

gen-i



PAMETNA ENERGIJA

Na papirju je svet rožnat oz. kako iz lepe vizije v PPT prideš v realni svet

Prehod v cloud – lessons learned

Sistemiški gradniki

INTELLIGENT

Preko podatkovnih tokov naprave in sisteme **medsebojno povežemo** v enoten ekosistem. S kombinacijo centralnih in lokalnih algoritmov zagotavljanje **optimalnega usklajenega delovanja** naprav in sistemov v realnem času **brez posredovanja** uporabnikov.

SMART

Algoritmi (v napravah³ in sistemih) na podlagi obdelave podatkov **samostojno neposredno vplivajo** na delovanje teh naprav in sistemov v realnem času brez posredovanja uporabnikov.

INSIGHT

Ugotovitve, pridobljene na podlagi obdelave (analitike) podatkov, **delimo** z drugimi uporabniki (strankami in službami) ter s tem **posredno vplivamo** na njihovo delovanje in obnašanje.

VISIBILITY

Podatke, ki jih zajemamo iz razpoložljivih virov (senzorji¹ in sistemi²) **zberemo** in katalogiziramo ter jih s tem **naredimo dostopne in vidne**.

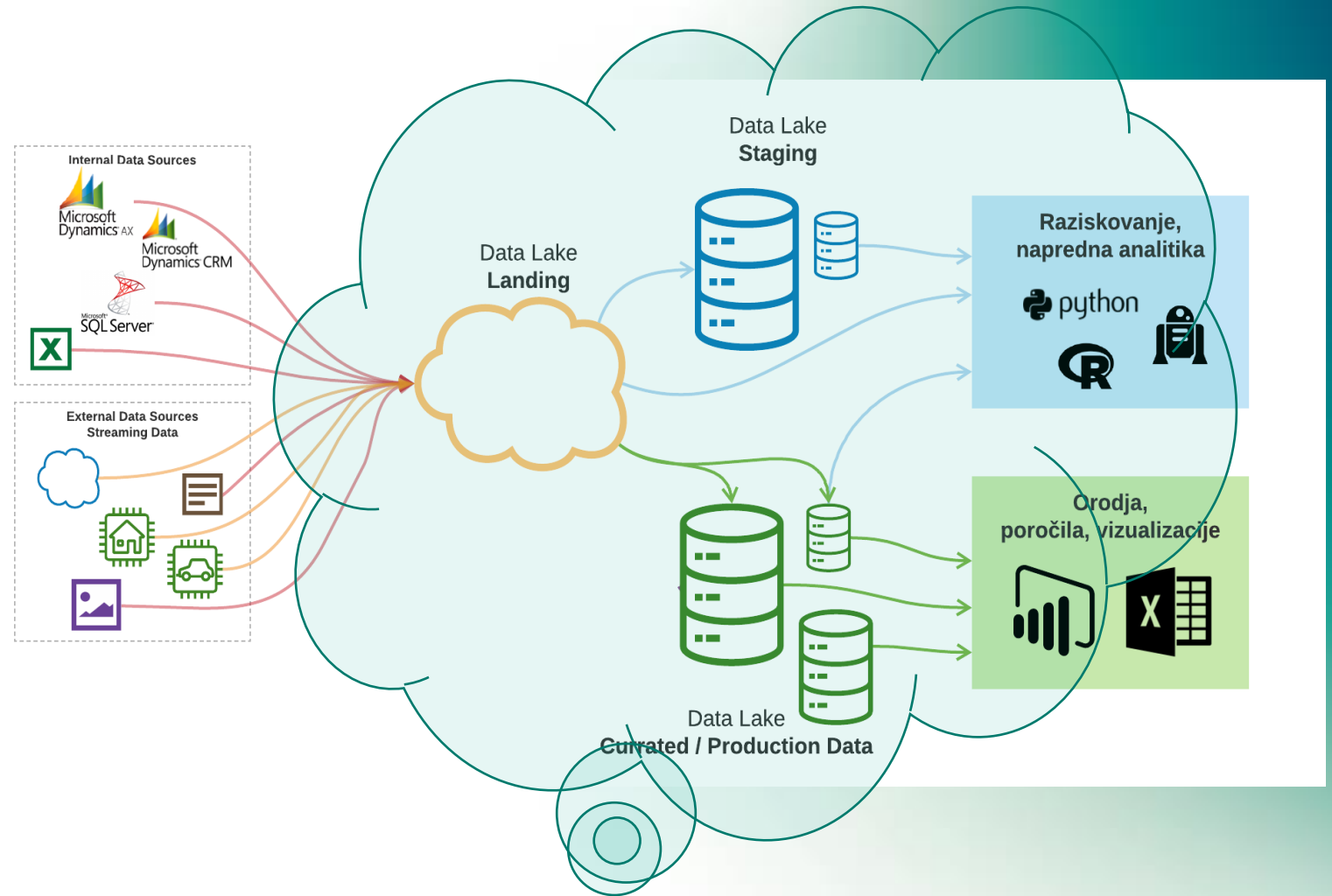
¹ Senzor kot oddaljena, samostojno delujoča naprava z možnostjo komuniciranja podatkov, namenjena digitalizaciji fizičnega sveta (npr. števec za električno energijo, razsmernik SE, ...)

