**THE ULTIMATE LOCAL AI SETUP IS HERE: n8n, Ollama & Qdrant - Installation Guide**

**Transcrito por** [**TurboScribe.ai**](https://turboscribe.ai/pt/?ref=docx_export_upsell)**.** [**Atualize para Ilimitado**](https://turboscribe.ai/pt/subscribed?ref=docx_export_upsell) **para remover esta mensagem.**

Depois de seguir este guia de instalação, você terá instalado o N8n, Quadrant, Olama e PostgreSQL localmente no seu PC. O N8n é o melhor toolkit para aqueles que querem construir aplicações RAC e agentes AI sem ter de escrever uma linha de códigos. Enquanto o N8n oferece uma inscrição de cloud no seu site, a aplicação é totalmente aberta e você pode instalá-la no seu PC gratuitamente.

Mas apenas instalar o N8n no seu PC não é suficiente. Se você quer construir agentes AI, você também precisa de um database de fatores, como Quadrant e, é claro, Olama, para gerar modelos de língua maior no seu PC. Mas o melhor jeito de instalar todos esses componentes e deixá-los trabalharem juntos é usando uma aplicação chamada Docker.

E se você não trabalhou com Docker no passado, isso pode soar um pouco intimidante, mas confie em mim, se você seguir os passos exatos que estou mostrando para você, você terá N8n e todas as outras aplicações instaladas nos próximos 10 minutos. Então, vamos começar com a instalação de Docker. E se você está trabalhando em um pouco de computador, você também tem que ter certeza de ativar SVM no seu BIOS, que significa máquina virtual segura.

E se você já fez tudo isso, vamos continuar com o próximo passo, que é a instalação de Docker. Então, apenas navegue para docker.com e então vamos descarregar Docker para o desktop para nosso sistema de operação. Então, apenas comece o processo de instalação e passe pelos passos de instalação.

E uma vez que fizemos isso, podemos abrir o desktop do Docker. Aceita o acordo de licença. O Docker agora propõe criar um acounto, mas vou esquecer esse passo.

Esquecer esse também. E aí você verá aqui Docker no seu desktop. E se tudo está funcionando, isso significa que podemos continuar com o próximo passo, que é a instalação de Git.

Então, vamos minimizar o Docker. Então, eu vou abrir um browser, procurar por Git, e então navegar para o primeiro resultado do Google, que é esse, e isso deve lhe trazer para esse site Git-SCM.com. E aqui eu vou descarregar a versão de Git para o meu sistema de operação. Então, clique em Descarregar, e neste caso, eu vou descarregar para Windows.

Então, para Git também, vamos navegar para o processo de instalação. Há muitas opções, mas a opção de default é sempre a correta. Então, apenas clique em Próximo.

E agora vamos descarregar a última aplicação, que é o VS Code. Apenas procure por VS Code e então descarregue VS Code para seu sistema de operação. E uma vez que o VS Code foi descarregado, eu vou maximizar ele, e você provavelmente tem um diretório em algum lugar na sua máquina, onde você constrói todos os seus projetos.

Então, eu vou criar um diretório assim, direto no meu desktop. Então, Novos Projetos Folder. E então, dentro do VS Code, eu vou navegar para Files, Abrir Folder, e vamos abrir este folder.

Então, selecione o folder. Então, eu vou confiar nos autores, e então vamos navegar no GitHub para a página da AI Starter Kit, que é esta página. Você pode encontrar o link na descrição deste vídeo.

Pressione o botão de baixo, e então vamos copiar este comando. Abra um terminal, assim. Apenas copie o comando aqui, e então pressione o botão de entrar, para clonar o repositório GitHub.

No meu caso, o comando GitHub ainda não está reconhecido, e isso significa que eu tenho que reiniciar minha máquina. Ok, então eu apenas reiniciei minha máquina, então vamos tentar de novo. E eu verei que o repositório GitHub foi baixado para a minha máquina.

Então, eu vou navegar de volta para Files, Abrir Folder, e desta vez, eu vou abrir o folder dentro deste folder, o folder que acabou de ser criado. Então, selecione este folder, e então clique em Selecionar Folder. Este folder contém alguns fios, e o fio Docker Compose é o mais importante.

Então, eu vou abrir este, e aí você vê aqui o conteúdo do fio Docker Compose. E antes de lançar este fio, eu quero fazer um pequeno mudando. Então, eu apenas navego para a linha 51, e em algum lugar nesta linha, você deve ouvir ou ver o set-up do database de Postgres, mas, por defaulto, este database não está exposto.

Então, eu vou copiar desta linha 32, os portos, e então navegar para onde você vê Restart, na linha 54, paste os portos, e apenas faça com que você veja assim, e então eu vou mudar o número de porto, porque este porto é o porto que o Ulama já está usando, para 5432, 5432, como este. E agora, nós deveríamos ser capazes de lançar este fio, então vamos navegar de volta para o repositório Getter, para as instalações de instruções. Ok, então, neste momento, nós já executamos este comando, e então você vê aqui que há instruções diferentes para diferentes usuários em diferentes sistemas de operação, e com diferentes GPUs.

Por exemplo, para NVIDIA, há este comando, e então para os usuários de AMD, GPU ou Linux, há este comando. Eu quero manter este vídeo tão simples quanto possível, então nós vamos procurar aqui para todos os outros, e então você vai copiar este comando, bem aqui, navegar de volta para o VS Code, de novo, abrir um terminal, assegure-se de que você já salvou seu arquivo do Docker Compose, então salvá-lo bem assim. E então, antes de nós executarmos este comando, nós vamos fazer com que nós tenhamos o Docker aberto, então apenas abrimos a aplicação do Docker.

