

Apellidos, nome /Apellidos, nombre: Nishikawa Muñumer, Alfonso			
Título: Aplicación web de gestión de bases de datos NoSQL a través de Apache Gora, en ExtJS y servicios REST en Spring			Mención cursada: Ingeniería del Software

Resumo / Resumen:

En este proyecto de desarrolla una aplicación web que permite acceder a almacenes de datos NoSQL a través de Apache Gora. Con ello se supe una carencia existente entre la comunidad de software libre de Apache Gora, ya que hasta ahora no existía ninguna aplicación que permitiese gestionar los datos de manera genérica persistidos a través de este motor.

Apache Gora es una capa de acceso a datos, equivalente Hibernate, pero para bases de datos NoSQL fundamentado en un modelo de datos abstracto compartido por todos los almacenes de datos soportados. La aplicación desarrollada permite la manipulación de datos y la gestión de esquemas y mapeos de manera visual.

Adicionalmente, se ha extendido Apache Gora con un desarrollo que permite exponer información de la base de datos nativa que se esté utilizando, y se ha creado un módulo que permite realizar análisis en Hadoop mediante Apache Pig.

La página web de la aplicación está disponible en <https://www.goraexplorer.com>

Posibles aplicaciones / Posibles aplicaciones:

El objetivo principal de la aplicación es el acceso y mantenimiento cómodo de los datos almacenados a través de Apache Gora. Poder crear filas, modificar filas y borrar filas, y definir esquemas y mapeos mediante el ratón evitando tener que editar a mano complejos ficheros de configuración.

En las etapas de desarrollo futuro se contemplan las aplicaciones que involucran ejecutar trabajos en Hadoop como copias de datos, transformaciones, y procesos analíticos que posteriormente se muestren en paneles de mando, de manera que la aplicación realice la función de una gestión lo más integral posible en cuanto a la gestión de cada origen de datos.

Etapas para o seu desenvolvemento futuro / Etapas para su desarrollo futuro:

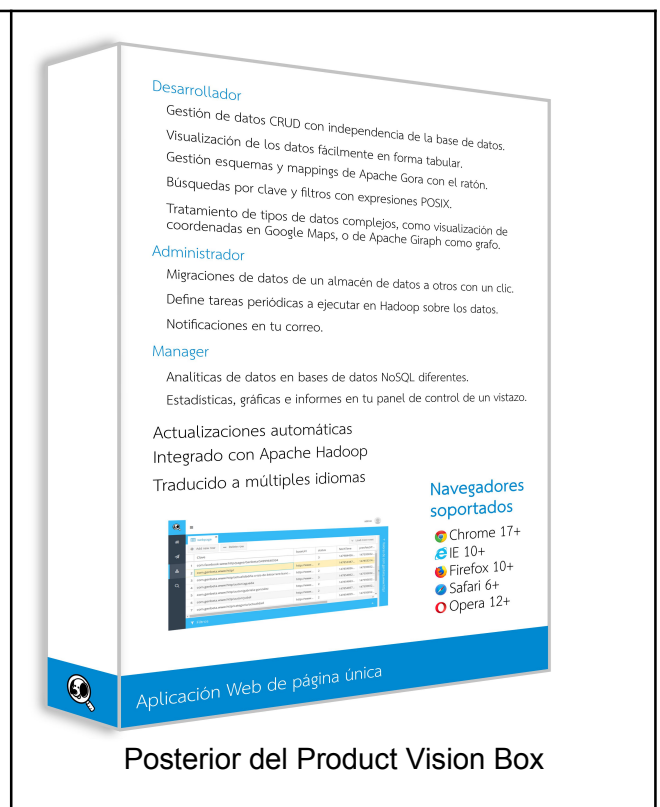
Actualmente el almacén de datos soportado es HBase, y se realizarán las modificaciones necesarias al core de Apache Gora para soportar el resto de almacenes de datos: MongoDB, Cassandra, Accumulo, y muchos otros.

