# Zajęcia 6

Zadanie 1

Flatten json - wczytaj 10 wybranych atrybutów z pliku brzydki.json (kolumna features).

Zadanie 2

Stwórz diagram draw.io pokazujący przykładowy proces ETL dla wybranego przez siebie przykładu do interview z Lab nr 5.

Podczas projektowania pokaż że zastosowałeś dobre praktyki z wykładów. (Wymyśl kilka transformacji, jakiś plik konfiguracyjny)

Zadanie 3

Poka Yoke

Napisz 5 metody, które mogą być użyte w Pipeline tak aby tyły odporne na błędy użytkownika, jak najbardziej „produkcyjnie”. Możesz użyć tego co już stworzyłeś i usprawnij rozwiązanie na bardziej odporne na błędy biorąc pod uwagę dobre praktyki.

Walidacja danych wejściowych:

Przed rozpoczęciem przetwarzania danych z czujników i systemów monitorujących produkcję żarówek, wykonujemy walidację danych wejściowych. Sprawdzamy, czy dane mają odpowiedni format, czy zawierają oczekiwane wartości, czy nie brakuje żadnych istotnych informacji.

Obsługa wyjątków:

W każdym kroku procesu ETL dodajemy mechanizmy obsługi wyjątków, które przechwytują i zarządzają błędami w sposób kontrolowany. Na przykład, jeśli dane z czujnika są niekompletne lub nieprawidłowe, obsługujemy ten błąd poprzez zapisanie go do logów i kontynuowanie przetwarzania z następnymi poprawnymi danymi.

Zapisywanie stanu pośredniego:

Regularnie zapisujemy stan pośredni procesu przetwarzania, aby w przypadku wystąpienia błędu można było wznowić przetwarzanie od ostatniego poprawnie ukończonego kroku. Na przykład, po każdym poprawnym kroku przetwarzania zapisujemy wyniki do bazy danych lub do plików w chmurze, aby można było je odtworzyć w przypadku potrzeby.

Monitorowanie i logowanie:

Dodajemy mechanizmy monitorowania i logowania, które umożliwią śledzenie postępu przetwarzania oraz rejestrowanie wszelkich błędów, ostrzeżeń i zdarzeń. Możemy monitorować parametry produkcji, takie jak temperatura, wilgotność, prąd itp., i rejestrować wszelkie anomalie lub odchylenia od normy.

Automatyczne testy jednostkowe i integracyjne:

Przed uruchomieniem procesu ETL, przeprowadzamy automatyczne testy jednostkowe i integracyjne, aby upewnić się, że wszystkie komponenty działają poprawnie i są odporne na różne scenariusze błędów. Testujemy poprawność przetwarzania danych, skuteczność algorytmów analizy anomalii oraz prawidłowość logowania i monitorowania.