
Data Infrastructure as Code

Building core data services in a small team

Michał Gancarski, Zalando SE

Introduction

Team Nucleo - Data Lake Core

- ❖ eight engineers, lead, producer and the product specialist



Ashwini



Ire



Max



Alberto

LEAD



Michal



Nikita



Onur



Kay

PRODUCER



Saurav



Viacheslav



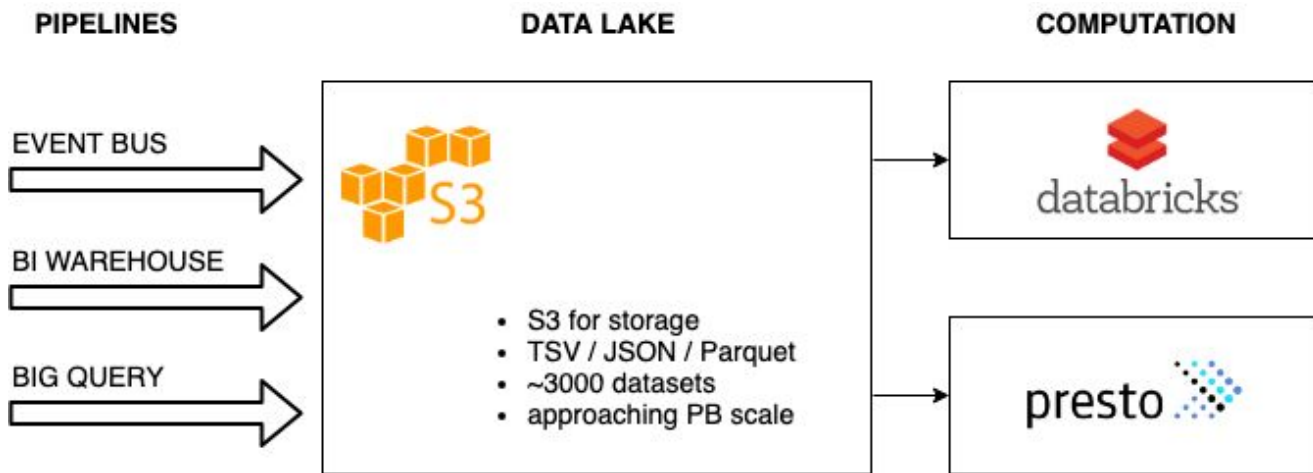
Akhil

PRODUCT SPECIALIST

Introduction

The Infrastructure

- ❖ thousands of datasets, hundreds of users, tens of clusters



Introduction

Main Challenges

- ❖ project overload
- ❖ GDPR
- ❖ compliance
- ❖ job market



Introduction

Value Multipliers and Cost Savers

$$RoW = \frac{V - C}{C}$$

RoW - Return on Whatever

V - value

C - cost

(+V) Reach Out

Be Friends with Stakeholders

- ❖ compliance
- ❖ security & governance
- ❖ IT architecture

- ❖ business intelligence
- ❖ product analytics



(+V) Reach Out

Offer Multiple Support Channels

- ❖ team email
- ❖ #datalake-users
- ❖ #databricks-users
- ❖ office hours
- ❖ guest developers



(-C) Outsource & Empower

Watch Users Solving It For You

❖ users are your community

Troubled User 1 16:40

Hi, when will be the data in dwh_insider.d_content_production updated?

Helpful User 2 16:42 • Edited

please be aware taht due to a delay in BI main ETL, also Data Lake exports are delayed

