



# Csaba Toth

- Full stack engineer
  - Director of Product Engineering at SportsBoard (startup)
  - CTO at ThruThink (startup)
- Interested in AR/XR, AI/ML, lately Gen AI
- GDG (Google Developer Group) Fresno lead, WTM (Women Techmakers) Fresno ambassador, tech meetup junkie



- ThruThink® is a business budgeting on-line app to create professional budgets and forecasts
- product of literally decades of experience and careful thought, and thousands of calculations
- Thru-hiking, or through-hiking, is the act of hiking an established long-distance trail end-to-end continuously
- There are no dedicated personnel for support chat agent roles, had a “classic” chat agent integration



# ThruThink Customer Chat Agent

- Invaluable help, given that
  - It stays relatively grounded
  - Won't hallucinate\* wildly
- Desired abilities:
  - Main goal: answer ThruThink software specific questions: “In ThruThink can I make adjustments on the Cash Flow Control page?”
  - Nice to have: answer more generic questions such as: “How much inventory should I have?”



# ThruThink Customer Chat Agent

Nice to have: “How much inventory should I have?”

The amount of inventory that you should have on hand depends on several factors, including your business type, industry, and sales volume. Here are a few considerations to keep in mind when determining the appropriate amount of inventory for your business:

**Sales volume:** If your business experiences high sales volume, you may need to keep more inventory on hand to meet customer demand. On the other hand, if sales are slower, you may be able to keep less inventory in stock.

**Lead time:** If it takes a long time to receive inventory from suppliers, you may need to keep more on hand to avoid stockouts.

**Seasonality:** If your business experiences seasonal fluctuations in demand, you may need to adjust your inventory levels accordingly.

**Storage space:** Consider how much physical space you have available to store inventory. You don't want to overstock and have inventory sitting unused and taking up valuable space.

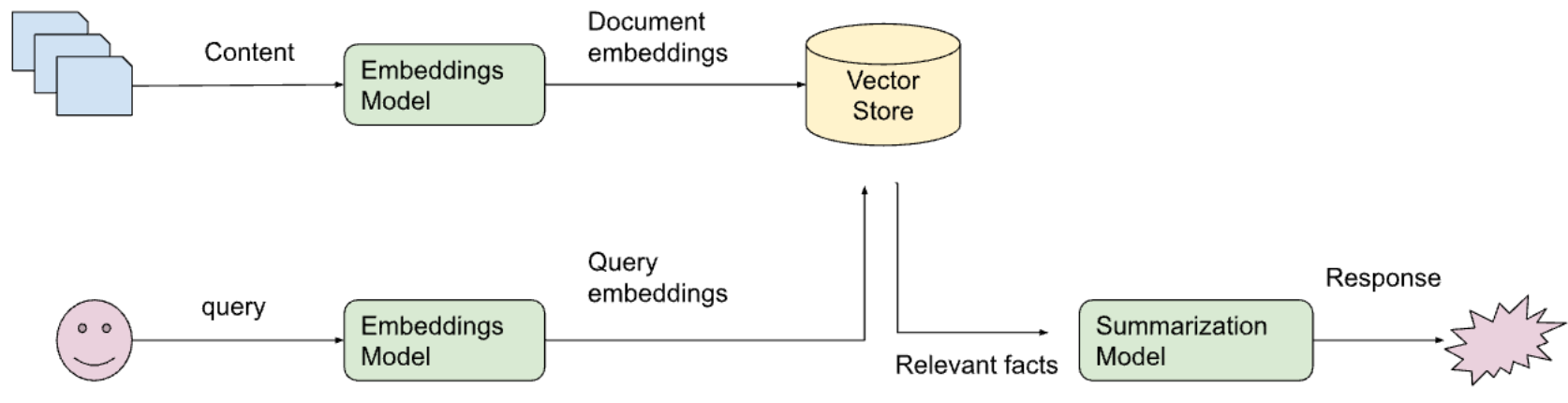
Ultimately, the goal is to have enough inventory to meet customer demand without tying up too much capital in excess stock. As a general rule of thumb, aim to keep inventory levels at a level that allows you to cover several months of sales, while also keeping a close eye on inventory turnover rates to ensure that you are not holding onto excess inventory for too long.



