



---

# ARGO

## Specifica Tecnica

*Gruppo Argo — Progetto ChatSQL*

### Informazioni sul documento

<b>Versione</b>	0.0.2
<b>Approvazione</b>	TODO
<b>Uso</b>	Esterno
<b>Distribuzione</b>	Prof. Tullio Vardanega Prof. Riccardo Cardin Gruppo Argo



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

## Registro delle modifiche

Ver.	Data	Redazione	Verifica	Descrizione
0.0.2	2024-07-25	Riccardo Cavalli	Tommaso Stocco, Mattia Zecchinato	Completata sezione scelte tecnologiche
0.0.1	2024-07-20	Riccardo Cavalli	Tommaso Stocco, Mattia Zecchinato	Prima stesura del documento

## Indice

<b>1</b>	<b>Introduzione</b>	<b>5</b>
1.1	Scopo del documento . . . . .	5
1.2	Riferimenti . . . . .	5
1.2.1	Riferimenti normativi . . . . .	5
1.2.2	Riferimenti informativi . . . . .	5
1.3	Glossario . . . . .	6
<b>2</b>	<b>Tecnologie utilizzate</b>	<b>7</b>
2.1	Backend . . . . .	7
2.1.1	FastAPI . . . . .	7
2.1.2	Python . . . . .	7
2.1.3	txtai . . . . .	8
2.1.4	SQLAlchemy . . . . .	8
2.1.5	JWT . . . . .	8
2.2	Frontend . . . . .	9
2.2.1	Vue.js . . . . .	9
2.2.2	Axios . . . . .	9
2.2.3	TypeScript . . . . .	10
2.2.4	PrimeVue . . . . .	10
2.3	Altri strumenti e tecnologie . . . . .	10
2.3.1	Docker . . . . .	10
2.3.2	SQLite . . . . .	11
2.3.3	Hugging Face . . . . .	11
2.3.4	ChatGPT . . . . .	12
2.3.5	LMSYS Chatbot Arena . . . . .	12
2.3.6	LM Studio . . . . .	12
2.4	Panoramica delle tecnologie . . . . .	13
2.4.1	Framework . . . . .	13
2.4.2	Linguaggi . . . . .	13
2.4.3	Librerie . . . . .	14
2.4.4	Strumenti . . . . .	15
<b>3</b>	<b>Architettura</b>	<b>17</b>

## **Elenco delle tabelle**

2.1	Framework utilizzati . . . . .	13
2.2	Linguaggi utilizzati . . . . .	13
2.3	Librerie utilizzate . . . . .	14
2.4	Strumenti utilizzati . . . . .	15

## **Elenco delle figure**

# 1 Introduzione

## 1.1 Scopo del documento

Il presente documento ha lo scopo di descrivere le scelte tecnologiche e progettuali alla base dello sviluppo del prodotto ChatSQL. Verranno quindi illustrati gli stili e i pattern architetturali adottati dal team.

## 1.2 Riferimenti

Il presente documento si basa su normative elaborate dal team, dall'ente proponente o da entità esterne, oltre a includere materiali informativi. Tali riferimenti sono elencati di seguito.

### 1.2.1 Riferimenti normativi

- *Norme di Progetto v1.0.0*;
- Regolamento del progetto didattico:  
<https://www.math.unipd.it/~tullio/IS-1/2023/Dispense/PD2.pdf>;
- Capitolato C9 - ChatSQL:
  - <https://www.math.unipd.it/~tullio/IS-1/2023/Progetto/C9.pdf>  
(Ultimo accesso: 2024-07-02);
  - <https://www.math.unipd.it/~tullio/IS-1/2023/Progetto/C9p.pdf>  
(Ultimo accesso: 2024-07-02).

