# 

# Univerzita Jana Evangelisty Purkyně v Ústí nad Labem (UJEP) Přírodovědecká fakulta (PřF), Katedra informatiky (KI)

**OLAP a Data mining (KI/ODM)**

# 

ds

P

**Seminární práce**

**OLAP a ClickHouse**

Ondřej Švorc (F23209)

Aplikovaná informatika  
OLAP a Data mining (KI/ODM)

xx. xx. 2025

## Zkratky

DBMS = Database Management System

OLAP = Online Analytical Processing

## Volba DBMS s OLAP podporou

Pro tento projekt jsem zvolil produkt ClickHouse Cloud, především díky jednoduchosti použití ve srovnání s ostatními DBMS s OLAP podporou. Před zvolením ClickHouse Cloud jsem si krátce prohlédl a vyzkoušel Azure Analytics Service, Apache Druid a SSMS.

## Prerekvizity

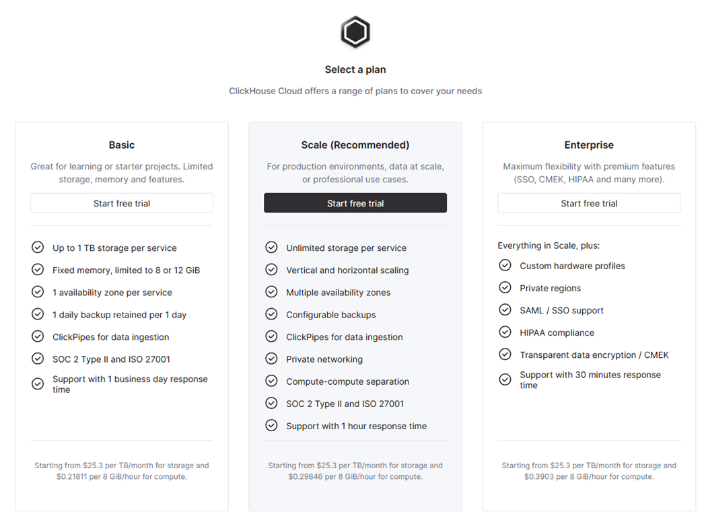
Díky tomu, že ClickHouse Cloud je webová aplikace, za jediné prerekvizity lze považovat:

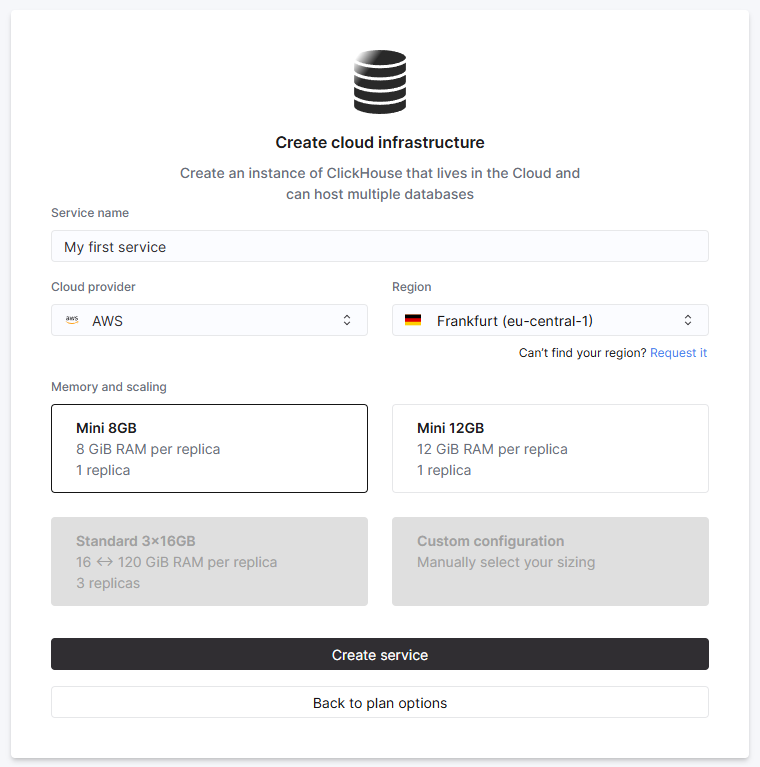
1. Založený účet
2. Zvolené předplatné
3. Vytvořená cloudová služba pro provoz

Osobně jsem zvolil možnost registrace pomocí **Google účtu** a předplatné **Basic**. Při vytváření cloudové služby jsem ponechal název **My first Service**, vybral poskytovatele **AWS**, lokaci ve **Frankfurtu nad Mohanem** (zatím nejbližší možná) a **8 GB RAM**.

