



# Comunidad Web Caché

Un Modelo de Caché Web Distribuido  
Sobre el protocolo HTTP y P2P

Kevin Moraga

“

*Sólo un tonto no tiene **miedo**.*

*El **valor** es ver el miedo y seguir adelante de todas formas.*

*Julian Assange*

# ● Agenda

○ Introducción

Fundamento teórico

Diseño de la CWC

Resultados Obtenidos

Conclusiones

Trabajo Futuro





CWC

# Introducción

¿Cómo nació la idea?

- ¿Qué es la Comunidad Web Caché?

- La CWC usca utilizar el concepto de comunidad en un entorno P2P para poder potenciar la transferencia de datos desde un servidor web.

A thin vertical line runs down the left side of the slide, with a small open circle positioned at the same vertical level as the main text.

# 2,923 Millones

Usuarios de Internet en el 2014

● Distribuidos alrededor del mundo

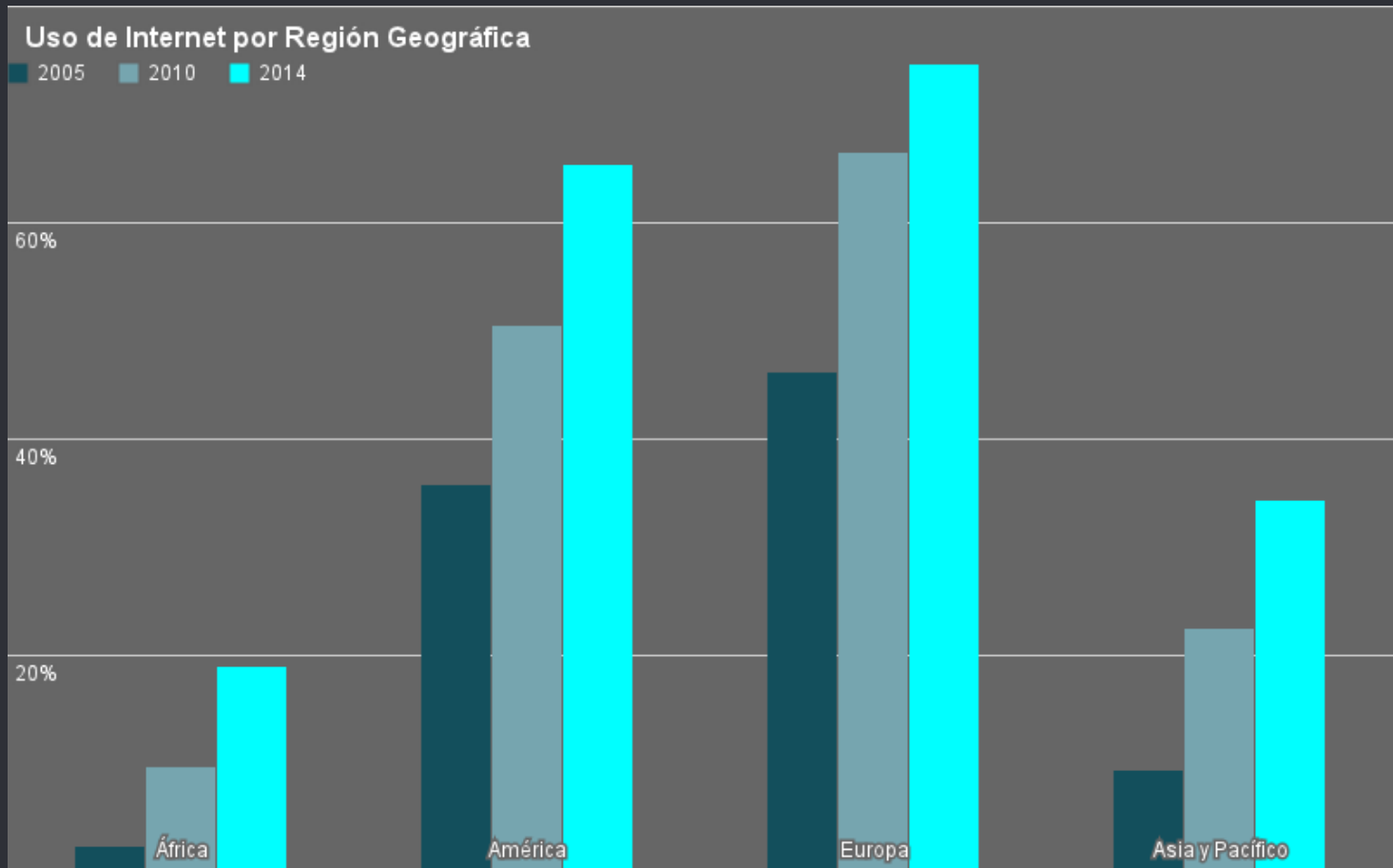


## ● Uso de Internet por región geográfica

	2005	2010	2014
África	<b>2.4%</b>	<b>9.8%</b>	<b>19.0%</b>
América	<b>35.9%</b>	<b>50.5%</b>	<b>65.5%</b>
Europa	<b>46.3%</b>	<b>66.6%</b>	<b>74.8%</b>
Asia y Pacífico	<b>9.4%</b>	<b>22.5%</b>	<b>34.4%</b>



