

DeX

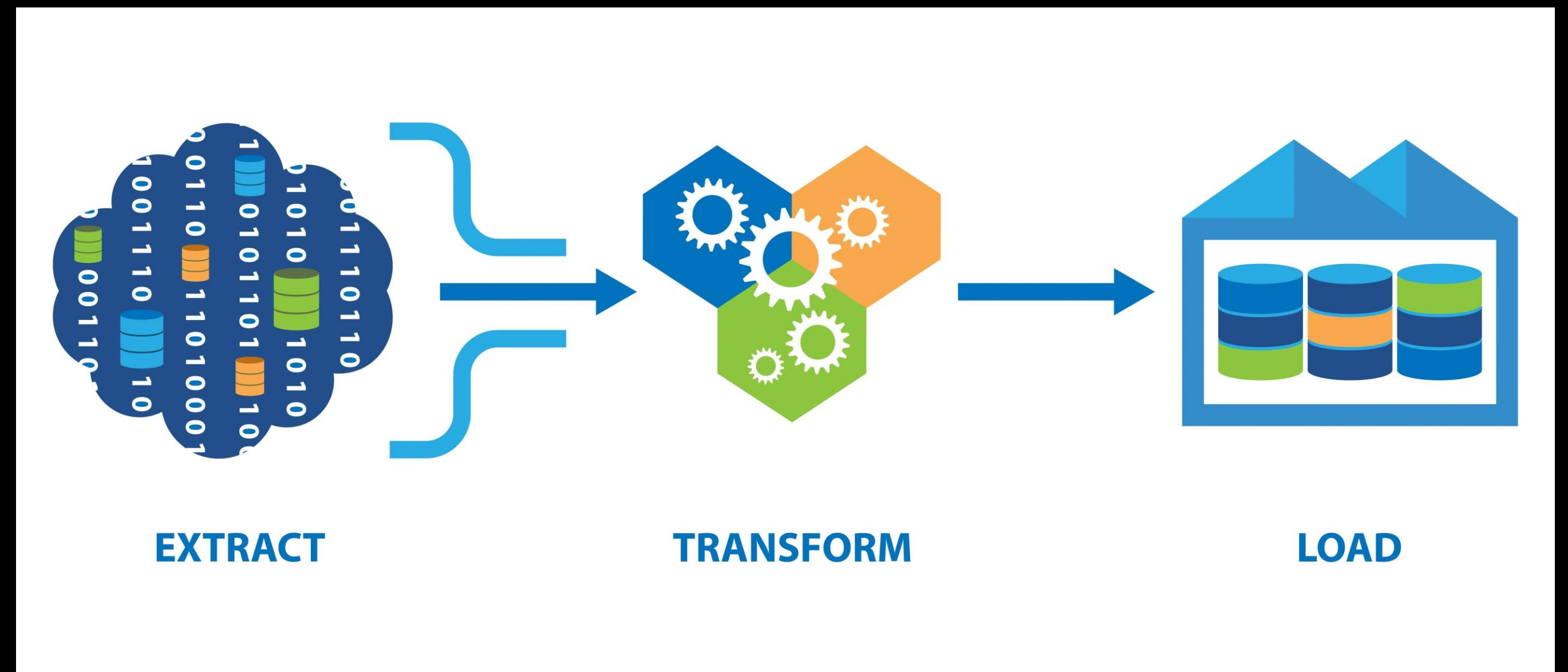
Automated data extraction from financial reports

jovan@mqsoft.rs Technobank 2023

Data Extraction

definition of data extraction

- process of retrieving data out of data sources for further data processing
- usual data workflow: extract -> transform -> load #ETL



PDF as data source

problem with tabular data representation in PDF

- **P**ortable **D**ocument **F**ormat
- designed as an output format with objective to make it **look** the same at any device
- defines instructions to place characters at precise (x, y) coordinates
- has no internal representation of a table structure
- can be text based or image (aka. scanned) based

Tools

how to extract data from PDF

- copy (from PDF) and paste (to Excel)
- PDF (to Excel) converters (for example: Adobe Acrobat)
- Direct import to Excel
- programmatic PDF table extraction tools (for example: Tabula, Camelot)

