

Laboratorio 2: Sistemas Distribuidos

Profesores: Jorge Díaz

Ayudantes de Lab: Iñaki Oyarzun M. & Javiera Cárdenas

Octubre 2023

1 Objetivos del laboratorio

- Aprender acerca de la comunicación en sistemas distribuidos.
- Familiarizarse con los conceptos asociados a **Naming y Coordinación** en Sistemas Distribuidos
- Profundizar el uso de **Golang y gRPC**

2 Introducción

Dentro de los sistemas distribuidos, el manejo de información se realiza en la mayoría de las situaciones a partir de la centralización en una entidad o sistema que conoce y distribuye estos datos. Para estos casos, la entidad mantiene un registro de donde esta almacenada la información para en caso de que sea solicitada pueda buscar y saber a quién preguntar, esto se conoce como **naming**. Generalmente se observa también en sistemas como Hadoop.

Para poner esto en práctica, se propone el siguiente problema en donde haciendo uso de gRPC se aplicará un caso de naming y coordinación entre entidades para subir y obtener información.

3 Tecnologías

- El lenguaje de programación a utilizar es **Go**
- Para las comunicaciones se utilizará **gRPC**

4 Laboratorio

4.1 Contexto

¡Una nueva enfermedad muy infecciosa se esta descontrolando! y múltiples personas alrededor de todos los continentes se están contagiando debido a que es altamente infecciosa y a veces mortal, en vista de esto, la Organización Mundial de la Salud (OMS) empezó a recopilar datos alrededor del mundo para así saber como esta grave enfermedad esta afectando a la población mundial, lamentablemente se encuentran poco y nada de registros.

En vista de esto la ONU decide realizar múltiples campañas para generar conciencia y dar a conocer esta rara e infecciosa enfermedad a partir de los datos de la OMS, pero en vista de que estos son casi nulos. Se comienza una campaña previa en cada continente para obtener datos de su población infectada y las posibles muertes que hayan ocurrido y así generar una inmensa fuente de información para que la ONU pueda lanzar su gran campaña de conciencia hacia todos los continentes.

4.2 Explicación

4.2.1 Diagramas de sistemas

Para ayudar a la ONU, se le hace entrega de la siguiente información que indica como esta planteado el sistema:

- Los nuevos sistemas funcionan a partir de 4 nodos de información (llamados Continentes) los cuales almacenan datos de usuarios de manera desordenada de 2 tipos (Infectado o Muerto), teniendo una entidad común que se encarga de coordinar y a la vez conoce la forma en que se distribuye la información (llamado OMS o NameNode). Los computadores de la red mundial se conectan a este NameNode para subir sus archivos y este se encarga de distribuirla. (Revisar Diagrama adjunto para mayor detalle.)

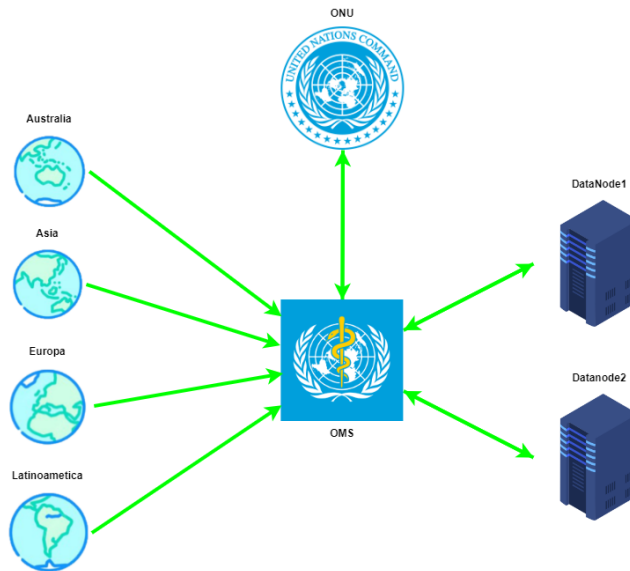


Figure 1: Estructura de los sistemas actuales

- Se requerirá establecer una conexión con el NameNode desde la ONU para poder consultar los datos correctamente ordenados, generando la distribución de conexiones previamente mostrada.

