

# Indexing and Retrieval - Part 1

## Instructors

Prashant Sahu

Manager - Data Science, Analytics Vidhya

Ravi Theja

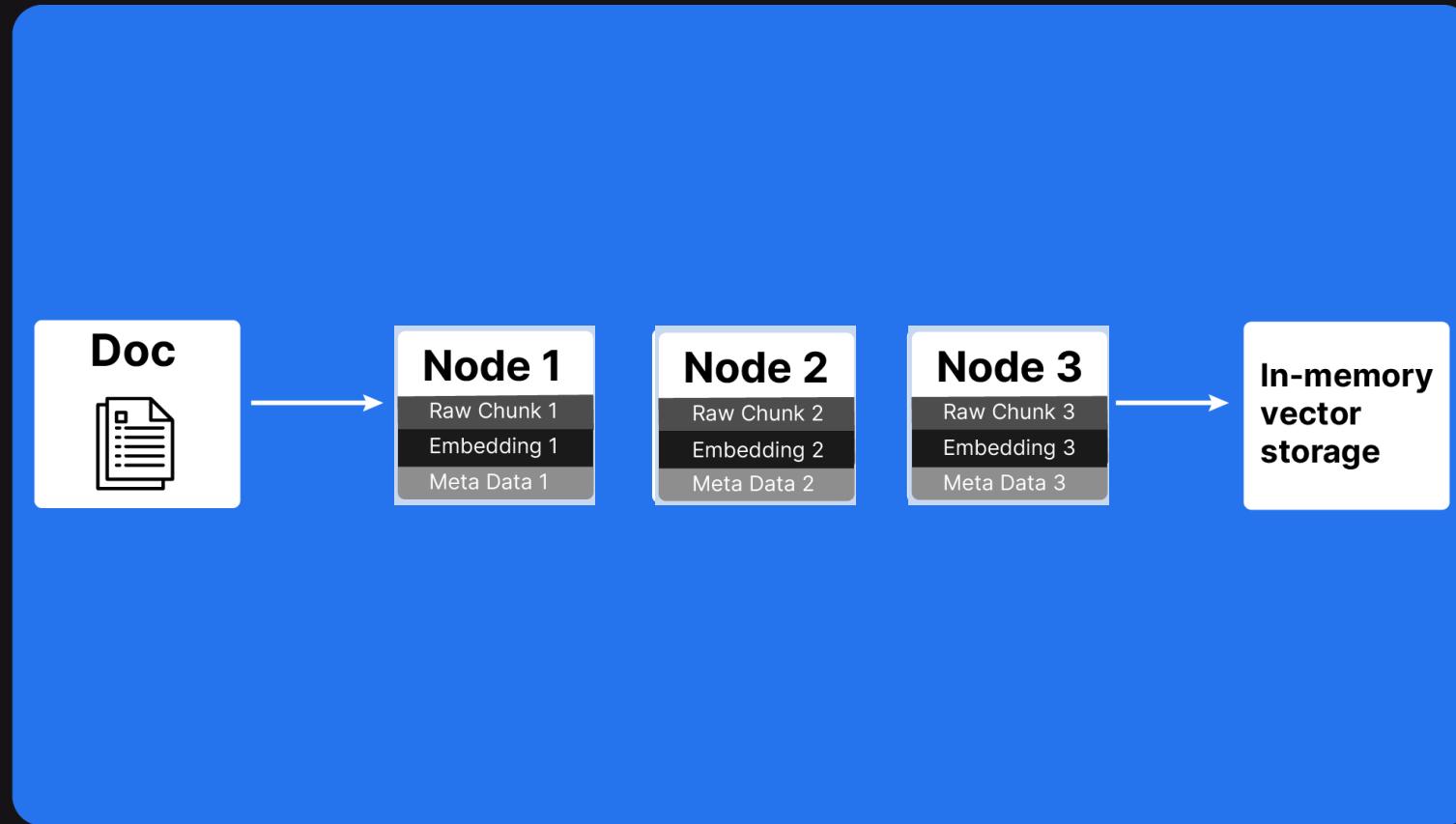
Developer Advocate Engineer, LlamalIndex



