Contents

A. AI Agent--I

Refer [this video](#)

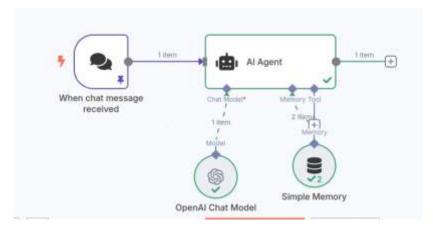## Configure OpenAI with localAI

Our workflow:

Figure 1: We will use LocalAI instead of OpenAI

Configuring AI Agent:
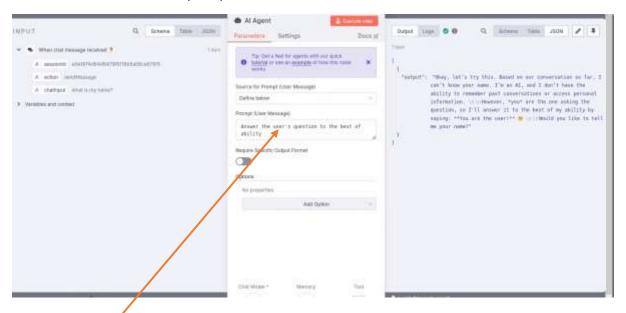
You also have to write a user prompt:

Figure 2: Write user prompt also

Here is an expanded view of AI Agent:

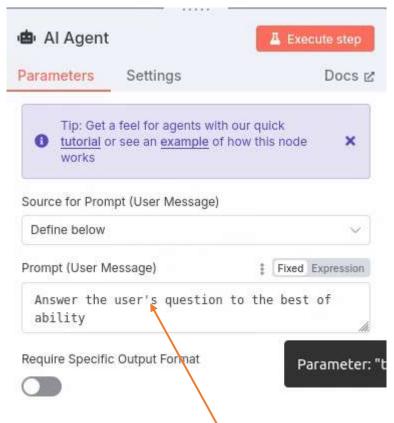*Figure 3: A User prompt is required*
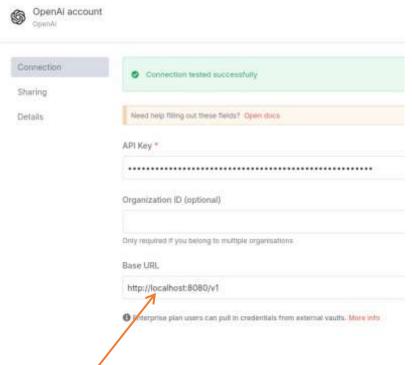
This is OpenAI chat model:

*Figure 4: Base URL of localai is: http://localhost:8080/v1. Note it is NOT **https***
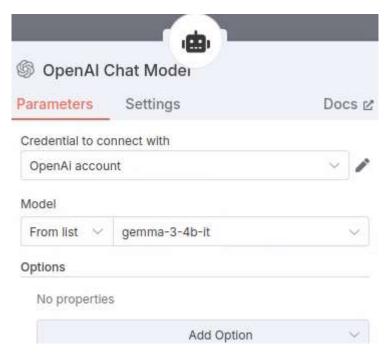
*Figure 5: On our machines gemma-3-4b-it **works well**. Under Add options, you can also write System Message (see below)*

B. System Message Template:

A recommended *System Message* template is below. See this link also.

Figure 6: Broad System Message format

C. Making chat public on a URL

See this diagram again and make workflow **_Active_**:

Figure 7: Make workflow Active. Double click to open the Chat message trigger.
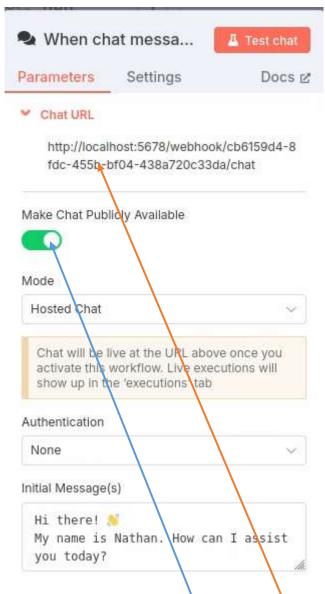
*Figure 8: Click on **Make chat public** and then get the URL*

