Safe Harbor Statement

The preceding is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle’s products remains at the sole discretion of Oracle.
Tax Fraud Detection Platform

Building Tax Fraud detection platform using Big Data Spatial and Graph Technologies

Wojtek Wcislo

Oracle EMEA Consulting
wojciech.wcislo@oracle.com

Redwood Shores, February 2nd, 2017
About us

- Oracle Consulting, EMEA based FMW competence center
- Over 13 years of successful history
- Focusing on BigData, FastData, IoT, S&G, Integration and Cloud
- Hundreds of customers
- Delivered projects on 5 continents
Business context

- Growing problem of tax frauds, tax evasions and tax avoidance
- Effectiveness of existing approach
- End-to-end VAT chain visibility
- Huge amounts of potential data
- Over 2 million end "customers"
- Challenging legislation
- Need for fast, highly scalable solution
- Shortening time to action

Business context

• Growing problem of tax frauds, tax evasions and tax avoidance
• Effectivness of existing approach
• End-to-end VAT chain visibility
• Huge amounts of potential data
• Over 2 million end „customers”
• Challenging legislation
• Need for fast, highly scalable solution
• Shortening time to action

Nature of SAF-T (Single Audit File for Tax)

• Official, international standard defined by Organisation for Economic Co-operation and Development (OECD)
• Used for electronic exchange of accounting data from organizations to a national tax authority or external auditors.
• Increasingly adopted within European countries
• XML format based, current rev. 2.0 with.xsds
• Up to 200Gb of size (eg. utilities with millions of customers)

One of „typical” VAT fraud scenarios

1. Company in other EU member country
2. Importer „fake” company
3. Buffer Reseller
4. Exporter company
5. Company in other EU member country
6. Tax authority
7. Tax authority
8. Tax authority

- Not paying VAT
- Paying VAT
- Getting VAT refund
- ?
One of „typical” VAT fraud scenarios

Company in other EU member country → Importer „fake” company → Buffer Reseller → Exporter company → Company in other EU member country

Not paying VAT → Paying VAT → Getting VAT refund

Tax authority → Tax authority → Tax authority

0% VAT BORDER
One of „typical” VAT fraud scenarios
What’s important?

- Time
- VAT
- Profile
- Contractors
- Stocks
- Resources
- Turnover
One of „typical” VAT fraud scenarios
Steps to detect middle man

• Identification of importers (including missing ones)
  
  $\text{SELECT * WHERE (n0:Company)-[e1:Sell]->(n2:Invoice with Net_Amount_NO_VAT > 0, !(Tax_ID_Buyer=~'CC*'))}$

• Identification of exporters
  
  $\text{SELECT * WHERE (n0:Company)<-[e1:Sell]-(n2:Invoice with Net_Amount_NO_VAT > 0, !(Tax_ID_Seller=~'CC*'))}$

• Simplification of graph
  – Company -> Invoice -> Company to Company -> Company
  – Aggregation of VAT invoice items to edge

• Running FattestPath algorithm
  
  $\text{analyst.fattestPath(PgxGraph graph, ID rootId, EdgeProperty<java.lang.Double> capacity)}$

• Selecting common Middle Man nodes
Architecture - Production

Large enterprises
SMB organizations
Micro organizations

Cloud to HDFS agent

3rd party cloud proxy

Decryption
Decompression
Validation
Persistence

Hadoop cluster

Graph Database (NoSQL)

Analytics (Hive+BDSQL)

Analytics (BD S&G)

Reporting Dashboards

REST services (WebLogic)
Metadata database (Database 12c)

In production
In testing
Architecture – Test based on Oracle Cloud

Hadoop cluster

3rd party cloud proxy

Large enterprises

SMB organizations

Micro organizations

Oracle Cloud

Oracle Storage Cloud

Oracle Database Cloud

Oracle Java Cloud

Cloud to HDFS agent

REST services (WebLogic)

Metadata database (Database 12c)

Reporting Dashboards

Decrypt → Decompress → Validate → Persist

Oracle and/or its affiliates. All rights reserved.
Graph Analytics

- Hadoop cluster
- Graph Database (NoSQL)
- Prefiltering
- PGX Cluster
- In-Memory analytics
  (Big Data Spatial & Graph)
- PGQL
- Java API
- REST services
  (WebLogic)
- Graph Dashboards

Other systems

Copyright © 2014 Oracle and/or its affiliates. All rights reserved.
How it looks?

Property Graph Visualization

PGQL Query Builder
Conclusions

• „Industry” knowledge
• Volumes and size of the files (sequencing, in-memory processing)
• Near real time confirmation of positive validation
• Very short peak window, like 2-3 days each month
• Accessing complex data structures stored in Avro
• Security and data privacy contraints
• PGX with Property Graph QL + build in alghoritms + API
• Collaboration essential for effective detection of new patterns
• User friedly data visualization