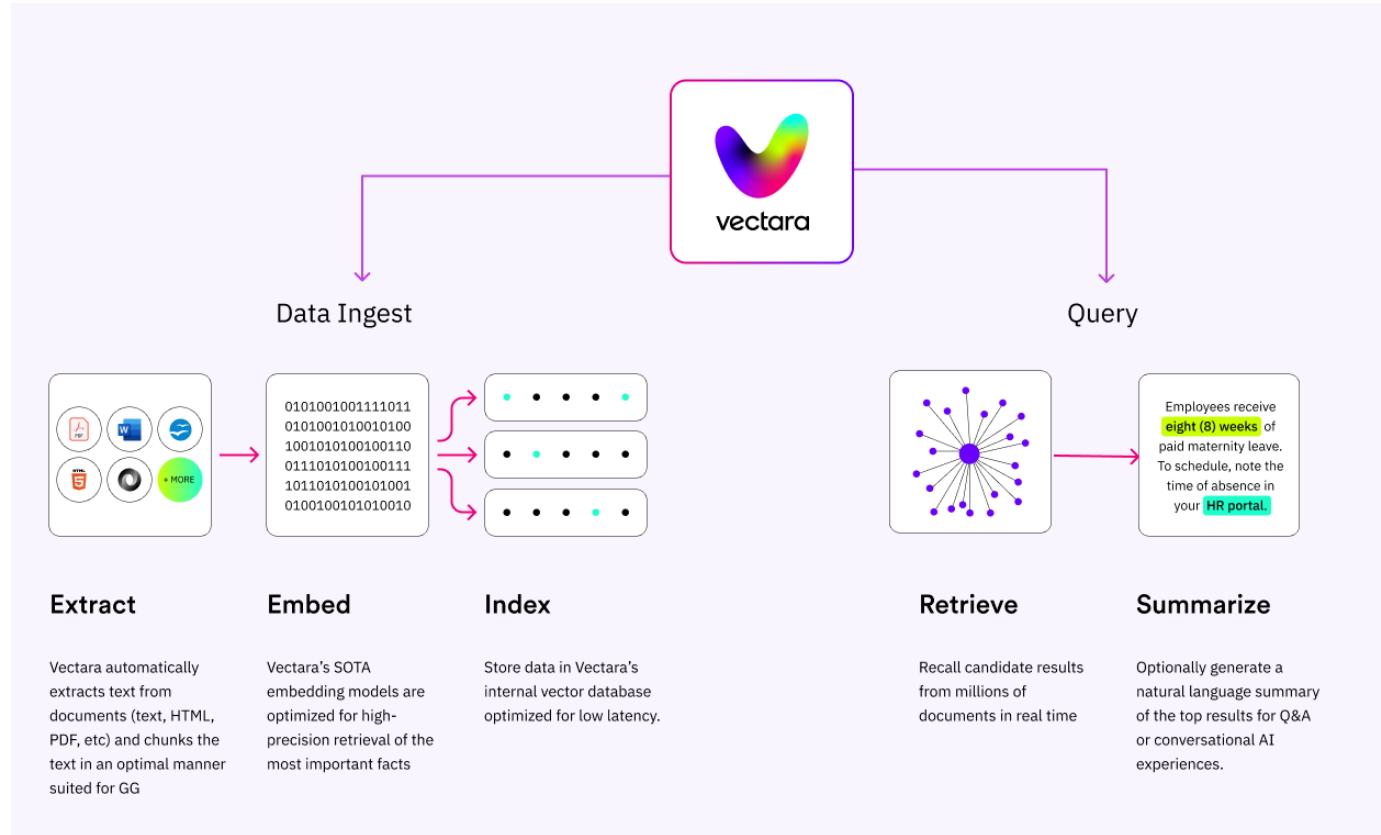
# ThruThink Customer Chat Agent

- The main goal needs grounding
- The secondary goal might fall back to the vast knowledge of an LLM
- Source documents: ThruThink's help database / documentation
  - 185 help topics
  - Organized into categories
  - Generated HTML, markdown formats before

