



PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERIA
DEPARTAMENTO INGENIERIA DE SISTEMAS
Proyecto Introducción a Sistemas Distribuidos

Período Académico 2022-01

Sistema de Medición de Calidad del Agua

Objetivos del Proyecto

- Desarrollar una solución a un problema de estructura distribuida
- Utilizar patrones de comunicación síncronos y asíncronos.
- Resolver problemas presentes en sistemas distribuidos, tales como fallas en los componentes.
- Reconocer elementos de desempeño y calidad asociados a la implementación de un sistema distribuido.
- Aplicar pasos relevantes del método científico para evaluar el rendimiento de algoritmos distribuidos



Contexto

El agua es uno de los recursos vitales en nuestro planeta. Con una población en aumento, es importante monitorear la calidad de este recurso para detectar cambios y tomar las acciones requeridas. La calidad de agua es una medida de qué tan adecuada es una fuente hídrica para un uso específico, por ejemplo, agricultura o generación de energía. Hay varios parámetros esenciales que nos ayudan a medir la calidad del agua, entre ellos el PH, la temperatura, la conductividad, el oxígeno disuelto y la turbidez (<https://www.hannacolombia.com/blog/post/190/guia-para-la-medicion-en-campo-calidad-del-agua>)

En este proyecto realizaremos un sistema distribuido sencillo y confiable (tolerante a fallas), donde se simula la medición de tres parámetros de una reserva de agua: PH, temperatura y oxígeno disuelto. Estos parámetros se enviarán a procesos monitores, que los almacenarán en forma adecuada y avisarán al sistema de calidad en caso de que se genere alguna alerta con los indicadores medidos.

Descripción del Sistema a Desarrollar

A continuación, se explica cada uno de los componentes de la arquitectura de software (ver figura 1):

Sensores: los sensores son simulados por procesos o hilos. Existirán tres tipos de sensores: **de PH, de temperatura y de oxígeno**. Al levantar cada proceso sensor, éstos deberán recibir 3 argumentos: tipo de sensor (temperatura, PH u oxígeno), cada cuanto tiempo (t) se va a enviar la medición a los procesos monitores y **un archivo de configuración**. En la tabla 1 se pueden observar valores válidos para cada uno de los parámetros de calidad. En el archivo de configuración se encontrarán tres probabilidades: probabilidad de producir valores dentro del rango, probabilidad de producir valores fuera del rango y probabilidad de producir valores erróneos (para el caso del proyecto serán valores negativos). Por ejemplo, si el archivo de configuración contiene, para la variable temperatura las siguientes probabilidades:

0.6 Valores correctos

0.3 Valores fuera del rango

0.1 Errores

De 10 medidas producidas por el sensor, 6 (60%) serán correctas, 3 (30%) estarán fuera del rango y 1 (10%) será invalida, por ejemplo:

68 69 -1 90 75 81 70 100 80 95

De acuerdo con estas probabilidades, los procesos sensores producirán un número aleatorio con valores dentro del rango, fuera del rango o erróneos, cada t unidades de tiempo. Esas mediciones se “publicarán” en el sistema para su consumo por parte de los monitores. Cada medición debe ir acompañada del tiempo (u hora) en el cuál se produjo.

Nota 1: el archivo de configuración puede tener otros valores de probabilidades distintos al del ejemplo.

Nota 2: dependiendo de su diseño y de lo que permita la librería, estos sensores pueden publicar los datos directamente o enviarlos a un proceso que sea quien publique la información a los monitores. Recuerde que pueden ejecutarse varios sensores al tiempo. El número de sensores debe ser como mínimo 6 (2 por cada variable a medir).

Monitores: son procesos que reciben las medidas de los sensores, validan que los datos no contengan errores y se encuentren en rangos razonables. Los monitores almacenan en un archivo o en una BD todas las medidas sin errores, aunque estén fuera del rango. Cada vez que encuentran una medida fuera del rango deben generar una alarma al **sistema de calidad**. **Un monitor puede ser suscriptor solamente de un tipo de indicador** (temperatura, PH u oxígeno)

El argumento que recibe este tipo proceso al iniciarse (vía argumentos del main) es el indicador al cual se va a suscribir.

Sistema de Calidad (SC): La única función de este proceso es esperar las posibles alarmas de los monitores e imprimir las alertas que llegan por pantalla. Cada vez que llega una alerta, debe imprimir el valor de Temperatura, PH u Oxígeno Disuelto que disparó la alerta. Al sistema de calidad deben usar solo los usuarios autorizados. **Implemente el acceso al sistema con usuario y contraseña utilizando una función de hash.**

Tabla 1: Rangos de los parámetros de calidad.

