

Introduction

IT landscapes have developed into a highly complex combination of various ERP systems, databases etc and very often with fragmented and duplicate data. This creates serious data analysis problems that could lead to poor decision making and costing companies millions each year. Many benefits can be achieved with “clean” data as can be seen in the slides that follow.

Some Benefits Of “Clean” Data

Cost Savings

- Repetitive "Free Text" spend can be identified and such items can be added to your stores catalogue with contract prices (even at zero stock levels). This leads to a reduction of off-contract spend as end users are aware of the existence of such items in the stores catalogue at normally far better prices.
- Reduced cost of placing orders and holding of inventory.
- Reduction of costly incorrect decision making.

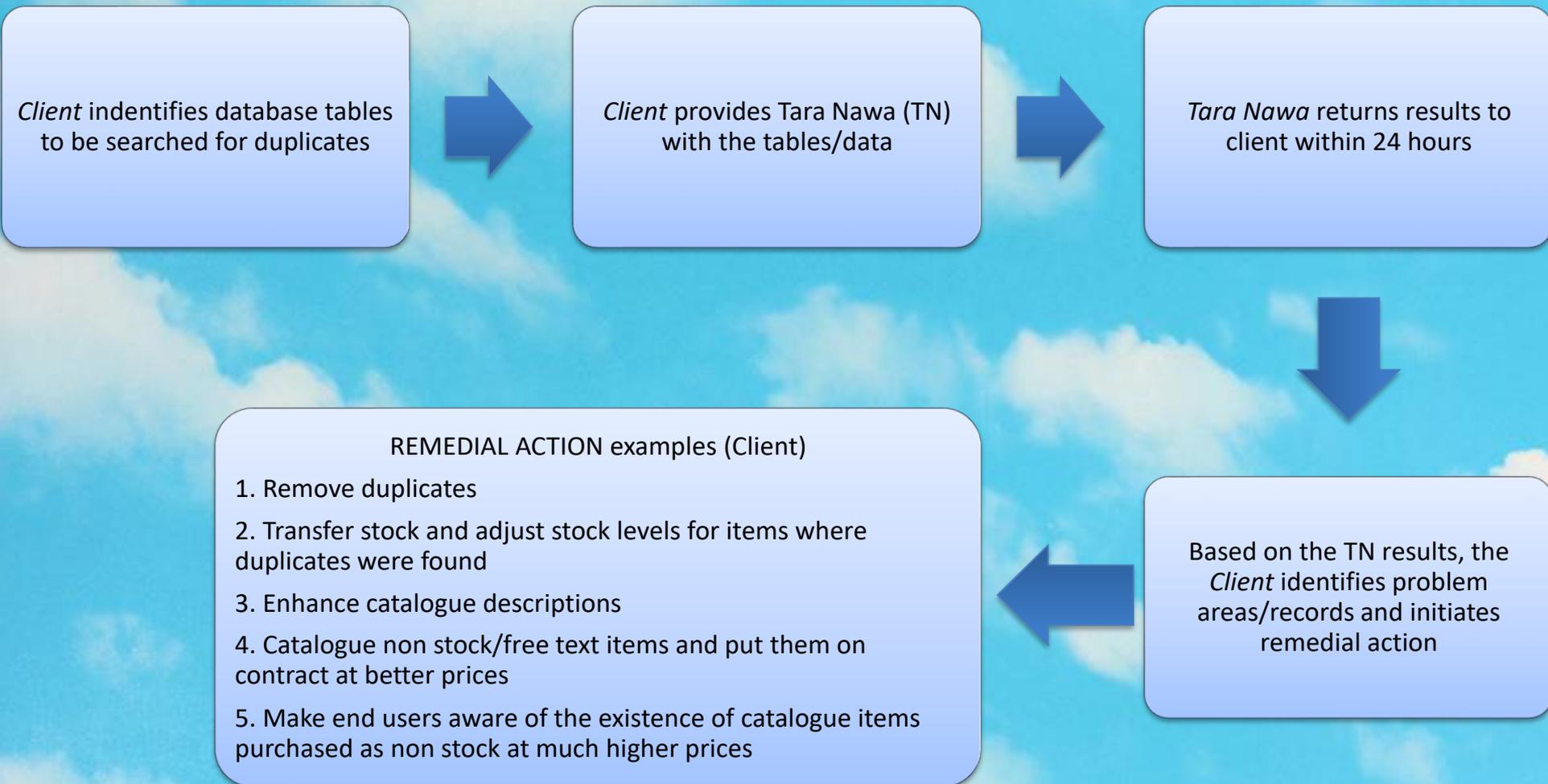
Efficiency Improvement

- Reduction of purchasing transactions
- Fewer items to manage in the warehouse
- Elimination of duplicate stockholding
- Optimization of stockholding

