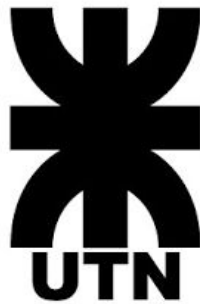


Ingeniería de Software de Fuentes Abiertas

Informe de comunidad de software libre



Profesor: Dr. Ricardo Medel

Alumno: Bustos Gustavo, 42843

Curso: 5K4_

Año 2017

Introducción	2
Historia	2
Comunidad Elastic	3
Descripción	3
Directorio o Toma de decisiones	3
Estructura	3
Modelo de Negocio y financiamiento	4
Otros financiamientos son Inversores de Capitales de riesgo	4
Herramientas	5
Clientes de productos de Elastic	6
Elastic Search - Software elegido	6
Mi participación	7
Links de Interés	11
Conclusión	12

Introducción

El presente informe trata sobre mi participación y colaboración en una comunidad de software libre llamada Elastic.co el cual es una comunidad que trabaja en el desarrollo y mantenimiento de varios productos o aplicaciones entre ellas las más importantes son: Elastic Search (de ahora en mas simplemente ES), Kibana, Logstash, etc. la idea es describir esta comunidad y dar a conocer sus productos que están siendo usados por muchas compañías importantes en el mundo entre ellos eBay, Mercado Libre, IBM, compañías de videojuegos como NVidia, Activision; medios de comunicación como el New York Times, BBC, etc; y compañías que se dedican al negocio de las acciones o bancos importantes como Goldman Sachs. El proyecto elastic brinda desde los productos antes mencionados con posibilidad de instalarse en tu propia compañía con servidores propios simplemente bajando el código fuente y compilarlo en esos entornos, por otro lado brinda servicios cloud el cual tienen un costo por hora y este se ve incrementado de acuerdo a la cantidad de nodos, memoria ram, almacenamiento en disco, etc. Otros servicios brindados son cursos el cual son pagos y se dictan en diferentes partes del mundo.

Por último describiré mi participación y respectivas conclusiones de mi experiencia en esta comunidad.

Historia

Elastic comenzó allá por el año 2010 cuando Shay Banon escribió las primeras líneas de ES e inmediatamente después liberó el código fuente, básicamente es su principal producto y da inicio a la comunidad. Dos años después en 2012 fue fundada la compañía Elastic en Amsterdam por gente que estaba detrás de los proyectos ES y Apache Lucene, de hecho ES está basado en este último.

El proyecto surge como una visión de que la búsqueda puede resolver una gran cantidad de problemas de datos entre ellos la búsqueda en tiempo real. Con la llegada de la computación en la nube, cambios en la infraestructura de IT y la demanda constante de nuevos tipos de requerimientos como el manejo de grandes volúmenes de datos, es entonces que Shay foresaw junto a Steven Schuurman, Uri Boness y Simon Willnauer compartieron la visión original de Shay uniéndose al equipo para crear la compañía Elastic. Tiempo después se crea la familia de productos llamada Elastic stack el cual se incluyen Kibana, Logstash, Beats, etc; esta familia incluye empleados alrededor de 32 países y la comunidad puede alcanzar a más de 100 países en el mundo.

Fuente: <https://www.elastic.co/>

Comunidad Elastic

Descripción

La comunidad es muy amplia y como se mencionó con anterioridad con gran variedad de productos sumandose mas al Stack actual, el lenguaje de programación usado en la mayoría de los productos es java y bajo licencia Apache 2.0 compatible con licencia GPL versión 3.0. Los creadores son Amsterdam de los Países Bajos, pero parte de sus directivos son de diferentes lugares del mundo.

Directorio o Toma de decisiones

Los miembros de la junta directiva son los siguientes:

1. Shay Banon, Fundador y CEO
2. Steven Schuurman, Fundador
3. Peter Fenton, General Partner, Benchmark Capital
4. Rod Johnson, Creador del Spring Framework, cofundador de SpringSource (ahora VMware)
5. Mike Volpi, Compañero, Index Ventures
6. Chetan Puttagunta, Socio general, New Enterprise Associates (NEA)
7. In Memoriam: Harry Weller, General Partner, New Enterprise Associates

Estructura

Es una comunidad muy grande y algunos de los números de la comunidad se pueden ver a continuación:
Son un total aproximadamente de 35.000 miembros.
Entre ellos hay: 15 administradores, 160 moderadores.