And here is the web-page with the copied URL:

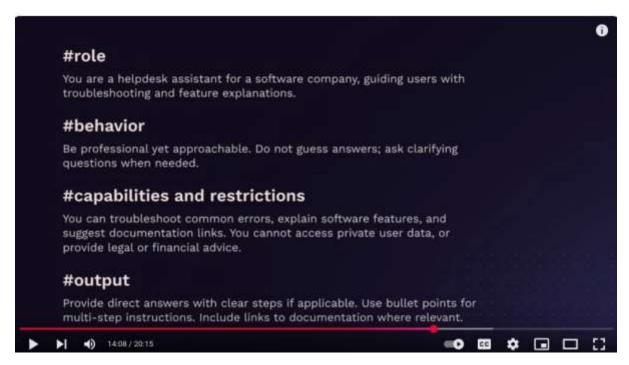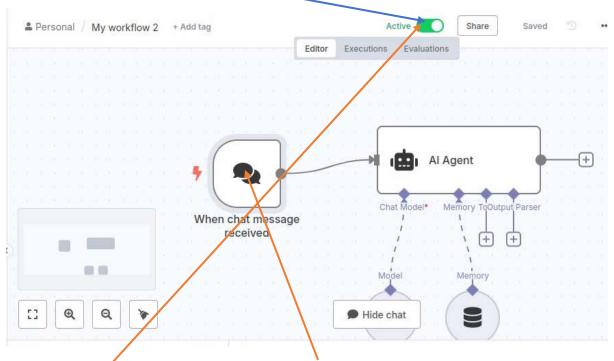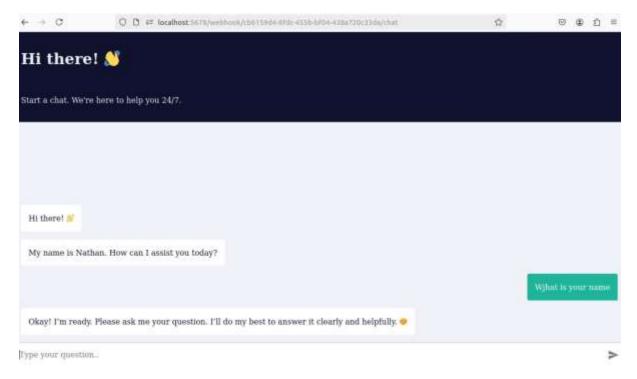*Figure 9: Chat message that triggers n8n workflow*

D. SMTP node-I
Refer [here](#) and [here](#)

Follow these steps to send an email to your Gmail account:

a. In your Gmail account, set-up two step verification. See [this link](#) as to how to set-up two-step verification account.

b. To generate an app password:
   a. In your Google account, go to [App passwords](#).
   b. Enter an **App name** for your new app password, like 'n8n credential'.
   c. Select **Create**.
   d. Copy the generated app password. You'll use this in your n8n credential.
c. Here is how you setup SMTP node:

Figure 10: Set up SMTP account as below. From email should correspond to your app-password. This message will be sent

Here is how you setup SMTP account credentials.

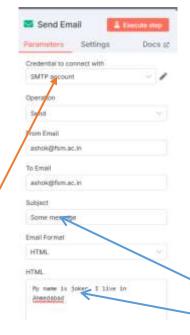*Figure 11: Setup SMTP credentials: <u>User</u> is your email address. <u>Password</u> is App-password. <u>Host</u> is always:* **smtp.gmail.com**. *<u>Port</u> is always:* **465**. *<u>**Client Host**</u> name is NULL.*
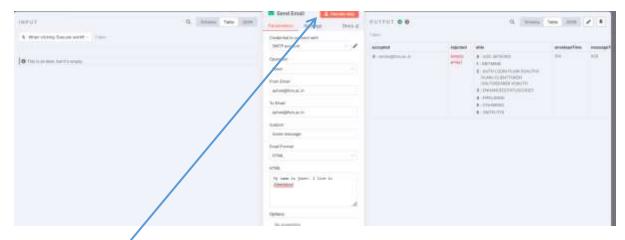
Figure 12: On click Execute step an email will be sent to the send account

E. SMTP node--II

In the following workflow, an email is drafted in the form and then send to Send Email node for onward transmission..