Data Quality Improvement

- Standardization and enhancement of catalogue descriptions
- Accurate/Reliable reporting and Business Intelligence

Typical Duplicate Finder Project





Examples Of Data To Be Searched For Duplicates



Repetitive Non Stock or “Free Text” purchases



**Non Stock (Free Text) Possible
Stock/Catalogue Items**



Stock/Catalogue Duplicates



Vendor Duplicates



Asset Register



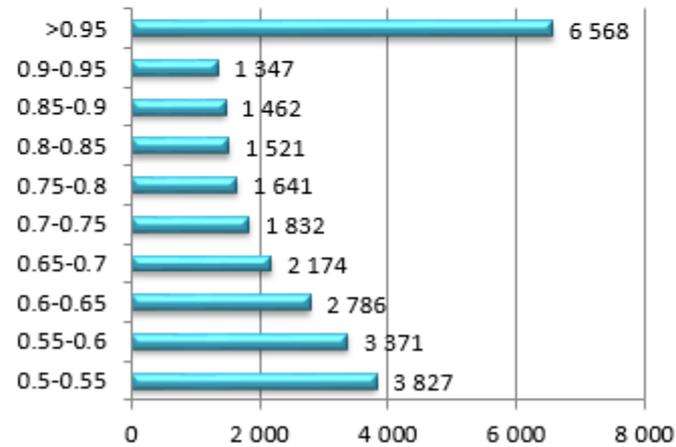
Summary Of Duplicate Finder Project

The client is provided with details and a summary of the project result as per example below, including the number of record combination comparisons made.

Task	Table Name	Table Description	Field	Records	Record Comparisons
Search for duplicate stock items	Not applicable	SQL - StockItemDetails	Description	37 436	700 708 330
Totals				37 436	700 708 330

Stock Lines	
Similarity Score	Total
0.5-0.55	3 827
0.55-0.6	3 371
0.6-0.65	2 786
0.65-0.7	2 174
0.7-0.75	1 832
0.75-0.8	1 641
0.8-0.85	1 521
0.85-0.9	1 462
0.9-0.95	1 347
>0.95	6 568
Grand Total	26 529

