



Data Engineer

Alchemia danych



Mariusz Słapek

{Data, Machine Learning, Software} Engineer
Trener w infoShare Academy

Senior Data Engineer z 7-letnim doświadczeniem w branży IT. W przeszłości pracował jako Data Scientist oraz Machine Learning Engineer. Specjalizuje się w inżynierii danych, inżynierii oprogramowania oraz uczeniu maszynowym. Jego pasją są informatyka i matematyka.

Podczas pracy kładzie duży nacisk na tworzenie skalowalnej, bezpiecznej i stabilnej architektury systemów, a także czystego i czytelnego kodu. Chętnie dzieli się wiedzą i doświadczeniem z obszaru Data Science, co daje mu dużo satysfakcji.



Plan webinaru

- Kluczowe umiejętności wymagane od Data Engineera
- Narzędzia używane w codziennej pracy Data Engineera
- Przykłady przepływu pracy z danymi na podstawie realnych przypadków
- **Warsztat praktyczny**
 - proces scrapowania danych, ETL (extract, transform, load)
 - przykład zastosowania Airflow w zarządzaniu przepływem danych
- Odpowiedzi na Wasze pytania :)



Materiały
zostaną wysłane
po webinarze



Zadawajcie **pytania**

hasto Data Engineer –

data engineer
Wyszukiwane hasło

+ Porównaj

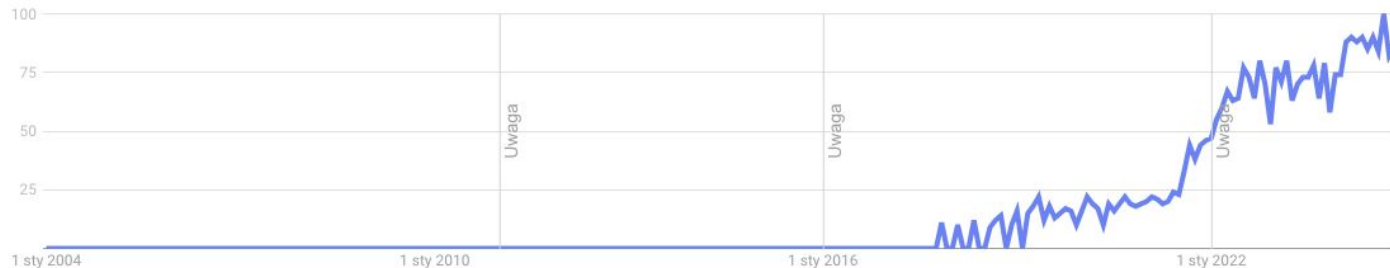
Polska ▼

2004 – dziś ▼

Wszystko ▼

Wyszukiwarka Google ▼

Zainteresowanie w ujęciu czasowym ?



hasło Data Engineer – Świat

• data engineer
Wyszukiwane hasło

+ Porównaj

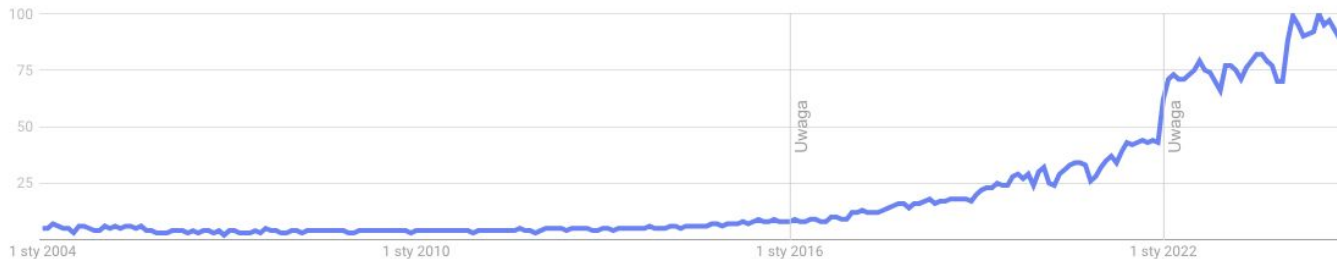
Cały świat ▼

2004 – dziś ▼

Wszystko ▼

Wyszukiwarka Google ▼

Zainteresowanie w ujęciu czasowym ?




