REAL PROJECTS

Apache Spark & Kafka - January 2021

EDEM
Escuela de Empresarios

1. Regulatory Banking Project
2. Transactions Notifications
3. Panama Papers
4. Markets Data Lake
5. Risk Reporting

1. Regulatory Banking Project 
2. Transactions Notifications
3. Panama Papers
4. Markets Data Lake
5. Risk Reporting

Regulatory Banking Project

- **Volcker Rule**
  - The **Volcker Rule** is a federal regulation that generally prohibits banks from conducting certain investment activities with their own accounts and limits their dealings with hedge funds and private equity funds, also called covered funds.
  - Very strong restrictions for banking
  - It's necessary to provide multiple reports to prove that the bank is aligned with these restrictions
  - Very high penalties

Regulatory Banking Project

- **Requirements**
  - Ingestion of a high volume of trades and positions in the system
    - CSV, JSON, Text
  - Once the data is in the system. It's necessary to apply certain metrics in order to extract the information necessary to generate regulatory reports
  - Reports are monthly
  - The system doesn't generate reports itself. It's enough only data.

Brainstorming

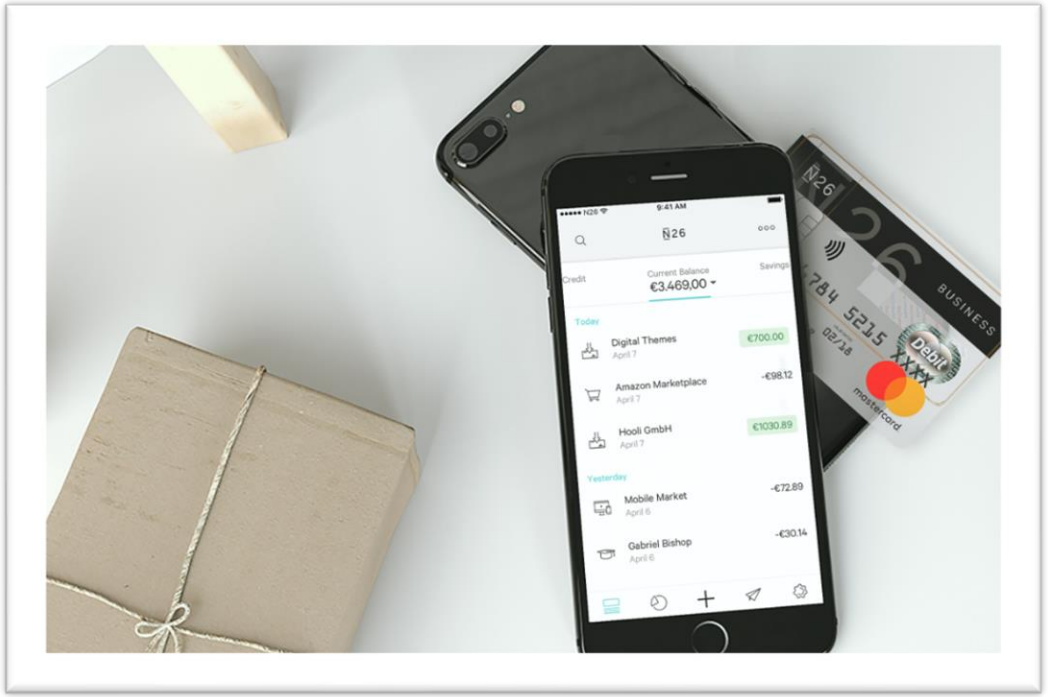
Hands on

- Read trades files (csv and json). Each file is a different position
- Normalize schema
- Calculate PnL % of the position considering open & close prices
- The metric is to calculate the best position for each day, showing the following columns (date, open, close, pnl, position)

1. Regulatory Banking Project
2. Transactions Notifications 
3. Panama Papers
4. Markets Data Lake
5. Risk Reporting

Transactions Notification

- Every transaction in the bank needs to be notified to the customer as soon as possible
- Clients and transactions are in a legacy system (**Oracle and DB2**)
- **High volume** of transaction in a bank
- It's necessary to notify to all customers for every transaction
- **Real Time** scenario



