

EZstack:

A Distributed Database with a Built-In Data Denormalization Platform

Ben Devore, Beshoi Genidy, Sam Given

Background

- Complicated data retrieval actions are extremely slow, especially at extremely high scale
- Joining and aggregating data are particularly inefficient
- Amazon found that every 100 milliseconds of latency cost them 1% in sales
- EZstack attempts to drastically reduce latencies by performing real-time data denormalization

Data Denormalization

- Data denormalization is process of trying to improve the read performance of a database by adding redundant copies of data that have been modified via aggregations, joins, or other operations that restructure the data in ways that optimize retrieval

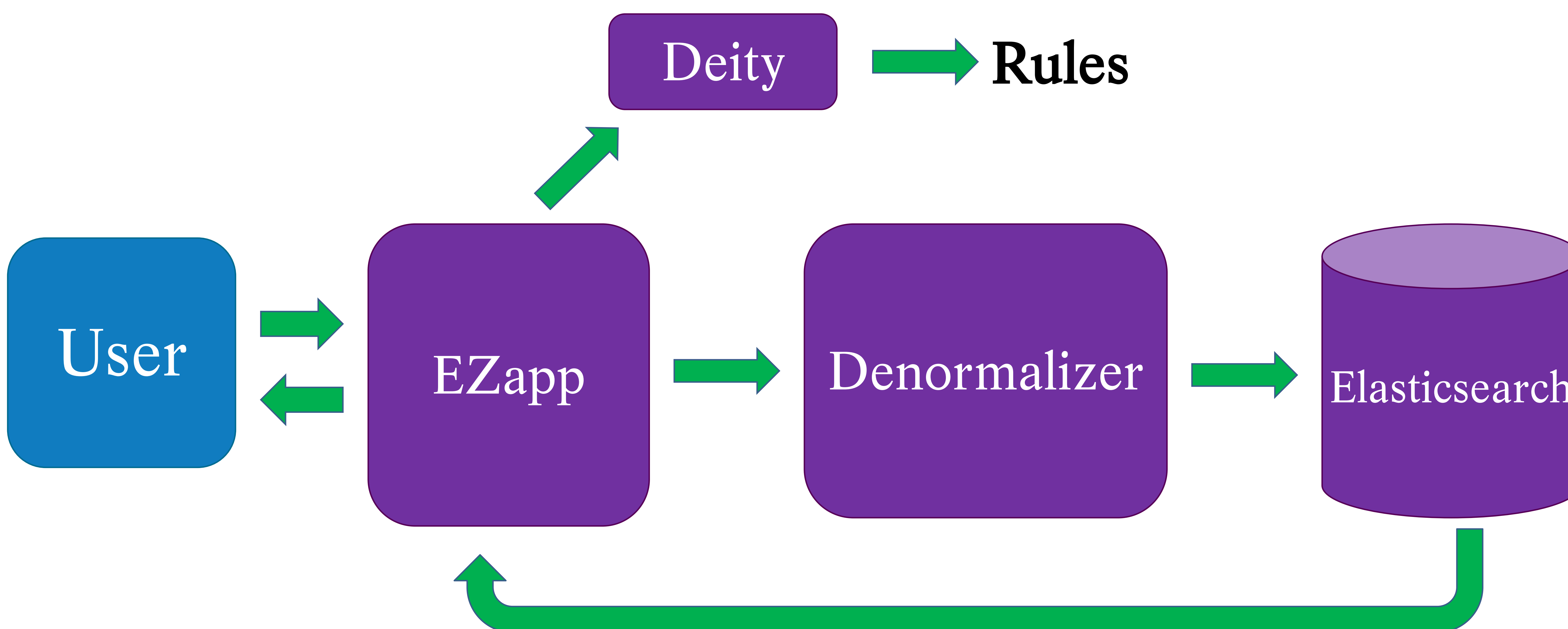
How It Works

- **EZapp**: a RESTful app server that the user interacts with in order to read and write data. It also responsible for both providing the Deity with query metrics and providing data about existing rules.
- **Denormalizer**: a distributed Apache Samza application that receives updates from EZapp and performs denormalizations according to existing rules in real time.
- **Deity**: receives query metrics from EZapp and analyzes them in real time to determine what rules would give the best performance increases.
- **Rules**: The set of denormalizations that the denormalizer is instructed to perform to make data retrieval more performant. Stored in Zookeeper.
- **Elasticsearch**: a distributed, sharded search index that EZstack uses for persistence of its data.

Technologies



System Architecture



Results

- As you can see in the graphs below, the results of our system show performance increases over **90%** when rules are used