justjoin.it #1 Job Board for tech industry in Europe


Job offers Top Companies Geek Post a job Sign in PLN


data-engineer x Location


Offers with salary All offers 461 offers Subscribe Remote Default


Work: All offers - 461 offers


**Senior Data Engineer**
Sigmoidal Wrocław, +1 Location Fully remote
24 000 - 31 000 PLN New
Snowflake Airflow AWS


**Data Engineer with GCP**
Lingaro Warszawa Fully remote
Undisclosed Salary New
BigQuery GCP Python


**Data Engineer (Java, Python, Azure)**
emagine Polska Warszawa Fully remote
25 000 - 27 700 PLN New
SQL Python Java


**Data Engineer**
GreenMinds Sp. z o.o. Warszawa Fully remote
26 000 - 29 000 PLN New
Python ETL Azure Synapse


**Data Engineer AWS+Snowflake**
Lingaro Warszawa Fully remote
Undisclosed Salary New
Snowflake AWS ETL


**Senior Data Engineer / Analyst**
Planner SD Wilno, +4 Locations Fully remote
17 223 - 21 528 PLN New
BigQuery SQL Python


**Data Engineer in Product Analytics**
ING Hubs Poland Warszawa
Undisclosed Salary New
SQL Apache Airflow Data Pipelines

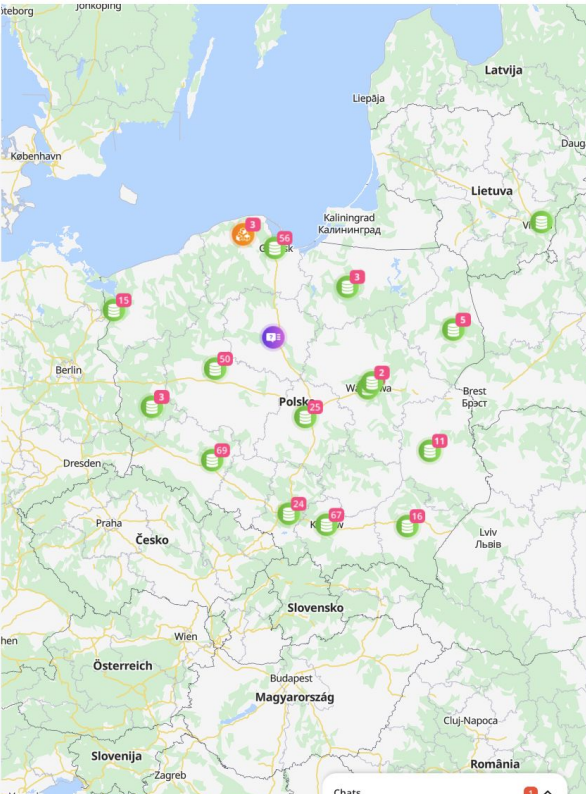
**Data Engineer**
InPost Poznań, +9 Locations Fully remote
Undisclosed Salary New
PySpark Python SQL

**Data Engineer**
ALTEN Polska Kraków, +4 Locations Fully remote
Undisclosed Salary New
Snowflake ETL Data

**Senior Data Engineer**
STX Next Poznań, +4 Locations Fully remote
23 520 - 26 880 PLN New
ETL NoSQL Python

**Data Engineer (ETL+GCP)**
PKO BP Finat Warszawa
Undisclosed Salary New
Oracle Kafka ETL

**Data Engineer (Senior & Mid) w/ Databricks**
ITLT Warszawa Fully remote
20 000 - 34 000 PLN New
ETL Databricks Python

A map of Europe with various countries labeled, including Poland, Germany, Czech Republic, Slovakia, Hungary, Romania, Lithuania, Latvia, and others. Numerous green and red pins are scattered across the map, indicating job locations. Some pins have numbers next to them, such as 1, 2, 3, 5, 6, 11, 15, 16, 24, 25, 26, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100.

Tu znajdziesz nową pracę. Sprawdź zarobki i aplikuj.