² Sistem kot interni transakcijski sistem, računalniški program, namenjen izvajanju poslovnih procesov (npr. ERP, ETRM, CRM, Portal, Mobilna aplikacija...)

³ Naprava kot kombinacija senzorjev in aktuatorjev (stikala, preklopniki, ...).

Gradniki in identifikacija vrzeli – pogled skozi tehnične gradnike

- Leta 2019 smo bili v fazi raziskovanja primerne okolja za realizacijo predstavljene vizije.
- Oblikovana iniciativa za službo IT „**VSE – TAKOJ**“ kot splošna usmeritev pri razvoju in integraciji aplikacij in naprav.
- Konkretni infrastrukturni gradniki okolja (**vir**i, **baze**, **povezave**, **ponori**...) so bili identificirani.
- Potreba po tehnološki agnostičnosti bodoče rešitve je bila identificirana in jasno razumljena - **povezljivost z različnimi tehnologijami**.
- Oblačna tehnologija je glede na potrebe **največ obetala** zato se je glavna raziskovanja osredotočila v to smer.



Arhitektura platforme

Analitika in obdelava

- Sodobni pristopi
- Neomejeni viri za obdelavo z lastnimi algoritmi
- Visokonivojske storitve (ML, AI)

Shramba podatkov

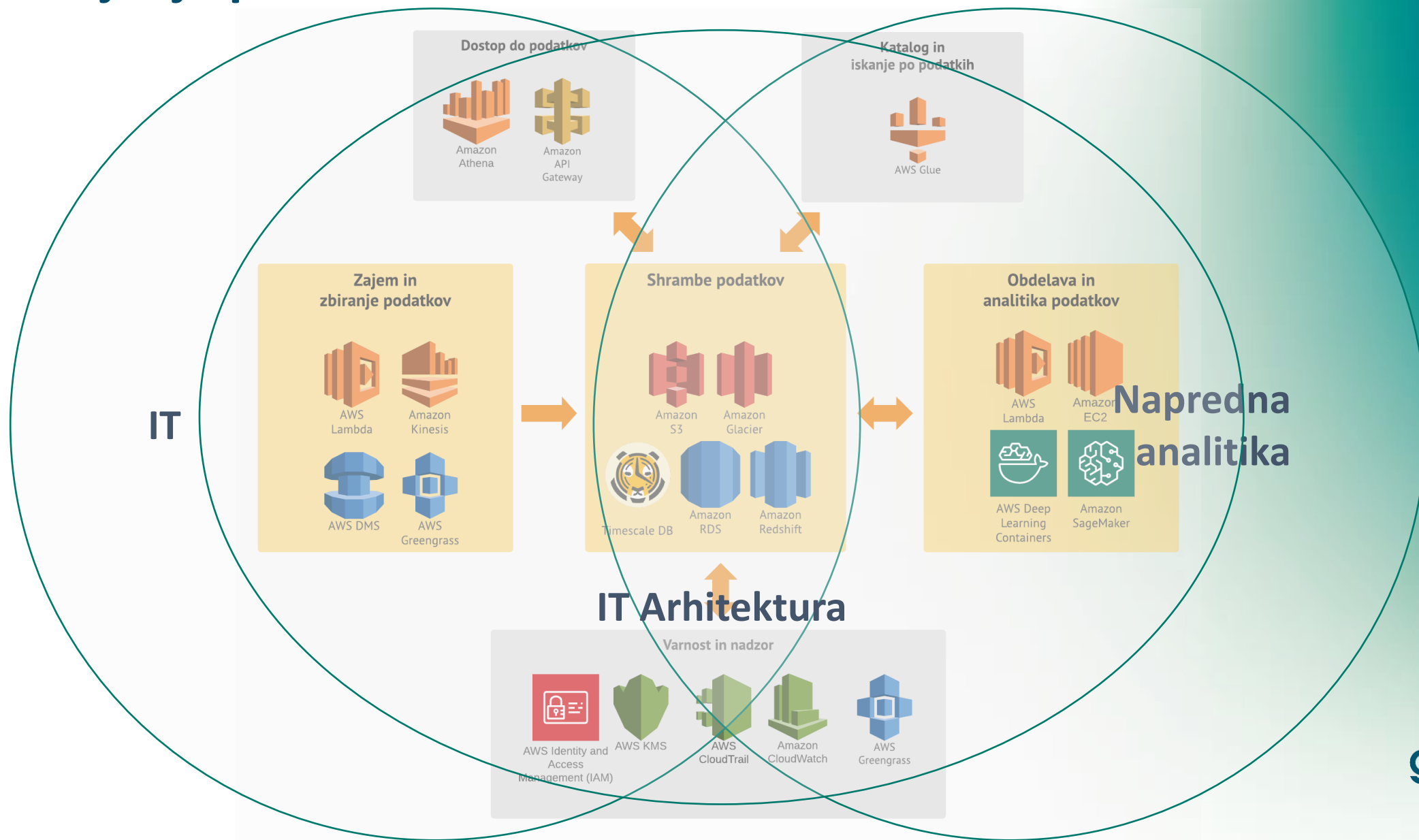
- Neomejena velikost in trajanje podatkov
- Za različne tipe podatkov

