	✓
brazilsouth	-	-	✓	-	✓
canadaeast	-	-	✓	-	✓
eastus	✓	✓	✓	-	✓
eastus2	✓	✓	✓	✓	✓
francecentral	-	-	✓	-	✓
germanywestcentral	-	-	✓	-	✓
japaneast	-	-	✓	-	✓
koreacentral	-	-	✓	-	✓
northcentralus	✓	✓	✓	-	✓
norwayeast	-	-	✓	-	✓
polandcentral	-	-	✓	-	✓
southafricanorth	-	-	✓	-	✓
southcentralus	✓	✓	✓	-	✓
southindia	-	-	✓	-	✓
spaincentral	-	-	✓	-	✓
swedencentral	✓	✓	✓	✓	✓
switzerlandnorth	-	-	✓	-	✓
uaenorth	-	-	✓	-	✓
uksouth	-	-	✓	-	✓
westeurope	-	-	✓	-	✓
westus	✓	✓	✓	-	✓
westus3	✓	✓	✓	-	✓



COME INTERROGO UN LLM? – STRATEGIE DI PROMPTING

Prompting Diretto

- Indica chiaramente il compito o la domanda.
- Esempio: «Riassumi questo articolo in 2 frasi.»

Few-Shot Prompting

- Fornisci esempi per guidare il modello.
- Esempio: «Traduci: 'Hello' → 'Ciao'.»

Zero-Shot Prompting

- Nessun esempio, solo il compito.
- Esempio: «Scrivi un haiku sulla primavera»

Chain-of-Thought Prompting

- Chiedi un ragionamento passo per passo.
- Esempio: «Risolvi: $25 + 30$, passo per passo»

Role-Play Prompting

- Assegna un ruolo per fornire contesto.
- Esempio: «Sei un tecnico informatico. Spiega il protocollo HTTP»

Interactive Prompting

- Raffina il risultato con follow-up.
- Esempio: «Rendilo più breve»



COME UTILIZZO UN LLM IN .NET?

- **REST API Call verso Modelli Hostati Online**
- **LLamaSharp** - libreria che permette di eseguire modelli Llama localmente, senza necessità di connessione internet, garantendo maggiore controllo e privacy.
- **OpenAIClient o AzureOpenAI Client**
- **Semantic Kernel** - Framework per costruire applicazioni AI che combinano modelli di linguaggio con altre tecnologie semantiche, migliorando la comprensione e l'elaborazione del linguaggio naturale.
- **ChatGptNet: Libreria .NET per integrare ChatGPT nelle applicazioni.**
- **Microsoft.Extensions.AI** - Set di estensioni per .NET che facilita l'integrazione di servizi AI nelle applicazioni.



MICROSOFT.EXTENSIONS.AI

Libreria .NET per semplificare l'integrazione dei servizi AI nelle applicazioni C#.

- Caratteristiche Principali:
 - Unified API: Interfacce coerenti per diversi servizi AI.
 - Astrazioni Standardizzate: Facilita l'interoperabilità tra modelli di linguaggio.
 - Flessibilità: Passaggio semplice tra fornitori senza compromettere la compatibilità.
- Vantaggi:
 - Facilità d'Uso: Sperimentazione con soluzioni AI mantenendo un'API unica.
 - Componentizzazione: Aggiunta di nuove funzionalità senza complicazioni.
 - Supporto per Modelli di Linguaggio: Compatibile con SLM e LLM, inclusi OpenAI e Azure AI.



PROBLEMA NUMERO 2









Image To Text?

In maniera abbastanza facile?

- **Azure AI Vision (OCR):**
Utilizza tecniche di riconoscimento ottico dei caratteri (OCR) basate sull'apprendimento automatico per estrarre testo stampato o scritto a mano da immagini come poster, segnali stradali e etichette di prodotti.
- **Document Intelligence read model**
Come Azure AI Vision, ma è focalizzato sui documenti
- **Modelli LLM multimodali che supportino immagini**



SCELTA DEL MODELLO LLM IN AZURE OPENAI

	GPT-4	GPT-4o	GPT-4o-Mini
Supporto alle immagini	 Deployment con vision		
Costi	In: \$30/MT Out: \$60/MT	\$2.50/MT \$10/MT	\$0.15/MT \$0.60/MT
Qualita image to text		  	
		Scelto per image to text	Scelto per text extraction



QUANTO MI COSTA PROCESSARE UN'IMMAGINE IN TOKENS?