Filtry

ITMarketingObsługa klientaE3 SprzedażHRInżynieriaPozostałe

FILTRY

Wyczyść

Wynagrodzenie (PLN)

Min

Max

Typ umowy

Lokalizacja

Szukaj

☐ Praca zdalna

☐ Praca hybrydowa

☐ Praca w terenie

Doświadczenie

☐ Stażysta

☐ Junior

☐ Mid

☐ Senior

☐ Expert

Język pracy

☐ polski

☐ angielski

+ Więcej

Benefity

☐ Budżet szkoleniowy

☐ Prywatna opieka

ZNALEZIONE OFERTY PRACY (1002)

Zapisz wyszukiwanie i otrzymuj oferty na maila

Sortowanie: trafność

Senior Data Engineer with Python

DataPythonSQLAWS

GFT PolandWarszawa

21 000 – 26 880 PLN

Lead Data Software Engineer (Java, GCP)

DataJavaSQLETL

EPAM SystemsZdanie

26 500 – 31 500 PLN

Senior Customer Integration Engineer (Data Engineer)

BackendPythonFlaskGit

Focal SystemsZdanie

40 974 – 47 803 PLN

Senior Software Engineer, Data Infrastructure

DevOpsAWSPythonSQL

AsanaWarszawa

29 000 – 47 000 PLN

ML engineer / Data Scientist

DataPythonMLOpsUI

ITFS Sp. z o.o.Zdanie +1

20 160 – 21 840 PLN

Data Engineer

DataSQLAzure StoragePower BI

emagine Sp. z o.o.Zdanie

28 560 – 31 920 PLN

Senior Data Engineer

DataSQLData engineeringCloud

PragmileWarsaw

27 720 – 32 760 PLN

kurs **Data Engineer**

10 h

prework – szacowany
czas na przygotowanie do
kursu

111 h

zajęcia z trenerem
(zdalnie na żywo)

30 h

szacowany czas na
indywidualną pracę
między zajęciami

Technologie i narzędzia, które opanujesz na kursie Data Engineer



AWS



AWS S3



SQL



Kubernetes



Python



Pandas



NumPy



Polars

(REST API)

API



Hadoop



Spark



DynamoDB



Kafka



Amazon
Redshift



Apache
Airflow



Databricks



Snowflake



kurs Data Engineer

Moduł 0: Prework



Moduł 1: Wprowadzenie do Data Engineering



Moduł 2: Wprowadzenie do AWS



Moduł 3: Database i zaawansowany SQL:



Moduł 4: Zaawansowany Python dla Data Engineering



Moduł 5: Wprowadzenie do Big Data (PySpark)



Moduł 6: Extract, Transform, Load (ETL)



Moduł 7: Praca z bazami NoSQL (DynamoDB)



Moduł 8: Streaming processing (Apache Kafka)



Moduł 9: Data Warehouses



Moduł 10: Data Lakes



Moduł 11: Automatyzacja procesów (Apache Airflow)



Moduł 12: Databricks



Moduł 13: Snowflake



Moduł 14: Data Mesh





kurs Data Engineer

7400 zł ~~8900 zł~~

przy zgłoszeniu do 3 grudnia

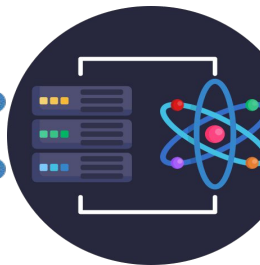
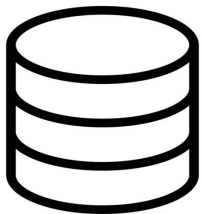
start: 25 stycznia 2025

<https://infohareacademy.com/kurs/kurs-data-engineer/>



Kluczowe umiejętności

Storage	ETL	ELT	Data Science / MLOps	Business Intelligence	Streaming
---------	-----	-----	-------------------------	--------------------------	-----------





Kluczowe umiejętności

DATA ENGINEERING ROADMAP

8. Data Governance & Security

- Data Masking / Encryption
- Data Policies / Compliance
- Authentication & Authorization

6. Data Processing

- Batch Processing: Spark
- Delta Lake & Delta Live Tables with Spark SS
- Stream Processing: Kafka

4. Storage

- Data warehousing
- Lakehouse
- OLAP & OLTP
- Storage Tools: S3/Azure Blob
- Warehousing Tools: Snowflake/BigQuery/Synapse

