



Valsts un pašvaldību NESTRUKTURĒTO DATU identificēšana un analīze

Leons Mednis



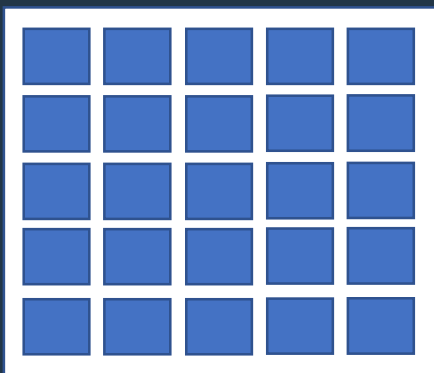
PEDAB

Dienas kārtība

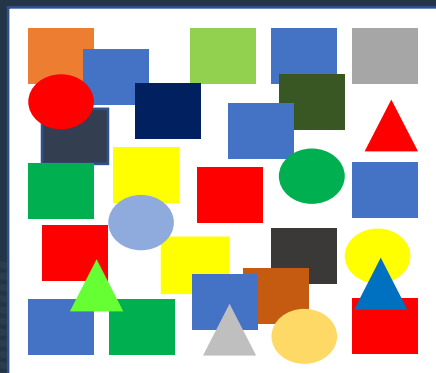
- Kas ir NESTRUKTURĒTIE DATI?
- Esošā situācija
- Tipiskākākie izaicinājumi nestrukturēto datu glabāšanā
- Praktiskie pielietojumu veidi
- Kā tas strādā?
- Kā mēs varam palīdzēt?
- Demo

Kas ir nestrukturētie dati

Strukturēti dati



Nestrukturēti dati



Nestrukturētie dati:

- glabājas dažādās formās (teksts, foto, skaņa utt.)
- nav raksturīga rinda-kolonna struktūra
- grūti analizējami izmantojot tradicionālos rīkus, piemēram, lai atrastu noteiktu objektu (vārdu, frāzi, datumu, numuru utt.)

Kas ir nestrukturētie dati



Piemēri:

- Word un PDF dokumenti (līgumi, iesniegumi, klienta vēsture utt.)
- Excel faili (katrā darba lapā iespējama sava datu struktūra, pastāv tekstu saturoši lauki, apraksti utt.)
- e-pasts (ietver strukturētu un nestrukturētu informāciju, kā arī pielikumus)
- PowerPoint prezentācijas
- tērzēšanas ieraksti (čati)
- bildes (pases, ID, personas foto u.c.)
- Utt.

Esošā situācija

80% un vairāk. Nestrukturēto datu apjoms no visiem organizācijas datiem

70% No glabātajiem nestrukturētajiem datiem nav vērtības

60% un vairāk. Par tik pieaug nestrukturēto datu apjoms ik gadu



Glabājamo datu apjoms aug ātrāk, nekā samazinās datu glabāšanas izmaksas

Esošā situācija

Vīzija, sākums...



... rezultāts.



[E-pasts jautājumiem - tiesraide@lps.lv](mailto:tiesraide@lps.lv)

 data
technologies

PEDAB

Tipiskākākie izaicinājumi nestrukturēto datu glabāšanā

- Uzkrātais informācijas apjoms failu veidā ļoti ātri pieaug
- Stratēģija «Glabājam visu! Gan jau noderēs!»
- Datu glabāšanas instrukciju un procedūru ievērošana
- Kā atbrīvoties no novecojošas un neaktuālas informācijas
- Kā atrast un nepazaudēt uzkrātās zināšanas un vērtīgu informāciju

Kādi ir galvenie uzdevumi nestrukturēto datu analīzei?