Parámetro	Valor mínimo	Valor máximo
Temperatura	68°F	89°F
PH	6.0	8.0
Oxígeno Disuelto	2 Mg/L	11Mg/L

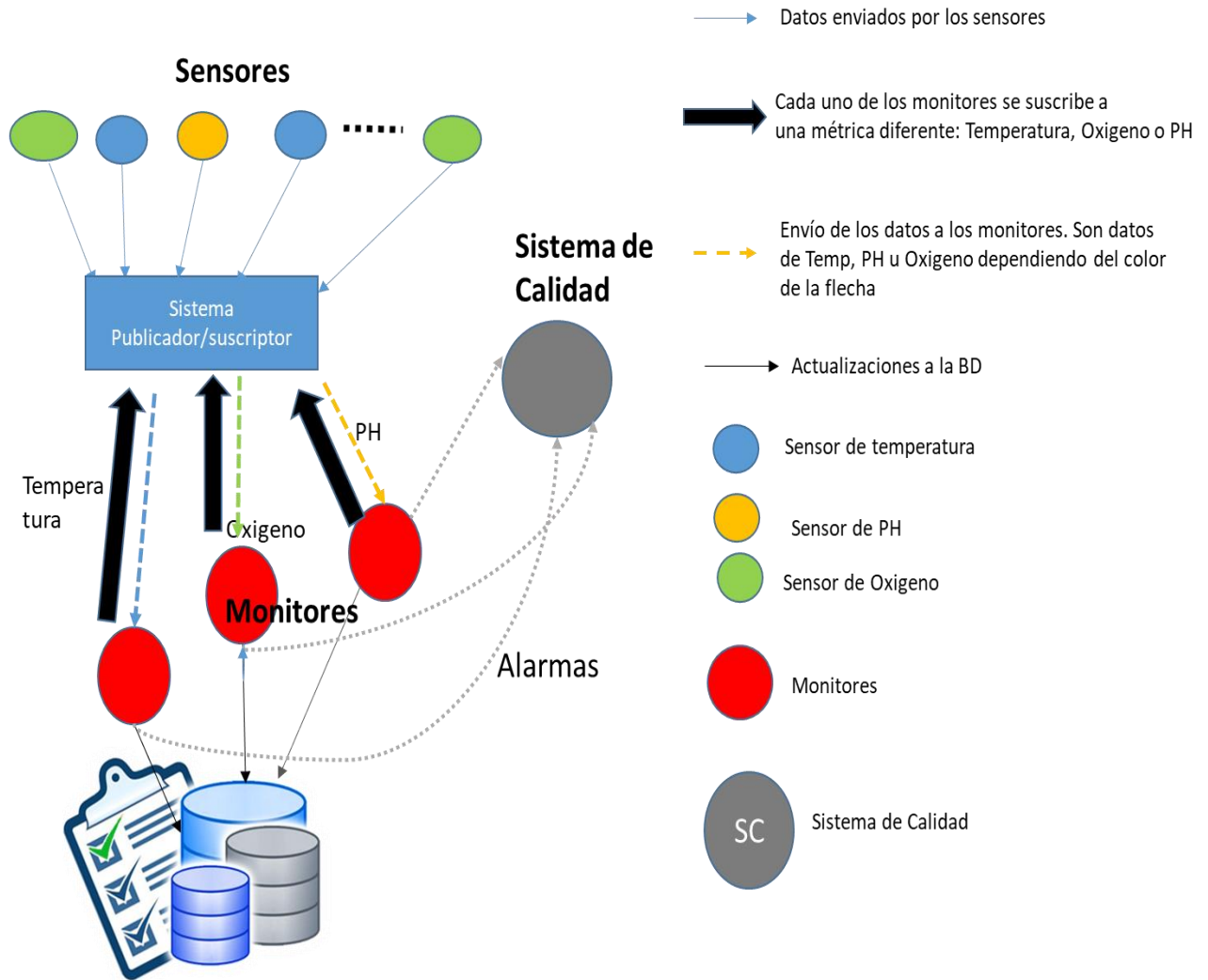


Figura 1: Arquitectura del Sistema

Detalles de Implementación

Comunicación: la comunicación entre los **sensores y monitores se debe hacer con el mecanismo publicador/suscriptor**. La comunicación entre los monitores y el sistema de calidad se puede realizar utilizando cualquiera de los otros patrones que ofrezca la librería ZeroMQ y que el equipo considere adecuado para la notificación.

Fallas: debe implementar una arquitectura tolerante a fallas, considerando una posible falla del componente monitor. Una forma de implementación es tener un componente réplica que realice las mismas funciones. **Debe existir también otro proceso encargado de chequear que todos los monitores están o no en funcionamiento (health check)**. Cuando se detecte que uno de estos procesos deja de funcionar, el proceso réplica debe asumir la carga. La reconexión del nuevo proceso y distribución de cargas debe hacerse de forma automática, es decir, será transparente para el cliente; las operaciones deben continuar en el punto en que se habían dejado y no iniciar desde el principio.