1. Programming

- Python, R, Java or Scala
- SQL
- Linux Commands

9. Infrastructure as Code

Kubernetes,
Terraforms & CI/CD

7. Big Data Technologies

- MapReduce
- Massive Parallel processing
- Distributed Systems
- Tools: Apache Hadoop, Apache Spark

5. Data Pipelines

- ETL Pipelines / Tools
- Scheduling & Orchestration
- Tools: Apache Airflow, Azure Data Factory, Dagster, Apache NiFi

3. Cloud Fundamentals

- Microsoft Azure
- Google Cloud
- AWS

2. Databases

- Relational & NoSQL Databases (MySQL, MongoDB, etc)
- Data modelling (Fact & Dimension Tables)

źródło:

https://x.com/tech_queen/status/1716380140001697996/photo/1



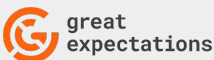
Narzędzia używane w codziennej pracy Data Engineera

INGEST	DATA LAKE	DATA VERSION CONTROL	COMPUTE ENGINES	PIPELINES	PRACTITIONER APPS		GOVERNANCE
Ingest Tech 	Object Storage 	 METADATA Metastores Open Table Formats 	Distributed Compute Analytics Engines 	Orchestration Data Quality / Observability 	MLOps End-to-End Modular Data Centric AI/ML ML/AI Observability & Monitoring Vector DBs Notebooks Analytics Workflows 		Data Catalog / Governance

źródło: <https://lakefs.io/wp-content/uploads/2024/05/SoDE24-state-of-data-engineering.png>



Narzędzia używane w codziennej pracy Data Engineera





Przykład przepływu pracy z danymi na podstawie Meta

- Instagram Analytics
- Recommending Content and Users
- User Based Advertising





Przykład przepływu pracy z danymi na podstawie otify

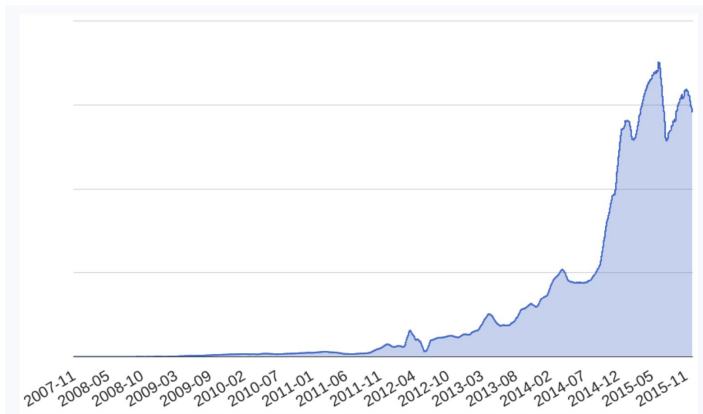


Figure 2. Amount of Spotify's Delivered Events over time

- 456 milionów miesięcznie aktywnych użytkowników (2023)
- 180 milionów subskrypcji
- 1 bilion zdarzeń na dzień (1,000,000,000,000)
- 38,000 zaplanowanych pipelines
- 220+ projektów ML
- 4 billion playlist stworzonych przez użytkowników
- 20,000 batch data pipelines zdefiniowanych w ponad 1,000 repozytoriach należących do ponad 300 zespołów

pipelines on-premise, operating the largest Hadoop cluster in Europe



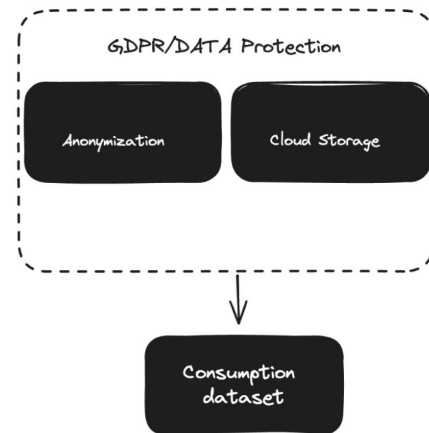
Cloud



Przykład przepływu pracy z danymi na podstawie Spotify

Ponad 100 inżynierów pracujących nad budową platformy danych Spotify

Datasets are encrypted with access control and retention set, as part of our compliance strategy

