reifier.

Fuzzy Matching With Spark

About Us

ALICE: This is impossible!

THE MAD HATTER: Only if you believe it is.

reifier.

The problem

According to Gartner, businesses are losing upto 25% potential revenue due to lack of holistic multichannel view of data.
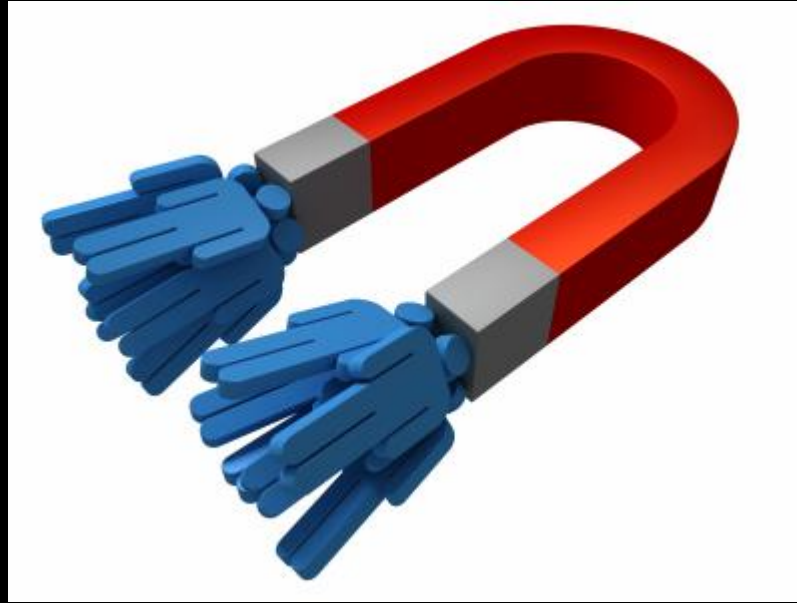
reifier.

The problem

© Nube Technologies

Challenges

- Quadratic nature of the problem
- No standard notion of similarity
- Omissions, typos and other issues

reifier.

Use case - Cross and Upselling

reifier.

Lead Generation

reifier.

BFSI

Personal Credit Ratings

Fraud detection

reifier.

Other Use Cases

Yellow Pages

Catalog and Inventory Management

reifier.

Wishlist

Works with any kind of data

Scalable

No manual configuration of rules or algorithms

reifier.

Spark Advantages

- Distributed
- Scalable
- In memory
- Machine Learning
- Sampling
- No need to orchestrate multiple jobs

reifier.

Reifier - Label

Are these duplicates?(Y/N)

hannagan	mikhayla	rupp lace	brentwoo dvlge	penshurst	2283	vic
mikhvyla	hannagan	windradyen street	brentwood vlge	penshurst	2257	vic
mikhayla	hannaan		brentwoodvlge	penshurst	2257	vic

reifier.

Reifier Output

A	B	C	D	E	F	G	H
1	rees	cooper	trott place	safari l odge	mayfield	6061	vic
1	cooper	rees	trott place	safari lodge	mayfield	6061	vic
2	oscat	kuhndt	kirkcaldie circuit	the stables		3163	wa
2	oscar	kuhndt	kirkcadic circuit	the stables		3163	wa

reifier.

Thank You !

reifier.