- Atrast informāciju, kurai ir nozīmīga vērtība no organizācija vai risku skatpunkta
- Identificēt sensitīvu informāciju
- Auditēt saglabātās informācijas atbilstību izstrādātajiem procesiem, procedūrām un atbilstības prasībām
- Atbrīvoties no novecojušas un nevērtīgas informācijas
- GDPR



Visbiežākie praktiskie pielietojumu veidi

PĀRVALDĪBA

Palielināt efektivitāti un skaidrību visās nestrukturētās informācijas glabāšanas vietās ORGANIZĀCIJĀ:

- Analizē nestrukturēto informāciju
- Uzkrāt zināšanas vienuviet
- Identificēt vērtīgu informāciju un to nepazaudēt
- Izanalizēt jaunus nestrukturētos datus apvienojoties

Visbiežākie praktiskie pielietojumu veidi

RISKI

Samazinot riskus:

- Identificēt informāciju, kuras glabāšana dod mazāku vērtību ORGANIZĀCIJAI, nekā tās glabāšana izmaksas vai riski informācijas noplūdes gadījumā
- Identificēt sensitīvas informācijas noplūdi
- Nepazaudēt vērtīgu informāciju

Visbiežākie praktiskie pielietojumu veidi

AUDITS

Dot vērtīgus atklājumus auditam un atbilstības nodrošināšanai:

- Atklāt neatbilstošas un nedrošas vietas, kurās tiek glabāta sensitīva (personu vai organizācijas) informācija

Visbiežākie praktiskie pielietojumu veidi

GDPR

Atrisināt GDPR izaicinājumus:

- Atrast personu datus nestrukturētajos datu avotos (failu serveri, e-pasta serveri, desktop utt.)
- Izzanalizēt esošo situāciju un identificēt neatbilstības

[E-pasts jautājumiem - tiesraide@lps.lv](mailto:tiesraide@lps.lv)



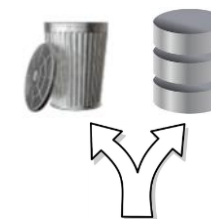
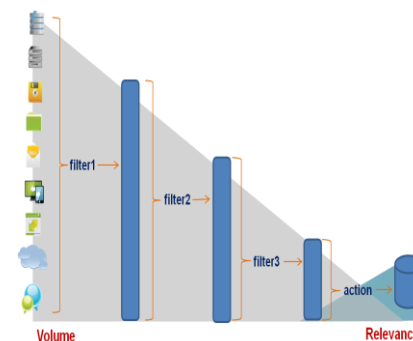
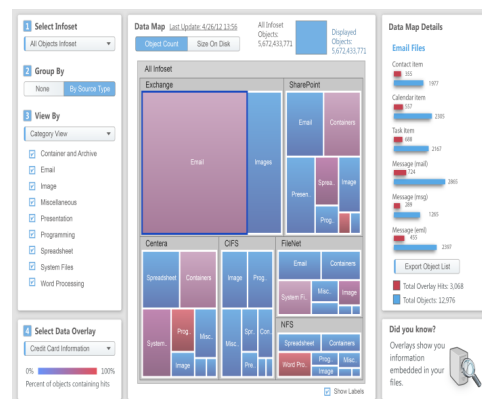
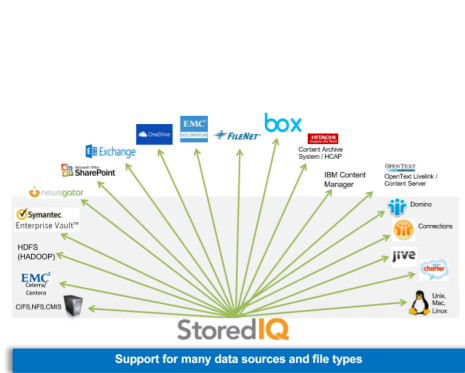
A close-up photograph of a person's hands holding a tablet computer. The scene is bathed in warm, golden light, likely from a window in the background. The person is wearing a dark blue long-sleeved shirt. The tablet is held in the right hand, and the left hand is resting on the wooden surface of a table. The background is softly blurred, showing a window and a small potted plant.

1

Kā tas strādā?