Necesario para la evaluación: el día de la sustentación es importante que se pueda observar:

- Operaciones que van realizando cada uno de los procesos y resultado de la operación, si aplica.
- Información guardada en la BD

Rendimiento del Sistema

Variables para medir el rendimiento del sistema

1. Defina variables que permitan medir el rendimiento de su sistema, por ejemplo: tiempo de almacenamiento de cada medición (tiempo en llegar desde el sensor a la BD), tiempo que tarda la llegada de cada alarma al sistema de calidad (desde el monitor que la detecta, hasta que se imprime), utilización de los procesadores, etc.
Debe definir al menos dos tipos de variables o métricas de rendimiento.
2. Establezca el procedimiento para tomar el valor de dichas variables, por ejemplo, va a instrumentar el programa?, va a usar ciertas herramientas especiales de monitoreo, cuáles serán?, etc.

Elementos que afectan el rendimiento del sistema

Existen muchos elementos que afectan el desempeño del sistema, por ejemplo: la carga de trabajo, los patrones de comunicación utilizados, la congestión en la red o en los servidores, etc.

La siguiente actividad que Ud. va a realizar es incorporar un elemento a su sistema que afecte su desempeño. Llamaremos **sistema A** al sistema sin este elemento (el sistema inicial que implementó) y **sistema B**, aquel que tienen un elemento que afecta su desempeño. A continuación les damos ejemplos de factores que pueden afectar el rendimiento:

- **Aumentar la carga:** disminuir el tiempo de generación de medidas desde los sensores o aumentar el número de sensores.
- **Patrones de comunicación:** en lugar del patrón publicador-suscriptor usar otro patrón como el push/pull de la librería (<https://programmerclick.com/article/915097237/>) para hacer llegar las mediciones de los sensores a los monitores.
- **Congestión en los servidores:** genere trabajos dummy adicionales en los servidores donde se ejecutan los monitores.

Comparación de los sistemas

Una vez que el grupo decida las variables a medir, cómo las va a medir y qué elemento va a incorporar para afectar el rendimiento del sistema, realice los siguientes pasos del método científico para comparar el sistema A con el sistema B

- Plantee una hipótesis.

- Realice medidas (experimentación) para comparar el sistema A con el B
- Análisis sus resultados
- Exponga las conclusiones

A continuación algunos enlaces que explican los diferentes pasos del método científico:

<https://www.todamateria.com/pasos-del-metodo-cientifico/>,

<https://www.universidadviu.com/int/actualidad/nuestros-expertos/cuales-son-las-fases-del-metodo-cientifico>.

PRIMERA ENTREGA

La primera entrega se realizará en la semana 10 a través de la plataforma. Los profesores avisarán el día y hora de sustentación.

La primera entrega consta de un informe donde se debe especificar:

- **Diseño del sistema:** Diagrama de clases y diagrama de secuencia. Este diseño debe incluir el o los componentes para enmascarar las fallas del sistema.
- Modelo de Interacción, de fallos y de seguridad
- El protocolo de pruebas que utilizará para la entrega final
- Describa detalladamente cómo afectará o modificará el rendimiento de su sistema y cómo va a comparar el rendimiento de los sistemas A y B: hipótesis, variables a medir y herramientas de medición.
- Implementación inicial: deben tener implementada al menos la generación de mediciones desde los sensores y su recepción por parte de los procesos monitores usando el patrón publicador/suscriptor.

Se deben implementar los requerimientos de acuerdo con el enunciado y la rúbrica.

El día de la sustentación, cada equipo tendrá como máximo **15 minutos** para mostrar sus resultados. Posteriormente responderá las preguntas que tengan los profesores.

SEGUNDA ENTREGA

La entrega se realizará en la semana 17 a través de la plataforma. Las sustentaciones se harán los días que correspondan para cada curso. El día de la sustentación los integrantes del equipo deben **mostrar la funcionalidad del proyecto**. Deben estar presentes todos los integrantes del grupo.

La entrega se compone de:

- En un archivo .zip código fuente de los programas que conforman el sistema y un archivo Readme donde indique cómo ejecutarlo.
- Se debe complementar la documentación de la primera entrega

- Un video de máximo 10 minutos donde muestra la topología implementada y explicar los siguientes aspectos de su proyecto:
 - a. Distribución de componentes en máquinas.
 - b. Librerías y patrones usados
 - c. Cómo se implementó la arquitectura confiable (tolerante a fallas)
- Un informe de máximo 5 páginas donde explican: hipótesis, experimentos realizados (especificaciones de hw y sw donde se realizaron las medidas, variables e instrumentos de medición), resultados obtenidos (si es posible acompañados por gráficos o tablas), conclusiones.

Equipos de Trabajo

El proyecto se realizará en grupos de trabajo de máximo tres personas

No puede existir replicación de documentos ni de código fuente entre grupos, lo cual se consideraría plagio.

MC/RG