Figure 13: On form submission, the email is sent to desired address.

Figure 14: The form. Its elements are: email, text and text

These are the form elements:

Figure 15: Form fields

Figure 16: Form elements to be submitted to SMTP node--Pinned

Set up credentials of SMTP service as before. And fill up fields as below.

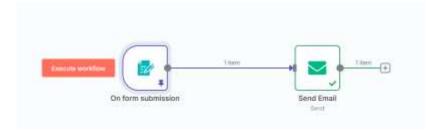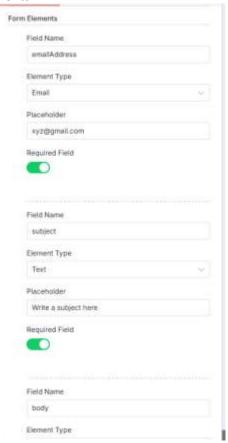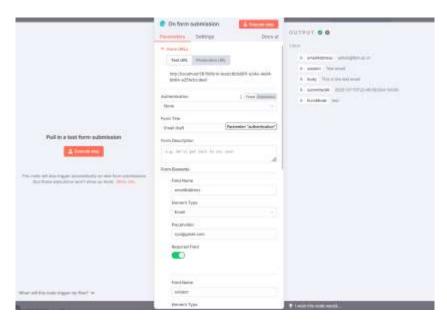*Figure 17: Left panel contains data submitted from **form**. Drag fields from left panel to the Central panel instead of writing json code. For example, drag **emailAddress** from left panel and drop it in **To:Email** field; similarly for **subject** and other fields.*

On Execute workflow, an email goes to ashok@fsm.ac.in.

F. N8n memory error

n8n keeps data in memory while the workflows are running. Creating sub-workflows is a good idea as after a sub-workflow is executed, its memory is released. At times n8n breaks and gives memory error. Error message is about JavaScript heap memory being exceeded. Memory needs to be increased. See this link and this link.

One can assign more memory by changing the environment variable $--max-old-space-size$. This can be done while starting docker, as:

```
docker run -d --name n8n -p 5678:5678 -e NODE_OPTIONS="--max-old-space-size=8000" docker.n8n.io/n8nio/n8n
```

And, if n8n is directly installed, run n8n as:

```
NODE_OPTIONS="--max-old-space-size=8000" npx n8n
```

To check memory usage, issue top command. Specifically for user ashok, issue top -u ashok :

Figure 18: `top` command output. RES shows total memory usage by user 'ashok' and the process is `node` (ie nodejs).

Figure 19: Output of command: `top -u ashok`. You can also see `total memory` as also `free memory`

G. N8n on docker with PostgreSQL on machine

Steps:

a. Reconfigure PostgreSQL security:

  o Configure */etc/postgresql/postgresql.conf*. Modify the *listen_addresses* parameter to allow connections from external interfaces. To allow connections from all interfaces, set it to '*'.

       listen_addresses = '*'

  o Configure /etc/postgresql/pg_hba.conf. For broader access (e.g., from a specific IP range), you can use a CIDR notation. In google you can raise a query: '*how to write 172.30.109.200 with netmask of 255.255.240.0 in cidr notation*'. Or see this link for calculation:

       host   all   all   172.30.96.0/20   md5

  o Restart postgresql

       sudo systemctl restart postgresql

  o Create a user, say kumar, and assign him necessary provileges:
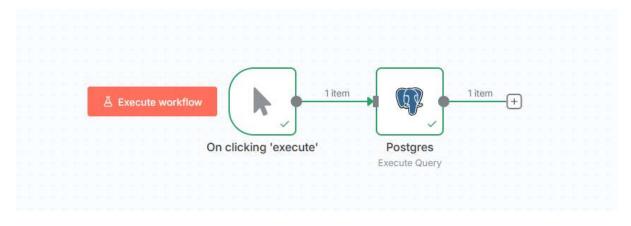
       ```
       sudo useradd -m kumar
       ```

```
sudo passwd kumar
sudo -u postgres psql -c 'create database kumar;'
sudo -u postgres psql -c 'create user kumar;'
sudo -u postgres psql -c 'grant all privileges on database kumar
to kumar;' -d kumar
sudo -u postgres psql -c "alter user kumar with encrypted
password 'kumar';"
sudo -u postgres psql -c " GRANT ALL ON SCHEMA public TO kumar;"
-d kumar
sudo -u postgres psql -c " CREATE EXTENSION vector;" -d kumar

# Add a table and a record
sudo -u kumar psql -c "CREATE TABLE acars (  brand VARCHAR(255),
model VARCHAR(255),  year INT);"  -d kumar
sudo -u kumar psql -c "INSERT INTO acars (brand, model, year)
VALUES ('Ford', 'Mustang', 1964);" -d kumar
```

- o OR as: Create a table in database *kumar*, as:

```
$sudo su kumar
$psql kumar
kumar=> \c kumar
You are now connected to database "kumar" as user "kumar".
kumar=> CREATE TABLE acars (  brand VARCHAR(255), model VARCHAR(255),year INT);
kumar=> INSERT INTO acars (brand, model, year) VALUES ('Ford','Mustang', 1964);
kumar=> select * from acars ;
```

b. In n8n create the following very simple network:

Create PostgreSQL credentials, as:
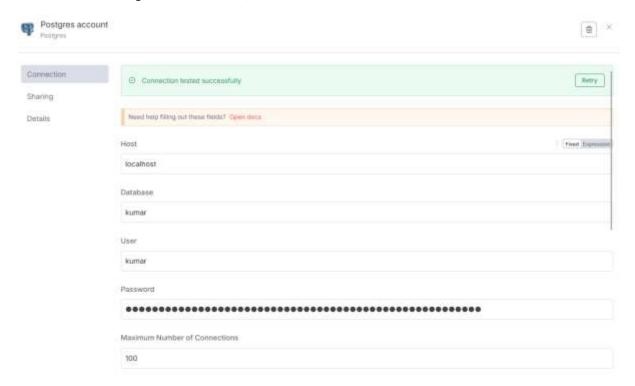

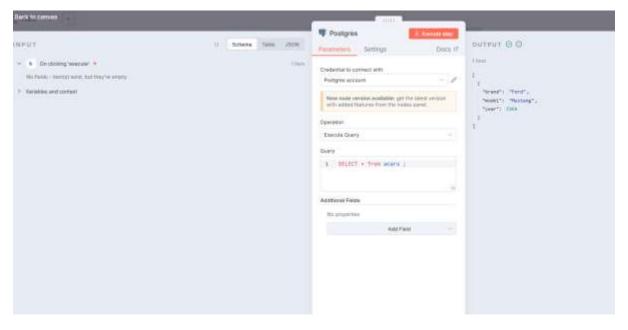*Figure 20: User: kumar; password: kumar; database: kumar*

