

DATASHEET

Deepgram On-premises



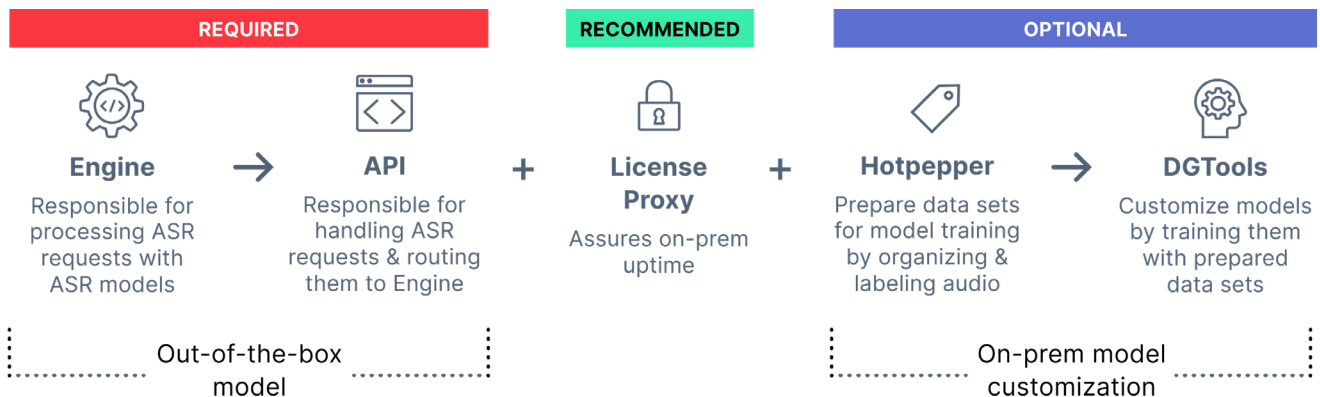
Brief overview of key benefits and components

- Security & data privacy compliance**
 Protect sensitive information, reducing risk, maintaining customer trust, and ensuring compliance.
- Deployment Flexibility**
 Deploy on bare-metal servers or in a virtual private cloud, with NVIDIA GPU-acceleration or traditional CPU compute.
- Massive Scalability**
 Container-based distribution supports horizontal scaling, enabling cost savings and efficiency through elastic deployments with load balancing that can respond to changes in demand.
- Low latency**
 Faster response times, improved user experience, and co-located data processing for time-sensitive applications.
- Low cost**
 Greater affordability and accessibility of products and services, leading to increased adoption and usage.

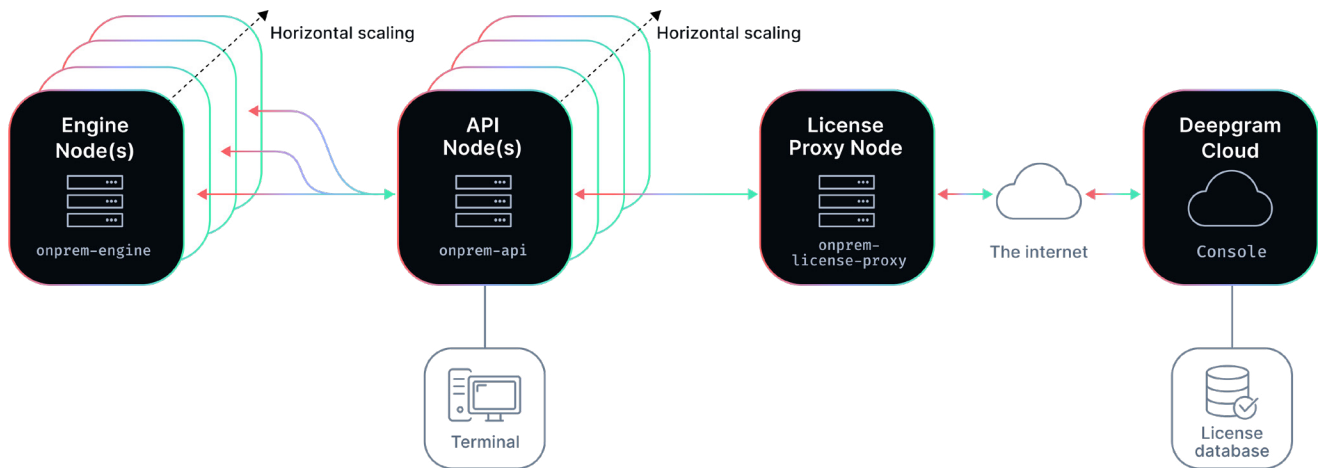
What does on-prem mean?

Simply put, on-prem is the software that is deployed. Specifically, it's a user-hosted automatic speech recognition (ASR) solution which is updated on a monthly basis.

On-prem deployment components



Deployment outline



On-prem node types

- API node
- Engine node
- License proxy node
- On-prem model customization node*

Environment

Deepgram On-prem is typically deployed to a virtual server in the cloud, but it can be deployed to a baremetal server as well.

*Not pictured



Deepgram supports a variety of deployment methods, including an on-premises (on-prem) offering, which is an isolated service deployed to customer-requisitioned cloud instances or customer data centers.

From the [On-Prem Deployment guide](#)

About Deepgram

Deepgram is a foundational AI company on a mission to understand human language. We give any developer access to the most advanced speech AI transcription and understanding with just an API call. Our models deliver the fastest, most accurate transcription alongside contextual features like summarization, sentiment analysis, and topic detection. Learn more at deepgram.com or [contact us](#).

Deepgram

Essential Building Blocks for Language AI

BLACK BOX®

855-324-9909 | BLACKBOX.COM