Troubled User 1 16:43

okay thank you

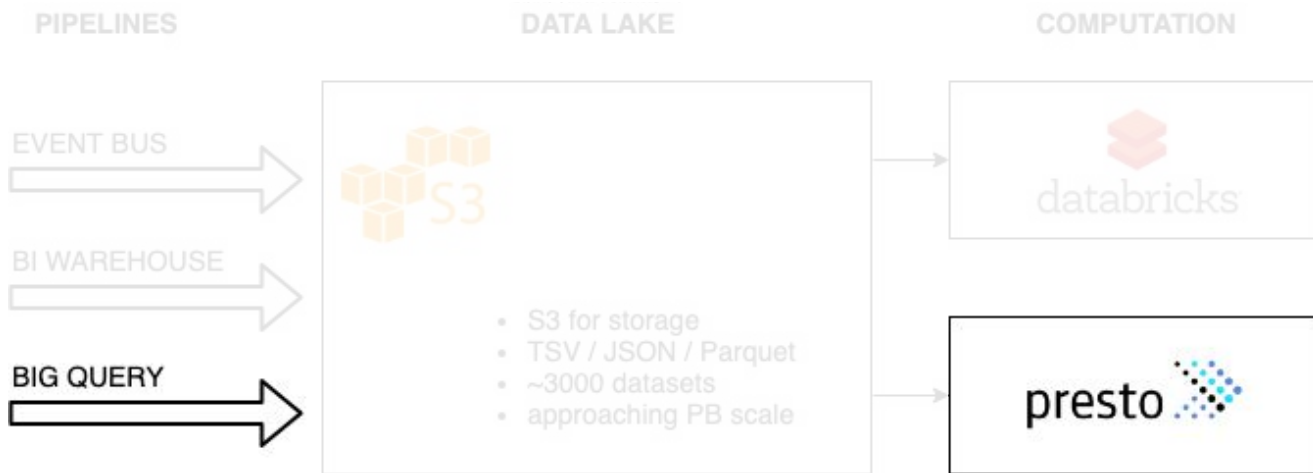
Helpful User 2 16:44

yw

(-C) Outsource & Empower

Let Others Do Your Work

❖ innersourced Big Query pipeline & Presto proxy



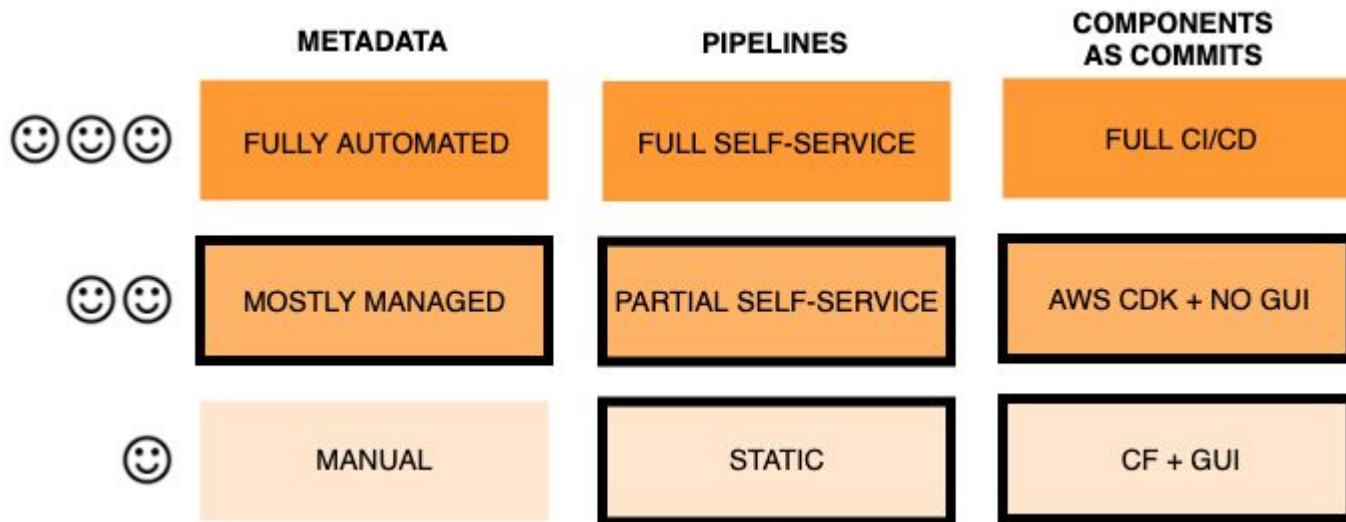
(-C) Outsource & Empower

Lean on Your Vendors. Hard.