# Recap of RAG Framework



# Indexing



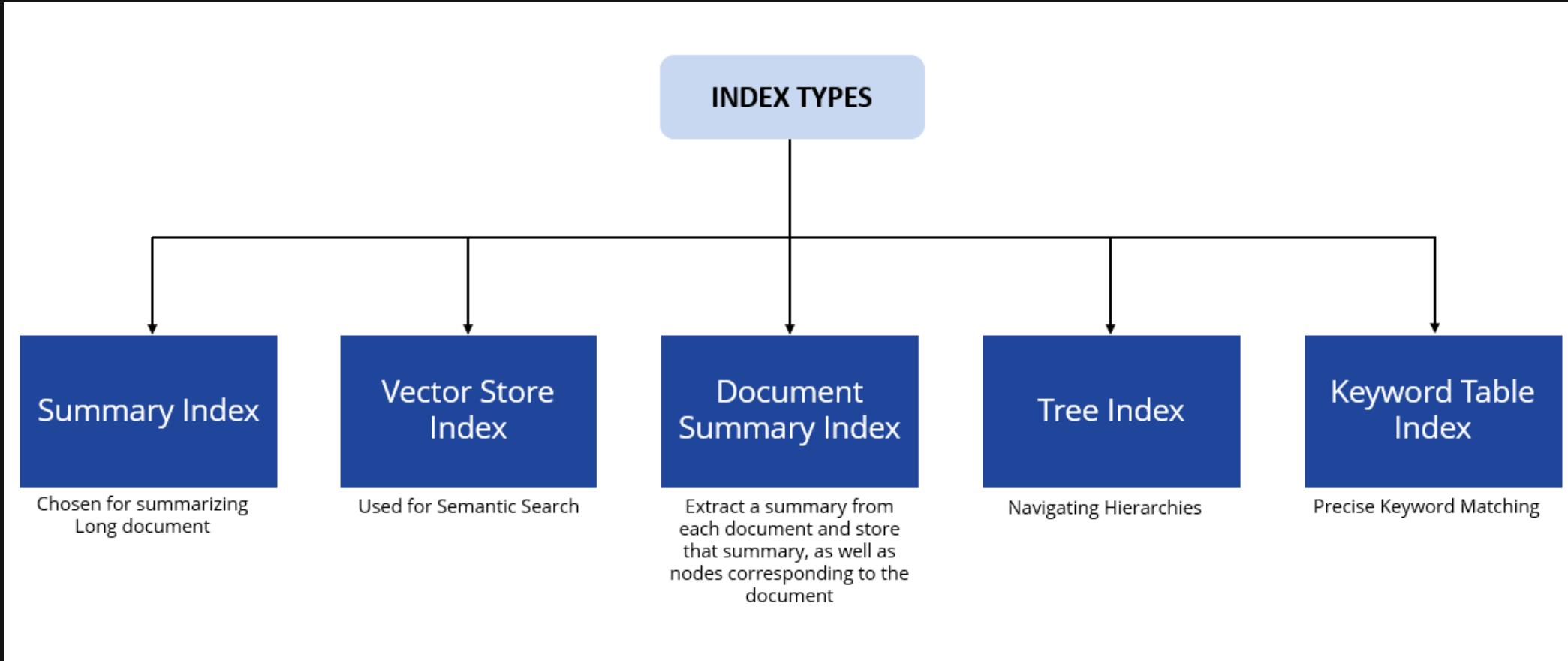
# What is an Index?

- Structured dataset enabling fast and efficient information retrieval.
- Allows quick location and retrieval of relevant information.
- Improves the generation process by accessing a large corpus efficiently.

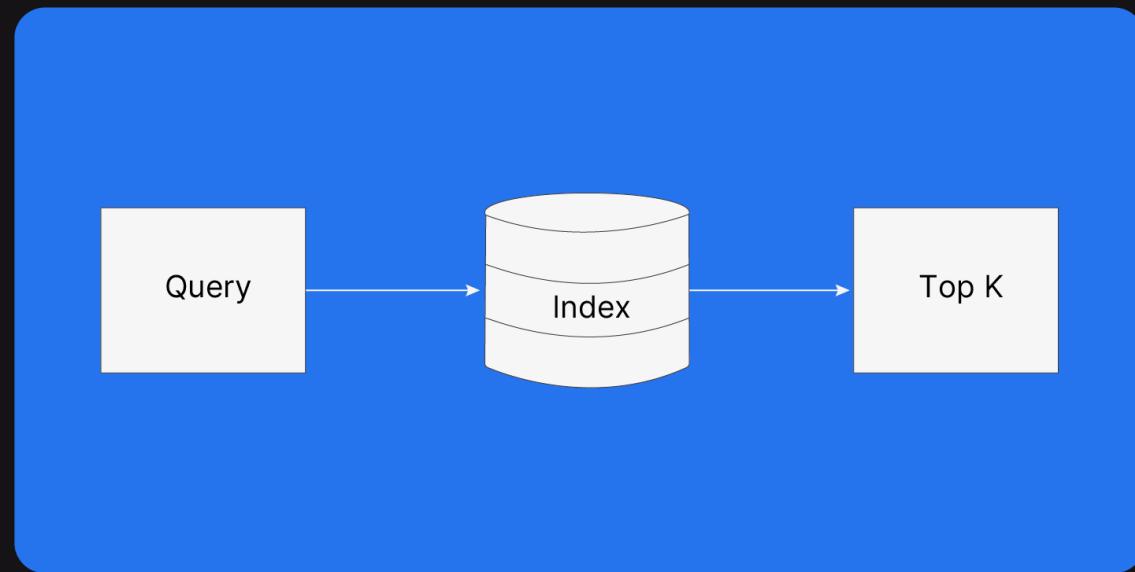
# Why do we need Index in a RAG system?

