A high volume, low latency cloud-based data collection and reporting system

Gyuhyong Jeon, gxj16103@utdallas.edu; Kha Nguyen, Kha.Nguyen1@utdallas.edu; Chase Cloutier, cxc155430@utdallas.edu; Lam Phan, ltp140030@utdallas.edu

Abstract
The goal of this project is to implement a cloud-based data capture (logging) system that receives and stores incoming JSON data messages and provides both automatic rules processing and user initiated queries. The JSON messages will be received from an Azure Event Hub. Queries, rules and reports will be managed via a web page.

Architecture

Impact
Currently, SkyKick is paying roughly $1500-$1800 a month to store and search their log messages from a multitude of Cloud-based servers and the clients. We plan to provide a similar solution to store and query the data in a fraction of the cost. Based on our estimates, our service will cost anywhere around $350-$450 a month.

Results

Performance
Performance was one of the biggest concern while working on this project. The requirement was to index 500 logs/sec. We used 4 cores 32GB Ram VM and was able to index 6000 logs/sec. The CPU was sitting around 50~60%, so there are some room to grow in the future.

Summary
We learned a lot about development process. Software Engineering does not only include coding, but also include a lot of communication, planning, and decision making. We had to draw the high level diagram, demonstrate a Proof of Concept, and plan the long-term goals based on the software requirements. It was a really good chance to learn real-world software engineering experience and work with our sponsor (customer).