**INSTRUCTIVO PARA LA REALIZACIÓN DE LA PROPUESTA DE PASANTÍA**

**Diseño e implementación de mecanismos de Extracción, Transformación y Carga de imágenes desde *Oracle 10g* a *Cassandra 2.1.8***

Firma Firma

Genesis Martinez Pedro Bonillo

30/03/2016

**Pasantía Presencial: x ó Pasantía Virtual:**

**Datos del Estudiante:**

Apellidos, Nombres: Martínez Jaspe, Génesis Carolina

C.I.: 20.175.751

Teléfonos:

Celular: 04242350783

Hab: 0212-2643572

Correo electrónico: genesismartinez.ucab@gmail.com

**Datos de la Empresa:**

Nombre de la empresa: PhD 2014 Consultores, C.A.

Dirección de la empresa: Avenida Principal de Chuao, Calle Los Araguaneyes con Calle Santa Cruz. Quinta N°46. Caracas, Miranda

Teléfonos:  +58 (212) 991.8911 / 993.0325 / 919.8816

**Datos del Tutor Empresarial:**

Apellidos, Nombres Tutor Empresarial: Bonillo, Pedro

Teléfono Tutor Empresarial: +58- 4241287373

Correo Electrónico Tutor Empresarial: pbonillo@gmail.com

Tutor Empresarial es Prof. UCAB: X Tutor Empresarial es (fue) tutor TEG ó TIG:

**Fecha de Inicio y Fin de la Pasantía:** 14 / 03 / 2016 al 06 / 05 / 2016

(verifique que su pasantía tenga una duración mínima de 6 semanas y máximo 8)

**Introducción**

El problema del Servicio Autónomo de Registros y Notarías (SAREN) surge del hecho de que se generan muchas transacciones al día en la mayoría de sus oficinas. Las cuales no están al tanto de los cambios o transacciones hechas por las demás oficinas, lo cual lleva a que cuando se quiera registrar una nueva empresa, pueda tardar muchas semanas o meses. Además de estos problemas también se tiene que la cantidad de datos generados durante el día requiere de equipos cada vez más costosos. SAREN se maneja con dos esquemas principalmente los cuales son el esquema público y el esquema mercantil, sin embargo, algunas tablas son comunes en los dos esquemas. SAREN almacena una gran cantidad de imágenes generadas por sus transacciones diarias, ya sean documentos, firmas, sellos, entre otros, por lo que sus servidores están al máximo de su capacidad.

Se propone la implementación de una arquitectura Big Data, de manera que la generación masiva de datos por día se encuentre en una arquitectura escalable, que permita realizar consultas eficientemente para los reportes necesarios de la organización, además de que permite la centralización de todos los datos de todas sus oficinas a nivel nacional.

**Objetivo General**

Implementar y probar mecanismos de *ETL (*Extact-Transform-Load) necesarios para migrar datos de las tablas que contengan imágenes de la base de datos *Oracle 10g* a *Cassandra 2.1.8*.

**Objetivos específicos**

1. Diseñar los mecanismos de *ETL*
2. Implementar los mecanismos de *ETL*
3. Probar dichos mecanismos en ambiente local
4. Ejecutar las transformaciones en ambiente de desarrollo del cliente