Se você não vê ele abrindo, é muito possível que você o veja aqui no cartão do sistema, então apenas clique aqui. Então, você vê que o Docker está aberto agora, então nós vamos navegar de volta para o VS Code, abrir um novo terminal, de novo, assegure-se de que seu arquivo do Docker Compose já foi salvado, então copie o comando bem aqui e então apresse o botão de entrar. E então, se todos os paquetes já foram instalados, nós podemos abrir o Docker, e então você vê aqui o kit de iniciativa de AI self-host, e se você expandir, você vê todos os diferentes contêineres que nós criamos e que agora estão funcionando, exceto esses dois, esses dois não deveriam estar funcionando.

Ok, então este já foi um bom começo, mas em muitos casos, você também quer começar o Docker uma vez que você iniciou seu computador. Então, para isso, eu vou navegar para o ícone de ação, e então você vai ativar o Start Docker Desktop quando você se inscreve no seu computador, este aqui, e então clique em Aplicar e Restart. E em alguns dos meus próximos vídeos, onde nós estamos construindo uma aplicação de geração aumentada de retrievo, por exemplo, nós também vamos usar um modelo de embedding, e para instalar isso, eu vou navegar para o contêiner que contém o Lama, então apenas clique neste, e então navegue para Executar, e nesta página, você pode instalar novos modelos de língua-alemã.

Então, primeiro de tudo, vejamos uma lista de todos os modelos atualmente instalados, então eu vou apenas escrever Lista do Lama e apertar o botão de entrar, e então você vê que automaticamente o modelo Lama 3.2 já foi instalado, então isso é realmente bom. Então agora vamos continuar e também instalar um modelo de embedding, então navegue para o Lama.com, você pode encontrar o link na descrição, então a aplicação do Lama te permite gerar modelos de língua-alemã localmente no seu computador, o Lama já foi instalado como parte do Starter Kit Local AI, mas agora nós só queremos expandi-lo com mais modelos, então neste momento nós já temos um modelo instalado, que é o Lama 3.2 modelo, este é um modelo realmente bom, e é o modelo que eu vou usar na maioria dos meus vídeos, mas se você quiser instalar, por exemplo, o modelo QN3, você pode apenas clicar neste modelo, então este modelo com apenas 600 milhões de parâmetros é um modelo bem pequeno, mas ele vai funcionar em quase qualquer computador, mas se você então quiser instalar, por exemplo, este modelo, este é um modelo realmente grande, e ele não vai funcionar em nenhum dos nossos computadores consumidores, então você pode apenas instalar qualquer modelo que você quiser, o modelo que eu quero instalar, porque é realmente importante para os próximos vídeos nesta série de tutoriais, é um modelo de embedding, e é este, MXBAI Embed Large, então apenas clique aqui, e agora você vai copiar este comando, navegar de volta para o Docker, ter certeza que você está no botão Execute, paste-o, e então pressione Enter para instalar este modelo. Então, uma vez que a instalação está finalizada, eu vou navegar de volta para a janela anterior, Containers, e então nós vamos abrir o N8n Application, então apenas pule por esta linha aqui, e então clique neste ícone para abrir a aplicação.

Isto vai iniciar seu browser, e então a primeira coisa que temos que fazer é instalar nosso aconto de proprietário, então eu vou apenas proporcionar meus credentos aqui, clique em Próximo, e agora estamos na aplicação do N8n, onde podemos criar workflows, então vamos clicar aqui em Criar Workflow, então vamos clicar aqui em Advanced Step, e então vamos clicar aqui em On Chat Message, então isto vai trigar um fluxo tão rápido quanto uma nova mensagem de chat está sendo recebida, então clique aqui, então vamos apenas navegar de volta, porque queremos expandir o fluxo com uma ação que está sendo executada, então clique neste, clique em AI, e então vamos ir para uma linha básica de LLM, e isto nos permite falar com um modelo de linguagem grande, claro que também temos que adicionar aqui o modelo que queremos falar, então vamos clicar no ícone de Plus, e aqui você pode, porque é N8n, você pode basicamente falar com qualquer modelo, então se você quiser falar com OpenAI, você pode fazer isso totalmente, mas porque este tutorial está focado em construir agentes de AI locais, nós, é claro, vamos trabalhar com o LLM, e vamos escolher para o modelo de LLM Chat, então aqui você pode escolher entre os modelos que já instalamos, este é um modelo de embedição, então não podemos realmente falar com ele, mas podemos falar com o LLM 3.2, então vamos tentar isso, vamos voltar, e então clique em Save, e então vamos abrir um chat para falar com o modelo, você vê que nossa mensagem está sendo enviada para o modelo de LLM, e o modelo de LLM agora vai falar com a gente, dependendo das especificações de hardware do seu PC, você vai receber mensagens mais rápido ou mais devagar do que eu, você vê aqui que recebemos o texto, e isso significa que tudo está funcionando como esperado, agora nós temos sucessivamente instalado o N8n Toolkit localmente no seu PC, espero que você esteja pronto para ir para o próximo passo, mas um dos casos mais populares de uso com a AI Generativa é construir um chatbot que pode usar seus próprios dados, e se você está pronto para construir este chatbot junto comigo, vamos fazer exatamente isso no próximo vídeo, espero vê-lo lá.

**Transcrito por** [**TurboScribe.ai**](https://turboscribe.ai/pt/?ref=docx_export_upsell)**.** [**Atualize para Ilimitado**](https://turboscribe.ai/pt/subscribed?ref=docx_export_upsell) **para remover esta mensagem.**