

Almacenamiento y procesamiento de un grafo de renta fija chilena en tiempo real

Estudiante: Matías Villegas
Profesor guía: Gonzalo Navarro

1. Introducción

Con el desarrollo de internet y la creciente necesidad en el almacenamiento y procesamiento de datos de forma efectiva cada vez se utilizan más las alternativas NoSQL como bases de datos [4].

Con estas alternativas NoSQL existen trade-offs en cuanto a la consistencia de la base de datos. En particular existen las bases de datos de grafos, que responden de manera eficiente cuando el problema a resolver posee una topología de nodos conectados entre ellos.

Con esto último, se presenta la posibilidad de modelar en forma de grafos los bonos de renta fija chilena. Cada nodo puede corresponder a un instrumento con una cierta duración, riesgo, emisor, etc. y cada arista puede representar la similitud entre distintos bonos.

En Chile, en la Bolsa de Comercio de Santiago (BCS), instrumentos financieros son transados diariamente. En la Bolsa es posible indicar que deseo ser un comprador (bid) o un vendedor (ask) por algún monto. Una vez que existe una pareja de comprador-vendedor que ofrecen un precio compatible se realiza una transacción del instrumento en cuestión. Esta información puede ser obtenida en tiempo real para su análisis.

Dada la posibilidad de comprar y vender a través de internet surgen nuevas estrategias en el mercado con la ayuda de la computación, como lo es la negociación de alta frecuencia (HFT) [3]. Esta última ha generado un crecimiento importante en las cantidades y volúmenes transados en países como Estados Unidos [2] o Inglaterra [1].

Para que estrategias de este tipo puedan ser implementadas es necesario tener acceso a más información, como el valor por los que los instrumentos transarían en tiempo real.

Para el caso de Chile, si bien existen transacciones diariamente, algunos instrumentos transan muchas veces (instrumentos líquidos) y otros podrían no tener ninguna transacción.

Esto hace difícil conocer el valor de los instrumentos en tiempo real, por lo que estrategias como HFT no pueden ser implementadas en la actualidad, privando al mercado chileno de lo que el beneficio de poder tener precios en tiempo real puede traer.

2. Estado del Arte

2.1. Cálculo de precios para instrumentos chilenos que no transan en la actualidad

Si bien en un mercado en que todos los instrumentos transan diariamente obtener sus valores en tiempo real es más simple, en Chile hay muchos instrumentos que no poseen liquidez y existen sólo dos empresas que realizan el cálculo de precios: LVA índices

y RiskAmerica. Este cálculo es realizado sólo una vez al día. No existe entidad ni trabajo previo de un sistema de precios en tiempo real en Chile.

2.2. Cálculo de precios por LVA Índices

En LVA índices, para el cálculo de precios que se realiza al final del día, el mercado es modelado como un grafo, en donde cada nodo corresponde a un instrumento de renta fija (con su duración, emisor, riesgo, monto, etc) y las aristas presentan pesos indicando la similitud entre un instrumento y otro. Esta información es almacenada en una base de datos relacional SQL Server. Dada la forma en que es almacenada esta información, en un momento del día se procesan todas las transacciones que han ocurrido en él y, instrumento por instrumento, son calculadas las aristas del grafo para poder aplicar el modelo de asignación de precios.

2.3. Transacciones diarias en la bolsa de comercio

Las transacciones ocurren en la BCS. Para el caso de renta fija chilena existen aproximadamente 6000 instrumentos que transan diariamente y varias transacciones por minuto. Estas transacciones pueden ocurrir en días hábiles entre las 9:00 y 16:00 hrs.

3. Problema a Resolver

Dado que existen instrumentos con mayor liquidez que otros, no es posible conocer el precio de todos los instrumentos con sólo observar las transacciones. Para esto se pueden emplear modelos matemáticos que, dada la transacción de un instrumento, son capaces de propagar un cambio en el valor de los instrumentos que no han transado. Esta propagación se realiza desde el instrumento transado hacia el resto de instrumentos que posean algún nivel de similitud, como lo sería poseer el mismo emisor, una duración cercana o mismo nivel de riesgo.

El problema a resolver consiste en encontrar una solución para procesar en tiempo real un modelo del mercado de renta fija chileno, definiendo un modelo matemático a utilizar, permitiendo la actualización de precios en el momento en que ocurren transacciones, teniendo así precios en línea. Esto aprovechando la utilización de threads y paralelismo.

3.1. Preguntas de Investigación

1. ¿Cuál es el modelo matemático adecuado para realizar los cálculos necesarios?
2. Dado que se busca obtener precios en tiempo real. ¿Cómo se implementa paralelismo en el momento que ocurran transacciones para realizar una propagación más veloz?
3. Para aquellos instrumentos que no poseen transacciones. ¿Es posible que la forma en que se procese el algoritmo de propagación permita guardar la explicación de cómo se llega al precio?
4. Tomando en cuenta que esto utilizaría paralelismo en el grafo. ¿Es necesario sacrificar la calidad del modelo en pos de obtener la información en tiempo real?
5. Si una estrategia permite calcular precios en tiempo real. ¿Existe alguna desventaja al comparar con un modelo procesado una sola vez al día?
6. ¿Existe algún trade-off al buscar el procesamiento en tiempo real? Por ejemplo: ¿Hay algún problema de concurrencia entre threads?

4. Hipótesis

A través de una base de datos que modele un grafo, como Neo4j [5], es posible almacenar los instrumentos del mercado a la vez que todas las relaciones que existen entre ellos.

Como el modelo a aplicar es de propagación, una base de datos de este tipo será lo suficientemente rápida como para obtener valores consistentes. Así mismo, será capaz de guardar una explicación del cálculo de precio al vuelo.