My requests		Requests I'm CC'd on		
Search requests		Status: Any		
Id	Subject	Created	Last activity	Status
22407	effects of OPTIMIZE and VACUUM to downstream	2 days ago	5 minutes ago	SOLVED
22493	Autotermination timeout	2 hours ago	16 minutes ago	OPEN
22157	My email address is always removed from my requests	9 days ago	3 hours ago	SOLVED
21288	Issue with users API	1 month ago	1 day ago	AWAITING YOUR REPLY
22436	question about s3-sqs as streaming source	1 day ago	1 day ago	SOLVED

(+V) Automate and Generalize

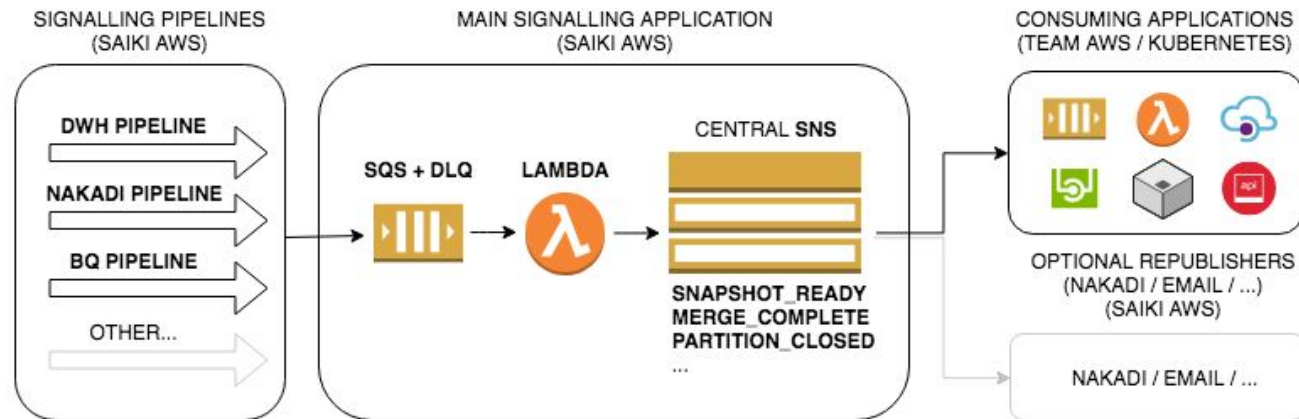
Climb The Ladder of Automation



(+V) Automate and Generalize

Treat Use Cases as Future Services

- ❖ pilot with one team
- ❖ parametrize and turn into a service



(-C) Protect and Simplify

Build The Wall



Alberto

**ARCHITECTURE
OFFICE POLITICS**



Kay

PROCESS & FACILITATION



Akhil

ROADMAP ALIGNMENT

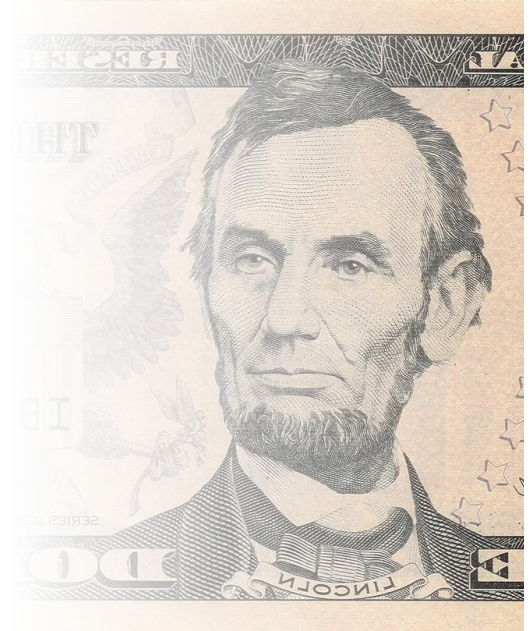
(-C) Protect and Simplify

Maintain Identity and Vision

"I strongly believe that the future prosperity of the American people depends on how well each data infrastructure team understands what NOT to build."

#bigdata #yolo"

Abraham Lincoln in a letter to Congress, 1845



Thank you!

Michał Gancarski, Zalando SE
