

Procesamiento HPC en Sistema de Silla Inteligente

Ajaya, William. Andres, Martin. Serrano, Alex. Trejo, Ian. Velazquez, Julian.

¹Universidad Nacional de La Matanza,
Departamento de Ingeniería e Investigaciones Tecnológicas,
Florencio Varela 1903 - San Justo, Argentina
williamajaya@gmail.com, Martinfernandoandres1996@gmail.com,
alexherrero.unlam@gmail.com, ianntrejo96@gmail.com, julycabj22@gmail.com.

Resumen. El objetivo de esta investigación es buscar una manera de exprimir todo el potencial de la computación en paralelo para agilizar el procesamiento de las grandes cantidades de datos recolectadas en tiempo real por medio de la aplicación móvil “Silla Inteligente” para realizar análisis estadísticos, y así poder obtener información relevante para varias áreas de competencia (Principalmente el área de salud).

Palabras claves: Big Data, HPC, OpenCL, Paralelismo, Sedentarismo.

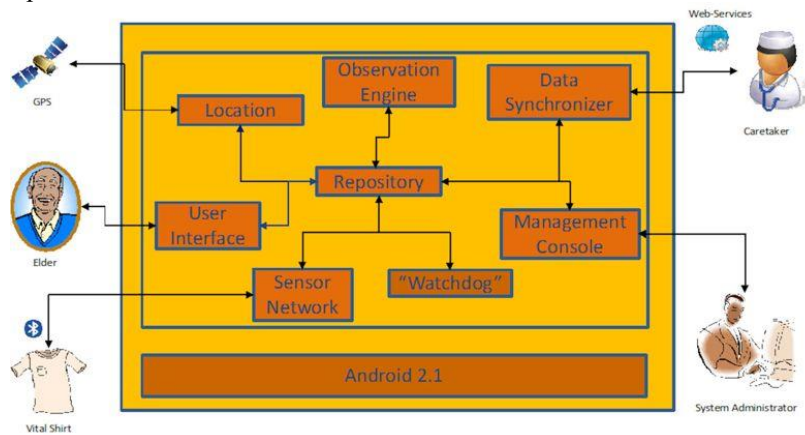
1 Introducción

El ser humano, en la actualidad, pasa una gran parte del tiempo sentado durante el transcurso de su vida, ya sea por motivos de salud, laborales, académicos, y un largo etcétera. El sistema “Silla Inteligente”, que consta de una parte de software compatible con Android, y una parte de hardware compuesta por un sistema embebido y varios sensores, busca reducir el impacto en nuestra salud que origina el estar sentado mucho tiempo al día.

Varios estudios académicos (Patología degenerativa de la columna lumbar. Rev. Soc. Esp. Dolor vol.14 no.3 Madrid abr. 2007) concuerdan en que en las últimas décadas, los problemas asociados al sedentarismo se han incrementado, entre los que se encuentran patologías degenerativas de la columna lumbar, diabetes, osteoporosis, sobrepeso, pérdida de visión a temprana edad (por la larga exposición a la pantalla del ordenador o la televisión) y una larga lista de daños colaterales, que podrían mitigarse con una silla en buenas condiciones, manteniendo una buena postura al sentarse, y evitando el estar mucho tiempo sentado (de forma continua) realizando ejercicios de estiramiento muscular, tratando de acumular 30 minutos de ejercicio diarios.

La aplicación mobile propia del sistema, inicialmente busca generar conciencia en la persona que la utiliza. Sin embargo, gran parte de estos datos podrían analizarse mediante BigData para encontrar patrones de comportamiento en diferentes áreas geográficas, rangos de edad y sectores ocupacionales, para generar reportes estadísticos que puedan servir como fuente para investigaciones académicas, o para que los ministerios de salud estén al tanto de las conductas de su población y así poder, por ejemplo, iniciar campañas de concientización.

Las aplicaciones orientadas al cuidado de la salud son una tendencia que van en crecimiento (registrando un crecimiento del 156,6% en Android, durante el último año). Tomando por ejemplo la fundación CETEMMSA, que tiene en proyecto la aplicación eCAALYX, que dispone de una camiseta con sensores, capaces de monitorizar el ritmo cardíaco, ritmo de respiración y la temperatura corporal. Ésta se va a comunicar de forma directa con los profesionales de la salud correspondientes en tiempo real.