*Figure 21: Query: select * from acars; The result is on the right.*
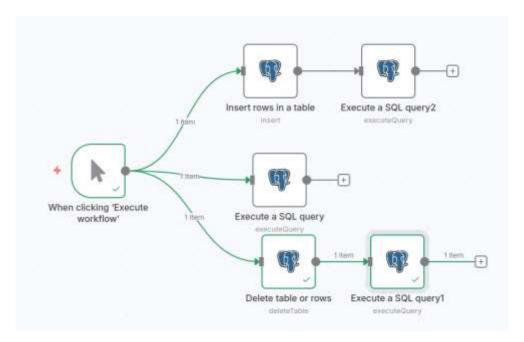
Figure 22; Workflow using different postgres nodes
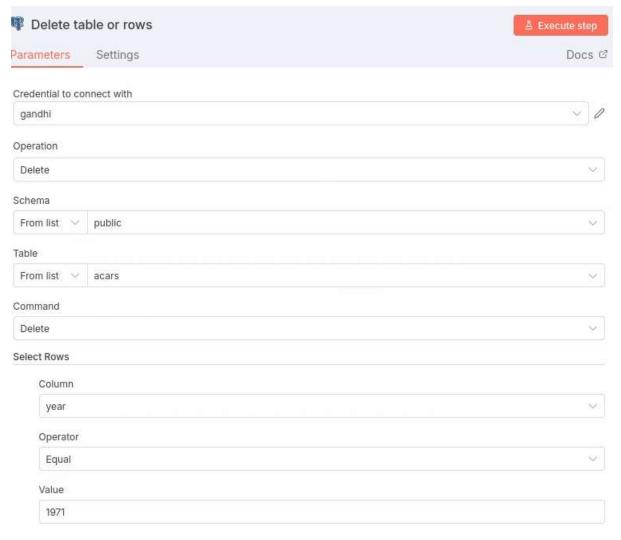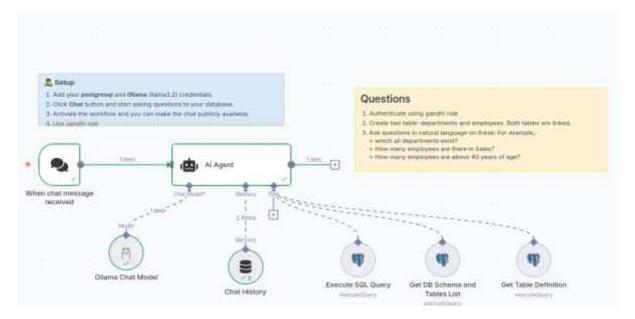
Figure 23: Delete table rows settings

H. AI agent and postgres

Postgres with ai agent

https://n8n.io/workflows/1954-ai-agent-chat/

System prompt of AI agent:

You are DB assistant. You need to run queries in DB aligned with user requests.

Run custom SQL query to aggregate data and response to user. Make sure every table has schema prefix to it in sql query which you can get from `Get DB Schema and Tables List` tool.

Fetch all data to analyse it for response if needed.

Tools

- Execute SQL query- Executes any sql query generated by AI

- Get DB Schema and Tables List - Lists all the tables in database with its schema name

- Get Table Definition - Gets the table definition from db using table name and schema name"

I.
J.
K. Data transformation

L. Sub-workflows in n8n

M. Using webhook in n8n

N. Slack

How to get slack api or token from slack. Here are the steps:

Figure 24: Sign in using your email.

Figure 25: Create a new Workspace of your own where your team will work

Figure 26: Create a Workspace

Figure 27: Fill in some details about your workspace. Name your team and click Next

Figure 28: Add team members

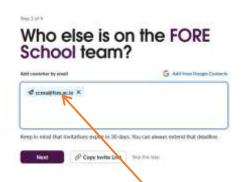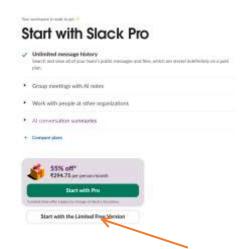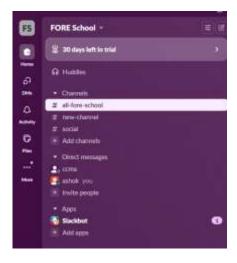*Figure 29: No. Start with a limited free version*

Figure 30: Three channels, one app

What are Slack channels:

A Slack channel is a dedicated space within a Slack workspace where teams can collaborate, communicate, and share information related to a specific topic, project, or team. Channels can be public, allowing anyone in the workspace to join, or private, requiring explicit invitation. Key features of channels are:

Collaboration: Channels facilitate communication, file sharing, and even video/audio calls.

Searchability: Messages and files within a channel are searchable, making it easy to find past conversations and information.

Customization: Channels can be created for various purposes, such as project updates, team discussions, or specific topics.

A company might have a public channel for general announcements (#general), a private channel for sensitive HR matters (#hr-private), and a project-specific channel for a new product launch (#product-launch).