Obsah obrázku text, snímek obrazovky, software, Operační systém

Obsah vygenerovaný umělou inteligencí může být nesprávný.





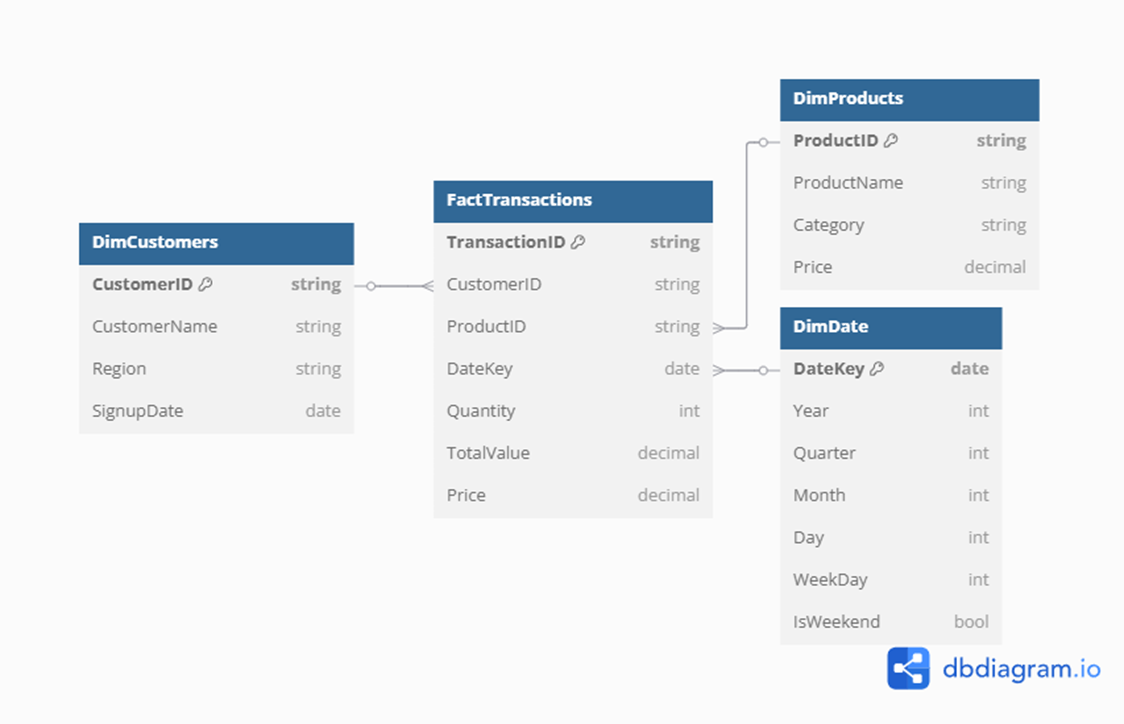
## Datová sada

Zvolil jsem datovou sadu ze stránky **Kaggle** fiktivního online obchodu **EverMart Online**.[[1]](#footnote-1)

Datová sada ve formátu CSV obsahuje data o zákaznících, produktech a provedených transakcí.

* Customers.csv
* Products.csv
* Transactions.csv

Transakce jsou duplicitně zapsána ve dvou souborech s mírně odlišnými formáty (Transactions.csv, EverMart\_Online\_Transactions.xlsx). CSV formát je vhodnější pro strojové zpracování, a proto nemá smysl XLSX soubor zahrnovat.



## Čištění datové sady

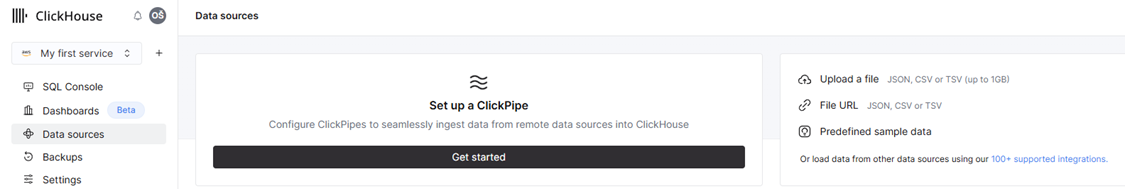
Datová sada byla již po stažení vyčištěná, proto není nutné provádět její čištění.

## Datová kostka

xxx.

## Import datové sady

x



Obsah obrázku snímek obrazovky, text

Obsah vygenerovaný umělou inteligencí může být nesprávný.

## Zdroje

1. WAMBLES, Chad. ECommerce transactions. *Kaggle* [online]. 3 January 2025. Available from: <https://www.kaggle.com/datasets/chadwambles/ecommerce-transactions> [↑](#footnote-ref-1)