Nestrukturēto datu analīze ar StoredIQ



Izvēlies un pievieno
Datu avotu

Sākotnējā Datu izpēte

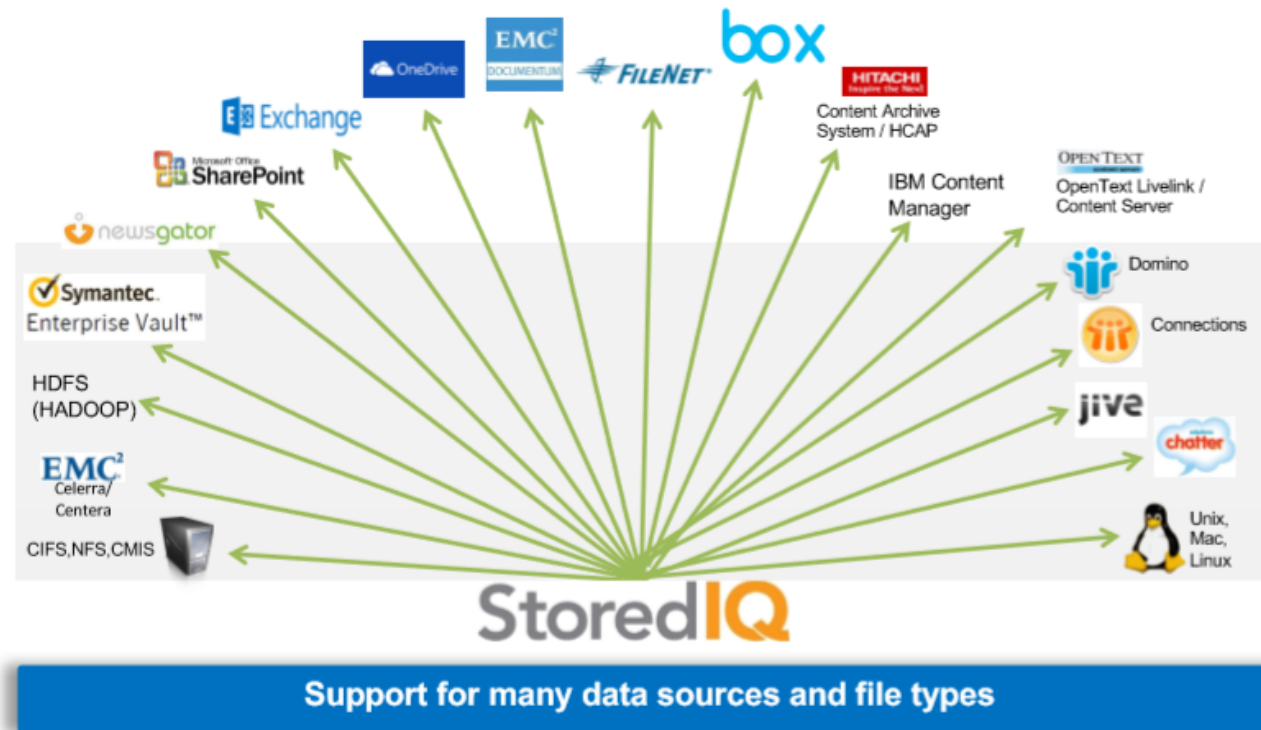
Analīzē - meklē,
attīri un precizē

Ziņo, dalies ar
secinājumiem

Rīkojies

Izvēlies un pievieno Datu avotu

Identificē un analizē datus vairāk kā 75 datu avotos un vairāk kā 450 failu veidos