4.2.2 Explicación entidades presentes

Esta sección explicará el funcionamiento de cada una de las entidades a crear para desarrollar el laboratorio.

- **Continentes:**
 - Conocen y se conectan solamente la IP de la OMS (NameNode)
 - Tendrá acceso a un archivo TXT con nombres de personas para poder obtener y utilizar para informar a la OMS.
 - Los continentes obtendrán un nombre al azar del archivo TXT para luego aplicar una probabilidad de determinar el estado de la persona (**55% de probabilidades de que esté infectada, 45% de probabilidades de que esté muerta**).

- Se conectarán de manera síncrona a la OMS para ingresar información acerca de los nombres de las personas y sus estados (**Muerto o Infectado**)
 - El formato del mensaje a enviar a la OMS **puede ser implementado de la forma que deseen los grupos**, sin embargo. **Solo pueden estar el Nombre Apellido y estado de la persona.**
 - Al iniciar el programa los continentes deben subir 5 datos de personas al iniciar, para luego subir un dato cada 3 segundos.
- **OMS (NameNode):**
 - Conoce y se conecta a la IP de los DataNodes, Continentes y la ONU.
 - Al recibir un mensaje de parte de los Continentes, el NameNode seleccionará a partir de la inicial del apellido de la persona un DataNode para entregarle la información (**si el apellido inicia con las letras entre la A y la M se irá al DataNode 1. Si el apellido inicia con la letra N hasta la Z irá al DataNode2**), luego de elegirlo, registra en una nueva línea dentro de un archivo llamado **DATA.txt** un ID que será un valor único otorgado por el NameNode para identificar esa persona, junto con el número del DataNode en donde fue almacenado el dato y el estado que tenía esa persona. Luego de ello le envía el mensaje al DataNode seleccionado con el ID que le asignó el NameNode, el nombre y el apellido para que sea almacenado.
 - Para cuando la ONU desee consultar los nombres de las personas muertas o infectadas, el NameNode deberá acceder a revisar su archivo **DATA.txt** para identificar cuáles son los IDs que contienen el estado y en cuál DataNode se encuentran para solicitarles los nombres de esas personas. Cuando el NameNode obtiene todos los nombres de las personas en ese momento infectadas o muertas deberá enviarlos como respuesta a la ONU.
- **DataNodes:**
 - Conocen y se conectan solamente a la IP del NameNode
 - Almacenan la información entregada por el NameNode en un archivo local, llamado **DATA.txt**
 - Cuando el NameNode lo solicite según el ID, los DataNodes deben entregar el nombre y apellido de la persona que coincida con el identificador.
- **ONU:**
 - Conocen y se conectan solamente a la IP del NameNode
 - Poseen una consola para consultar al NameNode por la información de los nombres de personas muertas o infectadas.
 - Cuando se muestre por pantalla cada nombre obtenido del NameNode, será por cada línea separado.

5 Restricciones

- Todo uso de librerías externas que no se han mencionado en el enunciado debe ser consultado en aula.
- La implementación de los formatos de mensajes a intercambiar entre entidades queda a libre implementación de los grupos, siendo obligatorio que solamente se utilicen las variables que se menciona en la sección anterior para cada intercambio. (Es decir, si NameNode / OMS solo puede informar a DataNode los nombres y el ID, sólo pueden trabajar en ese mensaje esas dos variables, no el estado).