Transactions Notification

- **Requirements**
  - Data is ingested from legacy system (Oracle and DB2) to a new Data Lake
  - Legacy system needs to be synchronized with new Data Lake
  - Notifications can contain marketing data, with different priorities
  - It’s possible to schedule notifications
  - It’s necessary to manage alerts and metrics in the system

Brainstorming

Hands on

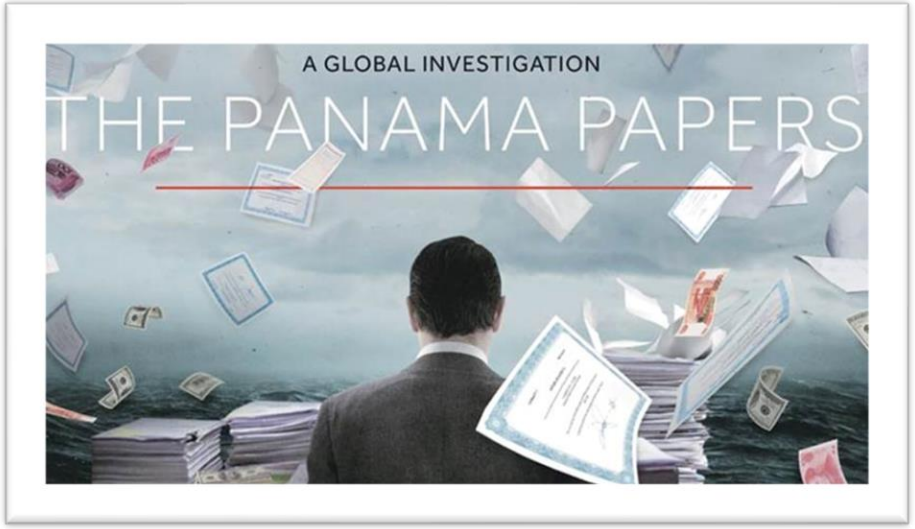
- Read from Kafka transactions topic
- Apply a fixed-window over the stream with 1 minute duration aggregating by "timestamp"

1. Regulatory Banking Project
2. Transactions Notifications
3. Panama Papers

4. Markets Data Lake
5. Risk Reporting

Panama Papers Project

- **Panama Papers**
  - 2.6TB of data (11.5 million documents) leaked that detail financial and attorney-client information for more than 214.488 offshore entities
  - The documents contain personal financial information about wealthy individuals and public officials
  - Companies created in Tax haven with low taxes.

