



**UNIVERSIDAD DE LA FRONTERA
FACULTAD DE INGENIERÍA Y CIENCIAS
DEPTO. CS. COMPUTACIÓN E INFORMÁTICA**

**“ETAPA DE PRE-PROCESAMIENTO DE MINERÍA DE DATOS
PARA REGISTROS DE ALMA COMMON SOFTWARE”**

**JAVIER IGNACIO FUENTES MUÑOZ
2016**



**UNIVERSIDAD DE LA FRONTERA
FACULTAD DE INGENIERÍA Y CIENCIAS
DEPTO. CS. COMPUTACIÓN E INFORMÁTICA**

**“ETAPA DE PRE-PROCESAMIENTO DE MINERÍA DE DATOS
PARA REGISTROS DE ALMA COMMON SOFTWARE”**

**TRABAJO PARA OPTAR AL TÍTULO
DE INGENIERO CIVIL INDUSTRIAL MENCIÓN INFORMÁTICA**

PROFESOR GUÍA: PATRICIO ABEL GALEAS ECHEVERRÍA

PROFESOR CO-GUÍA: GLORIA MILLARAY JULIA CURILEM SALDIAS

**JAVIER IGNACIO FUENTES MUÑOZ
2016**

**“ETAPA DE PRE-PROCESAMIENTO DE MINERÍA DE DATOS
PARA REGISTROS DE ALMA COMMON SOFTWARE”**

JAVIER IGNACIO FUENTES MUÑOZ

COMISIÓN EXAMINADORA

PATRICIO ABEL GALEAS ECHEVERRÍA
Profesor Guía

GLORIA MILLARAY JULIA CURILEM SALDIAS
Profesor Co-Guía

Dr. CARLOS FERNANDO CARES GALLARDO	Mg. JORGE ALBERTO HOCHSTETTER DIEZ	Otro profesor connotado
Profesor Examinador 1	Profesor Examinador 2	Profesor Examinador 3

Nota trabajo escrito :

Nota examen :

Nota final :

dedicatoria

RESUMEN

El presente informe muestra el desarrollo del trabajo de título “Etapa de pre-procesamiento de minería de datos para registros de ACS”. El Ataca Large Millimeter/Submillimeter Array (ALMA) es actualmente el observatorio astronómico más grande del mundo y como tal tiene un grado de complejidad muy alto. Los sistemas de ALMA son controlados por un software llamado Alma Common Software (ACS), sistema que se encarga de todas las operaciones relacionadas a la observación. Este trabajo se enfoca específicamente en el proceso de observación, en el cual, es de interés poder monitorear y analizar su comportamiento en base a los registros generados por ACS. Este trabajo consistió en la producción de una herramienta capaz de interpretar los registros de ACS e identificar los diferentes estados observables dentro de los procesos. Por otra parte también se generó una herramienta para descubrir secuencias en la ejecución de las observaciones, estas secuencias permiten analizar.

Índice general

1	Introducción	1
1.1	Descripción del problema	2
1.1.1	Exposición general del problema	2
1.1.2	Situación actual del problema	2
1.2	Objetivos	3
1.2.1	General	3
1.2.2	Específicos	3
2	Antecedentes Generales	4
3	Descripción de Actividades Realizadas	5
4	Resultados y discusión	6
5	Conclusiones	7
	Referencias	9
	Anexo A Tabla de Puntajes	10
	Anexo B Lista de publicaciones	11

Índice de figuras

Índice de tablas

CAPÍTULO 1

INTRODUCCIÓN

1.1. Descripción del problema

1.1.1. Exposición general del problema

ACS es un software distribuido altamente complejo. Como toda infraestructura de software, este presenta fallas en su funcionamiento las que pueden llegar a provocar la detención completa del sistema, provocando retrasos en las tareas de observación. Gran parte del funcionamiento de ACS es permanentemente registrado en archivos de registros, los cuales suman diariamente grandes cantidades de información sobre la operación y eventos asociados a diferentes niveles de alerta. Sin embargo, dada la gran cantidad de registros generados por este sistema, resulta complejo el poder analizar estos datos en forma eficiente y efectiva.

ALMA tiene una infraestructura altamente compleja, lo que se refleja en sus sistema de generación de logs, por lo que es natural encontrar problemáticas asociadas a numerosas causas y por este motivo es necesario buscar y analizar estas posibles problemáticas. El problema se plantea complejo tanto por la diversidad de los logs como por su cantidad.

Por otra parte, se ha desarrollado herramientas y metodologías de minería de datos que tienen por objetivo el apoyar el análisis de grandes volúmenes de datos de manera de poder extraer conocimientos nuevos y útiles. Estas herramientas son capaces de procesar grandes volúmenes de información y encontrar relaciones entre las múltiples variables que afectan los procesos. Para poder aplicar estas herramientas es necesario plantear objetivos de búsqueda y aplicar una metodología de minería de datos CRSIP.

Sin embargo, debido a la alta complejidad del sistema ACS, los objetivos de búsqueda que guían el análisis son múltiples, lo que requiere de una comprensión muy acabada de todo el proceso para poder modelar algunas partes a través de la minería de datos y buscar entonces las relaciones que permiten comprender por ejemplo, porqué y bajo qué circunstancias el sistema falla. El estudio y comprensión del sistema ACS permitirá la identificación de un problema específico para el cual es posible plantear una aproximación basada en minería de datos.

1.1.2. Situación actual del problema

Actualmente los registros de ACS son numerosos, el origen de estos registros provienen en su mayoría de todos los sistemas y subsistemas, desde operaciones de alto nivel hasta operaciones de bajo nivel. Todos estos registros son almacenados en un mismo formato y

lugar. Para acceder a estos registros es posible a través del servicio de Kibana, que permite realizar búsquedas y obtener estos registros. La idea principal es reducir el espacio generado por los registros de ACS en el contexto de un proceso de observación. Para este enfoque se plantea por parte de ALMA utilizar el concepto de máquinas de estado finitas (MEF) para reducir el conjunto de datos a estudiar, generando un modelo sobre el proceso de observación.

1.2. Objetivos

1.2.1. General

- Desarrollar una herramienta de pre-procesamiento para minería de datos basada en los registros de ACS.

1.2.2. Específicos

-

CAPÍTULO 2

ANTECEDENTES GENERALES

CAPÍTULO 3

DESCRIPCIÓN DE ACTIVIDADES REALIZADAS

CAPÍTULO 4

RESULTADOS Y DISCUSIÓN

CAPÍTULO 5

CONCLUSIONES

Conclusiones

Referencias

ANEXO A
TABLA DE PUNTAJES

ANEXO B
LISTA DE PUBLICACIONES