Zajem podatkov

- Varen, enostaven, hiter
- Skalabilen, ponovljiv



Upravljanje platforme



Prvi koncept

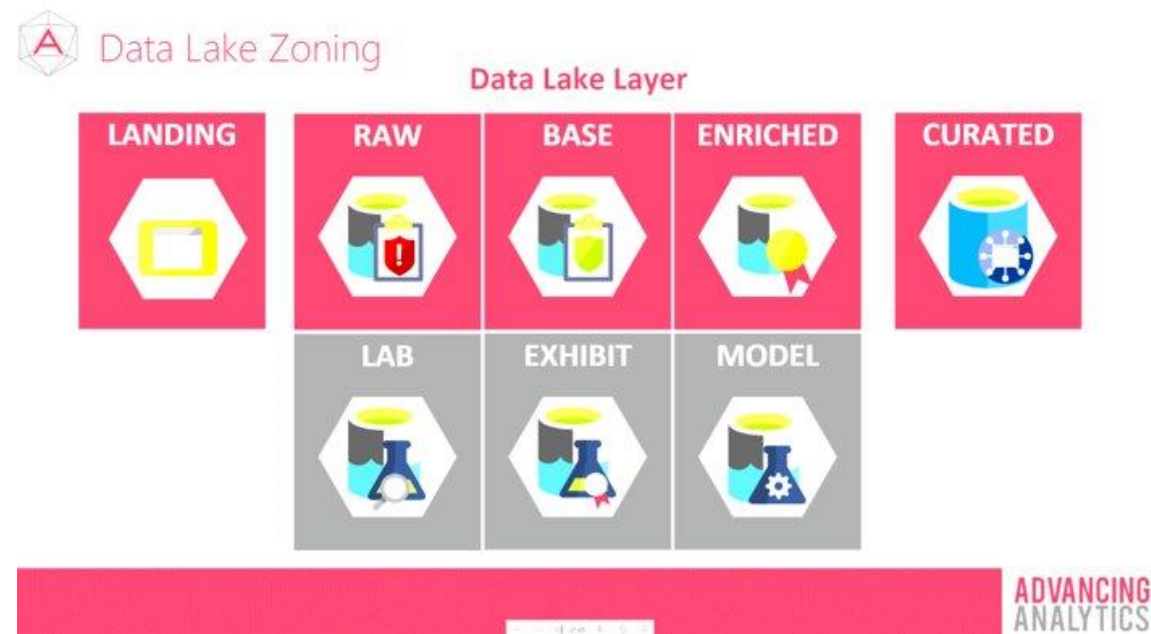
- Vzpostaviti framework za grajenje in orkestracijo toka podatkov od vira do končne destinacije v cloudu, s katerim zagotavljamo enotne