

Tehnična naloga

V podjetju želite vzpostaviti sistem za obdelavo dokumentov. Iz dokumentov želimo shranjevati naslednje podatke, ki so prepoznani s pomočjo storitve Typless:

- Številka dokumenta
- Datum dokumenta
- Tip dokumenta

Sistem bo obdeloval 1000 dokumentov na sekundo v PDF obliki. Povprečna velikost dokumenta je 1 MB.

Zahteve sistema so:

- Pobiranje dokumentov iz poljubnega zunanjega vira (lahko si izbereš svojega)
- Možnost urejanja podatkov na dokumentu
- Možnost brisanja dokumenta
- Možnost kreiranja dokumenta
- Možnost iskanja po številki, datumu in tipu dokumenta
- Možnost sporočanja v zunanji sistem po uspešni obdelavi dokumenta

Naredi skico arhitekture in napiši kodo za zgoraj omenjeni sistem, ki bo podpirala vse zahteve.

Skica mora vsebovati:

- Vse podatkovne modele in relacije med njimi
- Vse potrebne servise in način, kako komunicirajo med sabo
- Arhitekturo REST servisa z vsemi potrebnimi endpointi - Ta naj bo implementiran v poljubnem Python frameworku z delujočo kodo, lahko pa pomockanimi podatki

BONUS:

- Vzpostavi podatkovne modele tako, da bodo primerni za DynamoDB NoSQL bazo

Vprašanja:

- Kakšen bandwidth mora sistem imeti, da lahko podpira procesiranje takšnega števila obdelav na sekundo?
- Kako velik storage je potreben za shranjevanje dokumentov enega meseca?

V primeru kakršnihkoli vprašanj in dvomov pišite na klavdij.starman@algoritmik.si