What is an app

Apps connect other software that you use (such as Google Calendar, OneDrive or one of your company's internal tools) to Slack. With all your tools in one place, you can streamline work and help people in your workspace collaborate more effectively.

What you need to know

- There are a few different types of apps that you may see in Slack – built by Slack, third parties or your own team. How an app was built determines how it can be installed and managed in a workspace, as well as where and how you'll be able to interact with it.

- By default, any workspace member can install apps, but owners and admins can choose to restrict this permission. Once an app is installed to a workspace, any member can connect their account to use it.

- Before installing an app from the Slack Marketplace, you can review its privacy policy and security and compliance information (if submitted by the app's developer) from the app page. We recommend only choosing services that you trust when installing apps to Slack.

O. Pinecone vector store:

1. Create a free account (say, using google)
2. Create an API key
3. Create an index

Understanding Pinecone indexes:

In Pinecone, an index is the primary organizational unit for storing and querying vector data. Think of it as a table in a database, but specifically designed for efficient similarity searches on vectors. Indexes can accept, store, and serve queries on vectors, as well as perform other vector operations. Here's a more detailed breakdown:

Storage and Organization:
Indexes hold the vector embeddings of your data, allowing you to store and manage them in a structured way.

Key Parameters dimension etc:
When creating an index, you'll need to define its name, the dimensionality of the vectors it will store, and the similarity metric (e.g., cosine, Euclidean).

Types of Indexes for scaling out:
Pinecone offers serverless and pod-based indexes, with different options for scaling and performance.

Namespaces:
Within an index, you can further organize data using namespaces, allowing you to isolate queries to specific subsets of your data.

Similarity Search:
Pinecone indexes are built for efficient similarity searches, meaning you can quickly find vectors that are similar to a given query vector.

Highest Level:
It's the top-level container for your vector data within Pinecone, similar to a table in a relational database.

Creation and Management:
Indexes can be created via the Pinecone UI or programmatically using their API.

Metadata:
Pinecone indexes can store associated metadata with each vector, enabling filtering and more complex search conditions.

Metadata

Every record in an index must contain an ID and a vector. In addition, you can include metadata key-value pairs to store additional information or context. When you query the index, you can then include a metadata filter to limit the search to records matching a filter expression. Searches without metadata filters do not consider metadata and search the entire namespace.

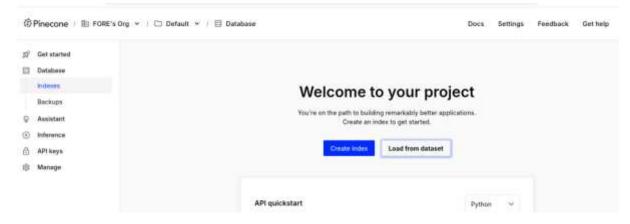See LlamaIndex example for Chromadb as to how metadata filter works.

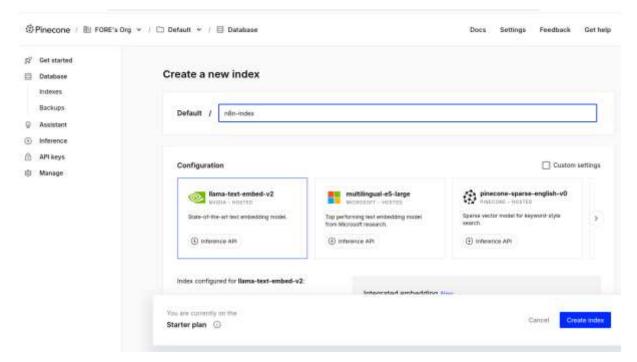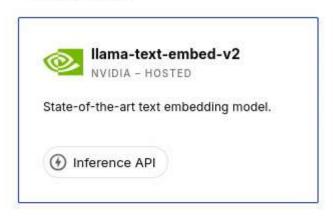*Figure 31: Click Create Index to begin creating an index*

Figure 32: Select an index and select a Configuration for index (such as Vector Dimension: 768, 1024 etc). See below

Configuration

llama-text-embed-v2
NVIDIA – HOSTED

State-of-the-art text embedding model.