## Uso de Internet por región geográfica



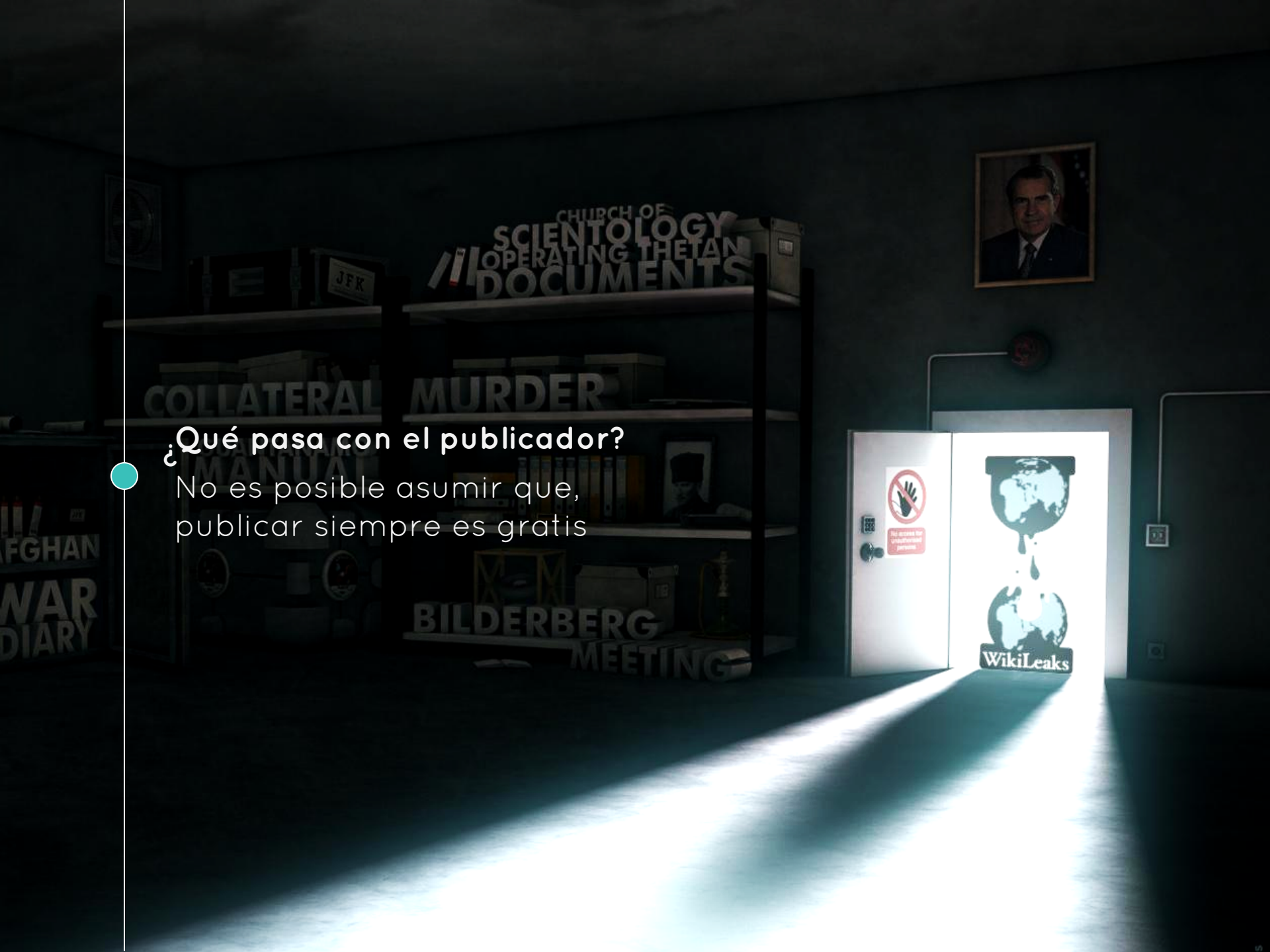


# Demanda y Recursos de Cómputo

¿Realmente se están utilizando  
eficientemente los recursos?

¿Qué pasa con el publicador?

No es posible asumir que,  
publicar siempre es gratis





Una organización con una comunidad amplia

Servidores en más  
de 40 países:

**\$200,000**

Días fuera de línea  
en 2010:

**10 días**

Seguidores en  
Twitter:

**2.1 Millones**



**-\$15,000,000**

Donaciones perdidas por bloqueos

**-\$1,000,000**

Defensa legal

**100% documentos**

Continúan en el dominio público



# Costo y Libertad

La libertad de expresión es muy cara



## Problema

Este se encuentra compuesto por al menos cinco puntos esenciales



## Problema

### Disponibilidad

El problema inherente de tener un sitio con una disponibilidad del 100%, soportando las solicitudes usuales de los clientes, las denegaciones distribuidas y las fallas en los servicios.

### Costo

Falta de recursos de quien publica contenido de calidad en Internet, esto implica un alto costo para poder atender un mayor número de clientes.

### Eficiencia

Un alto desperdicio de recursos en los clientes, entre estos recursos podemos citar:

- Ancho de Banda,
- Almacenamiento,
- Memoria y
- Procesamiento.



● Además...

### **Calidad de Servicio**

Una baja en la calidad del servicio conforme aumenta el número de clientes, tiempos de respuesta poco aceptables o bien se ve comprometida la continuidad del negocio.

### **Creciente demanda**

El número de usuarios en Internet se incrementa constantemente cada año, lo cual implica una creciente demanda de los servicios.



# Hipótesis

Objetivos del proyecto

## ● Hipótesis

○ El **desarrollo** de una caché web distribuida que permita:

- **compartir** de manera **transparente** y **eficiente**,

los recursos computacionales distribuidos ofrecidos colaborativamente, mediante:

- una **comunidad** de servidores,
- tendrá un impacto positivo en el **bajar** los *costos* y **mejorar** la *continuidad del negocio* para las comunidades virtuales.



# Objetivos

Objetivos del proyecto

- Objetivo Principal

- Desarrollar una **caché web distribuida** que permita compartir de manera transparente y eficiente los **recursos computacionales** distribuidos ofrecidos colaborativamente mediante una **comunidad** de servidores.



## ● Objetivos Específicos

### **Compilar**

y comparar trabajos relevantes relacionados con sistemas distribuidos.

### **Examinar**

las oportunidades de mejoramiento estratégico de las plataformas existentes.