**Alcance de la Pasantía**

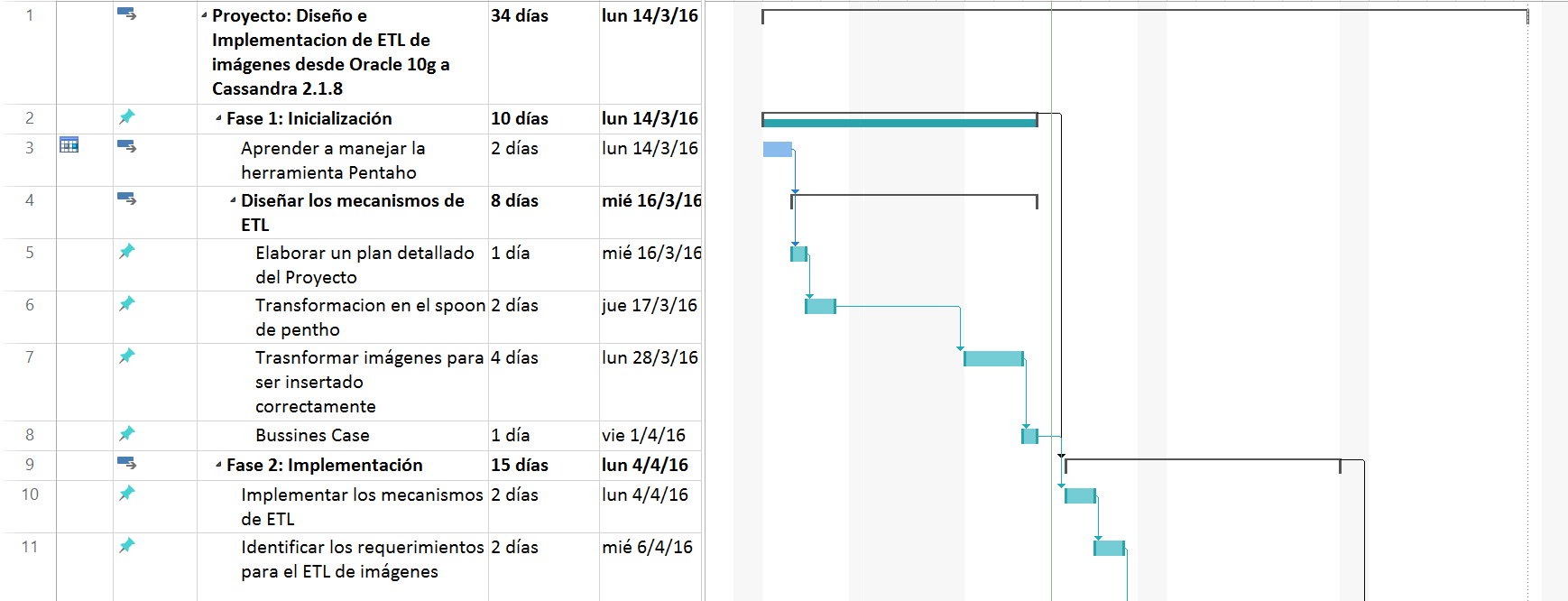
A continuación, se describirán el alcance de la pasantía:

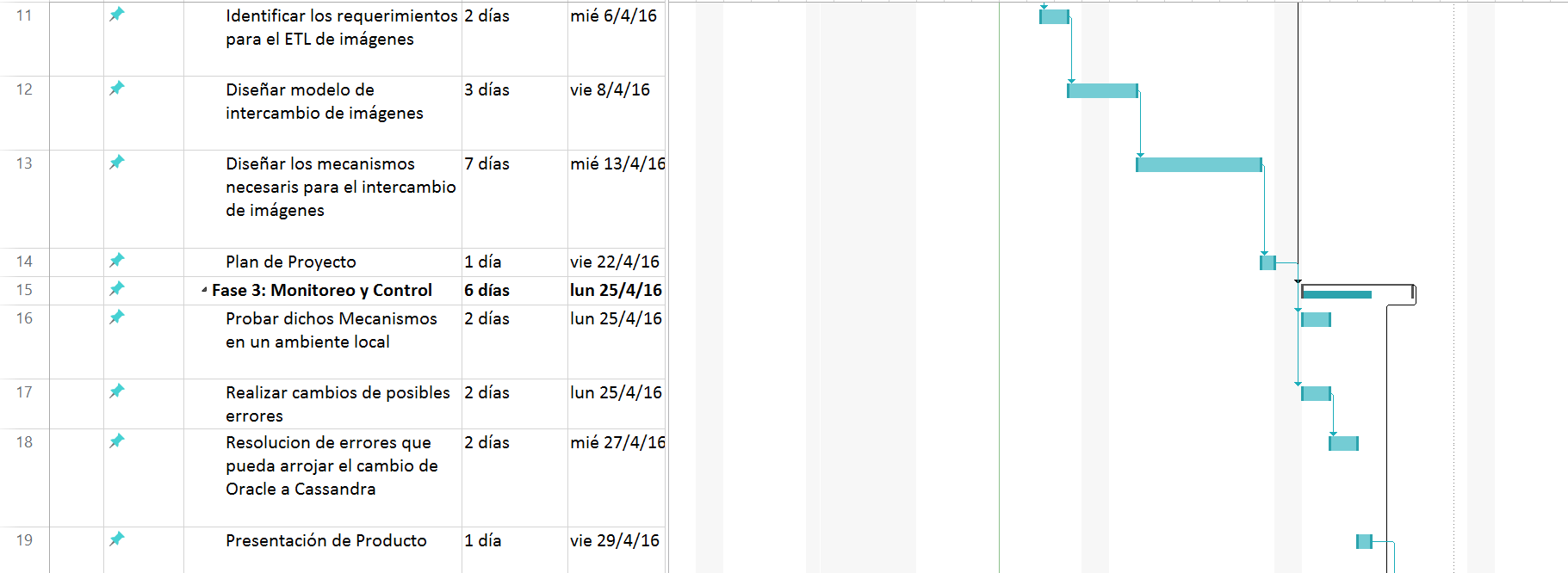
1. Diseñar los mecanismos de ETL:
   1. La transformación se realizara en el *spoon* de *pentaho* y consiste en realizar varios pasos, los cuales son:
      * Leer data Saren (*Table input*)
      * Seleccionar campos (*Select / Rename values*)
      * Limpiar campos (*Replacing in string*)
      * Limpiar clave (*String operations*)
      * Crear campos (*calculator*)
      * Crear Map y demás campos a insertar (*Script Values/ mod*)
      * Crear campos restantes (*calculator*)
      * Insertar en cassandra (*User Defined Java Class*)
   2. Se requerirá una modificación en él .jar de eclipse para agregar la función que inserta las imágenes en *Cassandra*.
   3. La imagen la cual está en formato de tipo blob requeríra de un tratamiento antes de ser insertada en *Cassandra*, por lo que se transformara al tipo *Bytebuffer* para que se insertara correctamente en la base de datos *Cassandra*. *(Apache*).
2. Se realizara una función en Java que tomara la imagen de *Cassandra* y la guardara en el disco duro, para su pronta revisión.
3. Para la realización de todas estas funciones se necesitara importar en el proyecto las librerías necesarias junto con los drivers de *Datastax* para *Cassandra*. (*Datastax*)
4. Para solucionar un posible problema con la memoria, la cual consiste en realizar un *job* en *pentaho* para que pasara un registro a la vez.