DeX

automated data extraction from tables in PDFs

- define columns manually
- do it only once per client report type
- define accounting codes and amounts columns
- process data in the background
- download populated excel template with extracted data

ART VISTA DOO
MRS 2020

Bruto bilans datum od: 01.01.2020. datum do: 31.08.2020. broj cifara: 3 klasa:

broj	opis	početno stanje		tekući promet		ukupan promet		saldo
		duguje	potražuje	duguje	potražuje	duguje	potražuje	
011	Kasneje, čekovi, bonovi, rotni i uložni marki	31.551.667,85				31.551.667,85		31.551.667,85
019	ISPRAVKA VREDNOSTI NEKRETNOSTI I IMOVINE		132.227,85				132.227,85	-132.227,85
022	Gradivinski objekti	38.963.912,05				38.963.912,05		38.963.912,05
023	Postrojenja i oprema	186.383.446,37		4.981.028,00	5.924.400,00	191.364.464,37	5.924.400,00	185.440.064,37
026	Nekretna, postrojenja i oprema u pripremi	440.590,00		1.835.536,67	1.166.426,67	2.685.146,67	1.166.426,67	918.730,00
027	ULAGANJA NA TUĐIM NEKRETNIMAMA, POSTROJENIMA I OPREMI	138.293.844,58		9.372.833,30		147.666.777,88		147.666.777,88
029	ISPRAVKA VREDNOSTI NEKRETNINA POSTROJENJA I OPREME		247.018.991,03	1.576.971,97	10.044.744,06	1.576.971,97	257.063.735,09	-255.486.763,12
048	Ostali dugoročni finansijski elementi	531.679,05				531.679,05		531.679,05
101	Materijal	2.060.182,07		-83.367,45	1.161.341,30	1.976.814,62	1.161.341,30	815.473,32
103	Alat i inventar	26.464.414,19		6.050.405,11	7.875,00	32.514.905,50	7.875,00	32.507.030,50
109	Ispravka vrednosti zaliha materijala		26.464.414,19		6.042.620,11		32.507.034,50	-32.507.034,50
110	Računi svrsni za materijal, rezervne delove i inventar u zemlji	846.752,05		111.653,90	9.466.187,14	10.897.978,17	9.579.841,04	1.318.137,13
113	Računi svrsni za rotnu u zemlji	165.291,36				165.291,36		165.291,36
204	Kupci u zemlji	20.500.766,77	0,00	46.157.571,41	55.413.298,35	66.658.941,18	55.413.298,37	13.245.642,81
205	Kupci u inostranstvu	2.694.394,05		5.982.308,28	7.095.233,98	8.676.601,53	7.095.233,98	1.581.369,35
206	Ispravka vrednosti potraživanja od prodaje		52.755,71				52.755,71	-52.755,71
218	Ostala potraživanja iz specifičnih poslova	4.573,94	4.573,94	19.660,00	19.660,00	24.233,94	24.233,94	0,00
221	Potraživanja od zaposlenih			13.535,71		13.535,71		13.535,71
222	Potraživanja od državnih organa i organizacija			2.326.108,00	1.973.855,00	2.326.108,00	1.973.855,00	352.253,00
223	Potraživanja za više plaćeni porez na dobitak	1.676.862,00			1.676.862,00	1.676.862,00	1,676.862,00	0,00
224	Potraživanja po osnovu preplaćenih ostalih poreza i doprinosa	270.991,34		7,32	13.848,70	270.998,66	13.848,70	257.149,96
228	Ostala kratkoročna potraživanja	1.459.856,21		-32,67	600.655,85	1.459.823,54	600.655,85	859.167,69

Izveštaj generisan pomoću edicije "Mali poslovni programi", proizvođača PS: IT (2003 - 2017), sva prava zaštićena.
Program je licenciran na: SANZA D.O.O., Potuljka 67a, lok. 4/07, Beograd. Štampano: 22.09.2020. 1 / 6

DeX: made with  by MQSoft

Thank You 😁