transformacije podatkov (standardiziran in poenoten proces):

oOd zajema: SOURCE->RAW (hramba as-is podatkov)

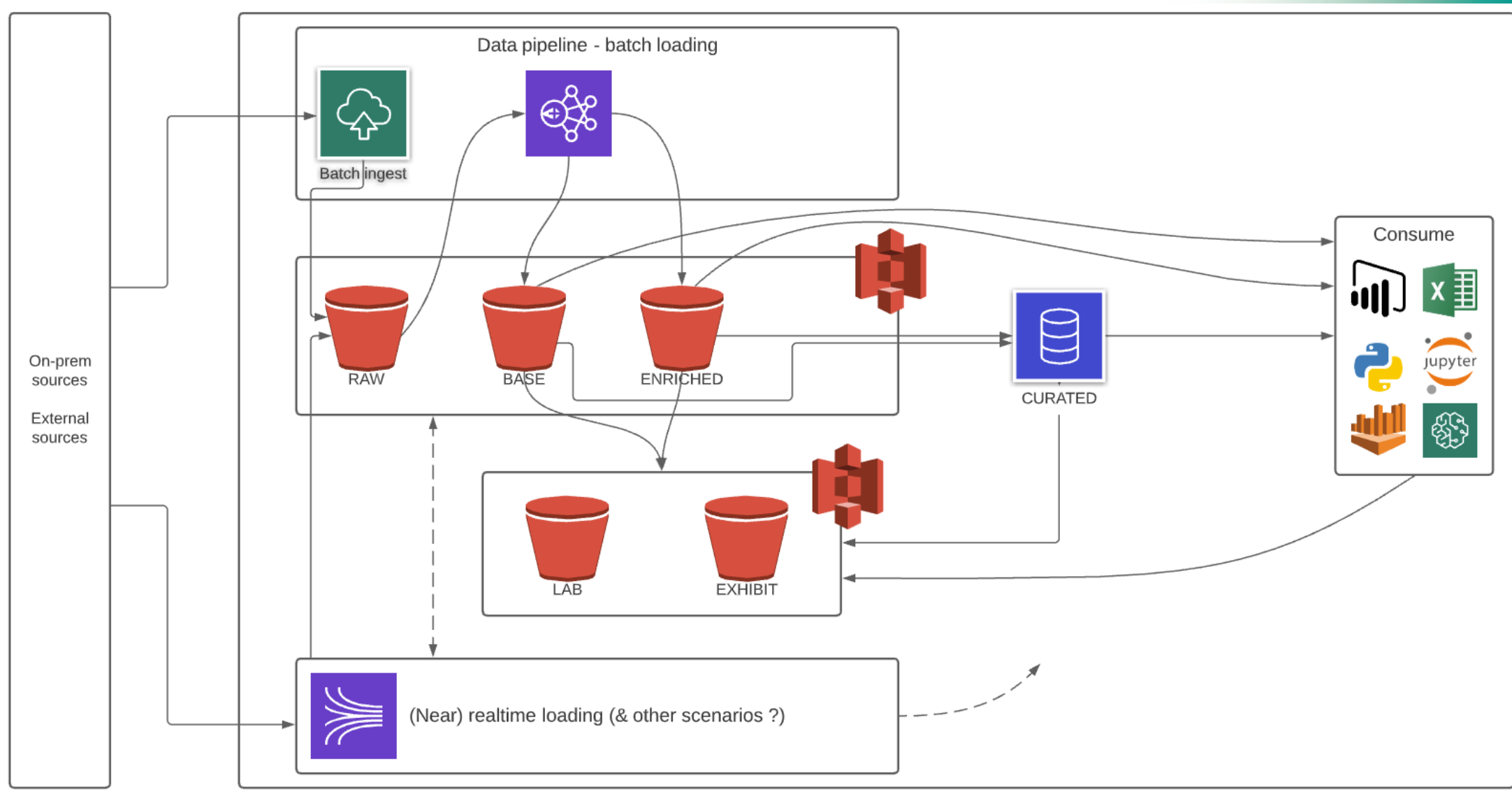
oSkozi data lake: RAW->BASE(->ENRICHED) (hramba generično poenotenih/obogatenih podatkov za različne vire; , podatki dobijo „ključ“)

