

Marcin Reed

Java Spark Developer

✉ java-spark-developer@przyklad.pl

☎ +48 600 202 626

📍 Olsztyn / hybrydowo

👤 Podsumowanie zawodowe

Inżynier platform danych budujący powtarzalne pipeline'y w Java Spark. Dbam o jakość danych, możliwość odtworzenia i koszty obliczeń.

📁 Doświadczenie zawodowe

- **Java Spark Developer**
Tidal Data Works 4/2021 - Obecny
 - Przepisałem zadania cech do Java Spark i przetwarzam 620 mln zdarzeń dziennie, skracając okno z 4 godzin do 76 minut.
 - Dodałem reguły jakości i partycje kwarantanny, zmniejszając błędy modeli z 5,4% do 0,7%.
 - Zaprojektowałem retry i zależności Airflow, oszczędzając zespołowi on-call 24 godziny analizy tygodniowo.
 - Utworzyłem 42 przypadki walidacyjne z 7 data scientistami, podnosząc śledzalność danych treningowych do 99,3%.
- **Programista Java**
SignalCraft Marketing 9/2017 - 3/2021
 - Przeniósłem 9 zadań MapReduce i Spark do EMR, zachowując kontrole wyjścia raportów marketingowych.
 - Dobrałem partycje i serializację, obniżając koszt obliczeń o 33% przy ukończeniu 98,8% zadań.
 - Wprowadziłem wersjonowanie schematów Hive, ograniczając awarie raportów o 58%.
 - Napisałem słownik danych i notebooki startowe, skracając uruchomienie pipeline'u dla 4 analityków o 45%.

<> Projekty

Jezioro cech zachowań klientów Powtarzalny pipeline zamieniający zdarzenia web, sklepowe i wsparcia w cechy dla modeli retencji.

Java
Spark
Hive
Airflow

🎓 Wykształcenie

Magister analityki

Uniwersytet Warmińsko-Mazurski

10/2014 - 6/2018

☆ Umiejętności

Umiejętności techniczne

Java, Apache Spark, Hadoop, Hive,
Airflow, AWS EMR

🇵🇵 **Języki**

Polski

Ojczysty

Angielski

Biegłość
zawodowa

🏆 **Certyfikaty** Cloudera Spark and
Hadoop Developer

Instytucja certyfikująca
2025