Commercial Opportunity and Social Impact

Commercial Opportunity

EZstack has an extremely vast potential user base. Potential users include large corporations, fast-growing startups, government agencies, researchers, and hobbyists. The product is engineered to work for customers at any scale, whether that be on a single server, or on thousands of servers. The purpose of EZstack is to provide a scalable backend for any data intensive application. Creating a data intensive backend is extremely expensive and time consuming to engineer, maintain, and deploy. This difficulty is precisely the problem that EZstack will solve, and it is a strong market driver for the software.

The business model for EZstack is very similar to those of Red Hat, MongoDB, and Elasticsearch. The software itself will be free and open source. This means that the public at large will be able to use the software free of charge. Additionally, the underlying source code of EZstack will be publicly available, which allows anybody to contribute modifications and improvements to the project. The developers of EZstack will monetize the software via two main revenue streams. The first main revenue stream is the sale of service and support to the more affluent users of the platform. Service and support is essential for enterprise customers of large software, as they require guarantees that they will be able to meet their service level agreements (SLA's). The second source of revenue for EZstack is the sale of the software as a fully hosted service. This means that the developers behind EZstack will completely manage the deployment of the software for customers that wish to use it, but don't wish to be responsible for maintaining it.

EZstack is an entirely managed backend as a service software. With entirely managed services there are many increased risks that are transferred from the operations team deploying the system to the developers developing the service. In the field of managed services, customers are to assume that the boxes deploying the service are black boxes that are well secured and are accurate in their execution. Due to this, during development significant ops work must be integrated into the product to enable secure deployment and access to data. Encryption, managed access, and role validation have to be build into EZstack from the outset, otherwise it will be very impractical for EZstack to guarantee secure usage of its services.

EZstack also has to account for industry standard security tools used by many enterprise customers that would be pursued. EZstack must have plugin support for systems such as Splunk and Q-Radar to ensure internal visibility and alerting can be integrated in an enterprise security management model. EZstack also has other risks relating to reliability, uptime, and data integrity. EZstack's managed service must have clear and high Service Level Agreements (SLA) to ensure the users will be focusing on developing their best product without having to worry about the underlying infrastructure.

EZstack is a unique product that is integrating many open source projects under the hood, which significantly increases the difficulties of ensuring data flow throughout the system. EZstack's system is using open source projects such as Kafka, Samza, Elasticsearch, Cassandra, various app servers, and many other services. Those services have to have a reliable and transparent communication system and data integrity needs to be verified at every stage of the system. This project takes on risks similar to those taken by cloud providers and Software as a Service (SaaS) companies; however reaping the many benefits cloud providers such as Azure and AWS have would be worth the risks accompanied by this project.

Societal and Global Impact

In the world of the data intensive applications, there is a very pervasive problem: data intensive backends are very complex, to the point where engineering them can become nearly impossible. This causes many potentially amazing applications to be very expensive to deploy and maintain. Those extreme expenses make it prohibitive for groups of people that do not have sufficient funds to break into the global software market. Even though large corporations have the capital required to engineer data intensive applications, there still exists many major problems that they cannot avoid. The most significant of these problems is that many of the existing systems are inflexible. It is nearly impossible to make alterations to an existing stack because most major alterations require the system to be almost entirely re-engineered, which is extremely expensive.

EZstack will provide a way of avoiding all of the aforementioned issues. It will enable smaller teams that had no previous method of developing data intensive applications an opportunity to do so, as well as allow large companies to be able to deploy and maintain systems that they otherwise would have had to spend large amounts of resources on. The system is designed to be plug-and-play, so it is usable by *anyone*. Additionally, EZstack is built to run at truly massive scale, so there will be no need to re-engineer the system to scale up.

The immediate beneficiaries from EZstack are engineers. With EZstack, the systems that engineers will be able to create will provide significantly more flexibility and scalability at a fraction of the cost of existing backends. In addition, EZstack will be far less complex to maintain, so the IT personnel tasked with assembling and maintaining systems built on top of it will benefit greatly. The agony of engineering case specific systems from the ground up will be over with the introduction of EZstack. EZstack should be a transformational innovation in the field of software engineering.

Although the most direct impact of EZstack is on those who will be implementing data intensive application, the greatest impact of the project is on the users of the applications created using EZstack. Because applications built on EZstack will be extremely simple to engineer, developers should have significantly more time to develop new products and new features, which ultimately benefit users. This impact will be global, as EZstack has no technical limitations.

The ethical and environmental implications of EZstack are relatively minimal. The existence of EZstack does not directly affect the core functionalities of the applications it powers, therefore it is difficult to assign any ethical or environmental accountability to it. Not unlike most open source technologies, EZstack can indeed be used for unethical purposes. However, the culpability for such unethical uses lies on the engineers and the applications they create, not on EZstack itself. For this reason, EZstack should not require any regulation.

EZstack will be providing an avenue for all groups of people to have access to a scalable software stack for data intensive applications; no longer will teams be limited by a small budget. This will lower the barrier to entry for many start-ups and small businesses, enabling more people to contribute great things to society. This will have a disparate impact on many different sectors of the economy. The virtue of a flexible system is that it can apply to anyone and everyone; there are no limitations on what can be built with EZstack. The technology used to avoid re-engineering costs is the same technology that enables the system to be used in a near-infinite amount of use cases. The usefulness of EZstack is not limited to domestic users. Many international governments, corporations, and research groups have the same needs as American entities, and EZstack will help solve their problems as well.

If widely adopted, it should be very easy for people to create connections between massive datasets in order to enhance the environment, provide useful statistics, or increase user connections. There are no existing systems that can comparably accomplish this. The utilization of EZstack could facilitate scores of new disruptive technologies.

Elevator Pitch

The Market Space

EZstack will have a great amount of use for many different groups of people. 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 are no systems that can provide the ease-of-use, horizontal scalability, and performance that EZstack will 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 EZstack will do exactly that. The overall achievements 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 are no systems 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.