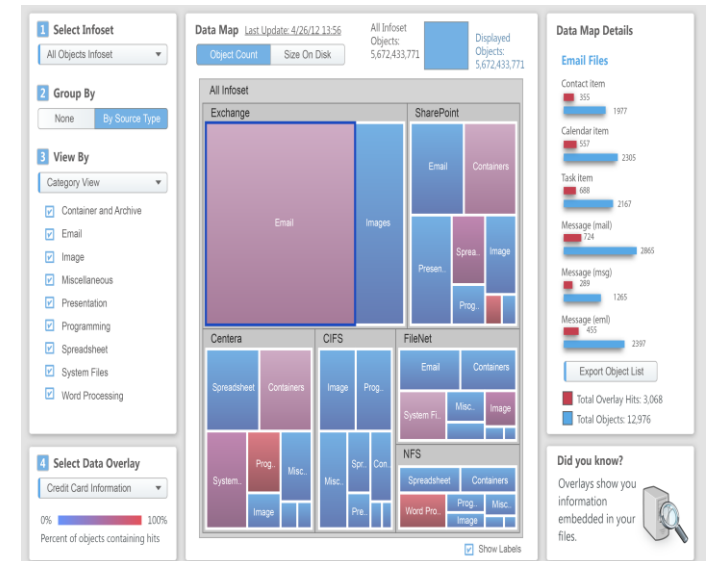
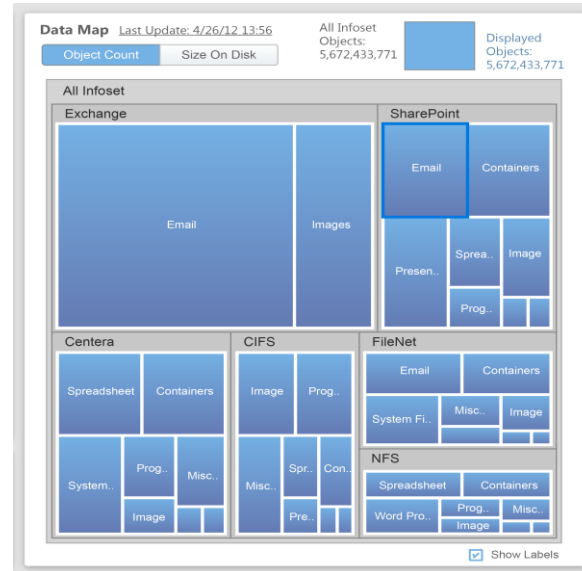


Sākotnējā Datu izpēt

SĀKOTNĒJĀ IZPĒTE - identificē, kur glabājas visvecākie, vislielākie vai vismazāk izmantotie dati utt.

DATU KARTE parāda, kādi failu tipi un kur glabājas (pēc vecuma, lieluma utt.)

HEATMAPS parāda dažādu filtru un meklēšanas kritēriju pārklāšanos.