Estadísticas de la comunidad

	Todo el tiempo	últimos 7 días	Últimos 30 días
Temas	59.6k	489	2.1k
Posts	281.1k	2.9k	12.1k
Usuarios	35.1k	319	1.4k
Usuarios Activos	—	2.4k	4.7k
Likes	8.4k	189	643

Modelo de Negocio y financiamiento

El modelo de negocio se centra principalmente en los siguientes aspectos:

Partners, se compone de tres tipos de socios:

- Referidos: ayudan a obtener clientes comerciales y como consecuencia de ello pueden obtener un bono de referido respecto al valor de una venta del primer año de suscripción.
- Premium: Completa las clases oficiales de capacitación técnica Elastic y realiza inversiones anuales de capacitación o desarrollo de mercado, revende suscripciones comerciales a los clientes finales.
- Socios profesionales revendedores, están conociendo Elastic y tienen un modelo de negocio orientado a la distribución, no tienen requisitos de formación inicial pero pueden revender la suscripción comercial a clientes finales

Soporte/Supporting: Brinda soporte de los productos que provee en la nube como ES, Kibana entre otros con un costo mínimo suscripción de u\$d 41 la hora e incrementándose de acuerdo a las necesidades de los clientes, por ejemplo permite elegir: número de nodos, memoria ram, almacenamiento en disco. Jugando un poco con las features se pueden visualizar costos mensuales de u\$d 4800 al mes y en algunos casos se pueden conseguir descuentos.

Cursos/Training: Obtiene ingresos de cursos que se dictan en diferentes partes del mundo algunos que he consultado están en un promedio de u\$d 1600

Otros financiamientos son Inversores de Capitales de riesgo

Benchmark Capital: es una empresa de capital de riesgo en fase inicial que se centra en empresas sociales, móviles, locales y en la nube que interrumpen diversas industrias. Fundada en 1995, la firma tiene oficinas en Menlo Park y San Francisco, California. Desde su fundación, Benchmark ha invertido en más de 250 nuevas startups con un valor de mercado superior a \$ 100 mil millones. Las inversiones han incluido eBay, Juniper Networks, MySQL, OpenTable, Yelp, Zillow, Friendster, JAMDAT, Instagram, Hortonworks, Dropbox, Uber, Twitter, Zipcar, Asana, Quora, Gaikai, Demandforce y DOMO.

Index Ventures: es una firma líder mundial de capital de riesgo activa en tecnología y biotecnología que invierte desde 1996. Desde sus oficinas en Ginebra, Londres y San Francisco, les encanta descubrir y apoyar a los equipos más emprendedores dondequiera que busquen construir mercados, Los negocios globales.

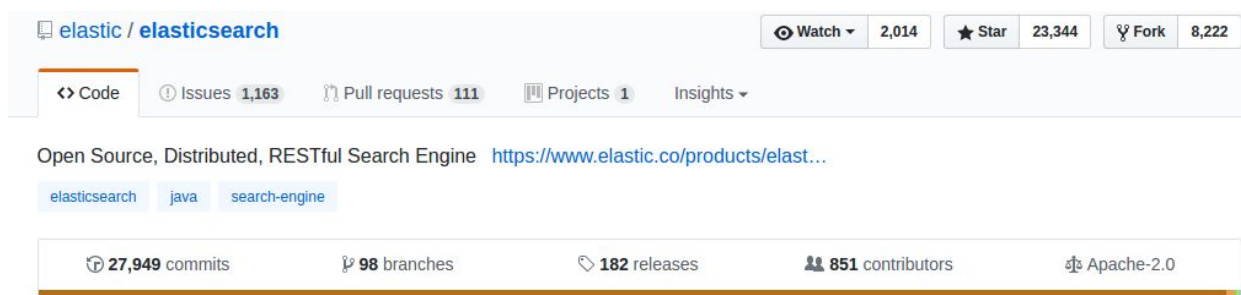
NEA: New Enterprise Associates, Inc. (NEA) es una empresa de capital de riesgo líder enfocada en ayudar a los empresarios a desarrollar negocios transformacionales a través de múltiples etapas, sectores y geografías. Con aproximadamente \$ 13 mil millones en capital comprometido, NEA invierte en tecnología de la información, cuidado de la salud y compañías de tecnología de energía en todas las etapas en el ciclo de vida de una empresa, desde la etapa de semillas hasta la OPI. El largo historial de inversiones exitosas de la firma incluye más de 175 OPIs de la compañía de la cartera y más de 300 adquisiciones.