Una solución en tiempo real poseerá una mejor calidad que una calculada al final del día pues, sin ser explicitado en el modelo, las transacciones que ocurren más cerca al cierre del día tendrán una propagación con menos ruido sobre el resto de instrumentos.

Es posible sea necesario sacrificar calidad del modelo dada la posibilidad de concurrencia entre threads, ya que el modelo deberá privarse de asumir que transacciones que ocurran en tiempos cercanos se procesen en el mismo orden que llegaron.

5. Objetivos

5.1. Objetivo General

Proponer, diseñar, implementar y evaluar un sistema que permita aplicar un modelo de propagación sobre un grafo en tiempo real.

5.2. Objetivos Específicos

- Almacenar la información del mercado de forma eficiente para su procesamiento en forma de grafo.
- Crear un modelo matemático de propagación para valores de instrumentos financieros
- Obtener un sistema que permita aplicar un algoritmo de propagación sobre un grafo aprovechando el paralelismo.
- Encontrar métricas útiles que permitan evaluar la calidad del grafo una vez realizada la propagación de una transacción.

6. Metodología

Para lograr los objetivos anteriores, el trabajo de tesis se dividirá en cinco fases:

6.1. Determinación de la forma de almacenamiento

En primer lugar será necesario determinar la forma de almacenamiento que permita aplicar un modelo de propagación sobre el grafo en tiempo real. Esto debe considerar la rapidez con la que se pueden realizar consultas y actualizaciones sobre la información.

Con este paso se procederá a exportar la información de los instrumentos de renta fija chilenos a la forma de almacenamiento.

6.2. Recepción de transacciones

Posteriormente se debe implementar un sistema que reciba y aplique una transacción sobre un instrumento. Este debe ser capaz de recibir del orden de 200 transacciones por minuto.

6.3. Propagación de la transacción en paralelo

El sistema debe realizar una propagación desde el instrumento transado hacia el resto del grafo. Esto con la utilización de threads para tener paralelismo con especial cuidado en evitar problemas de concurrencia y al mismo tiempo aprovechar la velocidad.

Aquí se medirá el tiempo que tarda una sola transacción en ser procesada. Una transacción tarda t en ser propagada por el grafo y se deben tomar en cuenta mecanismos para cuando existen dos transacciones en t_1 y t_2 tal que $t_2 - t_1 < t$.

Se evaluarán opciones como propagación completa en el grafo de forma aleatoria, propagación en prioridad y limitar la cantidad de nodos a los que es propagada la transacción.

6.4. Creación y evaluación del modelo de precios

Se buscará encontrar un modelo de precios que permita ser procesado en tiempo real y evaluar la validez de los precios obtenidos.

Este modelo será confeccionado desde cero pensando en la posibilidad de procesamiento en tiempo real y el trabajo previo (modelo de LVA) se usará sólo como punto de comparación.

Para evaluar y validar el modelo se realizarán al menos los siguientes experimentos:

- Un instrumento que transa, en el instante anterior a ser transado poseerá un precio calculado por el modelo. Se calcula el promedio de la diferencia entre el valor asignado por el modelo y el valor obtenido luego de la transacción. Este valor, para indicar un buen modelo, debiera ser lo más cercano a cero posible. Esto último puede evaluarse de forma independiente entre instrumentos líquidos e instrumentos poco líquidos, permitiendo una mayor diferencia promedio en aquellos que transan poco.
- La compra y venta de instrumentos representan inversiones. Dados dos instrumentos que posean duraciones similares y riesgos distintos, se espera ver que un instrumento de menor riesgo sea más caro que uno con menos probabilidad de pagar. Es decir, para dos bonos a y b :

$$\text{riesgo}(a) > \text{riesgo}(b) \Rightarrow \text{precio}(a) < \text{precio}(b)$$

Se realiza un conteo de los instrumentos que no cumplen esta condición dentro de buckets de duración y se busca minimizar la cantidad encontrada.

Como validación de calidad se compararán las métricas obtenidas en el modelo propuesto y los resultados actuales en LVA, aceptando un modelo en tiempo real que obtenga mejores métricas.

7. Resultados esperados

- Un método para procesar un modelo de propagación sobre un grafo con paralelismo.
- Precios para el mercado de renta fija chileno en tiempo real.

Referencias

- [1] Evangelos Benos y Satchit Sagade. «High-frequency trading behaviour and its impact on market quality: evidence from the UK equity market». En: (dic. de 2012). URL: <https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2012/high-frequency-trading-behaviour-and-its-impact-on-market-quality-evidence-from-the-uk-equity-market.pdf>.
- [2] Jonathan A. Brogaard. «HIGH FREQUENCY TRADING AND ITS IMPACT ON MARKET QUALITY». En: (nov. de 2010). URL: http://efa2011.efa-online.org/fisher.osu.edu/blogs/efa2011/files/MM_1_1.pdf.
- [3] Frank J. Fabozzi. «HIGH-FREQUENCY TRADING: METHODOLOGIES AND MARKET IMPACT». En: (dic. de 2012). URL: https://www.theifm.org/sites/default/files/inline-files/High%20Frequency%20Trading-pub2011_0.pdf.
- [4] Jing Han y col. «Survey on NoSQL database». En: *2011 6th International Conference on Pervasive Computing and Applications*. 2011, págs. 363-366.
- [5] Justin J. Miller. «Graph Database Applications and Concepts with Neo4j». En: (2013). URL: <https://pdfs.semanticscholar.org/322a/6e1f464330751dea2eb6beecac24466322ad.pdf>.