

Mejora de un Sistema de Auto-escalado para Sistemas Distribuidos

Trabajo de Fin de Grado

Grado de Ingeniería Informática

Autor: Miguel Alonso, Carlos

Tutor: Rampérez Martín, Victor

Febrero - 2022

ETSIINF - Universidad Politécnica de Madrid

Indice

1. Motivación del proyecto
2. Objetivos
3. Estado del Arte
4. Diseño e implementación de un generador de cargas de trabajo para sistemas pub/sub basados en contenido
5. Pruebas de rendimiento
6. Implementación de los modelos predictivos
7. Resultados y Conclusiones
8. Trabajo futuro

Motivación del proyecto

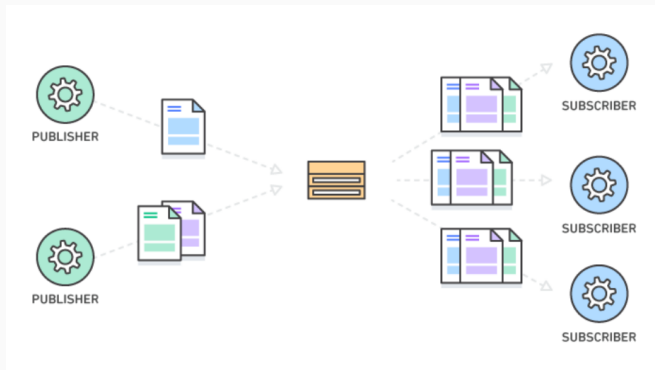
- Digitalización de la vida diaria
- Cloud Computing \implies elasticidad & *pay-as-you-go*
- Sistemas publicador/subscriptor auto-escalables
- Carga de trabajo de contexto real
- E-SilboPS

Objetivos

1. Estudio del estado del arte
2. Estudio de cargas de trabajo de contexto real para sistemas pub/sub basados en contenido
3. Diseño e implementación de pruebas de rendimiento que utilicen esta carga
4. Diseño e implementación de modelos predictivos sobre el rendimiento

Estado del Arte

Estado del Arte - Sistemas p/s



Paradigma publicador/subscriptor

Estado del Arte - Escasez de cargas de trabajo reales

Escasez de cargas de trabajo originadas de contexto real

Problemas...

- Privacidad de datos (usuarios y empresas)
- Intereses de los usuarios en las subscripciones

Cargas sintéticas basadas en parámetros estadísticos.

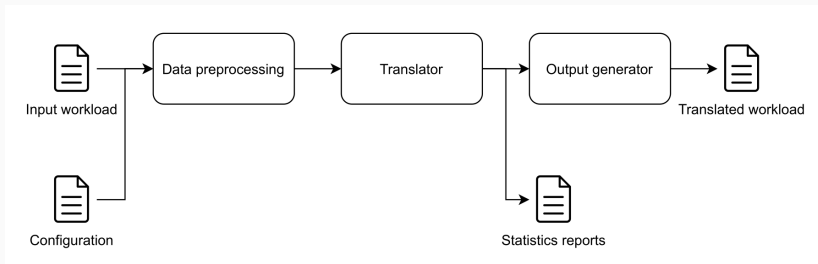
Estado del Arte - Auto-escalado

- Capacidad de adaptar recursos usados a la cantidad de trabajo recibida
- Tipos de sistemas auto-escalables
 - Reactivos (*reactive*)
 - Predictivos (*predictive*)

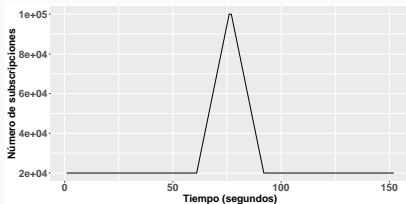
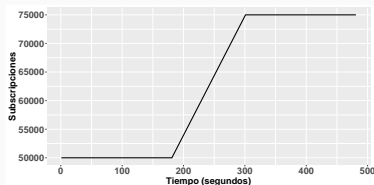
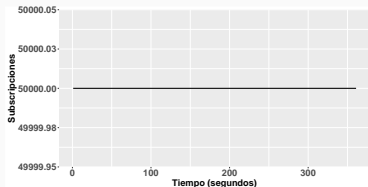
Diseño e implementación de un generador de cargas de trabajo para sistemas pub/sub basados en contenido