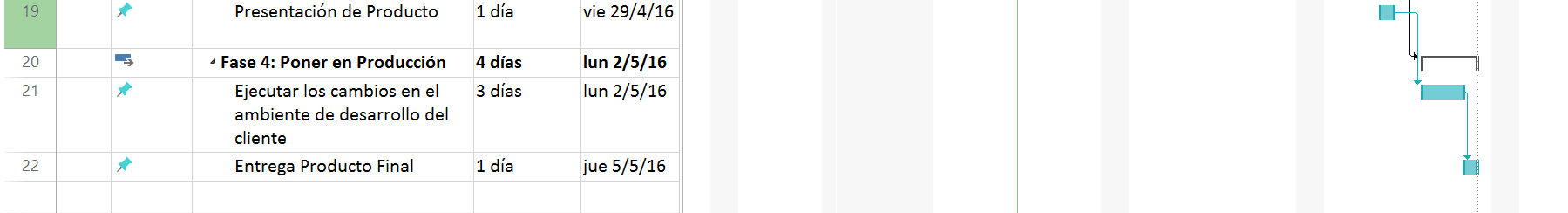
**Limitaciones de la Pasantía**

1. Debe migrarse a Cassandra V 2.1.8
2. Se utilizara Datastax 2.0, ya que es la que provee los drivers para Cassandra.
3. El paso de muchos datos puede traer problemas con la memoria, y es posible que se deba buscar una solución para dio problema.
4. Diseñar la ETL en la BD 10 g de un registro publico y un registro mercantil del Saren en el registro publico 390 y el registro mercantil 210 , los cuales tiene un aproximado de 6 millones de registros.
5. Esto se realizara en la oficina de PHD consultores 2014 con una conexión remota a las oficinas del Saren.

**Cronograma de la Pasantía**







**Carta de Aceptación del Tutor Empresarial**

Caracas, 30 de Marzo de 2016

Señores

Consejo de Escuela de Ingeniería Informática

Facultad de Ingeniería

Universidad Católica Andrés Bello

Presente.-

Por medio de la presente hago constar que estoy dispuesto a supervisar, en calidad de Tutor Empresarial, la pasantía propuesta por: Génesis Martínez C.I. 20.175751, para lo cual solicito la aprobación de ese Consejo de Escuela.

Así mismo hago constar que estoy conforme con las responsabilidades que me corresponde asumir, las cuales se indican a continuación:

* Darle una inducción al pasante para adaptarlo a la organización.
* Dirigir y asesorar al pasante durante su permanencia en la empresa.
* Brindarle al pasante todas las facilidades relacionadas con: ubicación, materiales, obtención de información de otras unidades organizativas, etc.
* Evaluar el informe final de la pasantía, el estudiante debe entregármelo a más tardar una semana posterior a la fecha de finalización de la pasantía y el tutor se compromete a devolvérselo firmado junto con la Planilla de Evaluación Individual del Tutor Empresarial a más tardar una semana después.

Atentamente,

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Firma

Dr. Pedro Bonillo