- **Efficiency**: Enables quick retrieval by organizing data for efficient searches.
- **Scalability**: Handles large-scale datasets, ensuring timely retrieval operations.
- **Accuracy**: Improves retrieval precision, enhancing response quality.
- **Real-time Performance**: Supports fast, real-time responses for applications like chatbots and virtual assistants.

# Index Types



# Retrieval



# Different Types of Retrievers

Retriever depends on the specific indexing technique

# Retrieval Modes for Different Indexes

## Summary Index

- Default: SummaryIndexRetriever
- Embedding: SummaryIndexEmbeddingRetriever
- llm: SummaryIndexLLMRetriever

## Document Summary Index

- llm: DocumentSummaryIndexLLMRetriever
- Embedding: DocumentSummaryIndexEmbeddingRetriever

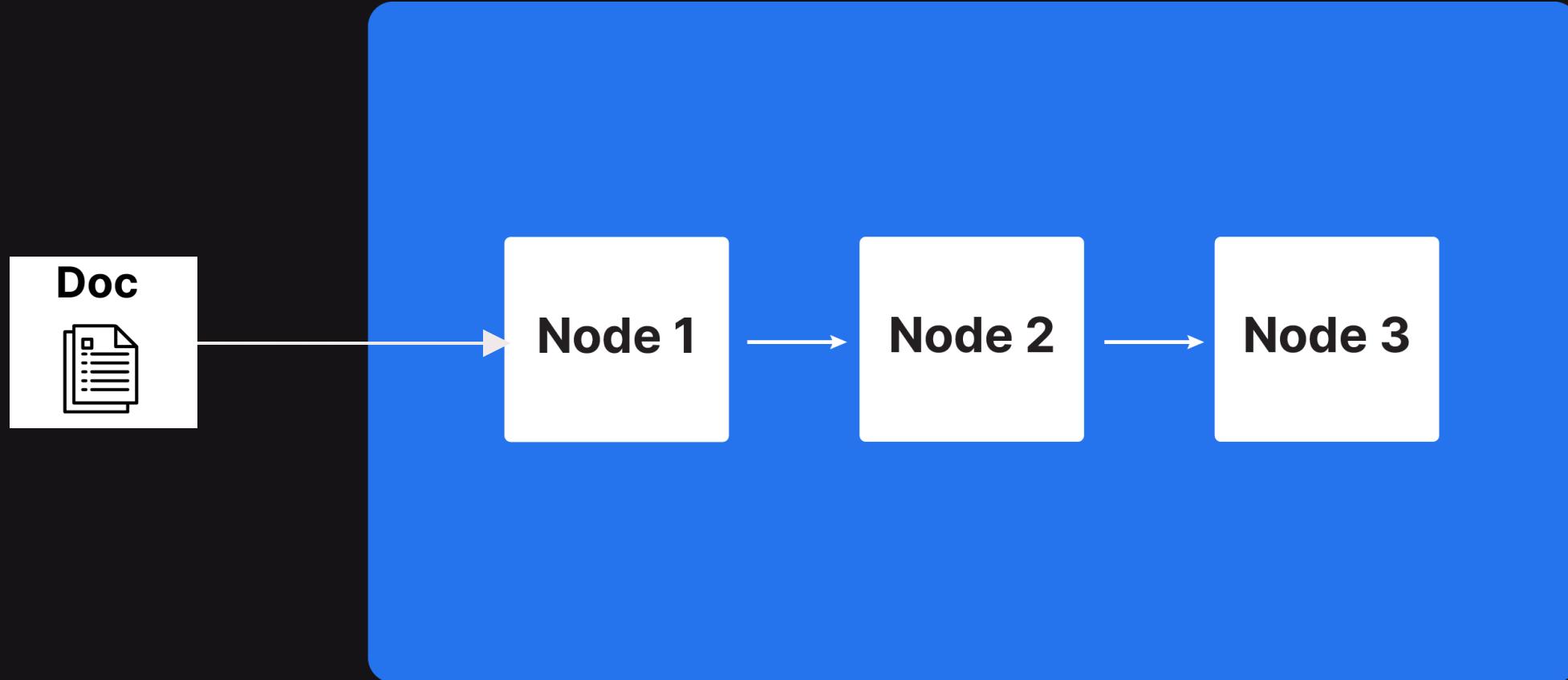
## Tree Index

- Select leaf: TreeSelectLeafRetriever
- Select leaf embedding: TreeSelectLeafEmbeddingRetriever
- All leaf: TreeAllLeafRetriever
- Root: TreeRootRetriever