oDo (potencialno) specializirane podatkovne destinacije (Timescale, Redshift, Snowflake, RDS...) (hramba podatkov za različne aplikacije, analitike, poročila...)

- ENRICHED->CURATED
- BASE->CURATED



The Flow of Data



GEN-I DATA PIPELINE

- Zagotoviti **enotno in fleksibilno** orkestracijo posameznih korakov in njihovih medsebojnih odvisnosti
 - Omogoča **dinamično prilagajanje** elastičnih performanc glede na količino podatkov za procesiranje
 - **Meta-data driven** (brez potrebnega znanja programiranja za kreiranje in konfiguriranje taskov)
 - Uporaba **DevOps** za GENI Data Pipeline
 - **laC**
 - Izvajanje taskov: po urniku, glede na dependencie (on success,...), ...
 - Logiranje izvedb opravil
 - Data lineage tracking in vizualna predstavitev toka podatkov
 - Fleksibilnost, hitro prilagajanje v primeru sprememb

GDP – Realni svet

- Super za generiranje kode:
 - **manj napak** v kodiranju
 - **hitrejši razvoj**
 - Na vseh okoljih **enaka koda**
 - Kolikor toliko enostavno **razširljiv**
 - Uporabnik ne potrebuje poznati Cloud storitev, pa jih lahko z lahkoto uporablja – **transformacije definira v json datoteki**
- Step funkcije za orkestracijo
 - **Počasne** pri „spoznavanju“, da je opravilo končano na EMR clustru (pri daljših opravilih 5 min)
 - Na začetku so bili izzivi s pisanjem definicij za state machine
 - EMR Serverless + stepfunkcije je hitreje
- **BASE, ENRICHED coni nikoli v resnici zaživele:**
 - **Athena počasna** (sedaj je verzija 3, ki je nekoliko hitrejša) v primerjavi z SQL; analitiki so se pritoževali (počasen dostop do podatkov)
 - Povezovanje različnih (ON-PREM) servicev na S3 – iskanje connectorjev ipd....
 - Končna destinacija za veliko virov je postal **Snowflake**
 - Več znanja v **relacijskih bazah, kot v Sparku**
- Dostop do podatkov na S3 (Power BI)
 - Preko **ODBC Simba Athena driverja**
 - Kar nekaj "predpogojev" v glue catalogu – da se pridobi pravilne podatke
 - Obvezno **particije**
 - Počasno za velike datasete (predobdelava podatkov za analitike)
- Potrebni so dodatni developerji, za vzdrževanje GDP
 - Podpora novim transformacijam, novim vhom/izhodnim formatom ipd.
 - Podpora novim vrstam okrestracije (gnezdenje step funkcij, izvajanje glede na nek pogoj ipd.)