6 Consideraciones

- **Prints por pantalla:** Para ver el desarrollo y a la vez como apoyo para el debugging dentro del laboratorio, se solicitará que se realicen los siguientes prints por pantalla como mínimo:
 - **(DataNode):** Mostrar por pantalla el mensaje recibido y respuesta enviada.
(Ej: Solicitud de NameNode recibida, mensaje enviado: X)
 - **(NameNode / OMS):** Mostrar por pantalla el mensaje recibido y respuesta enviada.
(Ej: Solicitud de Y recibida, mensaje enviado: X)
 - **(Continentes):** Mostrar por pantalla el mensaje al enviar información de una persona.
(Ej: Estado enviado: <NOMBRE> <APELLIDO> <MUERTO | INFECTADO>)
- Se utilizarán las 4 máquinas virtuales entregadas a cada equipo del laboratorio anterior.
- Los DataNodes, NameNode / OMS y ONU deben estar en máquinas diferentes y los continentes deben estar en cada máquina.
- Se realizará una ayudantía para poder resolver dudas y explicar la tarea. Será notificado por aula.
- Consultas sobre la tarea se deben realizar en Aula o enviar un correo a **Javiera** o **Iñaki** con el asunto **Consulta grupo XX - Lab 2**
- Las librerías de **Golang** permitidas son:
 - time
 - strconv
 - strings
 - math
 - rand
 - net
 - context
 - fmt
 - os
 - bufio
 - log
 - grpc
- **Sobre Github:** Cada grupo tendrá acceso a un repositorio privado (se ha creado uno nuevo dentro de la organización a la que fueron agregados) el cual le permitirá actualizar y entregar el código de su laboratorio. Para ello, el día de la entrega, deberá dejar en la rama principal (main) todos sus archivos finales.
- **Docker:** El uso de Docker es obligatorio para encapsular su código y ejecutarlo en cada una de las máquinas virtuales provistas.

7 Informe:

A partir de lo presentado en el Laboratorio de Redes deberán realizar un informe que responda las siguientes preguntas:

- ¿En este sistema, los datos se encuentran distribuidos equitativamente? Justifique.
- ¿Cuanto se demora su programa en realizar cada consulta desde la ONU?
- Si se agrega otro nodo que se encargue solo de recibir desde los continentes y subir la información a los DataNodes. ¿Aumenta o disminuye en el tiempo para realizar cada consulta? Justifique.
- A partir del sistema realizado, ¿Cual seria una forma mas eficiente de balancear la carga del sistema?

8 Reglas de Entrega:

- El laboratorio se entrega en **grupos de 3 personas**
- El laboratorio se debe presentar en el Laboratorio de Redes.
- La fecha de entrega es el día **23 de Octubre a las 23:55 hrs.**
- La tarea se revisará en las máquinas virtuales, por lo que, los archivos necesarios para la correcta ejecución de esta, deben estar en ellas. Recuerde que el código debe estar indentado, comentado, sin Warnings ni errores.
- Se aplicará un descuento de **5 puntos** al total de la nota por cada Warning, Error o Problema de Ejecución.
- Debe dejar un **MAKEFILE** o similar en cada máquina virtual asignada a su grupo para la ejecución de cada entidad. Este debe manejarse de la siguiente forma:
 - `make docker-ONU`: Iniciará el código hecho en Docker para la ONU.
 - `make docker-OMS`: Iniciará el código hecho en Docker para la OMS.
 - `make docker-datanode`: Iniciará el código hecho en Docker para los Datanodes.
 - `make docker-continentes`: Iniciará el código hecho en Docker para los continentes.
- Debe dejar un **README** en el repositorio, que indique las instrucciones de ejecución y lo asumido para el desarrollo del laboratorio.
- No se aceptan entregas que no se encuentren dentro del repositorio, tampoco de las máquinas y que no puedan ser ejecutadas desde una consola de comandos. Incumplimiento de estas reglas, significa **nota 0**.
- No se revisará de manera local los archivos, solo los almacenados en las máquinas virtuales.
- Cada hora o fracción de atraso se penalizará con un descuento de **5 puntos**.
- Copias serán evaluadas con **nota 0** y serán notificadas a los profesores y autoridades pertinentes.

9 Consultas:

Para hacer las consultas, recomendamos hacerlas por medio del foro del ramo en Aula. De esta forma los demás grupos pueden beneficiarse en base a la pregunta. **Se responderán consultas hasta 48 hrs. antes de la fecha y hora de entrega.**