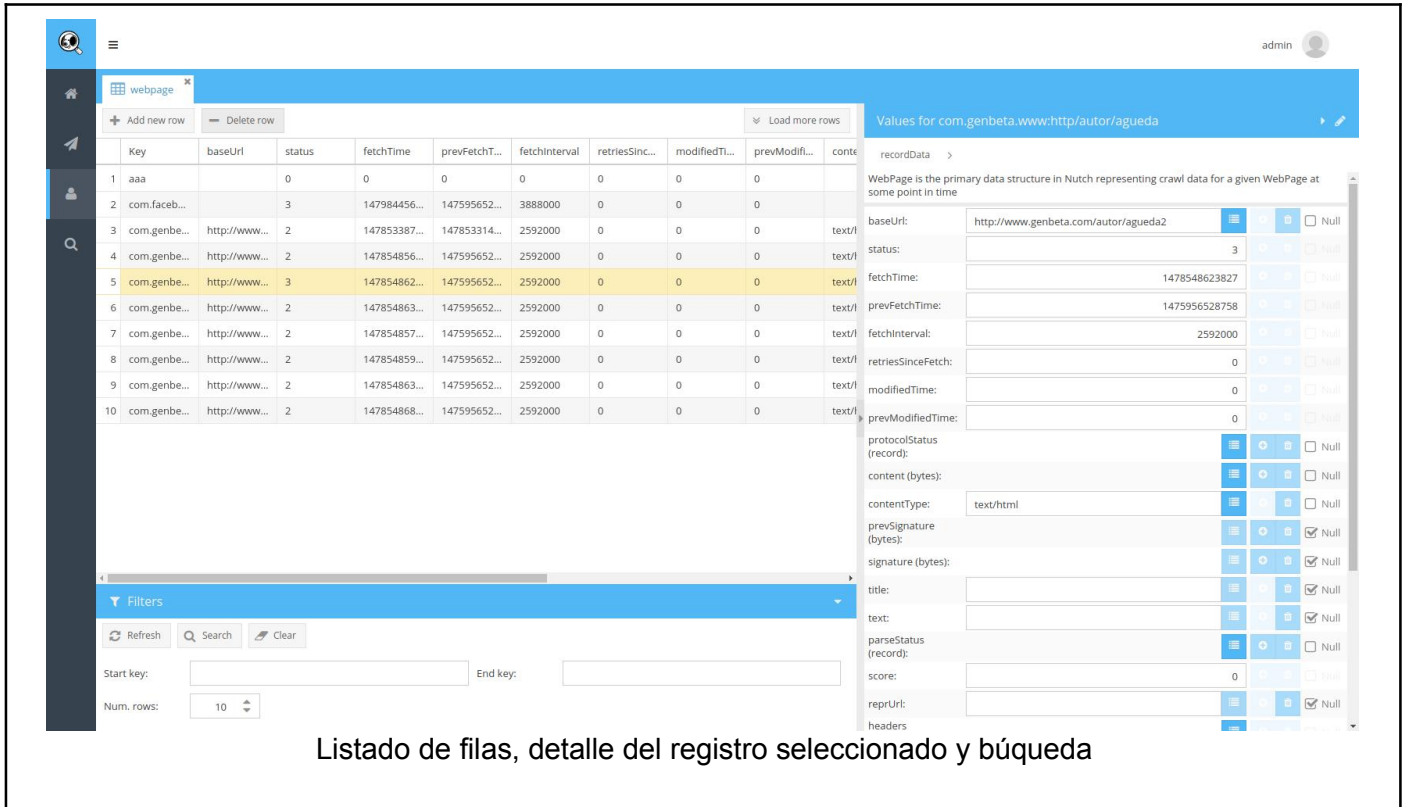
Aprovechando que Apache Gora se integra nativamente con los trabajos en Hadoop, se ha desarrollado un conector con Apache Pig para que se pueda acceder a los datos almacenados desde los scripts de Apache Pig. Esta funcionalidad permite realizar operaciones que requieren trabajos en Hadoop como el conteo de filas de las tablas de HBase.

Adicionalmente, se implementarán funcionalidades de panel de control para mostrar datos analíticos resultante de los procesos en Apache Pig.

Apache Gora también se integra con Apache Spark, por lo que también es una funcionalidad deseada el realizar los procesos analíticos también en Spark.

Imaxes representativas / Imágenes representativas:





The screenshot displays a web application interface for managing webpage records. It features a table with columns for Key, baseUrl, status, fetchTime, prevFetchTime, fetchInterval, retriesSinceFetch, modifiedTime, prevModifiedTime, and contentType. A detailed view of a selected record is shown on the right, and a search filter section is located at the bottom left.

Key	baseUrl	status	fetchTime	prevFetchTime	fetchInterval	retriesSinceFetch	modifiedTime	prevModifiedTime	contentType
1	aaa	0	0	0	0	0	0	0	
2	com.faceb...	3	147984456...	147595652...	3888000	0	0	0	
3	com.genbe...	2	147853387...	147853314...	2592000	0	0	0	text/html
4	com.genbe...	2	147854856...	147595652...	2592000	0	0	0	text/html
5	com.genbe...	3	147854862...	147595652...	2592000	0	0	0	text/html
6	com.genbe...	2	147854863...	147595652...	2592000	0	0	0	text/html
7	com.genbe...	2	147854857...	147595652...	2592000	0	0	0	text/html
8	com.genbe...	2	147854859...	147595652...	2592000	0	0	0	text/html
9	com.genbe...	2	147854863...	147595652...	2592000	0	0	0	text/html
10	com.genbe...	2	147854868...	147595652...	2592000	0	0	0	text/html

Filters

Refresh Search Clear

Start key: End key:

Num. rows:

Values for com.genbeta.www:http/autor/agueda

recordData

WebPage is the primary data structure in Nutch representing crawl data for a given WebPage at some point in time

baseUrl: http://www.genbeta.com/autor/agueda2

status: 3

fetchTime: 1478548623827

prevFetchTime: 1475956528758

fetchInterval: 2592000

retriesSinceFetch: 0

modifiedTime: 0

prevModifiedTime: 0

protocolStatus (record):

content (bytes):

contentType: text/html

prevSignature (bytes):

signature (bytes):

title:

text:

parseStatus (record):

score: 0

reprUrl:

headers

Listado de filas, detalle del registro seleccionado y búsqueda