### 1.2.2 Riferimenti informativi

- *Analisi dei Requisiti v1.0.1*;
- *Glossario v1.0.1*;
- Verbali interni:
  - 2024-04-03;
  - 2024-04-10;
  - 2024-04-16;
  - 2024-04-20;
  - 2024-04-25;
  - 2024-05-02;
  - 2024-05-07;
  - 2024-05-16;
  - 2024-05-23;
  - 2024-05-28;

- 2024-06-03;
  - 2024-06-14;
  - 2024-06-22;
  - 2024-07-06;
  - 2024-07-10;
  - 2024-07-18.
- Verbali esterni:
    - 2024-04-09;
    - 2024-05-06;
    - 2024-05-22;
    - 2024-06-07;
    - 2024-07-09.

### 1.3 Glossario

Allo scopo di evitare incomprensioni relative al linguaggio utilizzato nella documentazione di progetto, viene fornito un *Glossario*, nel quale ciascun termine è corredato da una spiegazione che mira a disambiguare il suo significato. I termini tecnici, gli acronimi e i vocaboli ritenuti ambigui vengono formattati in corsivo all'interno dei rispettivi documenti e marcati con una lettera <sub>G</sub> in pedice. Tutte le ricorrenze di un termine definito nel *Glossario* subiscono la formattazione sopracitata.

## 2 Tecnologie utilizzate

Nella sezione seguente sono documentate le scelte tecnologiche del team, la cui validità è stata dimostrata mediante lo sviluppo di un *Proof of Concept*. Per ciascuna tecnologia, il team riporta le motivazioni che ne hanno determinato la scelta.

### 2.1 Backend

#### 2.1.1 FastAPI

FastAPI è un *framework* web moderno per la creazione di *API* con Python.

##### Motivazioni

- **Flessibilità:** FastAPI offre ampia libertà agli sviluppatori e può essere facilmente integrato con librerie esterne; contrariamente a Django, non impone strutture rigide, consentendo agli sviluppatori di scegliere gli strumenti più adatti al loro caso d'uso;
- **Prestazioni:** FastAPI assicura prestazioni elevate grazie all'integrazione nativa con Starlette e Pydantic;
- **Orientato alla produttività:** FastAPI è progettato per massimizzare la produttività degli sviluppatori, riducendo il tempo necessario per la configurazione dell'ambiente e lo sviluppo delle API;
- **Documentazione API:** FastAPI genera automaticamente la documentazione interattiva delle API, a differenza di Flask e Django, che richiedono l'integrazione di librerie esterne per ottenere funzionalità simili;
- **Validazione automatica dei tipi:** FastAPI utilizza Pydantic per la validazione automatica dei dati;
- **Modernità:** FastAPI è un framework moderno, che sfrutta le funzionalità più recenti di Python, come le annotazioni di tipo.

#### 2.1.2 Python

Python è un linguaggio di programmazione ad alto livello ampiamente utilizzato in applicazioni di machine learning ed elaborazione del linguaggio naturale (NLP).

##### Motivazioni

- **Leggibilità:** Python ha una sintassi intuitiva e pulita, che agevola la leggibilità, la manutenzione e la collaborazione;
- **NLP:** Python è il linguaggio di riferimento per applicazioni di NLP (elaborazione del linguaggio naturale);
- **Librerie:** Python dispone di un'ampia gamma di librerie e moduli di terze parti;



- **Orientato agli oggetti:** Python è un linguaggio orientato agli oggetti; pertanto, consente di organizzare il codice seguendo i principi del design *SOLID<sub>e</sub>*;
- **Comunità:** Python dispone di una vasta comunità di sviluppatori che forniscono supporto e soluzioni a problemi comuni.

### 2.1.3 txtai

*txtai<sub>e</sub>* è un database di embeddings sviluppato in Python e progettato per ottimizzare la ricerca semantica.