Herramientas

Elastic invierte fuertemente en la construcción de una comunidad, ya sea interactuando con feedback de usuarios en IRC y listas de correo, mediante Solicitudes de Revisión o PR (Pull Requests en Github) o intercambiando historias en meetups y eventos.

A continuación están algunos de los recursos el cual es posible colaborar con la comunidad y los usuarios de la misma.

Github: Esta tool es tanto usada para versionado de código como también para gestión de issues a continuación una captura de las estadísticas del repositorio para producto ES.



Foros de discusión: el sitio de foros es una tool separada y en la cual use con más frecuencia, sin embargo los principales desarrolladores o más antiguos no usan tanto este medio, mas bien es usado para ayudar a otros programadores que no están tan involucrados en la comunidad y con usuarios que solo conocen el software como usuario final y quieren recibir alguna ayuda por parte de la comunidad, en la misma no solo estan los desarrolladores, sino que hay todo tipo de personas intercambiando información, se publican las noticias de nuevos productos el cual luego se envían en listas de correos automáticamente. Para obtener o escalar en niveles de confianza la comunidad aqui usa 5 niveles

Listas de email: Colaboración usando el correo electrónico.

IRC: Mensajería para intercambio mas ágil de información y colaboración mas en tiempo real.

Cientes de productos de Elastic

A continuación se muestran algunos de los clientes más importantes que usan productos de elastic, no se listan todos por ser una larga lista.

Dell.com: Soporte en la búsqueda de comercio electrónico para más de 60 países en más de 21 idiomas.

EBay: Búsqueda de 800 millones de anuncios en milisegundos con la creación de una plataforma Elasticsearch como una plataforma de servicio.

Netflix: Garantizar la entrega de mensajes y la excelencia operacional.

Facebook: ofrecer una mejor experiencia de ayuda para más de mil millones de usuarios.

NASA JPL: Impulsando la búsqueda de descubrimiento interplanetario.

Walmart: Obtener información sobre los patrones de compra de los clientes y las métricas de rendimiento de las tiendas.

Cisco: Reducir el tiempo de inactividad del sistema con ES en la base de la plataforma Cloud de Cisco.

Nvidia: Mejora de la experiencia del usuario mediante el procesamiento de más de mil millones de eventos todos los días.

Uber: Agregando métricas de negocio para controlar comportamientos críticos del mercado

Tinder: Conectando a personas alrededor del mundo.

CERN: Hacer posible el descubrimiento de la historia.

Microsoft: Proporcionar búsqueda en Azure y potenciar Dinámica Social.

Mozilla: Potenciando los esfuerzos de MozDef para proteger sus sistemas.

IBM: Provee el motor de análisis de logs operacional para Bluemix Apps.

GitHub: Acelerar el desarrollo de software.

MercadoLibre: Llevando la plataforma de eBay al mercado sudamericano.

McGraw-Hill: Transformar los libros de texto en expedientes de aprendizaje digital.

Elastic Search - Software elegido

Es un motor de búsqueda distribuido RESTful construido para la nube.