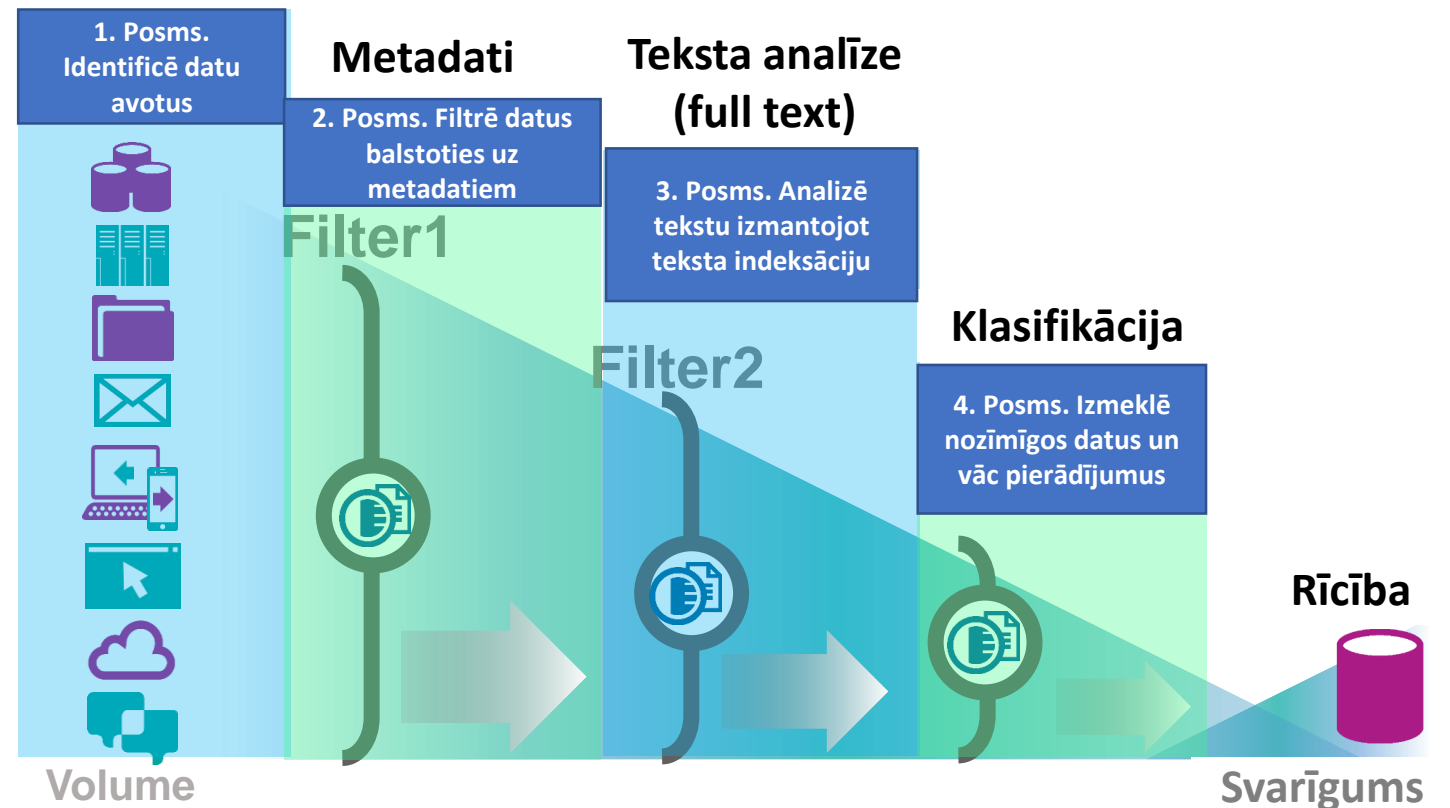


Analizē - meklē, attīri un precizē

Identificē mēklējamo informāciju, piemēram, personas kodu, ar meklēšanas kritērijiem vai sagatavotiem filtriem

Izveido pats savus filtrus atbilstoši uzdevumam

Izmanto 'cartridges' lai paplašinātu StoredIQ iespējas (tas ļauj izmantot dažādas tehnoloģijas, sākot ar vienkāršu regex sintakses izmantošanu un beidzot ar sarežģītiem valodas analīzes un mašīnmācīšanās algoritmu izmantošanu).



Izmanto dažādu atlasē kritēriju, filtru un mašīnmācīšanās tehnikas, lai atrastu un klasificētu atrasto informāciju