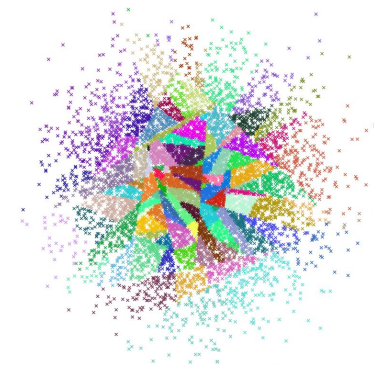


źródła: <https://engineering.atspotify.com/2024/04/data-platform-explained>, <https://spotify.github.io/#pytho>, <https://spotify.design/article/designing-data-science-tools-at-spotify>



Przykład przepływu pracy z danymi na podstawie otify

- Recommender Systems
- Audio analysis
- Discover Weekly, Wrapped annual
- Spotify's Fan Study



PEDALBOARD

źródła: <https://engineering.atspotify.com/2024/04/data-platform-explained>, <https://spotify.github.io/#pytho>,
<https://spotify.design/article/designing-data-science-tools-at-spotify>



Pytania i odpowiedzi



Anuszk:

Jak najczesciej optymalizuje sie zapytania SQL. Jakie najczesciej tworzy sie procedury i funkcje?



Magdalena:

Czy przebranżowienie się z Analityka Danych,
używającego na co dzień języka SQL oraz
czasami języka Python, to długi i ciężki
proces?



Konrad:

Jakie są narzędzia AI wspierające pracę
inżynierów danych?

Jakie są predykcje dla zawodu inżyniera
danych w ujęciu innowacji w obszarze AI?



Adam:

Czy doświadczenie w pracy z bazami danych SQL, w środowisku MATLAB, w programowaniu w języku C oraz w implementacji algorytmów sztucznej inteligencji (optymalizacji dyskretnej) może być przydatne w zawodzie Data Engineer?



Wioletta:

Jak powinno wyglądać portfolio osoby
aplikującej na stanowisko junior data
engineera?



Adam:

Czy będzie również poruszony temat
wykorzystania AI do wspomagania pracy
Inżyniera Danych?



Jakub:

Znalezienie pracy jako Junior Data Scientist jest dziś dużym wyzwaniem. Czy perspektywa dla Junior Data Engineera jest korzystniejsza? Jeśli tak to dlaczego? :)



PG

PG czy na kursie będzie np ćwiczenie request do API i zapis do bazy danych? na jaki poziom oceniliby Państwo zakres materiału oraz ćwiczeń w zakresie narzędzi wchodzących w skład kursu mid czy junior?

kurs **Data Engineer**

10 h

prework – szacowany
czas na przygotowanie do
kursu

111 h

zajęcia z trenerem
(zdalnie na żywo)

30 h

szacowany czas na
indywidualną pracę
między zajęciami

Technologie i narzędzia, które opanujesz na kursie Data Engineer



AWS



AWS S3



SQL



Kubernetes



Python



Pandas



NumPy



Polars

(REST API)

API



Hadoop



Spark



DynamoDB



Kafka



Amazon
Redshift



Apache
Airflow



Databricks



Snowflake



kurs Data Engineer

Moduł 0: Prework



Moduł 1: Wprowadzenie do Data Engineering



Moduł 2: Wprowadzenie do AWS



Moduł 3: Database i zaawansowany SQL:



Moduł 4: Zaawansowany Python dla Data Engineering



Moduł 5: Wprowadzenie do Big Data (PySpark)



Moduł 6: Extract, Transform, Load (ETL)



Moduł 7: Praca z bazami NoSQL (DynamoDB)



Moduł 8: Streaming processing (Apache Kafka)



Moduł 9: Data Warehouses



Moduł 10: Data Lakes



Moduł 11: Automatyzacja procesów (Apache Airflow)



Moduł 12: Databricks



Moduł 13: Snowflake



Moduł 14: Data Mesh





kurs Data Engineer

7400 zł ~~8900 zł~~

przy zgłoszeniu do 3 grudnia

start: 25 stycznia 2025

<https://infohareacademy.com/kurs/kurs-data-engineer/>

Kontakt

(+48) 530 100 686

kontakt@infoShareAcademy.com