### **Determinar**

los requerimientos de la caché web distribuida basada en HTTP.

## ● Objetivos Específicos (2)

### ○ Diseñar

una arquitectura de una caché web distribuida bajo un conjunto de principios tales como:

- transparencia de ubicación de los recursos
- solución de bajo costo
- calidad de servicio

### Desarrollar

un prototipo de la caché web distribuida.

### Analizar

el rendimiento de velocidad y el costo del entorno de CWC.



CWC

# Fundamento Teórico

Marco teórico del proyecto

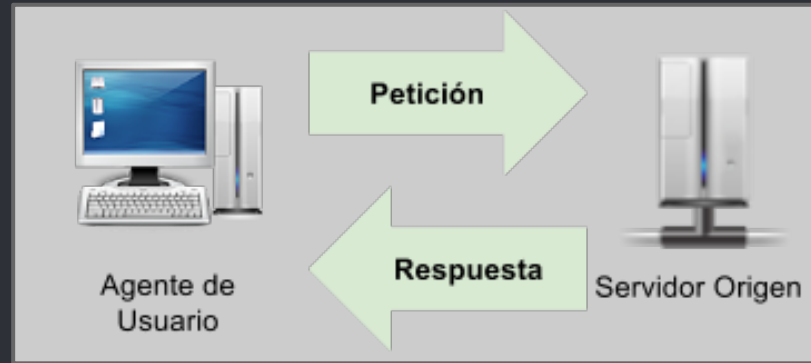




# Protocolo HTTP

Marco teórico del proyecto

## Protocolo HTTP



Es un modelo cliente-servidor, que además:

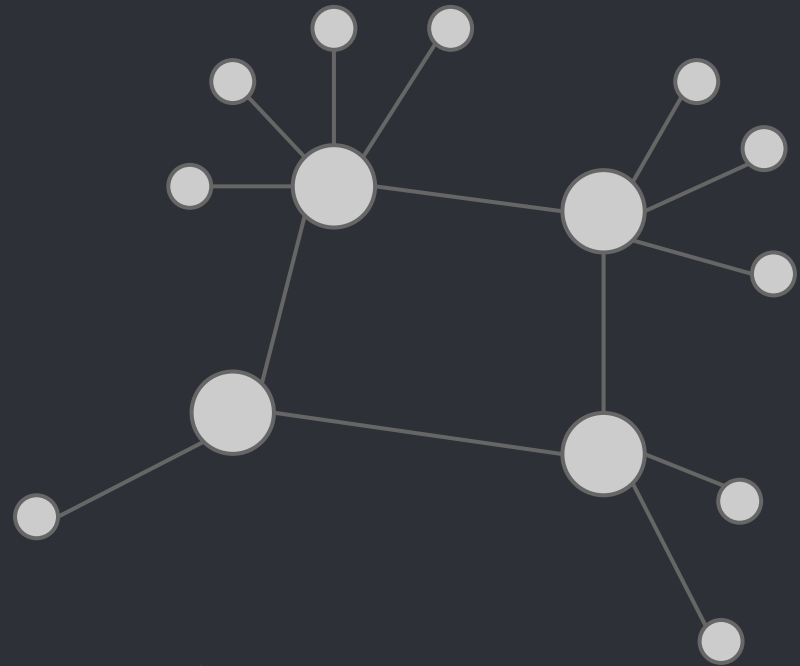
- Posee un conjunto reducido de **métodos**.
- El resto del mensaje se basa en **campos de encabezado** y el **cuerpo**.



# Protocolo P2P

Marco teórico del proyecto

## Protocolo P2P



El protocolo P2P puede clasificarse:

- Directorio Centralizado
- Directorio Distribuido
- Directorio Híbrido



CWC

# Diseño de la CWC

Requerimientos, Arquitectura y Diseño



# Requerimientos

Diseño de la CWC



## Actores

### Servidor Web

El servidor web es el encargado de servir los archivos que son solicitados por el usuario a través de un navegador web.

### Agente

El agente cliente es un navegador web, el cual cuando se hace una petición por parte de éste, se verifica si el servidor web puede utilizar el protocolo CWC.

### Usuario

Entre las actividades que tendrá el usuario tendrá: la instalación del cliente, establecer la velocidad de conexión, iniciar o detener el cliente y administrar los parámetros de la caché local.

### Administrador

Entre las actividades que tendrá el administrador: la instalación del cliente, establecer la velocidad de conexión, configurar todos los parámetros del módulo del Servidor Web y Administrará el módulo mediante la consola de administración.