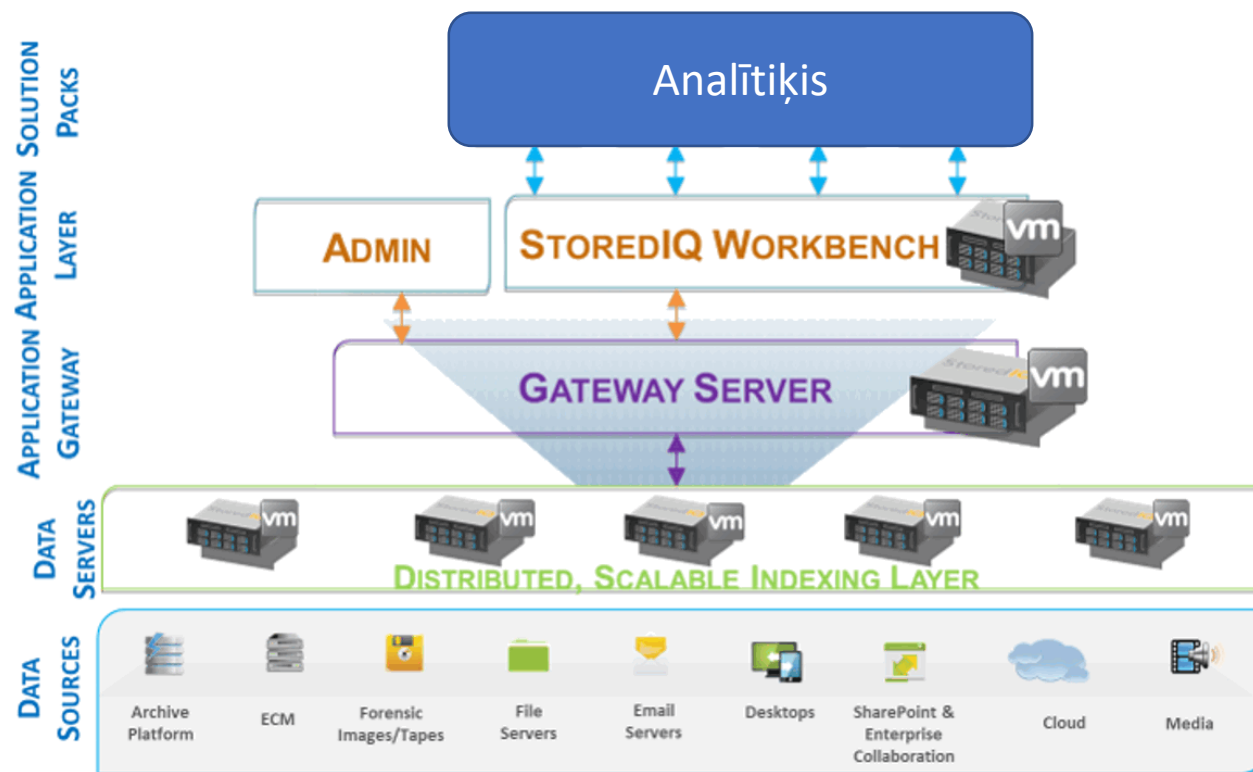
Rīkojies (pārvieta, kopē, dzēs, utt.)

Kā tas tiek realizēts?

StoredIQ satāv no 3 komponentēm(OVAs):

- Application stack
- Gateway server
- Data server

VMware vSphere - ismantojam vSphere Client lai izvietotu 3 virtuālās mašīnas uz ESX servera.



Kāpēc StoredIQ

- Nav nepieciešama papildu datubāze
- Analizē visus nestrukturētos datus no vienas vietas
- Uzkrāj zināšanas vienuviet
- Izmanto atkārtoti un paplašini izmantotos atlasē kritērijus un filtrus analizējot visus nestrukturētos datu avotus
- Rīkojies – pārvieto, kopē, dzēs vai veic citas darbības
- Viens risinājums – dažādi uzdevumi: pārvaldīt un satīrīt, auditēt, samazināt riskus un GDPR uzdevumi
- Labs Analītiķis var izdarīt gandrīz visu. Nepieciešama neliela IT palīdzība.
- Mērogojamība (scalability)

Kā varam palīdzēt?



1. Nodrošinot ar nestrukturēto datu analīzes risinājumu
2. Veicot apmācības darbam ar risinājumu
3. Analītiķa darbu un konsultācijas

Varam veikt visu no A līdz Z!

[E-pasts jautājumiem - tiesraide@lps.lv](mailto:tiesraide@lps.lv)



Risinājums

Platformas noma + Analītikas eksperts

60 dienu nomas maksa, 2 datu avotu veidi, 4 datu tipu veidi

- <2TB – 3500 eur
- Katrs nākamais TB – 800 eur

90 dienu nomas maksa, 5 datu avotu veidi, 7 datu tipu veidi

- <2TB – 6000 eur
- Katrs nākamais TB – 800 eur

120 dienu nomas maksa - pēc vienošanās

Jautājumu vai
neskaidrību
gadījumā, sazinies
ar mums!

Leons.Mednis@pedab.lv

+371 26336335

Jūsu jautājumi

[E-pasts jautājumiem - tiesraide@lps.lv](mailto:tiesraide@lps.lv)



DEMO >>>

Sazinies ar mums



LEONS MEDNIS

Vecākais analītikas eksperts

E-pasts: Leons.Mednis@pedab.lv

Telefons: +371 26336335

LinkedIn: [Leons Mednis](#)



JĀNIS TREIJS

E-pasts: janis.treijs@datatechnologies.lv

Telefons: +371 29233533

LinkedIn: [Janis Treijs](#)