

Abstract:

For my distributed systems and applications project this summer I have decided to build a distributed text analysis pipeline using the Ray framework in Python. This built system will take a large dataset of inputs (undecided on the type of inputs) and divide them into smaller chunks that will be processed in parallel. Each chunk segment will be sent to an LLM API like OpenAI or HuggingFace to perform a task like sentiment analysis or summarization. Overall I will measure the performance improvements compared to a normal baseline and analyze the systems's scalability and efficiency given different workloads.