Overview

Keywords: Data Denormalization, Stream Processing, Distributed Databases, Search Indexes, Bulk Processing

Modern data intensive software stacks consist of much more than a relational database. EZStack represents a new, innovative way to think about data denormalization. The goal of the project is to provide a highly scalable, highly efficient, eventually consistent data stack. This product is designed to benefit big-data companies in achieving an efficient, highly scalable data processing system with a simple interface, thus eliminating engineering difficulties within the data model. While most other big data stacks attempt to tackle these problems by investing large amounts of resources in developing their own custom solutions that is highly specific to their case, EZStack will attempt to provide a general solution that is applicable to many different use-cases, favoring ease-of-use and customization.

Intellectual Merit

This Small Business Innovation Research Phase I project strives to solve major challenges related to big-data systems, while simplifying their usage. High performance in this generic system will be achieved by leveraging a streaming data architecture. EZStack will have two major distributed datastores, a write optimized system of record and a search index. In between these two datastores will be a distributed denormalizer that transforms the normalized data in the system of record into more searchable formats. The transformations performed by the denormalizer will be determined by advanced query analysis. Additionally, EZStack will be extremely straightforward to deploy. Administrators will simply install an agent onto every node and EZStack will manage itself.

Broader Commercial Impact

Educators, government agencies, large corporations, and fast-growing startups have the same problem: They have a need to create connections between datasets in order to enhance our environment, provide useful statistics, or increase user connections. Big data is a large topic that has many issues that need to be tackled. Currently, there exists no system that can provide the ease-of-use, scalability, and performance boost that EZStack aims to achieve. EZStack should provide a plug-and-play solution for applications with many different data models, allowing engineering teams to easily and quickly deploy to production. If this project is successful, it will lower the barrier to entry of educators, government agencies and many other entities to develop data intensive applications.