Arquitectura de la plataforma móvil eCAALYX.

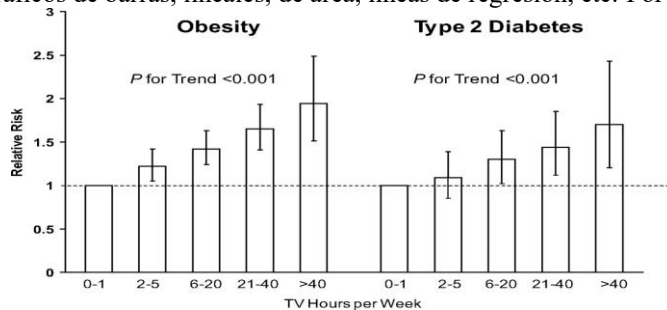
Fuente:

<https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/1475-925X-10-24>

Se elige BigData como base para el análisis de los datos recolectados, ya que sus características principales (Las cuatro “v”) coinciden con lo que se necesitaría para sacarle el mayor provecho a los datos provenientes de Silla Inteligente:

- Volúmen: Los volúmenes masivos de datos recolectados, producidos e intercambiados.
- Velocidad: La necesidad de un análisis continuo y en tiempo real.
- Variedad: Las muchas formas diferentes de datos que pueden ser difíciles de integrar.
- Veracidad: Los problemas relacionados con la incertidumbre y la confianza.

Una de las formas en las que puede obtenerse información estadística útil de los datos recolectados, es por medio de la creación de gráficos comparativos, para que sea más fácil para los expertos de salud interpretar los resultados de los análisis. Es decir, gráficos de barras, lineales, de área, líneas de regresión, etc. Por ejemplo:



Fuente: <http://hpsociety.info/news/type-2-diabetes-obesity.html>.

2 Desarrollo

Se solicita al usuario previamente su consentimiento para recolectar y utilizar sus datos de manera anónima, incluyendo tanto los datos propios obtenidos de los sensores durante intervalos de tiempo regulares, como los datos personales. Entre los datos con los cuales se debería empezar a trabajar en el proyecto, se encuentran los siguientes:

- Sexo.
- Edad.
- Ubicación.
- Distancia de la espalda a la silla.
- Temperatura del ambiente.
- Luminosidad del ambiente.
- Tiempo sentado

Si bien no hay una gran cantidad de campos presentes, la manera en la que pueden estar conectados entre sí, la relación entre esas conexiones y los datos ya existentes sobre estudios previos pueden dar lugar a una gran cantidad de combinaciones entre ellos, dando la posibilidad de encontrar patrones de comportamiento antes no estudiados. Mediante un consenso con profesionales e investigadores académicos orientados al área de salud, se construirá un conjunto de reglas (Base de conocimiento) que determinaran las posibles conexiones entre los datos a analizar.

Dando un ejemplo, si tomamos el campo de “Tiempo sentado” (Qué es calculado en la aplicación) y lo relacionamos con el campo “Distancia de la espalda a la silla” y “Ubicación”, podemos realizar un promedio para relacionarlo con las estadísticas existentes de enfermedades degenerativas del disco lumbar en un sector geográfico determinado.

Para realizar un análisis completo, deberán analizarse todos los campos por separado, y luego relaciones entre los mismos, por lo que el costo de procesamiento será bastante alto, ya que desde un inicio se requieren de utilizar cálculos estadísticos conocidos (Promedio, moda, desviación estándar, regresión lineal) sobre enormes cantidades de datos, por lo que se pretende utilizar el paradigma de HPC (Computación de altas prestaciones) conocido como GPGPU, más precisamente OpenCL.

OpenCL (Open Computing Language o lenguaje de computación abierto) consta de una interfaz de programación de aplicaciones y de un lenguaje de programación. OpenCL permite crear aplicaciones con paralelismo a nivel de datos y de tareas que pueden ejecutarse tanto en unidades centrales de procesamiento como en unidades de procesamiento gráfico. Esto puede reducir considerablemente los tiempos de procesamiento en los cálculos de inferencia estadística, aunque en gran parte seguirán dependiendo de la forma en la que será construida la base de conocimiento.

La implementación física se planéa realizar sobre un servidor Intel Xeon R4424 equipado con una AMD Fire Pro W9100, como inversión inicial, para luego plantear posibles expansiones en caso de que la demanda supere al rendimiento del equipo.