#### Motivazioni

- **Comunicazione con la Proponente:** *txtai* è stato proposto dalla *Proponente<sub>e</sub>*, sia nel capitolato che durante le riunioni esterne, per facilitare la comunicazione e le revisioni congiunte;
- **Query SQL-like:** *txtai* supporta query SQL-like per la ricerca semantica;
- **Configurazione minima:** *txtai* richiede una configurazione minima e fornisce una suite completa di strumenti di intelligenza artificiale, inclusa la traduzione automatica tra lingue;
- **Persistenza degli indici:** *txtai* semplifica il processo di memorizzazione e ripristino degli *indici<sub>e</sub>*;
- **Debug:** *txtai* mette a disposizione funzioni specifiche per il debug del processo di generazione di un *prompt<sub>e</sub>*.

### 2.1.4 SQLAlchemy

SQLAlchemy è lo strumento principale per l'interazione con database *SQL<sub>e</sub>* in Python.

#### Motivazioni

- **Portabilità:** SQLAlchemy supporta i principali *DBMS<sub>e</sub>* e agevola la transizione;
- **Performance e sicurezza:** SQLAlchemy massimizza l'efficienza e la sicurezza delle transazioni attraverso il pattern "Unit of Work";
- **SQL injection:** SQLAlchemy include una protezione integrata contro l'SQL injection;
- **Leggibilità e mantenibilità:** SQLAlchemy utilizza classi e oggetti Python per rappresentare le relazioni di un *database<sub>e</sub>*, rendendo il codice intuitivo e autoesplicativo.
- **Flessibilità:** SQLAlchemy consente di costruire query SQL "grezze" o di utilizzare un *ORM<sub>e</sub>* per interagire con il database.

### 2.1.5 JWT

JSON Web Token (JWT) uno standard web per lo scambio di dati.

## Motivazioni

- **Stateless:** JWT consente di autenticare le richieste senza necessità di sessioni sul server, migliorando la scalabilità e riducendo il carico sul server;
- **Sicurezza:** I JWT sono firmati digitalmente, il che garantisce l'integrità e l'autenticità del token;
- **Formato standard:** JWT è supportato da una vasta gamma di linguaggi e framework, tra cui FastAPI.

## 2.2 Frontend

### 2.2.1 Vue.js

Vue.js<sub>6</sub> è un framework JavaScript utilizzato per la creazione di interfacce utente reattive e dinamiche.

## Motivazioni

- **Component-based:** Vue.js è un framework basato su componenti, suddivisi in tre sezioni: template (HTML), stile (CSS) e logica (JavaScript);
- **Reattività:** Vue.js offre un sistema reattivo che permette di aggiornare automaticamente la vista dell'utente in base ai cambiamenti dello stato dell'applicazione;
- **Integrazione:** Vue.js può essere facilmente integrato con librerie esterne, come Axios e PrimeVue;
- **Prestazioni:** Vue.js implementa il concetto, introdotto da React, di DOM virtuale, riducendo i re-render non necessari e migliorando le prestazioni dell'applicazione;
- **MVVM:** Vue.js implementa il pattern MVVM (Model-View-ViewModel), con un focus sul livello ViewModel.

### 2.2.2 Axios

Axios è una libreria JavaScript utilizzata per connettersi con le API di *back-end*<sub>6</sub> e gestire le richieste effettuate tramite il protocollo HTTP.

## Motivazioni

- **Semplicità:** Axios è un libreria semplice e intuitiva per effettuare richieste HTTP;
- **Gestione dati JSON:** Axios converte automaticamente i dati da e in JSON;
- **Compatibilità e integrazione:** Axios è compatibile e facilmente integrabile con Vue 3 e la Composition API;
- **Promise-based:** Axios utilizza un'interfaccia Promise-based per gestire le richieste asincrone.

### 2.2.3 TypeScript

TypeScript è un linguaggio di programmazione che estende JavaScript, aggiungendo caratteristiche come i tipi di dato.