Generador - Generador de cargas de trabajo

- Carga extraída del proyecto *Mammoth*
- Análisis estático de la carga (datos estadísticos y compatibilidad)
- Necesidad del generador para traducir formato



Generador - Tipos de cargas de trabajo

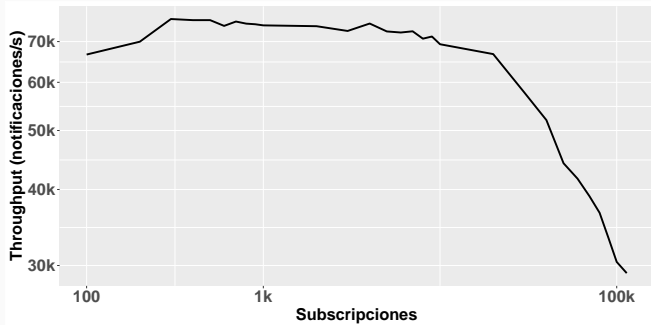


Pruebas de rendimiento

Pruebas - Pruebas implementadas

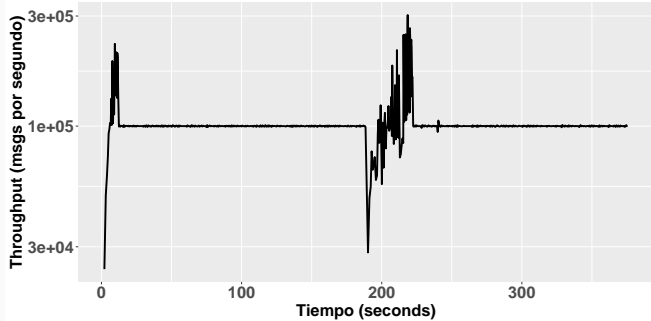
1. Toda la carga a diferentes input rates
2. Secuencia log. de subscripciones (input rate fijo)
3. Carga de subscripciones creciente (input rate fijo)
4. Carga de subscripciones estática a diferentes input rates
5. Carga de subscripciones con crecimiento puntual (input rate fijo)

Pruebas - Resultados de las pruebas



Throughput de la secuencia logarítmica de subscripciones

Pruebas - Resultados de las pruebas



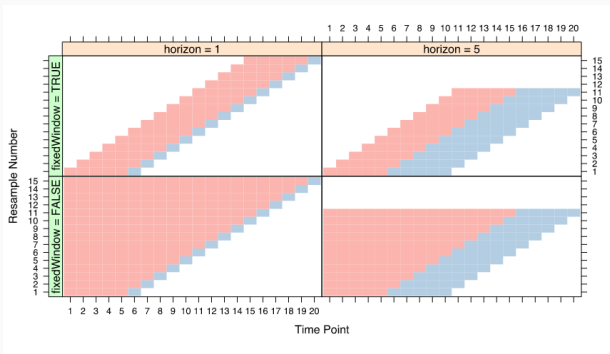
Throughput con un incremento puntual rápido del sistema

Implementación de los modelos predictivos

Modelos predictivos - Metodología

Rolling Forecasting Origin

- *initialWindow*: tam. inicial del set de entrenamiento
- *horizon*: tam. del set de test
- *fixedWindow*: series de entrenamiento



Modelos predictivos - Series Temporales y ML & DL

- Series Temporales
 - Modelos ARIMA
 - Modelos STL con ETS
- ML & DL
 - Regresión Lineal
 - Generalizados Aditivos
 - Generalizados Lineales
 - *Random Forest*
 - Redes Neuronales

Resultados y Conclusiones

- Diseño e implementación de un generador que traduce cargas de trabajo entre sistemas pub/sub
- Diseño e implementación de pruebas de rendimiento
- Implementados modelos predictivos para predecir el rendimiento del sistema
- Towards automatic evaluation and comparison of publish/subscribe systems performance improvements: A case study,
Journal of Web Engineering, JCR Q4

Trabajo futuro

- Prueba y aplicación de modelos predictivos
- Analizar la primera derivada de *throughput* y *tiempo de respuesta* para predecir la tendencia
- Analizar si añadiendo más métricas de rendimiento se mejoran las predicciones
- Aplicación de resultados al sistema de auto-escalado

Muchas gracias por su atención.