ElasticSearch CaseStudy

Business Problem

o Tokioo is a social utility which lets users socialize with friends by unifying other social & storage providers. It provides unified way to access storage files of users stored across different storage accounts like DropBox, 4Shared, SkyDrive and so on. It provides console to manage files across these providers, categorizes these contents to make it more user friendly, allows user to search across these storage accounts for specific keywords.

o Fetch huge amount of storage & contact information from various providers for millions of users

o Provide faceted search

o Search capability for binary files (.doc, .pdf, .xls, .txt, .rtf etc)

Business Solution

o High performance by using ElasticSearch - an OpenSource - indexing server

o Horizontal scalability achieved using Amazon cloud

o High availability achieved by Sharding & Replication provided by ElasticSearch

o Reliability - Fault-tolerance to partial and full cluster failure

Business Challenges

o The content is spread over the internet across

o High performance for real-time user requests

o Unified search capability across storage & social accounts

o Scalability of search server

Technical Solution

o Background processes fetch storage data from configured storage providers and it is indexed to ElasticSearch server clusters

o All the user requests for storage use-case are served by ElasticSearch index to provide high performance

o New index is created for each subscriber's storage accounts

o Indexes are divided into shards and replicated across cluster nodes. This provides high availability along with fault-tolerance to partial cluster failure

o Use write-behind to configured gateway to recover from full cluster failure