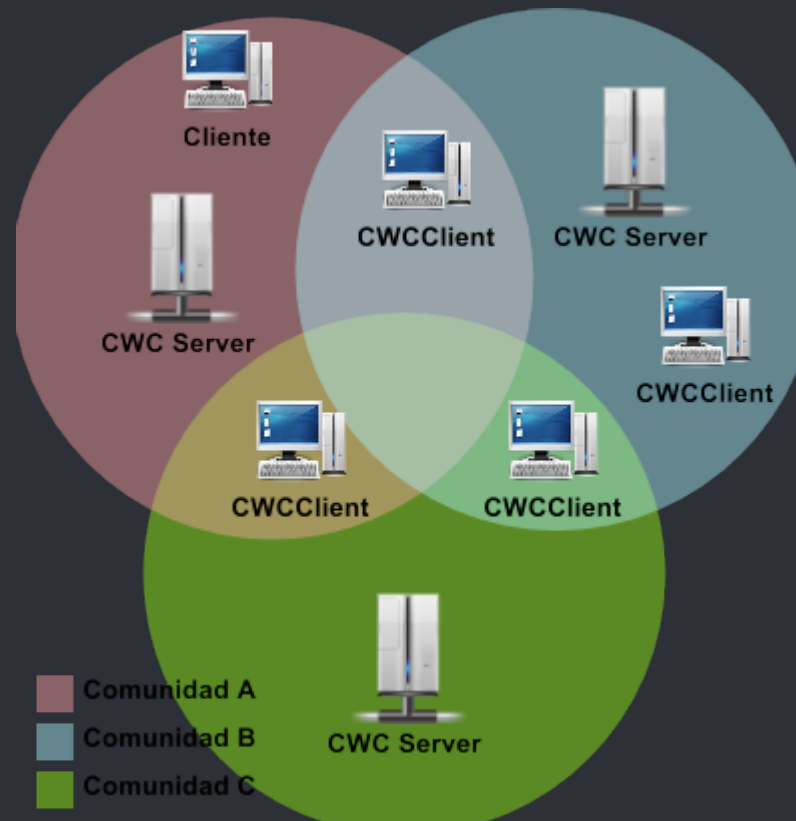


# Arquitectura

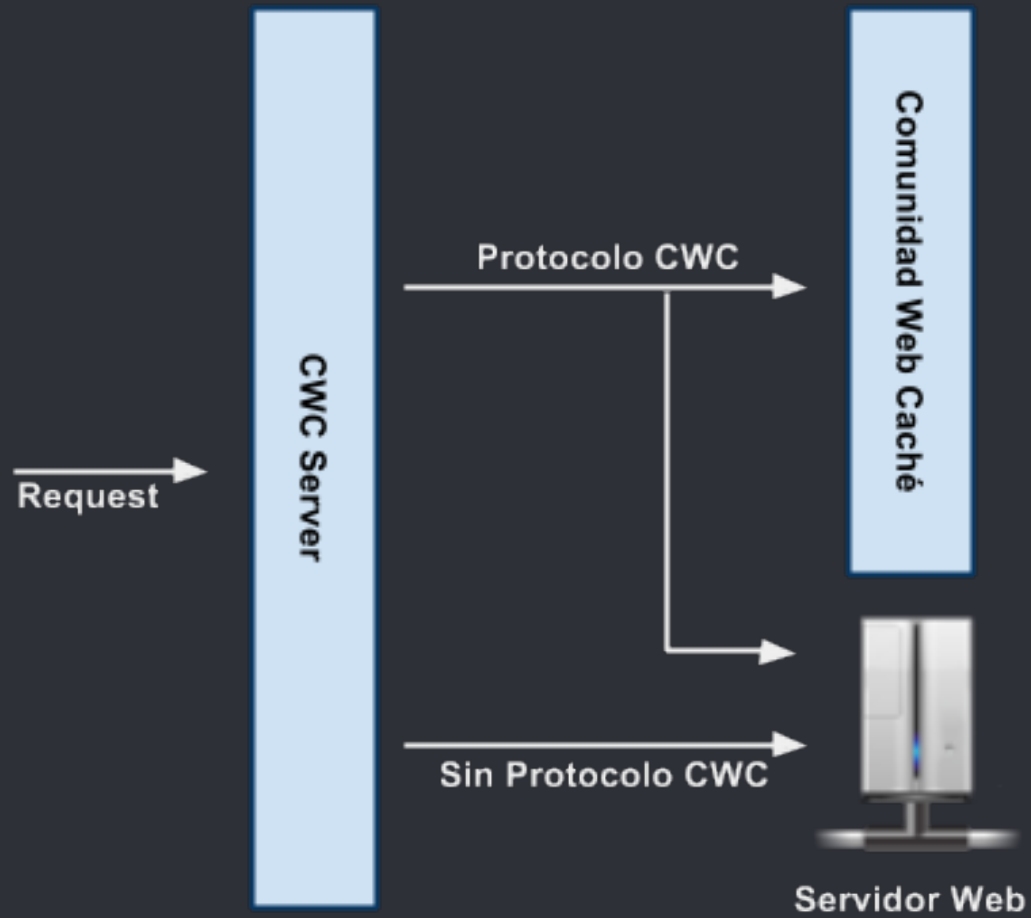
Diseño de la CWC



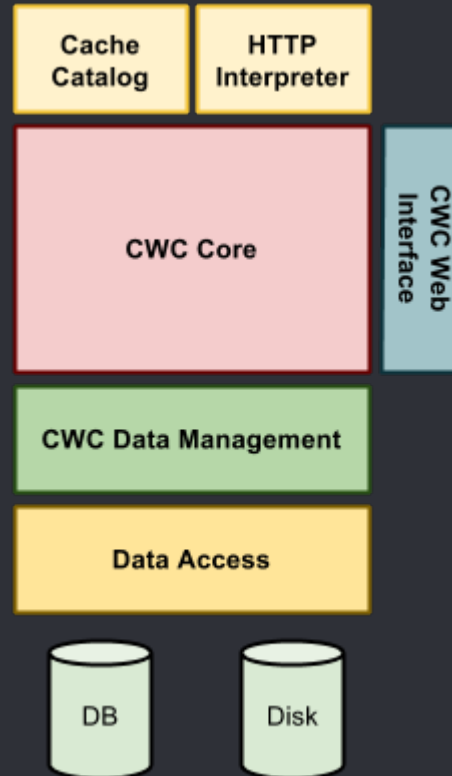
## Concepto de la CWC



## Comunicación del CWC



## Arquitectura del CWC Server

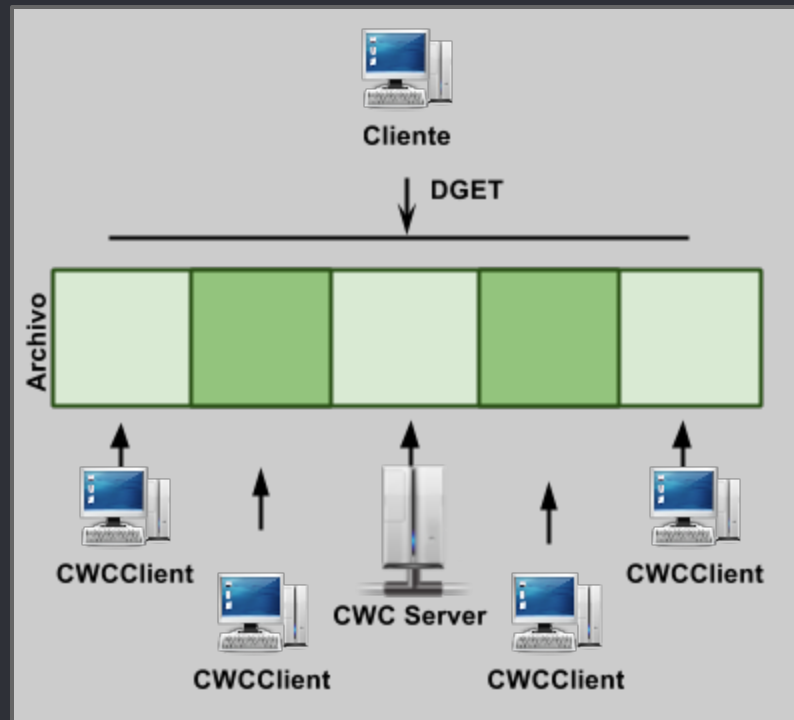




# Diseño del protocolo

Diseño de la CWC

## Funcionamiento del DGET



## Estructura del Mensaje



Los mensajes de operacion son **XML** que poseen:

- **Client ID** Identificador único
- **Req. Msg ID** número mensaje de envío
- **Instaladores** número de mensaje que se responde

## ● Algunas Operaciones

### ACK

Es un mensaje de Confirmación. Además puede ser enviado por cualquiera de las partes involucradas en la comunicación: Servidor o Cliente.

### GetClientID

Este mensaje se utiliza cuando un nuevo CWCCClient quiere formar parte de la comunidad.

### UpdateMembership

Este mensaje se utiliza para establecer el tipo de membresía, recursos que se están prestando a la caché y tipo de administración.

### En-Disable Object

En este caso el mensaje se utiliza para habilitar o deshabilitar un objeto que se encuentra en la caché.

### ModifyObject

Este mensaje es enviado por el servidor, en este caso se ha realizado un cambio del lado del servidor que debe ser propagado hacia los clientes.

### DeleteObject

Este mensaje es utilizado ya sea por los clientes o por el servidor, cuando un objeto se borra en alguno de los dos.



# Diseño del Caché Web Comunitario

Diseño de la CWC





## Funciones del CWC Server

### Administración de Objetos

La administración de Objetos en la caché local se refiere a cómo identificar cuáles objetos se pueden agregar a la caché del CWC y cuales no.

### Administración del CWC

La administración del CWC se refiere al manejo de los clientes que conforman la comunidad.

### Segmentación de Archivos

Cuando se va a servir un archivo desde la caché, éste se debe descomponer en un número de segmentos.

### Administración Remota

El servidor se encargará de decir cuando se debe tener un archivo en la caché (del lado del cliente) y cuando no se debe.

### Estadísticas

El principal objetivo es enviar la lista de cachés que tienen el objeto disponible, ordenada de tal manera que se presenten los que históricamente tiene mejores referencias de primero.