Similarity Score and Stock Lines



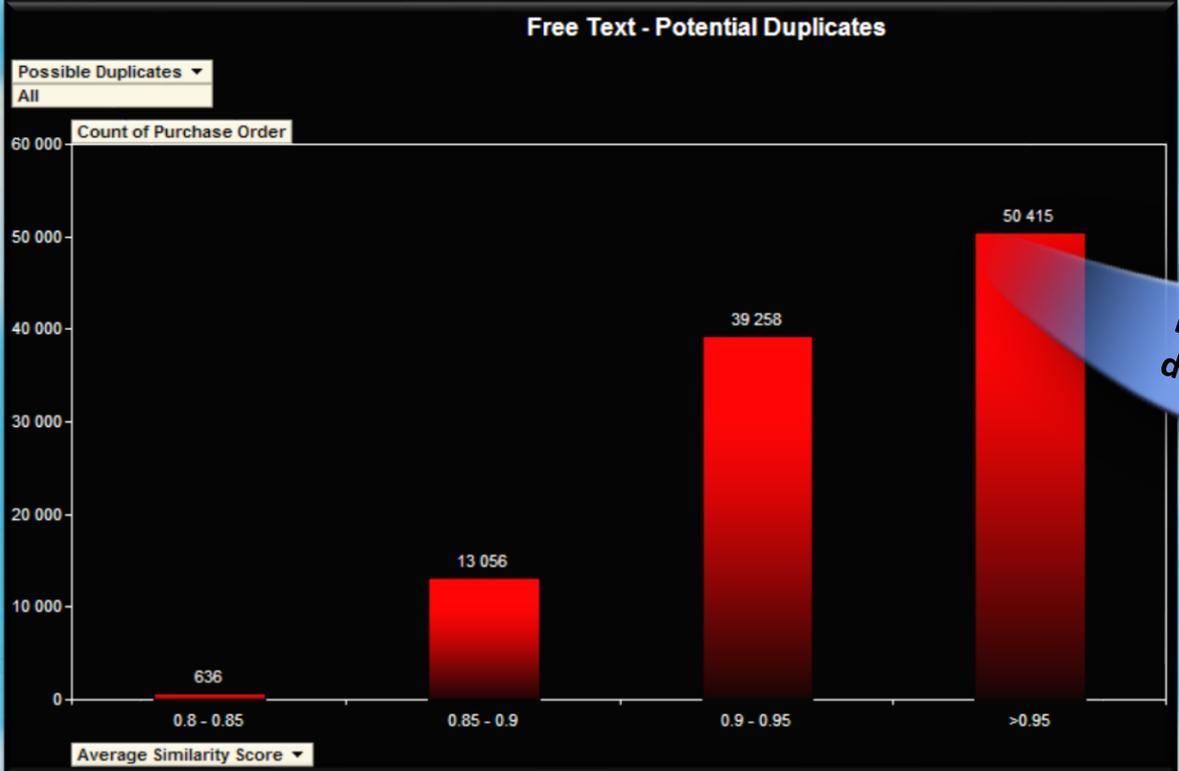
Non Stock (Free Text) Duplicates

The results will indicate the number of possible duplicates found together with a “Similarity Score”. The client may for example want to inspect only items with a “Similarity Score” of more than 85% and can filter for such items.

Non Stock Duplicates						
Duplicates	Record Number	Similarity Score	Description	PO Number	PO Line	Order Price
2	489	1	Semiskilled able to carry out service	4500114181	00020	37800
5	498	1	GWS 8-115 Bosch Angle grinder	4500139263	00010	570
5	498	1	GWS 8-115 Bosch Angle grinder	4500112599	00010	545
5	498	0.8541433	GWS 8 - 1152 Bosch Grinder	4500129348	00010	657
5	498	0.8541433	GWS 8-1152 BOSCH GRINDER	4500127104	00010	625
5	498	0.8541433	Gws 8-1152 Bosch Grinder	4500129480	00010	657
3	502	1	Pin Punches	4500135098	00120	117
3	502	1	Pin Punches	4500105798	00120	99
3	502	1	Pin punches	4500112600	00020	176
4	503	1	200mm Bacho shifting	4500112600	00030	144.15
4	503	0.975	Bacho Shifting 200mm	4500135120	00010	130
4	503	0.8494484	200mm bacho shifter	4500117107	00120	109
4	503	0.7908867	100mm bacho shifter	4500117107	00110	80.99

Duplicate Finder Results Based On Non Stock Or "Free Text" Purchases

The graph indicates that there were 50 415 order lines for which there were other orders with similar descriptions, **with an average similarity score of more than 95%**. The "Drill-down" reflects the purchase order number and the number of possible duplicates – thus for purchase order 3101400397, there were 969 purchase orders with similar descriptions.



Average Similarity Score	Purchase Order	Possible Duplicates
0.8 - 0.85		636
0.85 - 0.9		13056
0.9 - 0.95		39258
>0.95	3101400397	969
	3101347793	950
	3101358121	913
	3101277855	839
	3101364801	644
	4501270752	634
	500159397	478
	3101272801	474
	3101382379	450
	3101408131	431
	4500316920	419
	3300075143	409
	3101384475	400
	4500317716	381
	4500284412	366
	3101314338	347
	3200128854	343
	50415	
Grand Total		103365



Non Stock (Free Text) Possible Stock/Catalogue Items

Non Stock Possible Stock

PO Number	PO Line	PO Description	Material Description	Material Number	Similarity Score	Confidence Level	Order Price	Moving Ave Price
4500110423	00010	CLUTCH:3/4 TON LEVER HOIST	CLUTCH:3/4 TON LEVER HOIST	000000000991006738	100.000%	100.000%	45.00	
4500110467	00010	stainless steel hose	HOSE:STAINLESS STEEL CONVUTULED 1.5	000000000991015524	71.045%	35.812%	425.00	
4500110660	00010	3P2274	BOLT:3P2274	000000000991022667	87.487%	51.583%	15.79	
4500110728	00010	150mm Shifter Gedore	SHIFTING:155MM,MNFR: GEDORE	000000000991015377	70.189%	36.910%	72.82	1.00
4500110728	00020	250mm Shifter Gedore	SHIFTING:250MM, MNFR: GEDORE	000000000991015382	75.176%	30.092%	94.55	92.80
4500110728	00020	250mm Shifter Gedore	SHIFTING:450MM,MNFR: GEDORE	000000000991015380	70.080%	23.804%	94.55	213.00
4500110728	00140	scissors	SCISSORS:SURGICAL	000000000991012221	73.652%	25.132%	38.00	
4500110728	00140	scissors	SCISSORS:INDUSTRIAL	000000000991012224	79.136%	31.964%	38.00	32.00
4500110749	00010	Bearing 33110	BEARING:P/N:33110 Mnfr:SKF	000000000991004803	80.436%	44.551%	896.56	
4500110789	00010	220v Electrical engraver	ETCHER:ELECTRICAL ENGRAVER,220V	000000000991007504	82.089%	44.291%	125.90	180.00
4500110789	00010	220v Electrical engraver	ETCHER:ELECTRICAL ENGRAVER,220V	000000000991007504	82.089%	44.291%	125.90	215.00
4500110942	00010	FILE STATIONARY:OPTIPLAN FILES REF 430	FILE STATIONARY:OPTIPLAN FILES REF 430	000000000991007591	100.000%	100.000%	4.49	
4500111008	00010	Hicast super 2% wire	WIRE:HI CAST SUPER 2% S/S FIBRE	000000000991014462	70.166%	36.715%	4.05	
4500111050	00010	10mm wire rope	ROPE,WIRE:10MM,GALV STEEL	000000000991012162	78.898%	39.968%	8.50	510.00
4500111081	00030	slugger bit 26mm	BIT:SLUGGER ; SHORT SHANK 26MM	000000000991004963	77.486%	27.153%	172.00	216.00