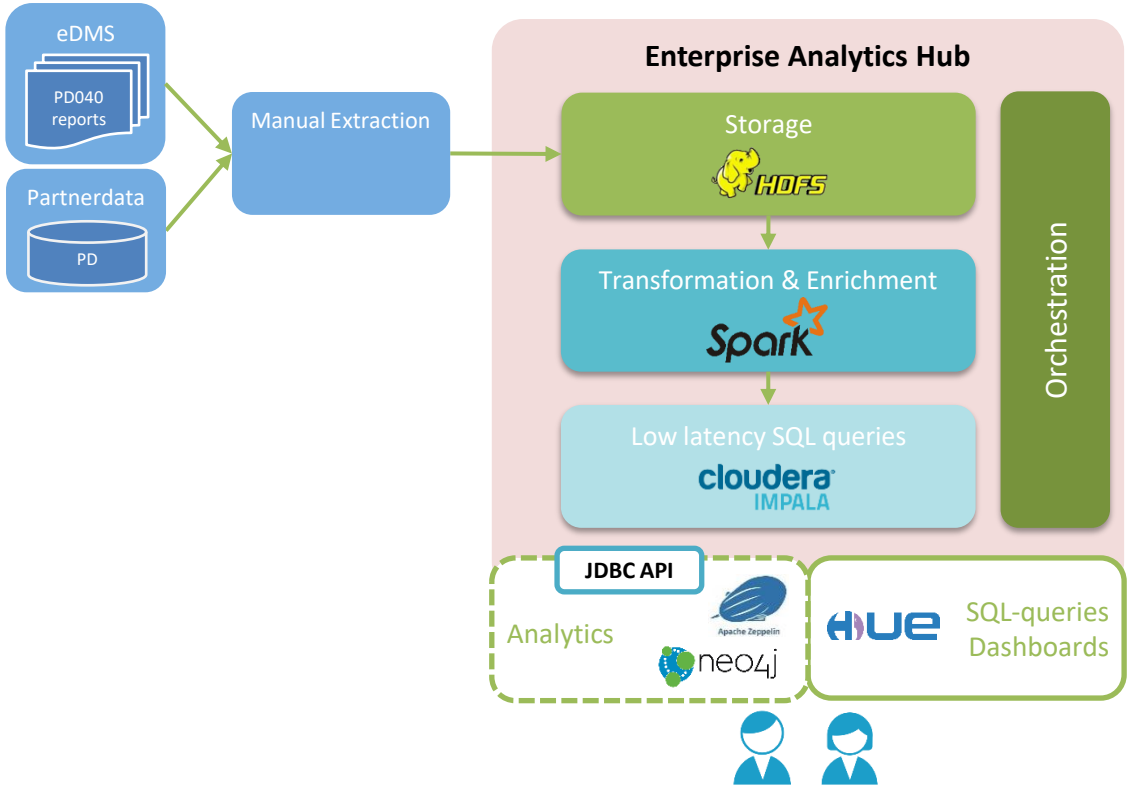


Panama Papers

- **Requirements**
  - Our client wants to know if a customer has been involved in Panama Papers
  - It's necessary to **extract from a Relational database** data form Panama papers into a Data Lake
  - Transformations are necessary to prepare data to be queried
  - It's necessary to **trace** a specific customer or organization in the database

Brainstorming

Architecture proposal



Hands on

- Trace “Spring Song International Co., Ltd” entity with Spark SQL using the Panama Papers dataset

1. Regulatory Banking Project
2. Transactions Notifications
3. Panama Papers
4. Markets Data Lake 
5. Risk Reporting

Market Data Project

- **Market data** is price and trade-related data for a financial instrument

- Delivery of price data from exchanges to users, such as traders is highly time-sensitive and involves **specialized technologies** designed to handle collection and throughput of massive data streams

Market Data Project

- **Requirements**
  - **Different sources** (SFTP, JMS, files, …), showing stock market in **real time**
  - **Batch processing will** be necessary to prepare data for different reports, apart from recalculations and other stuff
  - Different users will **consume data through API, reports and ad-hoc analysis**
  - It's necessary a platform to execute different algorithms (not necessarily real time here)
  - **Monitoring**, **logging**, **lineage** will be necessary
  - The solution needs to be **scalable**

Brainstorming

1. Regulatory Banking Project
2. Transactions Notifications
3. Panama Papers
4. Markets Data Lake
5. Risk Reporting

Risk Reporting Project

- **Interest rate risk in the banking book (IRRBB)** refers to the current or prospective risk to the bank's capital and earnings arising from adverse moments in interest rates that affect the bank's banking book positions
- Specific reports are necessary for regulators

Scenario	ΔEVE	ΔNII
Parallel up	✓	✓
Parallel down	✓	✓
Steepener	✓	✗
Flattener	✓	✗
Short rate up	✓	✗
Short rate down	✓	✗

Table 3. Impact in EV and net interest income (NII) both of stability caps/ pass-through floor and a shorter duration for NMDs.

EV volatility of NMDs.

	Parallel Up	Parallel Dn	Steepener	Flattener	Short Rate Up	Short Rate Dn
Internal Model	134	-102	47	-48	100	-79
BCBS Proposal	56	-21	4	-2	47	-21

EV volatility of fixed rate loans (*).

	Parallel Up	Parallel Dn	Steepener	Flattener	Short Rate Up	Short Rate Dn
Internal Model	-134	102	-47	48	-100	79
BCBS Proposal	-134	102	-47	48	-100	79

(*) Trying to make it simple, we assume that the bank prepayment treatment = prepayment treatment of the CP.

Global EV volatility of the bank.

	Parallel Up	Parallel Dn	Steepener	Flattener	Short Rate Up	Short Rate Dn
Internal Model	0	0	0	0	0	0
BCBS Proposal	-78	80	-43	45	-52	58

Risk Reporting Project

- **Requirements**
  - More than 130 feeds imported to Data **Lake through Oracle Database**
  - **Build a ETL** with data ingested to ingest, enrich and prepare data to be consumed
  - **Official reports** need to be in Excel and PDF format
  - **Portal web** will be necessary to manage the system and generate reports
  - **Monitoring and log analysis** is a must.

Brainstorming