## Keyword Table Index

- Keyword: KGTableRetriever
- Embedding: KGTableRetriever
- Hybrid: KGTableRetriever

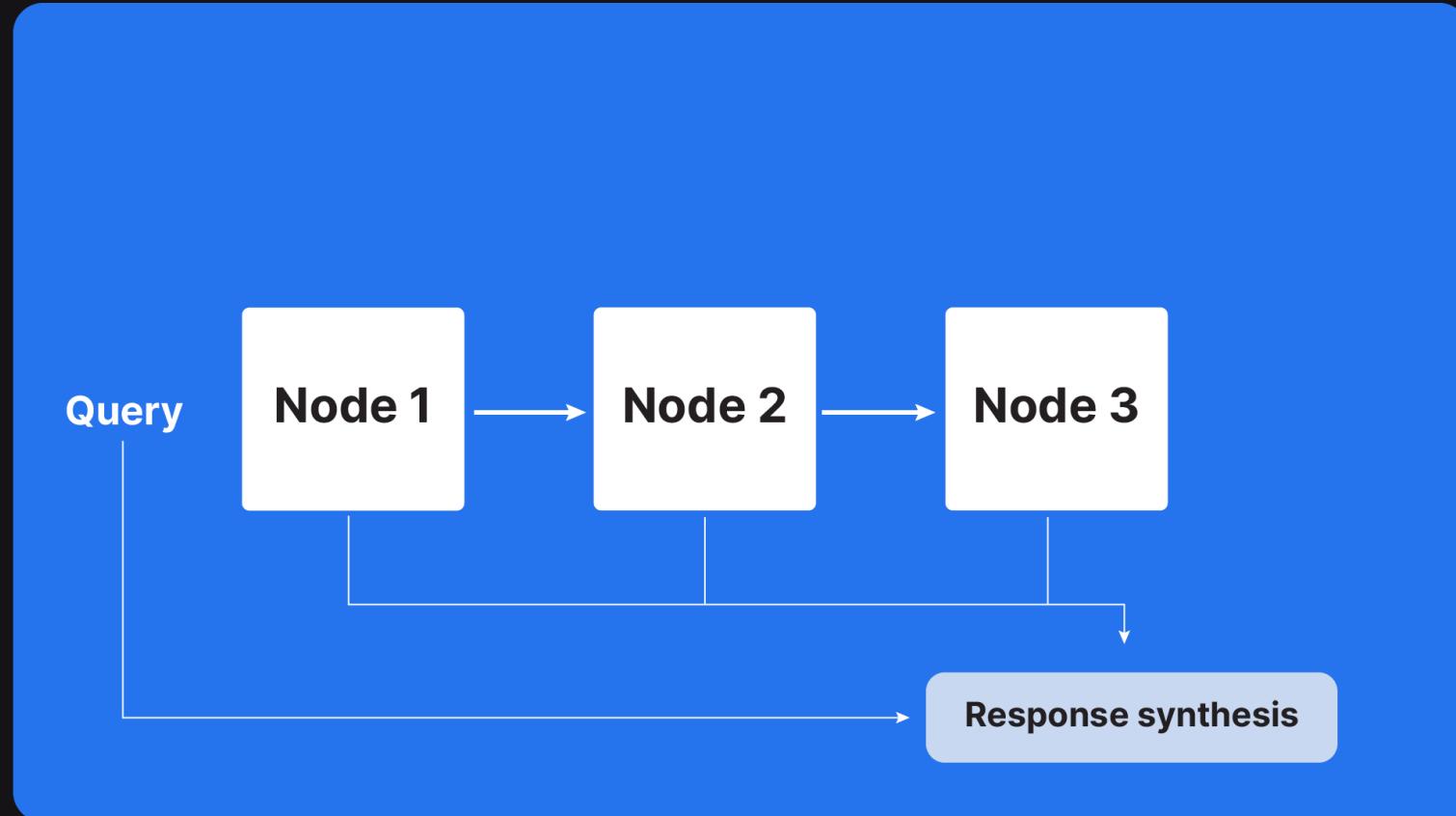
# 1. Summary Index



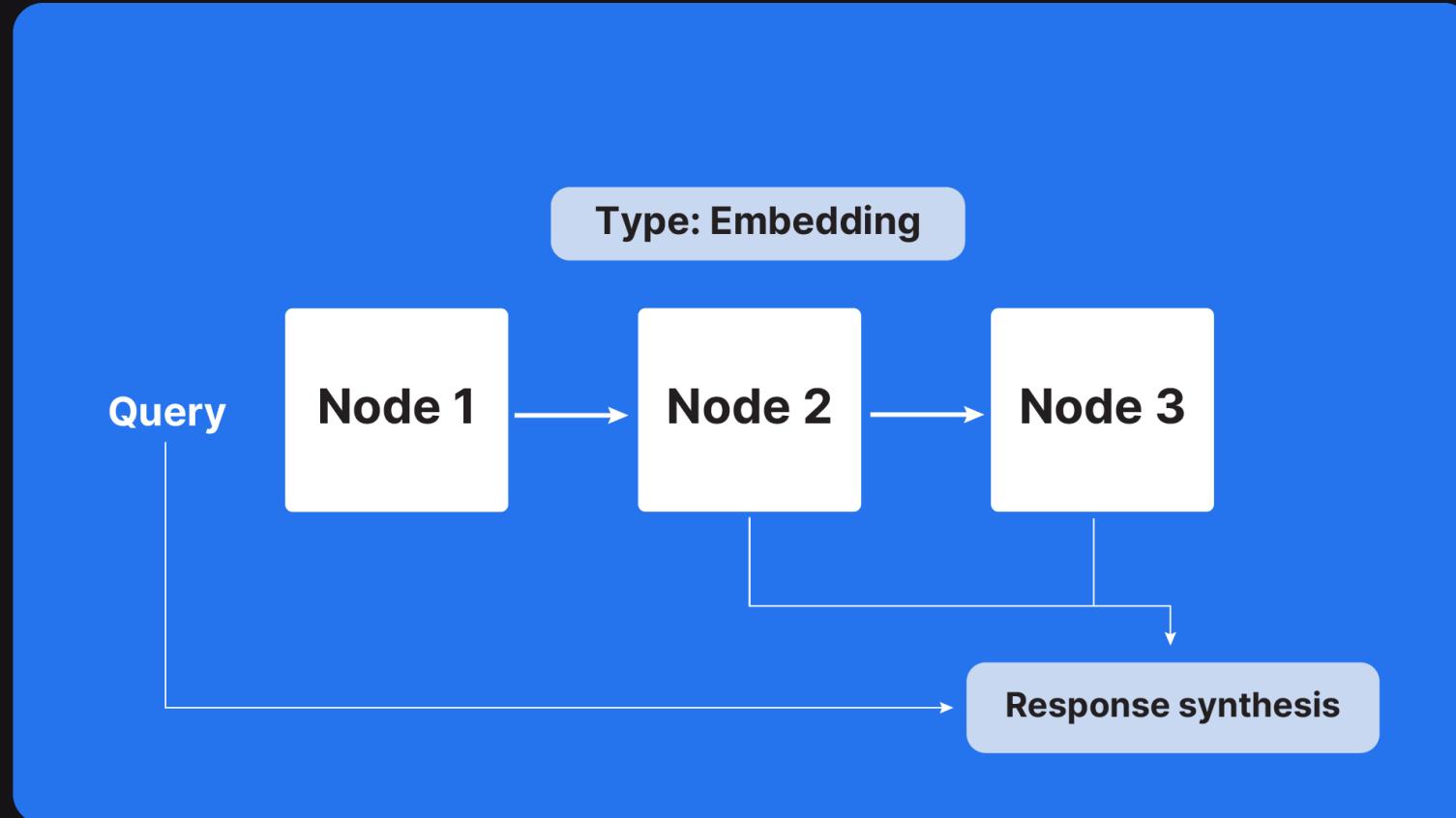
# Retrieval Modes for Summary Index

- **SummaryIndexRetriever**: Uses traditional keyword-based search techniques.
- **SummaryIndexEmbeddingRetriever**: Uses pre-computed embeddings for semantic similarity search.
- **SummaryIndexLLMRetriever**: Utilizes a language model for dynamic query interpretation and retrieval.