Spremembe, ki smo jih uvedli

- Opuščamo BASE in ENRICHED cono, kjer se je izkazala, da je analitiki ne uporabljajo (optimizacija: price – performancne – usability)
- Ker je končna destinacija Snowflake, smo se odločili, da bomo transformacije naredili v Snowflake
 - Ingest preko Snowflake objektov se je izkazal za hitrejšega (EMR obdelava 15 min vs. Snowflake 1 min).
 - DISCLAIMER:** EMR taski mogoče niso tako optimizirani kot bi lahko bili, 1 EMR za vsa opravila (kakšne ure so zelo zasedene, druge manj)
- Kjer potrebujemo EMR – spreminjamo na uporabo **EMR serverless (na voljo v Frankfurt regiji cca. 1 leto)**
 - Omogoče boljše taggiranje resurcev
 - Kaže da bo tudi hitrejši
 - Kadar destinacija ni (samo) Snowflake

Glavna sporočila

- Od 2019 z zametki iz še zgodnejših let poteka **intenzivno usmerjeno raziskovanje** s ciljem vzpostavitve robustne, skalabilne, univerzalne in fleksibilne skupne podatkovno analitske platforme.
- Ta predstavlja univerzalne temelje, na katerih lahko:
 - brez pomembnejših omejitev **zbiramo in ponujamo** (delamo vidne) relevantne podatke, tako tiste, ki so nam danes že dostopni kot tiste, ki jih še bomo potrebovali za delovanje naprednejših in kompleksnejših storitev v prihodnosti
 - s pomočjo analitike in analitskih ter vizualizacijskih orodij najdemo **dodano vrednost**, ki se skriva v teh podatkih
 - podatkovno povežemo in z algoritmi vzpostavimo usklajeno delovanje tehnološko in funkcionalno raznolikih naprav in sistemov
 - tovrstne rešitve **agilno** vzpostavljamo, skaliramo, spreminjamo in z njimi eksperimentiramo
 - prav zaradi enotne skupne platforme večkrat **ponovno uporabljamo že razvite** funkcionalnosti
- Platforma je preko skupnega delovnega okolja uskladila in povezala v enotno razvojno gonilo službe Napredne analitike, IT in IT Arhitekture na področju zajema, hrambe in obdelave podatkov.
- Platforma je več ali manj **nevidna za vire** in končne (poslovne) uporabnike, pri čemer pa je pomembno da se poslovni načrtovalci novih funkcij in storitev zavedajo(mo) njenega obstoja, pomena in zmogljivosti.
- Platforma je „živa“ in se prilagaja potrebam uporabnikov
 - Konstantno se **dogajajo spremembe, moramo slediti AWS novostim**

Lessons learned

- Platforma mora slediti uporabnikom in ne viziji!
- Spremembe platforme so koristne
 - Če prva verzija ni Ok, ne se ustrašit je podret in na novo postaviti
- Obvezno slediti novostim AWS – česar danes ni, je jutri na voljo
 - Ne smemo zaspati – konstanto iskanje možnosti za izboljšave
- Kaj je uporabno /dovolj dobro in kaj ne, povedo uporabniki



Pinboard

@Pinboard



The Programmers' Credo: we do these things not because they are easy, but because we thought they were going to be easy

10:15 PM · Aug 5, 2016



12.5K



47



Share this Tweet