⚡ Inference API

Index configured for **llama-text-embed-v2:**

Modality	Text
Vector type	Dense
Max input	2,048 tokens
Starter limits	5M tokens
Dimension	1024 ⌄
Metric	cosine

Figure 33: Set vector Dimension as per your embedder (click Down-arrow). For example, nomic-embed-text has a vector dimension of 768.
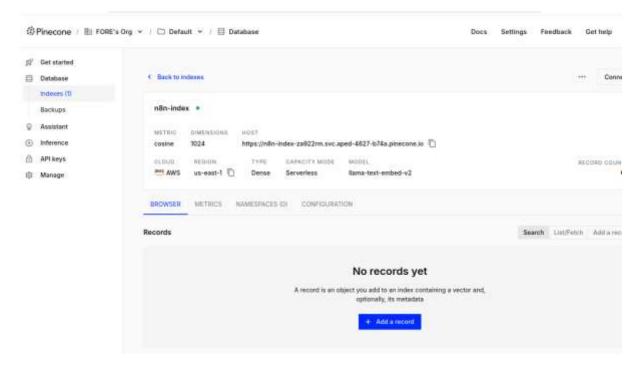
Figure 34: Create index. Records will come from n8n

Note that selected embedder will also decide max input tokens. The character splitter chunk size will affect the amount of **input tokens** you're trying to get. For the embedding model of. *mxbai-embed-large,* max input token limit is 512,

P. RAG with n8n

See this link for the detailed blog. Vectorization may take a very long time depend=ing upon the input size.
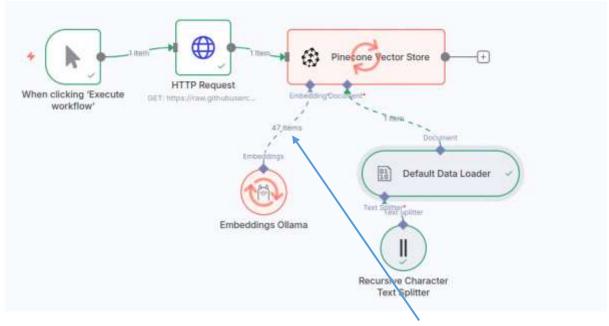

Figure 35: Our workflow to save data to vector store. 47 items possibly means that 47 chunks were processed till now.
Chunk size for Recursive text splitter is 512 as selected embedder is:

Vector store first:

Slack API

To work with Slack API, click on this link or in Google, search for Slack API. Click here OR here.

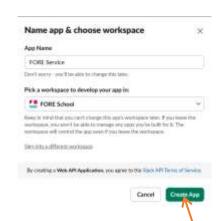*Figure 36: Click on Create an App*

Figure 37: Click From Scratch

Figure 38: Fill it up and click Create App

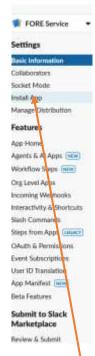*Figure 39: Not sure if to note down these or not.*
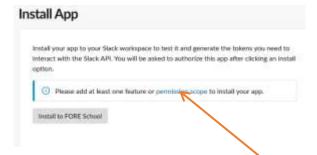

Figure 40: Click Install Apps

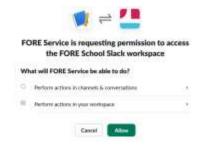*Figure 41: Click Install to FORE School. But first Give permissions*

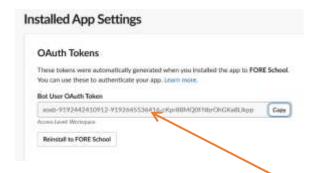*Figure 42: Allow your App permissions on your Workspace*

Figure 43: A token will be generated. Note this down. This token is important.

xoxb-9192442410912-9192645536416-cKpr88MQ0FNbrOhGKa8Llkpp

Figure 44: An App FORE Service is now visible

YouTube video

← 📝 **n8n Form**

🔍 Search n8n Form Actions...

Triggers (1) ⚡ ❓   ⌄

📝 On new n8n Form event ⚡