Las características incluyen:

- Motor de búsqueda distribuido y altamente disponible.
 - Cada índice está completamente fragmentado con un número configurable de fragmentos.
 - Cada fragmento puede tener una o más réplicas.
 - Operaciones de lectura / búsqueda realizadas en cualquiera de los fragmentos de réplica.
- Multi Tenants con múltiples tipos.
 - Soporte para más de un índice.
 - Soporte para más de un tipo por índice.
 - Configuración de nivel de índice (número de fragmentos, almacenamiento de índice, ...).
- Varios conjuntos de API

- API HTTP RESTful
- API Java nativa.
- Todas las API realizan el desvío automático de la operación del nodo.
- Orientado a documentos
 - No hay necesidad de una definición de esquema inicial.
 - El esquema se puede definir por tipo para la personalización del proceso de indexación.
- Confiable, Asíncrono escribir detrás para la persistencia a largo plazo.
- (Near) Búsqueda en tiempo real.
- Construido sobre Lucene

Mi participación

Encontré la comunidad gracias al sitio codetriage.com donde es un sitio de registro de repositorios en github o resumidamente un repositorio de repositorios de github, en el mismo se pueden encontrar proyectos escritos en variados lenguajes de programación y cualquier usuario puede filtrar por el tema de interés encontrando comunidades interesantes. Básicamente como soy programador java decidí filtrar por esta tecnología y buscar alguna comunidad de mi interés.

Me llamó la atención la comunidad de elastic por varias razones, una de ellas es que es una herramienta muy usada en desarrollos Big Data siendo esto uno de mis objetivos de aprendizaje para mi desarrollo profesional a futuro y a su vez la comunidad estaba en rojo por cantidad de issues, por lo cual decidí ser partícipe colaborando con la misma. Al comienzo decidí ir por ayudar con la documentación y corregirlos ya que en mi experiencia estas suelen tener muchos errores, pero luego de toparme con que se necesitaba de tecnologías desconocidas para mí entonces me moví a la revisión de código o intentar colaborar en los foros de discusión.

Para colaborar con documentación y ayudar con mejoras en el código fuente se debe firmar digitalmente un CLA y el cual hice siguiendo mi idea inicial, para ello se debe ingresar al sitio <https://www.elastic.co/contributor-agreement> y seguir los siguiente pasos:

- 1) luego de elegir la opción “An Individual Contributor” se procede con la firma digital de un CLA (Contributor License Agreement).
- 2) Llenar con tus datos personales el form en formato pdf que se abre, ver la imagen abajo:

Start

Signed

Signature:

Click here to sign

Email:

Enter your email address

Title

Enter your job title

Github Username

Date

Jun 21, 2017

3) Luego de dar el ok se enviará un email a tu cuenta personal con la confirmación de la persona que se está suscribiendo, a continuación el mail recibido con mis datos.

Individual Contributor License Agreement between elasticsearch BV and Gustavo Gabriel Bustos is Signed and Filed!



elasticsearch BV <echosign@echosign.com>
to me, CLA

May 2



elastic

POWERED BY
Adobe Sign



Individual Contributor License Agreement between elasticsearch BV and Gustavo Gabriel Bustos is Signed and Filed!

From: **CLA Committers** (elasticsearch BV)
To: Gustavo Gabriel Bustos and **CLA Committers**

Luego de realizados los pasos anteriores me registre en los foros de discusión y navegué por los mismos para entender la temática y buscar dónde colaborar con la comunidad el siguiente link para registrarse: <https://discuss.elastic.co/>.

Al comienzo obtuve varias respuestas automáticas porque al ser una comunidad muy grande hay que autoayudarse en cierto punto, luego de haber entendido el funcionamiento hice un post para ayudar con una configuración en uno de los productos aprovechando mis conocimientos en java, a continuación se pueden visualizar algunas capturas de pantallas de todas mis participaciones y una breve explicación de cada una de ellas.

a) Envío de mensaje privado para recibir respuesta de alguien sobre algún aspecto en la que se

pueda participar, el mismo fué como consecuencia de algunos mensajes automáticos que recibía:



ggbustos Gustavo Gabriel Bustos

May 20

Hello guys,

I hope this note find you well. As you know I am a new member on this community, I would like to participate as a collaborator and share with you some of my skills and experience in order to help you in some of the projects that you have, I know that is necessary to gain your trust, after see my skills then you can tell me in which project do you need some help as beginner in this community.