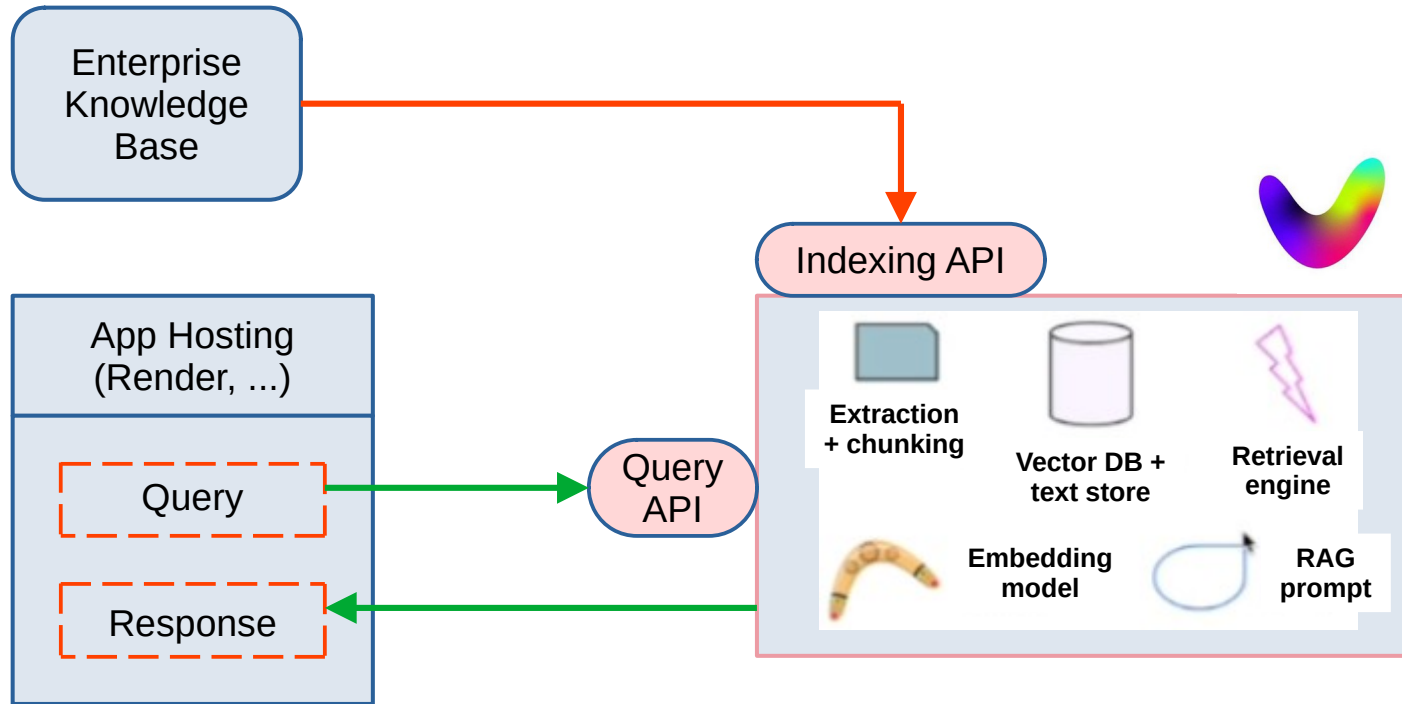
# Retrieval Augmented Generation



# Retrieval Augmented Generation

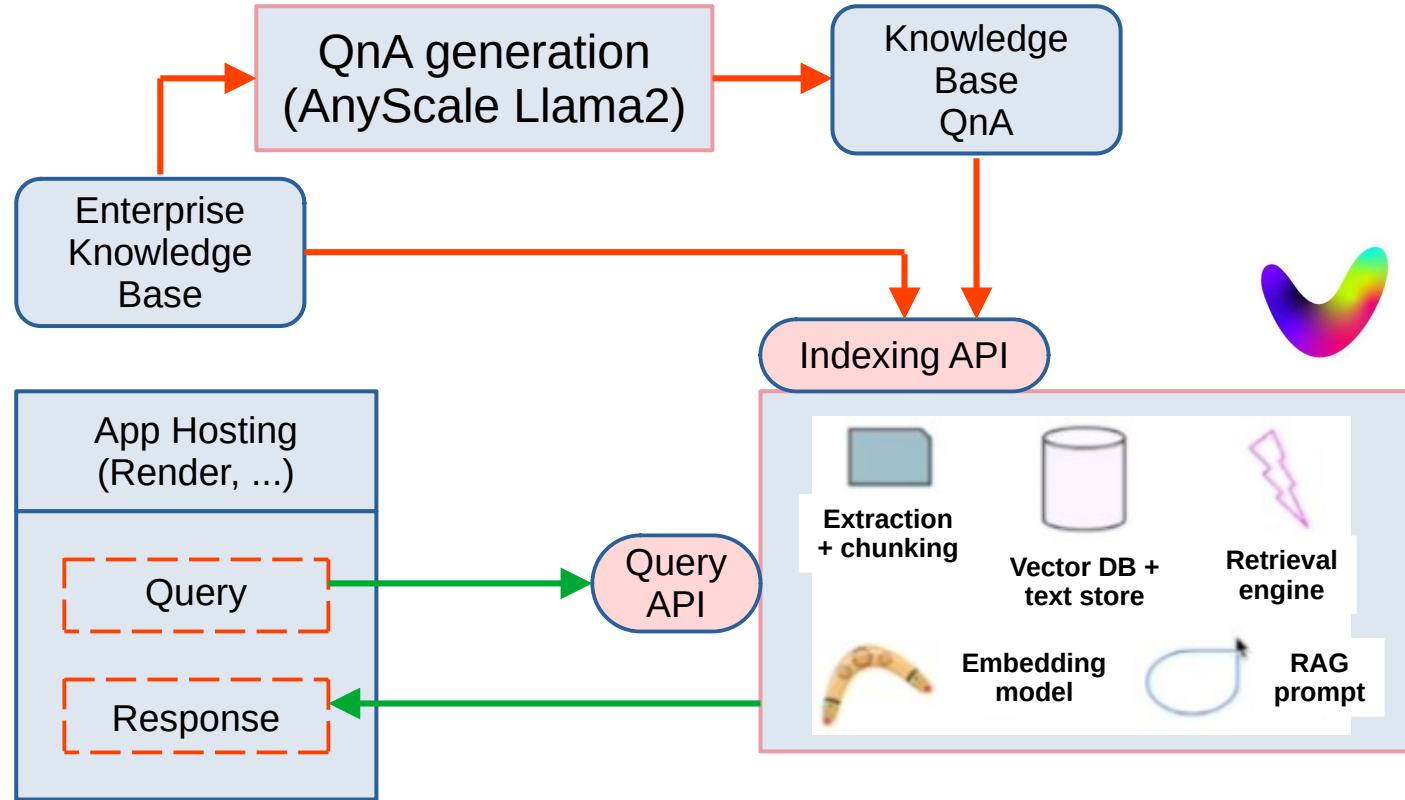


# Retrieval Augmented Generation

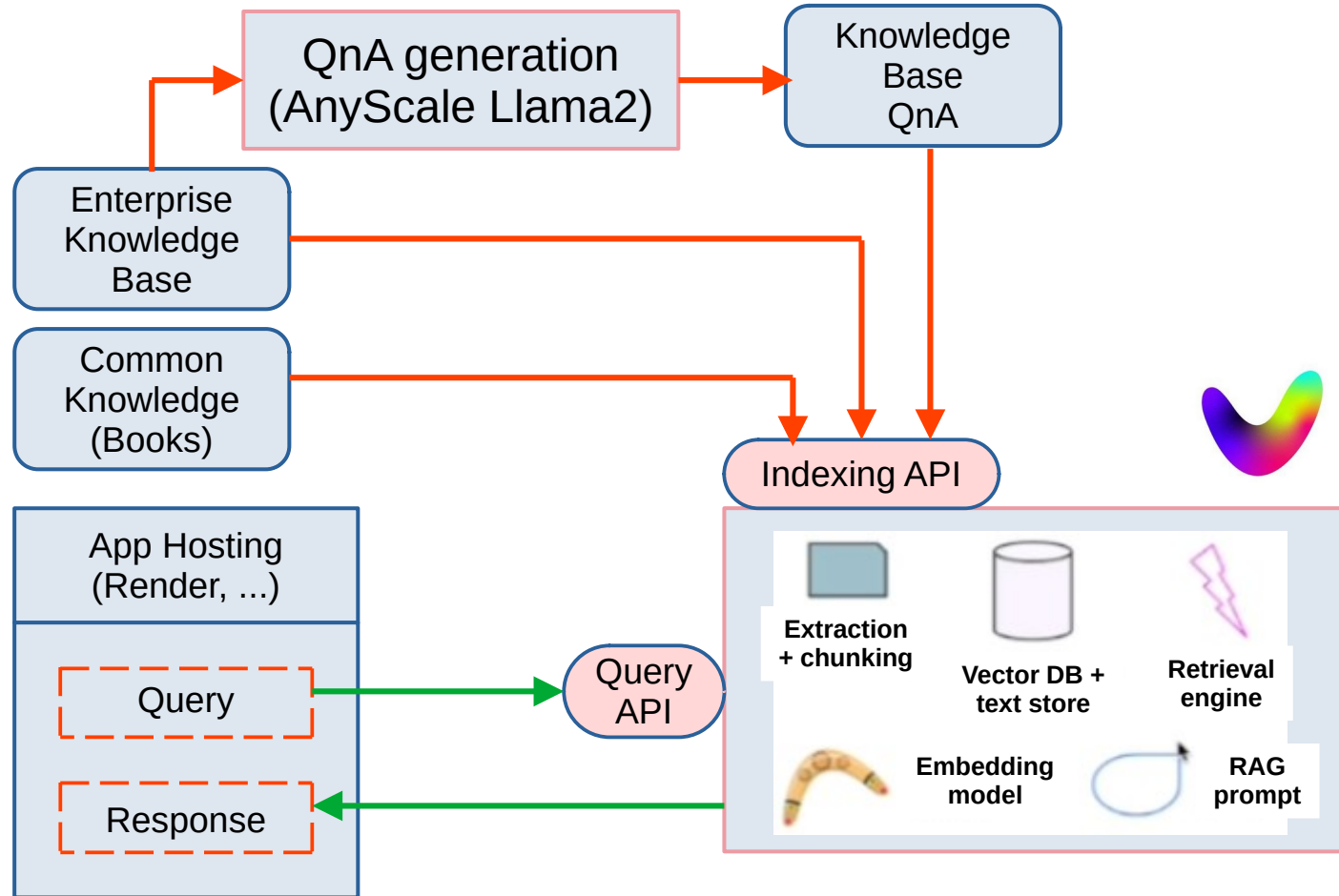




# QBRAG: QnA Boosted RAG



# QBRAG: QnA Boosted RAG





# Accomplishments

- Contributed back to [https://github.com/nestordemeure/question\\_extractor](https://github.com/nestordemeure/question_extractor) for the QnA generation
  - Check-pointing to resume generation (can take 6-7 hours)
  - Export AnyScale fine tuning format
  - Tone down rate limit hitting to not hammer the API as much
- Script to process `question\_extractor`'s `questions.json` to generate markdown QnA per help topic



## Accomplishments – cont.

- All of my document loading scripts (help markdown, QnA markdown, books, ...) are in my repository
- Modified Vectara React application for my needs, also open source
- Hosting the evaluation application on the Render platform (free tier might need some warm-up time)  
<https://thruthink-support.onrender.com/>



# Considerations

- Is fine tuning needed, is it beneficial? Test / measure if that would actually produce better results
- Improve supporting documents even more (for example policies: ToS, AUP, ...)
- Integrate the solution into ThruThink's application (closed source):
  - Will allow the invocation of help topic display when user clicks on certain document references