

Análisis de Cointegración IPSA y COLCAP mediante PySpark Batch + Spark Structured Streaming

Ricardo Herrera

December 3, 2025

Objetivo General

Aplicar técnicas avanzadas de análisis financiero utilizando:

- PySpark en modo Batch
- Spark Structured Streaming
- Tests de cointegración (Engle–Granger)
- Análisis de correlación y volatilidad en tiempo real

Datos utilizados

Archivo Excel: **Datos.xlsx**

- Fecha
- IPSA (Chile)
- COLCAP (Colombia)

Transformaciones:

- Precios → Log-precios
- Retornos diarios

Análisis Batch: Qué se realizó

- ① Lectura del archivo Excel
- ② Limpieza y orden temporal
- ③ Cálculo de:
 - Log-precios
 - Retornos
- ④ Test ADF en:
 - Niveles
 - Retornos
- ⑤ Test Engle–Granger para cointegración histórica

Código principal (Batch)

```
stat, p = adfuller(df['LogIPSA  
']) egstat, egp,= coint(df[ 'LogIPSA '], df[ 'LogCOLCAP '])
```

Spark Structured Streaming

Objetivo:

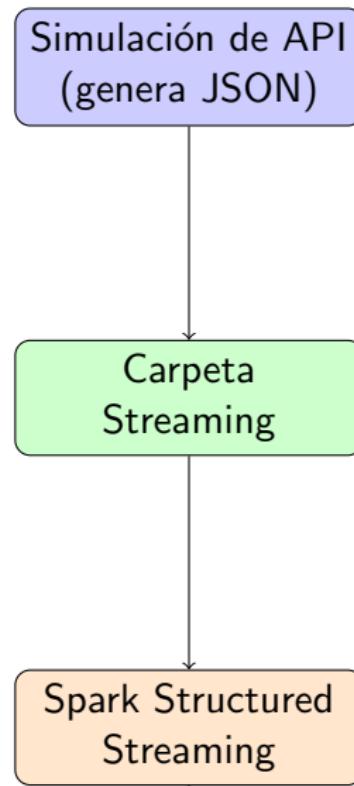
Simular una API financiera

Generar datos nuevos en tiempo real y enviarlos a Spark Streaming.

Características:

- Un archivo JSON nuevo cada 5 segundos
- Spark lo procesa en micro-batches
- Se recalculan:
 - Correlación rolling
 - Volatilidad reciente
 - Cointegración en ventana móvil

Arquitectura general



Código principal (Streaming)

```
query =  
    streaming_df.writeStream .foreachBatch(processMicrobatch) .option("checkpointLocation",  
    '5 seconds') .start()
```

Funciones de análisis en cada micro-batch

En cada llegada de datos Spark calcula:

- Correlación rolling (60 días)
- Volatilidad reciente vs. histórica
- Cointegración en ventana móvil (250 días)

Alertas:

- Correlación baja
- Volatilidad inusualmente alta
- Aparición/desaparición de cointegración

Resultados esperados

- Ver si IPSA y COLCAP están cointegrados
- Observar cómo cambian las relaciones cuando llegan nuevos datos
- Detectar cambios relevantes en:
 - Estructura de dependencia
 - Correlación
 - Cointegración

Conclusiones

- Se implementó un sistema híbrido Batch + Streaming
- La simulación de API permitió generar datos en tiempo real
- Los indicadores se recalcularon dinámicamente
- El enfoque permite monitorizar relaciones entre mercados de forma continua y automatizada

¡Gracias!