CWC

# Resultados Obtenidos

Análisis y Resultados



# Análisis de Rendimiento

Resultados Obtenidos

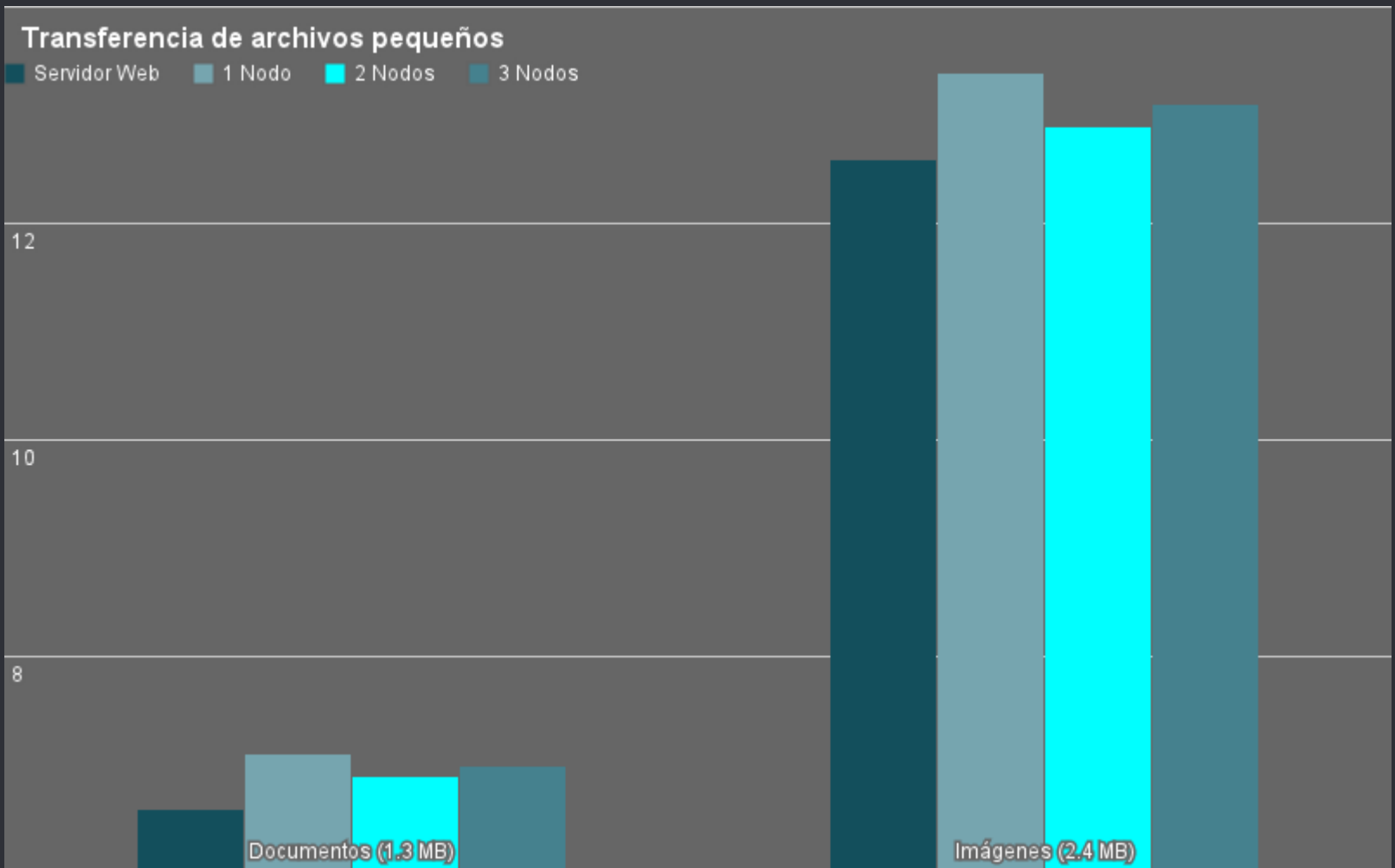
## ● Análisis de Rendimiento

○ Se realizó la medición de tiempo de transferencia para:

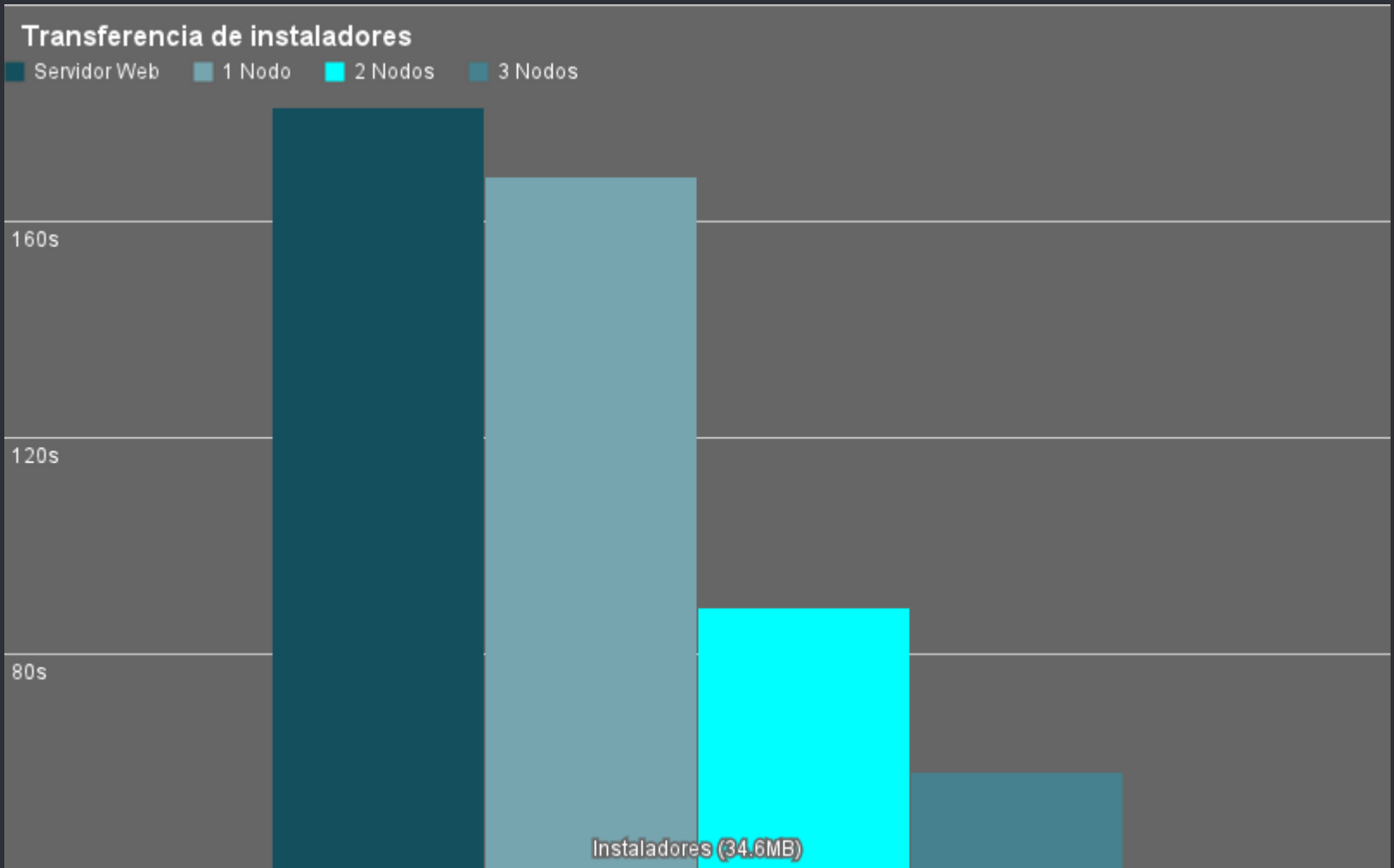
- **Imágenes** de tamaño 2.4 MB
- **Documentos** de tamaño 1.3 MB
- **Instaladores** de tamaño 34.6 MB
- **Videos** de tamaño 345.9 MB

Realizando la comparación entre HTTP y CWC, con distinta cantidad de nodos.