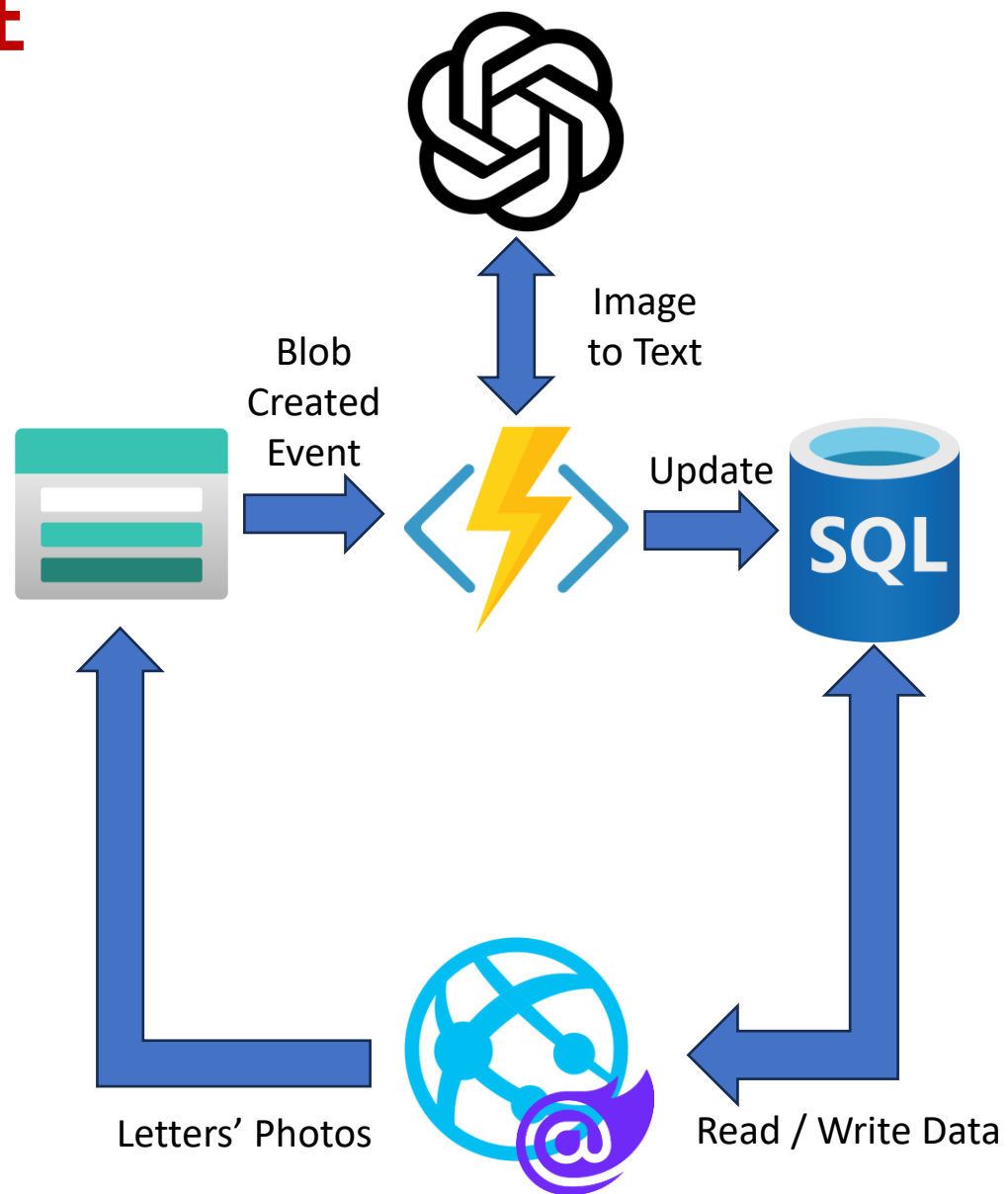
1. Ogni immagine è ridimensionata per adattarsi alla dimensione massima di 2048px, con un minimo di 768px sul lato più corto, mantenendo l'aspect ratio originale.
2. L'immagine ridimensionata viene suddivisa in tiles basati sulla dimensione della tessera del modello, pari a 512px x 512px.
3. Il numero totale di token viene calcolato moltiplicando il numero di tiles per il numero di tokens per tile (170) e aggiungendo 85 token di base:

$$\text{Tokens} = 85 + 170 \times \text{NTiles}$$

Questo tool calcola il costo in automatico →
<https://jamesmcraft.github.io/openai-image-token-calculator/>



PIANO D'AZIONE





PROBLEMI NUMERO 3 E 4

Tokens per minuto e tempi di elaborazione

Azure OpenAI mette un limite massimo di token al minuto per ciascun modello AI utilizzato.

Gli elfi sono numerosi e generano molte richieste contemporaneamente. Questa concorrenza può causare il rapido esaurimento del budget di token al minuto, rallentando o bloccando il sistema.

