

Método de prevención de fraude con bases de datos no-relacionales

Abstracto

Se ejemplificará un caso de uso de bases de datos no relacionales para detectar fraude de falsificación de identidad, haciendo uso de las ventajas (en tiempo de ejecución) que tienen las bases de este tipo para encontrar relaciones entre datos.

El tipo de fraude al que me referiré es aquel en el que una persona falsifica una o varias identidades para solicitar créditos a la institución sin tener la obligación de pagarlos, y dejando a la institución con dificultades para recuperar la deuda otorgada.

El arte de la falsificación

La falsificación de identidad puede generar grandes pérdidas a las instituciones bancarias. Una vez hecho el fraude, no habrá ningún individuo para hacer el cobro y el banco perderá tiempo buscando un fantasma. Pero se necesita un poco de organización.

Un robo de identidad puede ser un buen punto de partida. A veces todo lo que se necesita es un documento real como un INE, estados de cuenta. Equipado con esto, un defraudador puede falsificar una primera identidad, o bien, puede inventar un nombre, edad, fecha de nacimiento, dirección, número de teléfono, etc. El objetivo es crear una identidad (falsa) que tenga todas las características de una persona real. Los estafadores se asegurarán de que su identidad pueda pasar por alto los controles de seguridad de sus objetivos. Aquí es donde tener una pieza "real", verificable, de identidad (como los estados de cuenta de una persona) es útil.

Implementación

Primero generé una base de datos (relacional) con registros generados aleatoriamente (nombre, número de teléfono, número de tarjeta de crédito, dirección de correo electrónico, dirección iP), y después establecí las restricciones y relaciones para migrar esta base de datos a un esquema no-relacional.

	name	birth_date	email	cc	phone	ip
1	Joe Collier	1908-11-29	ting@ivu.wve	9751-3505-4992-9419	5741646512	209.3.56.91
2	Joaquin Sporer	1948-09-09	rtkvsznmaj@whagubd.row	9672-1203-3295-6360	8214766973	246.166.168.154
3	Oren Miller	1940-06-10	pdfn@whgbvl.nhg	2060-4282-7553-1885	5641388594	235.78.105.33
4	Brittanie Turcotte	1928-07-03	ujmze@wqzy.nex	4558-1150-6072-5013	5931422195	125.51.135.137
5	Krysten Sanford	1964-02-02	zelubavv@bgacmlstd.tsw	1688-3771-6791-6336	7195913678	185.236.158.1
6	Elvira Turner	1991-07-31	tk@tmoukcha.eth	8611-7963-3961-6011	7643166473	190.113.99.64
7	Jordon Nitzsche	1920-06-14	v@myv.glz	6028-4565-7400-4477	4752912496	241.92.99.73
8	Earleen Beer	1979-09-06	faju@ix.cml	4678-5236-7750-7739	6271986871	145.59.78.120
9	Betty Mante	1948-10-01	eraocnbt@vxjdhpfr.bu	6857-8167-8541-8492	7168126321	6.223.110.227
10	Bettye Smith	1903-02-03	fwlubbaxe@hgvn.pgd	9807-4230-2409-5601	7385263259	88.98.136.58
11	Louie Lindgren	1927-05-09	ohzadlqu@wiyelsamc.nju	1118-5833-6625-6106	7828242167	2.130.27.33
12	Stefany Purdy	1915-08-26	r@bcqsmshod.hwu	5351-6176-4001-8373	2836711423	238.65.206.150
13	Ellamae Klein	1902-02-21	dvs@vliba	7295-4742-4722-9675	3961624876	200.96.148.207
14	Rosendo Hickie	1944-12-04	z@vwin@ceqk.jqt	9079-9717-9801-9393	1572819231	2.106.233.139
15	Fernando Hilll	1919-07-30	zawhxptm.upi	7531-1386-3844-1046	5989439816	179.53.6.146
16	Yong Hansen	1945-05-22	zashwl@agywn.gzs	9836-5644-9756-1480	8625876713	245.53.150.186
17	Anabel Kessler	1997-02-10	awjkw@ymn.cpv	8343-4665-9843-4447	2346937358	19.141.138.27
18	Darrell Rowe	1914-09-21	wnfogdcxag@v.sux	2841-7402-4075-5563	9357856148	128.49.213.132
19	Maude Bartoletti	1997-12-11	ubktsf@cs.gxp	7951-9321-6719-4688	5863591346	159.30.84.100
20	Byron Stanton	2000-07-02	uidpg@dckeefng.rcd	8196-1396-3290-1749	6327831458	143.132.244.203

Showing 1 to 20 of 1,000 entries



Análisis

Por un lado puede obtenerse el máximo subgrafo (de información de registro) conexo a partir de un nuevo usuario y verificar las relaciones que pueda tener con otros usuarios, por otro lado puede hacerse una análisis más general del número y tamaño de anillos que forman usuarios que comparten información.

Rendimiento

La ventaja principal de la base de datos no-relacional para este caso es que las consultas o *'queries'* mencionados anteriormente emplean un tiempo de ejecución mínimo y que además son tareas que pueden llegar a ser imposibles de ejecutar en un esquema relacional.



Usuario con información de registro única.



Usuarios que comparten únicamente dirección IP.



Anillo de usuarios que comparten número de teléfono y dirección de correo electrónico.

Las consultas efectuadas fueron:

- El máximo subgrafo (de información de registro) conexo a partir de un nuevo usuario:

```
MATCH (root {name: "Desire Green" })-[rels*]-(b)
UNWIND rels AS rel
RETURN root,
  collect({start: startNode(rel),
    type: type(rel),
    end: endNode(rel)}) as component
```

- El número y tamaño de anillos que forman usuarios con información de registro duplicada (con tiempo de ejecución de 0.6 segundos aproximadamente):

```
MATCH (person:Person)-[]->(contactInformation)
WITH contactInformation,
count(person) AS RingSize
MATCH (contactInformation)-[]-(person)
WITH collect(person.name) AS Persons,
contactInformation, RingSize
WHERE RingSize > 1
RETURN Persons AS FraudRing,
labels(contactInformation) AS ContactType,
RingSize
ORDER BY RingSize DESC
```

FraudRing	ContactType	RingSize
[Jimmie Koss, Junita Lesch, Meghann Schaefer, Harley Fay, Ted Hermiston, Elbert Paucek]	[IP]	6
[Mertie Mitchell, Werner Marquardt, Guillermo Hegmann, Lane Rippin, Lillie Mraz, Sabrina Abshire]	[IP]	6
[Necole Dare, Marion Gutmann, Deedra Franecki, Dotty Hyatt, Bryon Bruen, Lorina Berge]	[IP]	6
[Markus Glover, Edward Quigley, Lindsay Konopelski, Gigi Bailey, Fausto Huels, Leonarda Haag]	[IP]	6

Bibliografía:

[1] CRAN, 2016, "Package 'RNeo4j'", vía internet:

<https://cran.r-project.org/web/packages/RNeo4j/RNeo4j.pdf>

[2] Linkurious, 2016, "How to detect bank loan fraud with graphs : part 2", vía Internet:

<https://linkurio.us/how-to-detect-bank-loan-fraud-with-graphs-part-2/>

[3] Nicole White, 2016, "Demo of RNeo4j part 1", vía internet:

<https://nicolewhite.github.io/2014/05/30/demo-of-rneo4j-part1.html>