# Guide RAG - Retrieval-Augmented Generation

## Introduction au RAG

Le RAG est une technique qui combine la recherche d'information
et la generation de texte pour creer des systemes d'IA plus precis.

## Architecture

1. Ingestion des documents : Les documents sont decoupes en chunks
2. Vectorisation : Chaque chunk est transforme en vecteur d'embedding
3. Stockage : Les vecteurs sont stockes dans une base vectorielle
4. Recherche : Les requetes sont vectorisees et comparees aux documents
5. Generation : Un LLM genere une reponse basee sur les documents trouves

## Avantages du RAG

- Reponses basees sur des sources fiables et a jour
- Reduction des hallucinations du modele
- Possibilite de citer les sources utilisees
- Adaptation facile a de nouveaux domaines