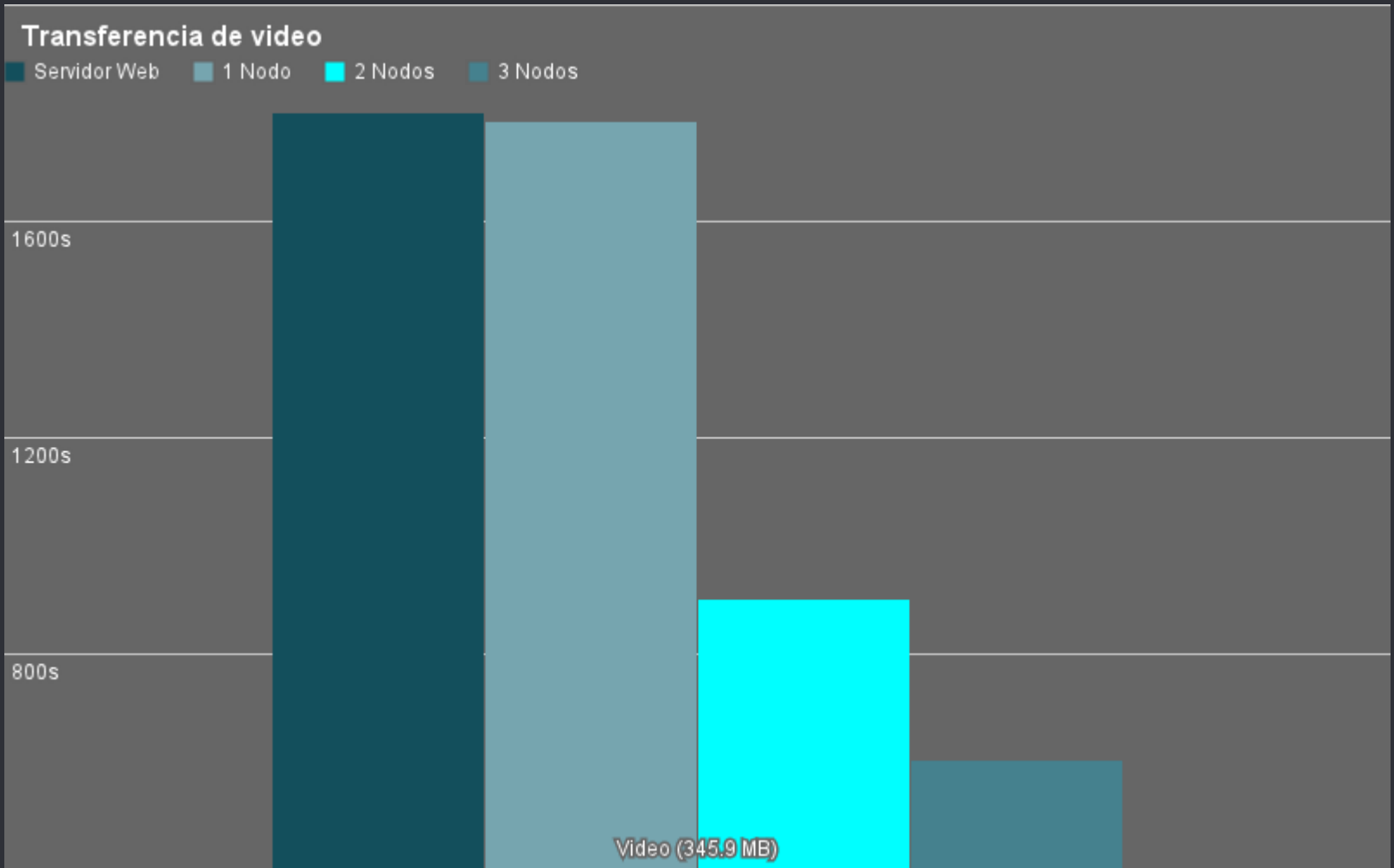
## Archivos Pequeños



## Archivos de Instalación



## Archivos de Video





# CWC y Eficiencia

La CWC utiliza eficientemente los recursos en comparación con el modelo tradicional de HTTP