#### Motivazioni

- **Compatibilità retroattiva** (backward compatibility): Qualsiasi programma scritto in JavaScript può funzionare in TypeScript senza richiedere modifiche;
- **Robustezza**: TypeScript fornisce un sistema di type-checking basato sulla tipizzazione statica, che permette di rilevare errori di tipo durante la fase di sviluppo;
- **Documentazione automatica**: La tipizzazione funge da documentazione incorporata, semplificando la manutenzione e la collaborazione;
- **Supporto IDE**: TypeScript è supportato dai principali editor, come Visual Studio Code, che forniscono funzionalità avanzate di debugging e formattazione.

### 2.2.4 PrimeVue

PrimeVue è una libreria di componenti per Vue.js.

#### Motivazioni

- **Integrazione**: PrimeVue è stato progettato appositamente per Vue.js e, pertanto, offre un'integrazione nativa con il *framework*;
- **Completezza**: PrimeVue offre una vasta gamma di componenti, temi e stili personalizzabili;
- **Funzionalità avanzate**: PrimeVue fornisce componenti per la creazione di interfacce utente complesse e interattive. Inoltre, supporta le caratteristiche e funzionalità avanzate di Vue.js;
- **Documentazione e supporto**: La documentazione e i materiali di supporto sono orientati all'integrazione con Vue.js.

## 2.3 Altri strumenti e tecnologie

### 2.3.1 Docker

*Docker* è una piattaforma open-source per la creazione, la distribuzione e l'esecuzione di applicazioni in contenitori leggeri e portabili.

#### Motivazioni

- **Isolamento delle applicazioni**: Docker consente di isolare le applicazioni in contenitori, garantendo che ciascuna applicazione abbia un ambiente di esecuzione indipendente;
- **Collaborazione**: La condivisione delle librerie e delle dipendenze semplifica la collaborazione e riduce il rischio di conflitti;

- **Portabilità:** I contenitori possono essere eseguiti su qualsiasi sistema che supporti Docker, indipendentemente dall'ambiente di sviluppo o di produzione;
- **Scalabilità:** Docker permette di aggiungere o rimuovere container in maniera rapida ed efficiente;
- **CI/CD:** Docker può essere utilizzato per automatizzare i processi di continuous integration e continuous delivery.

### 2.3.2 SQLite

SQLite<sub>®</sub> è una libreria di gestione di database relazionali integrata nella maggior parte dei linguaggi di programmazione.

#### Motivazioni

- **Leggerezza:** SQLite è estremamente leggero e richiede una quantità inferiore di risorse rispetto ai DBMS tradizionali;
- **Autosufficienza:** SQLite non richiede un processo separato per funzionare ed è progettato per essere integrato direttamente nelle applicazioni;
- **Compatibilità:** SQLite supporta la quasi totalità delle funzionalità dei database SQL standard;
- **Portabilità:** SQLite memorizza l'intero database in un singolo file, semplificando il trasferimento dei dati tra sistemi e piattaforme differenti;
- **ACID Compliance:** SQLite supporta le transazioni ACID (Atomicità, Coerenza, Isolamento e Durabilità), garantendo l'integrità dei dati anche in caso di mal-funzionamenti;
- **Performance:** SQLite è ottimizzato per gestire database di piccole e medie dimensioni.

### 2.3.3 Hugging Face

Hugging Face<sub>®</sub> è una piattaforma open-source che fornisce strumenti e risorse per lavorare su progetti di NLP (elaborazione del linguaggio naturale).

#### Motivazioni

- **Modelli pre-addestrati:** Hugging Face mette a disposizione una vasta collezione di modelli pre-addestrati;
- **Modelli open-source:** Hugging Face fornisce un'ampia selezione di modelli open-source;
- **Modelli locali:** Hugging Face offre la possibilità di scaricare i *modelli<sub>®</sub>* in locale per l'esecuzione offline;
- **Comunità:** La piattaforma dispone di una comunità attiva di sviluppatori e ricercatori che contribuiscono al miglioramento continuo dei modelli;

- **Supporto per diverse lingue:** Hugging Face offre una varietà di modelli multi-lingue e modelli per la traduzione.