Skills:

- I am a java developer (14 years of experience), I worked with Rest Services, Soap services, Web Applications Arquitecture in some projects. Technologies used (Backend): Java, EJB, Spring, Struts, WSO2 suite tools (ESB, MB, API Manager, DAS, DSS), JMS queue, etc...

Other technologies I worked (frontend): HTML, JSP, XML, JSON, JS frameworks, JavaScript, Ajax, SOA Integrations (current job), etc.

SO: Linux

Currently, I am learning Big Data technologies like Hadoop, Machine Learning (Stanford project), mongodb.

I can help you also with (I can start with this if you want): documenting, documents translation to spanish (my native language), reviewing java code in some PR, searching bugs or just learning about elastic search, etc.

Let me know if you need any additional information or just tell me where I can help you.

Regards,

Gustavo.-

- b) Intenté ayudar con un problema de configuración del producto Elasticsearch que al indexar documentos daba un error de OutOfMemoryError por haber llenado la memoria Heap Space de la JVM de java:

Heap space - OutOfMemoryError while indexing shapes

 Elasticsearch



tdanielczyk Tomek

16d

Hi,

While indexing shapes with high precision and no distance_error_pct elastic crashes on java.lang.OutOfMemoryError: java heap space.

After crash elastic is not restarted, my configuration is default with heap space 4gb on machine with 8gb. Memory is locked (bootstrap.memory_lock: true). There is single node in the cluster, default configuration of indices.

Is there any way how I can prevent from the crash without automatic restart? Something like indexing request timeout? Any solution without precision and distance error change that will prevent from ES crash will be appreciated.

Error Heap Space del usuario Tomek



ggustos Gustavo Gabriel Bustos

15d

Hi Tomek,

I am new on the community and also in the elastic stacks but am trying to help in some way, did you try by configuring the jvm.options in ES? probably you can try by increasing your current value e.g 4gb to 6gb?.

Let me know if you have any positive result.

Thanks,

Gustavo.-

    Reply

Mi respuesta, para ello tuve que instalarme el software y probarlo.

Link del thread: <https://discuss.elastic.co/t/heap-space-outofmemoryerror-while-indexing-shapes/88384>

- c) Revisión de código, para el código tuve que pasarme horas entendiéndolo y sinceramente fue demasiado complejo requiriendo mucho más tiempo de lo pensado, además para lograr colaborar en este aspecto se necesita un conocimiento más profundo de la arquitectura haciendo trainings, meetups, webinars, etc. por lo que solo me limité a hacer observaciones menores.



gustavobustos reviewed 17 days ago

[View changes](#)

core/src/test/java/org/elasticsearch/action/bulk/TransportShardBulkActionTests.java

```
...    ...    @@ -646,6 +655,199 @@ public void testPrepareIndexOpOnReplica() throws Except
646    655        closeShards(shard);
647    656    }
648    657
658    + public void testProcessUpdateResponse() throws Exception {
```



gustavobustos 17 days ago

Add Javadocs to the test method



jasontedor 17 days ago

Member

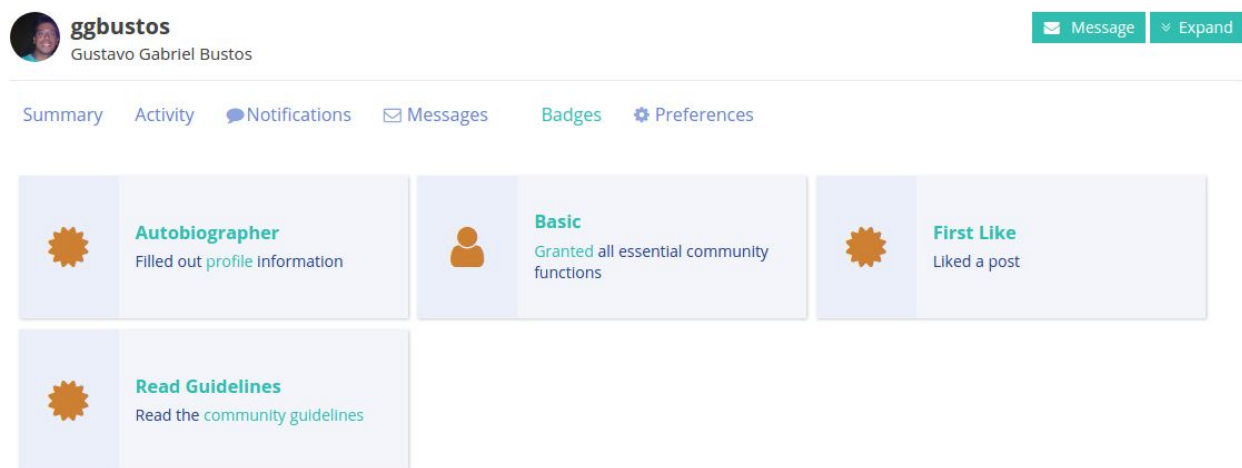
It's a test method which does not need Javadocs as it's not part of a public API. Sometimes we add explanatory comments to non-straightforward tests but this is not such a test.



