Intel Xeon processors are designed for enterprise and data center workloads, providing exceptional performance for AI and machine learning applications. These processors feature multiple cores, large cache sizes, and advanced instruction sets optimized for computational intensive tasks.

The Qdrant vector database is a high-performance vector similarity search engine built in Rust. It provides fast and accurate semantic search capabilities, making it ideal for retrieval-augmented generation (RAG) systems. Qdrant supports various distance metrics including cosine similarity, dot product, and Euclidean distance.

RAG systems combine the power of large language models with external knowledge retrieval. The process involves encoding documents into vector embeddings, storing them in a vector database, and retrieving relevant context based on user queries. This approach enables AI systems to access up-to-date information beyond their training data.

Text embedding models like BAAI/bge-base-en-v1.5 convert text into dense vector representations that capture semantic meaning. These embeddings enable similarity searches across large document collections, forming the foundation of modern information retrieval systems.

Reranking models improve search quality by reordering retrieved documents based on their relevance to the query. The BAAI/bge-reranker-base model provides state-of-the-art reranking capabilities for RAG pipelines.