#### 2.3.4 ChatGPT

*ChatGPT*<sub>o</sub> (Chat Generative Pre-trained Transformer) è un ChatBOT basato su intelligenza artificiale e apprendimento automatico, sviluppato da OpenAI.

##### Motivazioni

- **Popolarità:** ChatGPT è uno dei ChatBOT più diffusi nel campo informatico, e l'adequatezza delle sue risposte è stata dimostrata in numerosi contesti;
- **Integrazione:** ChatGPT offre API per l'integrazione con applicazioni esterne;
- **Supporto Multilingue:** I modelli alla base di ChatGPT supportano diverse lingue, tra cui l'italiano e l'inglese.

#### 2.3.5 LMSYS Chatbot Arena

LMSYS Chatbot Arena è uno spazio online dedicato al *benchmarking*<sub>o</sub> di LLM.

##### Motivazioni

- **Benchmark:** LMSYS Chatbot Arena permette di confrontare modelli diversi e di selezionare il miglior output in ciascuna iterazione;
- **Disponibilità:** LMSYS Chatbot Arena consente di testare i modelli più diffusi, inclusi ChatGPT, Gemini, Claude e LLaMA.

#### 2.3.6 LM Studio

*LM Studio*<sub>o</sub> è un'applicazione desktop che semplifica il processo di testing di modelli linguistici di grandi dimensioni (LLM).

##### Motivazioni

- **Semplicità:** LM Studio offre un'interfaccia intuitiva e user-friendly per testare i modelli LLM;
- **Test offline:** LM Studio consente di scaricare le versioni quantizzate dei modelli per l'esecuzione offline;
- **Debug:** LM Studio fornisce strumenti per il debug e l'analisi dei risultati;
- **Server OpenAI-like:** Le interazioni con il server locale di LM Studio seguono il formato API di OpenAI;
- **Prestazioni:** LM Studio fornisce un'opzione per abilitare l'accelerazione GPU e velocizzare i processi di inferenza dei modelli.

## 2.4 Panoramica delle tecnologie

Di seguito è fornita una panoramica generale delle tecnologie utilizzate dal team, suddivise nelle seguenti categorie:

- **Framework;**
- **Linguaggi;**
- **Librerie;**
- **Strumenti.**

### 2.4.1 Framework

Tabella 2.1: Framework utilizzati

Nome	Versione	Descrizione
FastAPI	0.110.0	Framework web moderno per la creazione di API con Python.
Vue.js	3.4.21	Framework JavaScript per la creazione di interfacce utente reattive e dinamiche.
Unittest (oppure pytest)	/	Framework di testing per Python, ispirato al modello di JUnit, un framework di unit testing per Java.
Jest	29.7.0	Framework di testing per JavaScript e TypeScript. Jest offre una suite completa di strumenti per il testing automatizzato.
Cypress	13.6.0	Framework di testing che può essere utilizzato per testare l'intera applicazione simulando l'interazione dell'utente.

### 2.4.2 Linguaggi

Tabella 2.2: Linguaggi utilizzati

Nome	Versione	Descrizione
HTML	5	Linguaggio di markup utilizzato per definire la struttura delle pagine web.
Continua nella prossima pagina		

Tabella 2.2: Linguaggi utilizzati (continua)

Nome	Versione	Descrizione
CSS	3	Linguaggio usato per definire lo stile e l'aspetto estetico delle pagine web.
TypeScript	5	Espansione del linguaggio JavaScript, progettata per migliorare lo sviluppo di pagine web dinamiche e interattive.
Python	3.10.12	Linguaggio di programmazione ad alto livello e orientato agli oggetti.