Reply...

Como respuesta obtuve un comentario de uno de los desarrolladores más importantes del producto ES: quién es **Jason Tedor**? es Ingeniero de Software de Elastic, es uno de los desarrolladores más importantes dentro de la comunidad de Elasticsearch el cual le encantan todo lo que sea tecnología distribuidas. Antes de unirse a Elastic, Jason fue un ingeniero de backend utilizando entorno de Hadoop para manejar uno de los conjuntos de datos clínicos más grandes del mundo, y construyó simulaciones de Monte Carlo para modelar carteras de préstamos comerciales.

Por último, al haber participado escalé en algunos niveles de confianza y gané algunos badges que me otorgaron, obtuve el nivel Basic (Trust Level 1) de los 5 niveles que tienen, la mayoría de los users obtienen este nivel, de todos modos vale aclarar que los nuevos miembros obtiene el New (Trust Level 0):



The screenshot shows a user profile for 'ggbustos' (Gustavo Gabriel Bustos). At the top right are 'Message' and 'Expand' buttons. Below the profile name is a navigation bar with 'Summary', 'Activity', 'Notifications', 'Messages', 'Badges', and 'Preferences'. The 'Badges' section displays four items: 'Autobiographer' (Filled out profile information), 'Basic' (Granted all essential community functions), 'First Like' (Liked a post), and 'Read Guidelines' (Read the community guidelines). Each badge has a sun icon.

Links de Interés

- Sitio con comunidades open source para colaborar del mundo entero, solo en github: <https://www.codetriage.com/?language=Java>
- Repositorio github de todos los productos de ES: <https://github.com/elastic>
- Sitio principal: www.elastic.co
- Cursos y clases virtuales: <https://www.elastic.co/training>

Conclusión

La experiencia obtenida como resultado de mi participación ha sido enorme e enriquecedora, personalmente reconozco que me he decidido por una comunidad de considerable tamaño y fue necesario mucha dedicación para entenderla, conocer las diversas herramientas y las más variadas que jamás haya visto, fue también complicado lograr respuesta de alguna persona ya que al principio solo veía respuestas automáticas, es decir, se debe tener predisposición por la autogestión y paciencia para coordinarse con personas en otras partes del mundo con el fin de obtener resultados.

Por otro lado mi idea original fué colaborar con revisiones de código y/o corregir algún bug, pero ahí fué necesario mucho tiempo para entender el núcleo del mismo, es un software de alta disponibilidad y tiene algoritmos muy avanzados, al ser de complejidad se requiere mucho más tiempo de aprendizaje y conocimiento de los productos muy profundos, por analogía comparable a un primer trabajo donde la curva de aprendizaje es elevada en los primeros meses de iniciado.

Finalmente concluyo que es una comunidad muy bien organizada para el tamaño de la misma, utilizando herramientas novedosas en el mercado y a la altura para un producto como ES, algunas de ellas como las meetups, webinars, IRC, github, blogs, entre diversos recursos, etc. Vale aclarar Github utiliza ES como Motor de búsqueda, sinceramente un software innovador y a la altura de las necesidades actuales del mercado siendo que es un proyecto Open Source.