3 Explicación del algoritmo.

De todos los algoritmos que pueden beneficiarse de HPC, será como dado el más básico y uno de los mayores beneficiados

Básicamente, cada campo de nuestra base de datos será procesado en paralelo, para obtener una relación entre el dato y nuestra base de conocimiento.

```
void mainFunction(Tupla *t)
{
    for(i=0;i<Ncampos;i++){
        evaluarCampos(t,i);

        For(j=0;j <t->campo[i].Nrelaciones ;j++)
            actualizarEstadisticas(t->campo[i].relacion[j]);
            actualizarGraficas(t->campo[i].relacion[j]);
    }
}

__kernel void evaluarCampos (__global Tupla **t, __global int TipoCampo)
{
    switch(TipoCampo){
        case DISTANCIA_SILLA: generarRelaciones(t, TipoCampo);
        case TIEMPO_SENTADO: generarRelaciones(t, TipoCampo);
        case TEMPERATURA: generarRelaciones(t, TipoCampo);
        case SEXO: generarRelaciones(t, TipoCampo);
        case EDAD: generarRelaciones(t, TipoCampo);
        case UBICACIÓN: generarRelaciones(t, TipoCampo);
        case LUMINOSIDAD: generarRelaciones(t, TipoCampo);
    }
}
```

4 Pruebas que pueden realizarse

-Implementar un sistema web que muestre como se actualizan los gráficos estadísticos en tiempo real, para que cualquier persona pueda consultarlos en cualquier momento del día, y poder solicitar su estadística personalizada.

-Verificar que verdaderamente hay una diferencia de rendimiento, podría implementarse el servidor en primera instancia sin OpenCL, paralelizando con OpenMP y realizar comparativas de performance.

-Probar el servicio con una enorme base de conocimiento y tests de fatiga para comprobar que el sistema responda de manera adecuada.

5 Conclusiones

En resumen, se ha propuesto la creación de una herramienta que combina el poderío de Big Data y HPC para encontrar nuevas formas de contribuir a la salud y al bienestar de las personas, siendo que normalmente son herramientas utilizadas para fines comerciales. Sin embargo, el éxito de la plataforma dependería parcialmente de la contribución de los usuarios y de la utilidad que los expertos de la salud puedan encontrar en ella, además de que la construcción de una base de conocimiento para poder relacionar los datos recolectados entre sí, ocuparía de la opinión de muchos expertos, para llegar a un acuerdo en cómo estas relaciones de datos podrían relacionarse con las distintas enfermedades y estudios previos existentes.

Encontramos en OpenCL una forma de optimizar la herramienta, ya que podría mantener actualizadas las estadísticas y gráficos en tiempo real, algo que sería complicado sin una técnica de paralelización, ya que por ejemplo, generar varios gráficos de regresión lineal al mismo tiempo con una gran cantidad de datos ingresando en todo momento, podría colapsar al servidor.

Para trabajos futuros, se recomendaría buscar una forma de obtener más datos, por ejemplo, añadiendo más sensores (Sin que esto impacte fuertemente en el precio del sistema Silla Inteligente) y mantenerse al pendiente de los estudios científicos relacionados a las enfermedades orientadas al sedentarismo.

6 Referencias

1. López, José - Neurocirujano.: Patología degenerativa de la columna lumbar. Rev. Soc. Esp. Dolor vol.14 no.3 Madrid abr. 2007
2. Cabrera de León, Antonio.: Sedentarismo: tiempo de ocio activo frente a porcentaje del gasto energético. :Revista Española de Cardiología Volume 60, Issue 3, March 2007, Pages 244-250
3. Boulos, M., Wheeler, S., Tavares, C., Jones, R. : How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX (2011). : <https://doi.org/10.1186/1475-925X-10-24>
4. Koumpouros, Yiannis.: Big Data in Healthcare : Technological Educational Institute of Athens, Greece (2014)
5. Venus Henry-Fuenteseca, Ing. :Paralelización del Algoritmo Expectación-Maximización Utilizando OpenCL : Instituto Superior Politécnico José Antonio Echeverría La Habana, Cuba (2014)
6. Guangbao Guo.: Parallel Statistical Computing for Statistical Inference : Shandong University of Technology (2012)