### 2.4.3 Librerie

Tabella 2.3: Librerie utilizzate

Nome	Versione	Descrizione
txtai	7.0.0	Database di embeddings sviluppato in Python e progettato per ottimizzare la ricerca semantica.
SQLAlchemy	2.0.31	Libreria open-source che fornisce un toolkit SQL e un ORM per le interazioni con database.
jsonschema	4.23.0	Libreria per la validazione di JSON Schema in Python.
Pydantic	2.8.2	Libreria Python per la validazione dei dati mediante le annotazioni di tipo.
PyJWT	2.8.0	Libreria Python per la gestione dei JSON Web Token.
parameterized	0.9.0	Libreria Python utilizzata per creare test parametrizzati, ovvero test che vengono eseguiti più volte con diversi set di parametri.
openai	1.37.1	Libreria Python per l'interazione con i servizi di OpenAI.
aiofiles	24.1.0	Libreria Python per eseguire operazioni di I/O su file in modo asincrono.
PrimeVue	3.53.0	Libreria di componenti per Vue.js.
Continua nella prossima pagina		

Tabella 2.3: Librerie utilizzate (continua)

Nome	Versione	Descrizione
Vue I18n	9.0.0	Libreria per la creazione di interfacce utente multilingue in applicazioni Vue.js.
Vuex (oppure Pinia)	3.10.12	Libreria di gestione dello stato in applicazioni Vue.js.
Axios	3.10.12	Libreria JavaScript utilizzata per connettersi con le API di back-end e gestire le richieste effettuate tramite il protocollo HTTP.
Vue Test Utils	2.4.6	Libreria per il testing di componenti in applicazioni Vue.js. Vue Test Utils supporta sia test di unità che test di integrazione.

#### 2.4.4 Strumenti

Tabella 2.4: Strumenti utilizzati

Nome	Versione	Descrizione
Git	2.45.2	Sistema di controllo di versione distribuito utilizzato principalmente nello sviluppo software.
pip	24.0	Sistema di gestione dei pacchetti per Python.
npm	10.7.0	Sistema di gestione dei pacchetti per JavaScript e per l'ambiente di esecuzione Node.js.
ESLint	9.5.0	Strumento di analisi statica che mira a individuare errori di programmazione, problemi stilistici e costrutti sospetti nel codice JavaScript.
Prettier	3.3.0	Formattatore di codice per JavaScript, TypeScript, JSON, HTML, CSS, Vue e YAML.
Pylint	3.2.6	Strumento di analisi statica per Python.
Vue Test Utils	3.0.0	Libreria per il testing di componenti in applicazioni Vue.js.
Continua nella prossima pagina		



Tabella 2.4: Strumenti utilizzati (continua)

Nome	Versione	Descrizione
Docker	/	Piattaforma per l'esecuzione di applicazioni in contenitori isolati. Docker può essere utilizzato in combinazione con GitHub Actions per automatizzare i processi di build, test e distribuzione.
SQLite	/	Sistema di gestione di database relazionali open-source, leggero e facilmente portabile.
Hugging Face	/	Piattaforma di hosting di modelli e set di dati di machine learning.
ChatGPT	/	ChatBOT basato su intelligenza artificiale e apprendimento automatico.
LMSYS Chatbot Arena	/	Spazio online dedicato al benchmarking di LLM.
LM Studio	/	Applicazione desktop per il testing di LLM in locale.



### 3 Architettura

Il prodotto ChatSQL è basato su un'architettura client-server. Il client è l'interfaccia attraverso la quale gli utenti interagiscono con il sistema, come ad esempio un browser web. In altre parole, il client è il componente che richiede risorse o servizi. Il server è l'applicazione che riceve ed elabora le richieste provenienti da uno o più client, fornendo risposte appropriate. Per migliorare lo sviluppo collaborativo, la modularità e la manutenibilità, il sistema è stato suddiviso in due componenti principali:

- **Front-end:** La componente di un sistema che l'utente visualizza e con cui può interagire;
- **Back-end:** La componente che gestisce la logica di business, l'elaborazione dei dati e la comunicazione con i database e altri servizi.