Stock/Catalogue Duplicates



Stock Duplicates

Duplicates	Record Number	Similarity Score	Material Number	Material Description	Moving Ave Price	Price Unit
4	1132	1	00000000456000400	STRIP,WEAR:BOTTOM RETURN RUNNER	172.03	1
4	1132	0.7890463	00000000456000662	RUNNER:RETURN BOTTOM		
4	1132	0.7245893	00000000456000654	STRIP,WEAR:GUIDE TOP RETURN RUNNER	118.69	1
4	1132	0.7245893	00000000456000303	STRIP,WEAR:GUIDE TOP RETURN RUNNER	105.5	1
3	1140	1	00000000456000565	GEARBOX:1036:1,REDLER CONVEY CHAIN	16095	1
3	1140	0.836661	00000000456000418	GEARBOX:REDLER CONVEY CHAIN	3743	1
3	1140	0.7274572	00000000456000264	GEARBOX:1113:1,REDLER CONVEY CHAIN	21126	1
3	1150	1	00000000456000743	ASSY:COLLECTOR PLT HAMMER	438.6	1
3	1150	0.7084796	00000000410901430	COLLECTOR:ASSY	3180	1
3	1150	0.7084796	00000000410901448	COLLECTOR:ASSY	1878.5	1
2	1151	1	00000000456000751	ASSY:HAMMER,DISCHARGE ELECTRODE	92.75	1
2	1151	0.7599669	00000000200710421	ELECTRODE:DISCHARGE	5421	1
2	1153	1	00000000456000785	INSULATOR:RAPPER,DIA 220 X LG 405MM	3750	1
2	1153	0.7003184	00000000162701011	INSULATOR:RAPPER	1305.1	1
2	1157	1	00000000456003014	SPRING:SHOCK ABSORBER	984.21	1
2	1157	0.7239414	000000000635010304	ABSORBER:	24.27	1
2	1158	1	00000000456003015	ASSY:DRIVE BEARING	3926.15	1
2	1158	0.8278202	00000000461920013	ASSY:DRIVE		
2	1165	1	00000000456003022	COUPL:FLUID DRIVE,IRREGULAR	11100	1
2	1165	0.8651574	000000000991014884	COUPL:FLUID DRIVE	24098	1
3	1186	1	00000000460300018	MODULE:CARD,BATTERY CHARGER	3628	1
3	1186	0.7735772	000000000991013614	BATTERY:CHARGER	48.9	1
3	1186	0.7535772	000000000635010123	CHARGER:BATTERY	9022.2	1



Vendor Duplicates

Vendor Duplicates

Duplicates	Record Number	Weighted Similarity Score	LIFNR	NAME1	Similarity Score (NAME1)
3	1845	100.00%	0000022002	TEMOC ENGINEERING cc	100.00%
3	1845	100.00%	0000080026	TEMOC ENGINEERING CC	100.00%
3	1845	95.59%	0000021842	TEMOC ENGINEERING	91.54%
2	1851	100.00%	0000022008	SIYAHAMBA INDUSTRIAL SUPPLIES (PTY	100.00%
2	1851	80.03%	0000080080	SIYAHAMBA INDUSTRIAL	82.73%
2	1855	100.00%	0000022013	MBONJENI TRADING ENTERPRISE cc	100.00%
2	1855	97.08%	0000080192	MBONJENI TRADING ENTERPRISE	94.45%
2	1865	100.00%	0000022033	PRESTIGE LABORATORY SUPPLIES cc	100.00%
2	1865	94.88%	0000022034	PRESTIGE LABORATORY SUPPLIES cc	100.00%
2	1882	100.00%	0000022062	MTS BELTING AND ACCESSORIES cc	100.00%
2	1882	97.18%	0000080193	MTS BELTING AND ACCESSORIES	95.87%

Contact Details

Contact Person: Elmo Erasmus
Mobile Phone: +264 (0) 81 128 6388
Telephone: +264 64 402 403
Email: elmo@taranawa.com