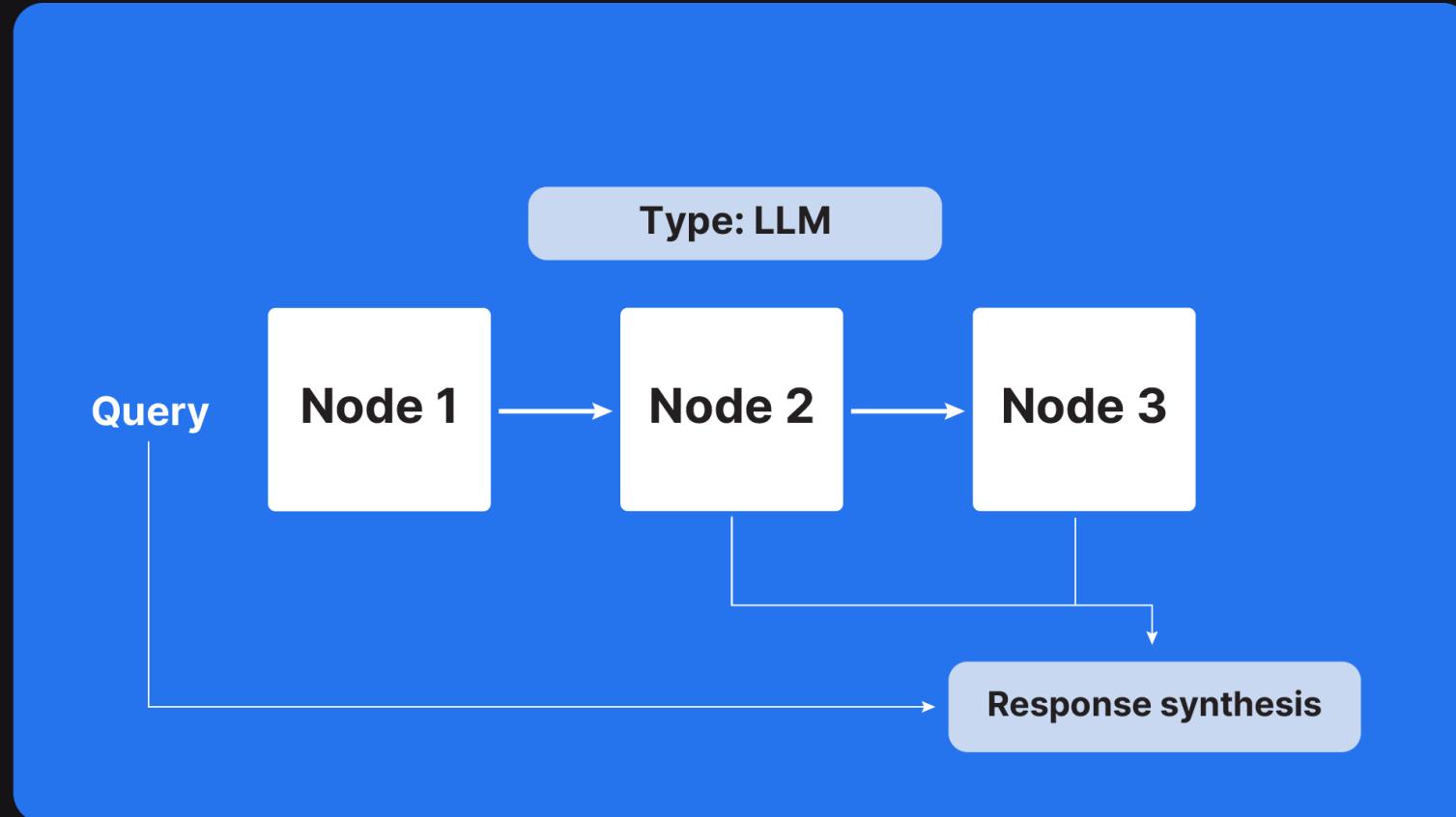
# 1A. Summary Index Retriever



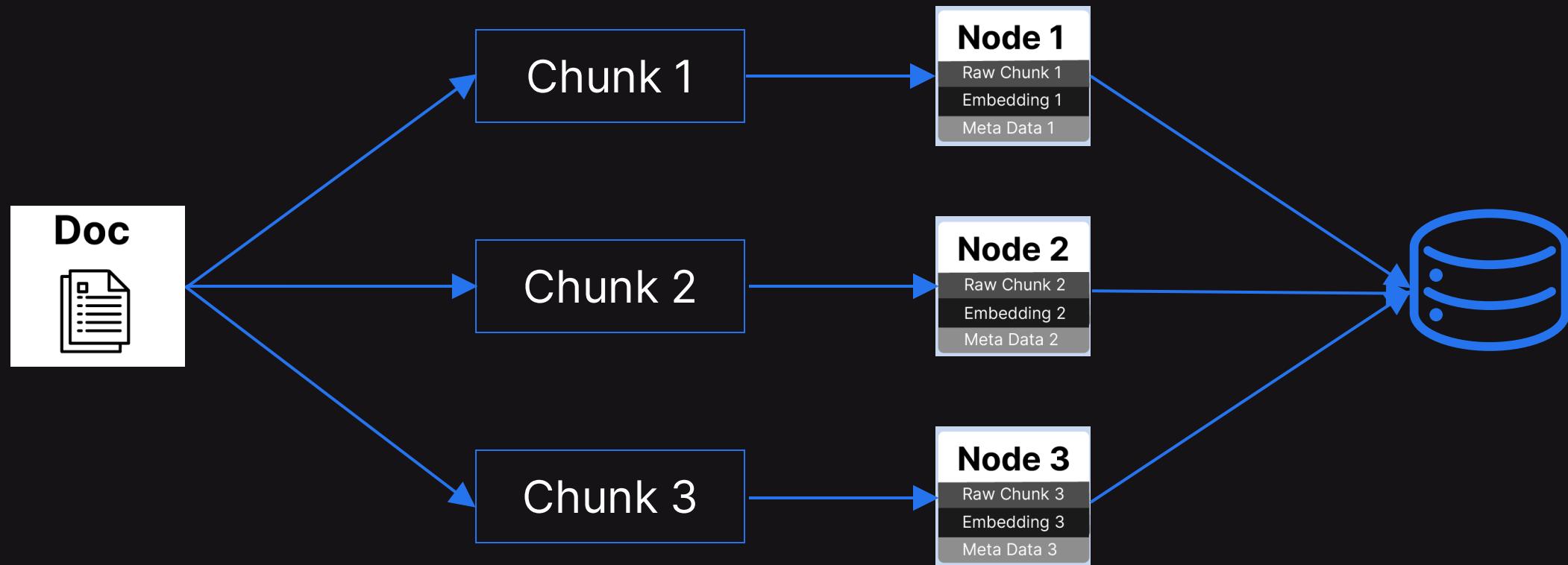
# 1B. Summary Index Embedding Retriever



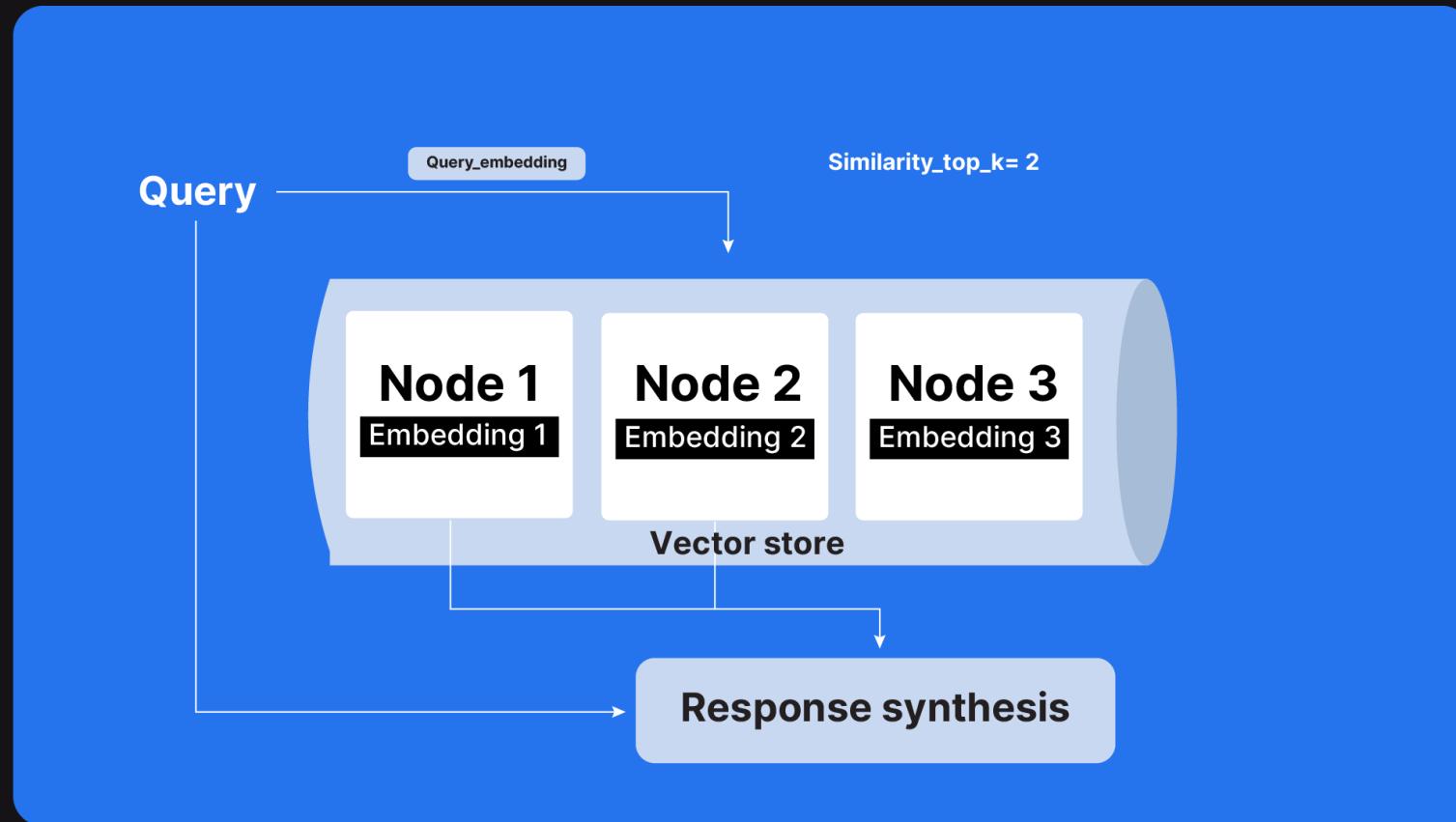
