TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Mechatronics, Informatics
and Interdisciplinary Studies

# TECHNOLOGIE PRO BIG DATA
# CVIČENÍ VI.
# APACHE SPARK II

*Lukáš Matějů*

*7.11.2023 | TPB*

# DNEŠNÍ CVIČENÍ

1. zjistěte průměrný počet kamarádů podle věku
   - k dispozici máte soubory *fakefriends-header.csv* (data) a *spark-sql-cv.py* (kód)
   - vhodně rozšiřte kód
     - pro řešení můžete použít SQL dotazy nebo funkce přímo nad DataFrame

```
userID,name,age,friends
0,Will,33,385
1,Jean-Luc,26,2
2,Hugh,55,221
3,Deanna,40,465
```

```
+---+-----------+
|age|friends_avg|
+---+-----------+
| 18|     343.38|
| 19|     213.27|
| 20|      165.0|
| 21|     350.88|
```

2. zjistěte celkovou výši objednávek pro každého zákazníka
   - k dispozici máte soubor *customer-orders.csv* (data)
   - vytvořte skript vracející pro každého zákazníka celkovou utracenou částku
     - skript pojmenujte *total-spent-by-customer-dataframes.py*
     - řešení postavte na DataFramech
   - BONUS: celkovou částku zaokrouhlete na dvě desetinná místa a uložte do sloupce total_spent, výsledný seznam vraťte setříděný podle celkové částky
     - tip: funkce agg

```
|     39|     6193.11|
|     73|      6206.2|
|     68|     6375.45|
+-------+------------+
```

# DNEŠNÍ CVIČENÍ

3. nalezněte všechny superhrdiny, kteří mají nejméně propojení
   - k dispozici máte soubory marvel-graph.txt, marvel-names.txt a heroes.py
   - vhodně upravte a rozšiřte kód
     - uvažujte 1 propojení jako minimum
     - skript pojmenujte most-obscure-superheroes.py a řešte pomocí DataFramů
   - BONUS: řešení navrhněte bez předpokladu minimálního počtu propojení 1
     - minimum propojení zjistěte algoritmicky
     - hrdinové bez propojení se nepočítají

| name |
|---|
| BERSERKER II |
| BLARE/ |
| MARVEL BOY II/MARTIN |
| MARVEL BOY/MARTIN BU |
| GIURESCU, RADU |
| CLUMSY FOULUP |
| FENRIS |
| RANDAK |
| SHARKSKIN |
| CALLAHAN, DANNY |
| DEATHCHARGE |
| RUNE |
| SEA LEOPARD |
| RED WOLF II |
| ZANTOR |
| JOHNSON, LYNDON BAIN |
| LUNATIK II |
| KULL |
| GERVASE, LADY ALYSSA |