Elevator Pitch

The Market Space

There is a very clear target customer for our application. Smaller software development teams that would normally not have the financial capabilities required to have access to big data management are very handcuffed in their capabilities. Access to big data management systems can be extremely beneficial for companies in almost every field, and currently it is often far too expensive for companies that are not large corporations to develop systems that can efficiently manage data. Additionally, large companies that are capable of creating big data systems must spend large amounts of resources on those systems. Additionally, if a system needs to be altered to match new data, significant new costs will often be incurred. The goal of EZstack is to address both of these problems with one system that is both easily deployable and economical for all customers, regardless of financial means.

The Value Proposition

Currently, there exists no system that can provide the ease-of-use, horizontal scalability, and performance 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 startup companies, research teams. and many other entities to develop data intensive applications. There is a significant need to create connections between large datasets in order to enhance our environment, provide useful statistics, and enrich user experiences. Big data is an expansive field with many issues that need to be tackled, and it is the goal of EZstack is to do exactly that. The overall goals of the EZstack system are guaranteed efficiency for both reads and writes, horizontal scalability, and simple deployability without the significant costs that such a system would normally require.

The Innovation

EZstack 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 read-optimized 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.

Project Summary

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 are 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.