# Análisis de Costo

Resultados Obtenidos

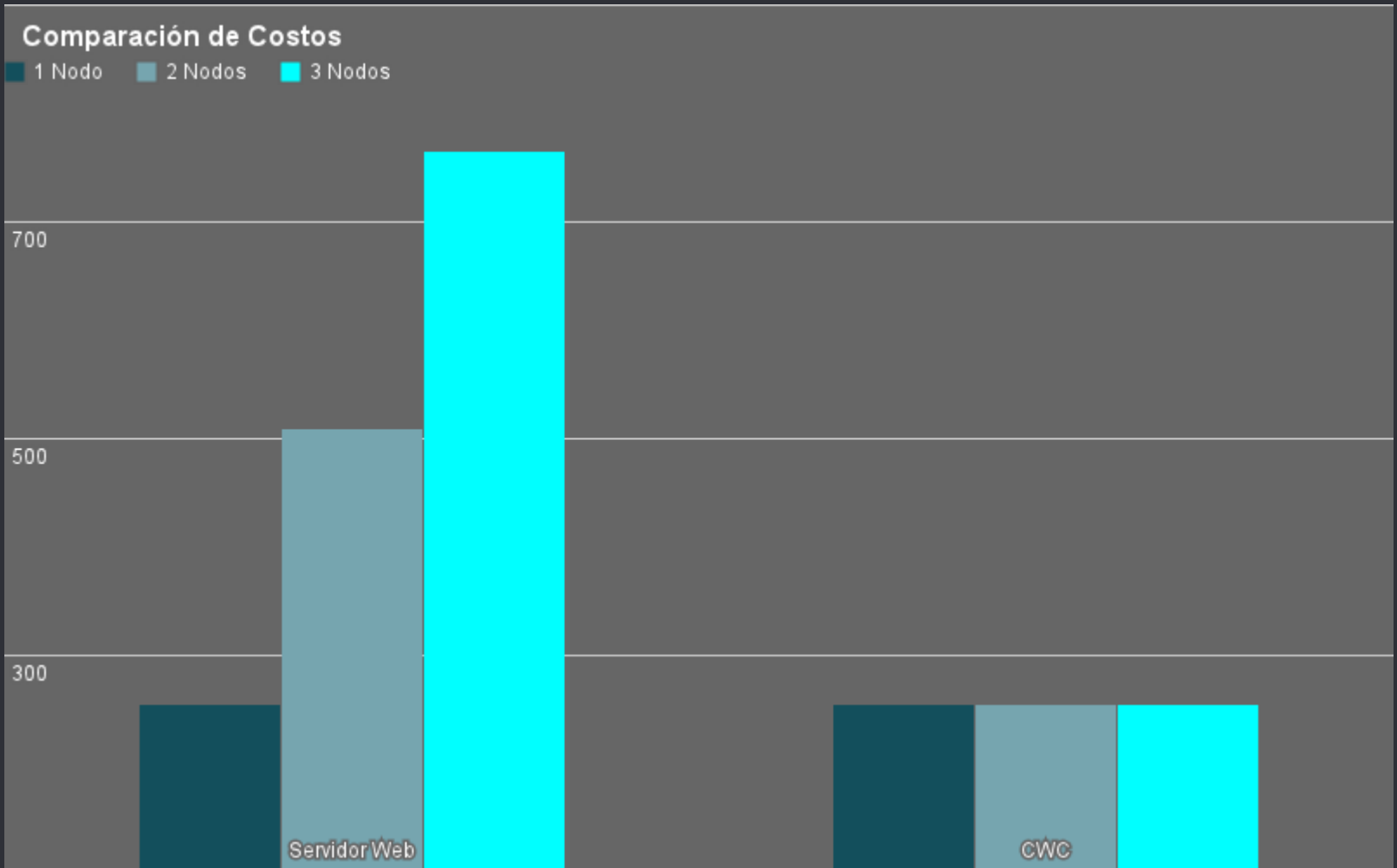
## ● Análisis de Costo

○ Se realizó un análisis de costos, con los siguientes componentes:

- **Memoria**, costo \$40
- **Procesador**, costo \$120
- **Almacenamiento**, costo \$55
- **Interfaz** de red, costo \$10

Realizando una comparación de costos, conforme aumentaba la cantidad de nodos

## Resultados de Costo





# CWC y Costos

La CWC tiene un impacto positivo en bajar los costos de una publicación en Internet



# Análisis de Continuidad de Negocio

Resultados Obtenidos

## ● Análisis de Continuidad de Negocio

○ Aunque sí es posible implementar continuidad del negocio con HTTP, **no es una tarea transparente.**

Pues:

- El **balanceo** de cargas no es nativo en HTTP
- La creación de servidores **espejo** no es parte del protocolo de HTTP



# Transparencia y Continuidad del negocio

La CWC mejora la administración de la continuidad del negocio de una manera transparente



CWC

# Conclusiones y Trabajo Futuro

Análisis y Resultados





## Conclusiones

### Trabajo Previos

Al momento de comparar los trabajos relevantes relacionados con sistemas distribuidos se encontró que muchos otros proyectos han aunado el área de web caching distribuido, pero cada uno expone su propia necesidad, y la mayoría de ellos tiene ambiciones relacionadas con las LAN.

### Investigación

En la especificación de los requerimientos de una caché web distribuida basada en HTTP, fue necesario la investigación de múltiples protocolos, modelos y análisis de distintas necesidades de mercado, para adaptar el diseño a un Internet en aumento de usuarios.



## Conclusiones de Diseño

### **Transparencia de recursos**

El proyecto pretende proveer un sistema libre de fallas, totalmente distribuido el cual ponga a disposición los recursos de la comunidad formada, a los miembros de la misma.

### **Solución de bajo Costo**

El proyecto impulsa el uso de software libre liberado bajo la licencia GPL de GNU. Además el uso de recursos subutilizados.

### **Continuidad de Negocio**

El protocolo CWC permite soportar caídas de los nodos y aún así seguir sirviendo el sitio web.

## ● Conclusiones

### ○ Mejoramiento Estratégico

En el momento de decisión de estrategias se optó por aprovechar cualquier oportunidad de mejoramiento estratégico de las **plataformas existentes**, por cual sólo se agregaron algunos comandos al estándar HTTP, sin dejar la **compatibilidad** hacia atrás con toda la infraestructura ya creada.

## Conclusiones



### Costos

Fue posible concluir que el diseño del protocolo CWC sustenta la hipótesis de decrecimiento en los costos de publicación de un sitio web, dado que conforme se aumentan los nodos, el costo de la infraestructura desciende.



### Rendimiento

Se encontró que el aumento de nodos a una infraestructura basada en el protocolo CWC, genera un impacto positivo en el tiempo de transferencia de archivos desde el sitio web hacia los clientes.



# Trabajo Futuro

Temas para futuras investigaciones

## Trabajo Futuro



### Servidor CWC

El **desarrollo** e implementación de **módulos** compatibles con un **servidor** web.



### Cliente CWC

El **desarrollo** de **plugins** para un explorador web que cumpla con la especificación del **cliente** CWC.



### Flujo multimedia

La inclusión de **algoritmos** que soporten la **transferencia** de flujos de audio, video o **multimedia** en general.



### Seguridad

La especificación de mayor seguridad dentro del protocolo CWC. Donde se tome en cuenta la transferencia de archivos de manera **confiable**, **auténtica** e **íntegra**.



### Sitios dinámicos

La especificación de algoritmos que sustenten el manejo de **transferencia** de páginas dinámicas, producto de la **programación** e **interacción** entre el usuario y un sitio web.

**iGracias!**

¿Preguntas?

@kmoragas

kevin@kmoragas.com