# 1C. Summary Index LLM Retriever



## 2. Vector Store Index



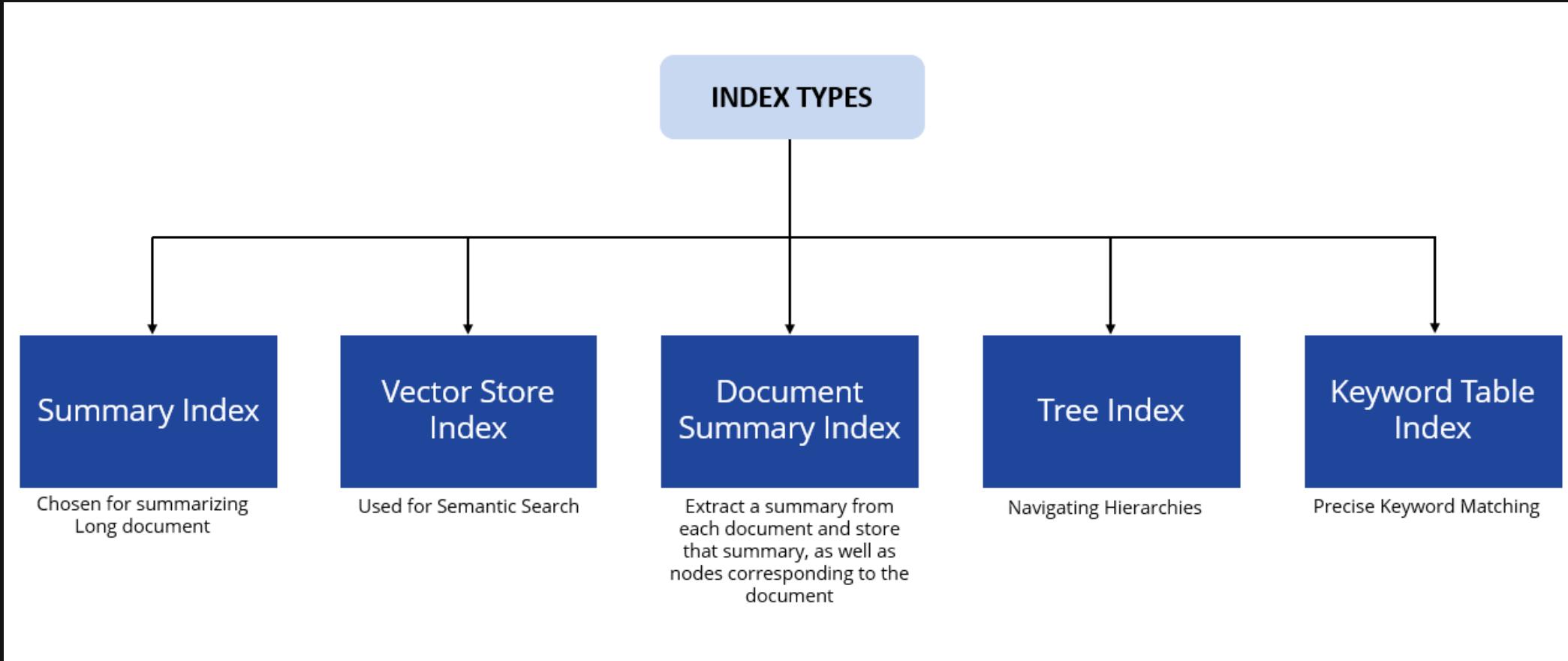
# Vector Store Index Retriever



# Vector Store Index vs Summary Index

- SummaryIndex stores all the nodes in the form of the sequence/list in the storage, unlike the vector storage index.
- Embeddings are created during the querying time rather than during index construction itself.

# Index Types



# Thank You