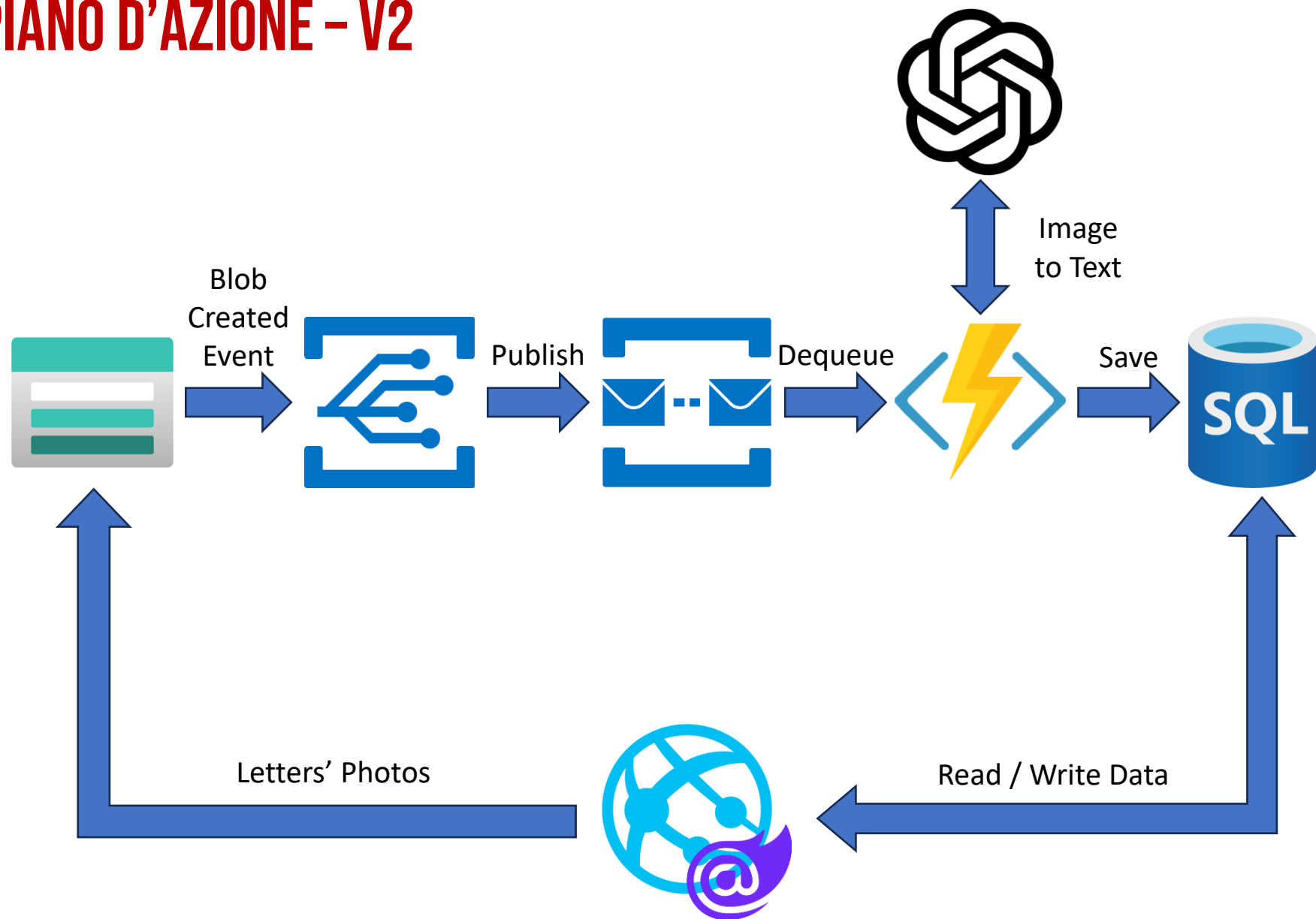
Per ottenere la risposta dall'LLM, servono circa 4-5 secondi tra Image to text e data extraction.

Soluzione:

Metto una coda e processo le richieste in sequenza



PIANO D'AZIONE - V2





PROBLEMA NUMERO 5

Le Azure Functions.

- Con .NET 9 è supportato solamente usando «isolated runtime»
- Su Visual Studio il template funzionava per le azure functions V1 con NetFramework.
- Tutte le classi per i vari bindings sono sotto al namespace **Microsoft.Azure.Functions.Worker** e non sotto a **Microsoft.Azure.WebJobs.Extensions**
- Il progetto ha l'output type **Exe**
- Microsoft.NET.Sdk.Functions **non** è da includere tra i nugets
- La function app ha un **Main()**
- *Se avete fatto dei custom bindings, sono da rivedere perché potrebbero non essere compatibili*



XMASDEV

<oh>oh</oh>



DEMO – LET'S BUILD IT!



UNA STORIA VERA...

Grazie alla nuova tecnologia, Babbo Natale ora non ha più bisogno di portare con sé l'intero pacco di letterine durante le consegne. Ogni notte di Natale, Babbo Natale apre l'applicazione e visualizza l'elenco dei regali da consegnare per ciascuna casa.

Con un sorriso, esclama: "Ora sì che posso concentrarmi su quello che è importante: portare gioia e magia in ogni casa!"

E così, con la combinazione perfetta di tradizione e tecnologia, il Natale diventò ancora più magico, dimostrando che anche Babbo Natale sa tenersi al passo con i tempi. **Fine.**





GRAZIE!

Nicola Paro

Solution Architect

beanTech



Let's Connect!



linkedin.com/in/nicolaparo



github.com/nicolaparo/xmasdev2024