

# Subiect preselecție internship

Cloud Security Analytics Team - Bitdefender

Să se implementeze în limbajul Python un script care îndeplinește următoarele cerințe:

1. Atunci când primește în linia de comandă parametrul **scan**:
  - a. Culege următoarele informații despre fișierele prezente în directorul de lucru curent (în *current directory*), construind pentru acestea un dicționar al cărui chei îl reprezintă denumirile fișierelor analizate:
    - i. **FileSize**: dimensiunea fișierului, exprimată în octeți
    - ii. **Changed**: data și ora ultimei modificări a fișierului, în formatul textual (string) *YYYY/MM/DD-hh:mm*, unde:
      1. *YYYY* reprezintă, pe 4 cifre zecimale, anul (de exemplu 2019)
      2. *MM* este luna din an, 01 însemnând Ianuarie iar 12 reprezentând Decembrie
      3. *DD*, cu valori între 1 și 28/29/30/31, reprezintă ziua din lună
      4. *hh* și *mm* sunt ora (00-23) și minutul (00-59)
    - iii. **Hidden**: marchează fișierele care au atributul de fișier (*file attribute*) *hidden* setat; valoarea va fi 1 dacă fișierul este ascuns sau 0 în caz contrar
    - iv. **ReadOnly**: valoare 0 sau 1 care specifică dacă fișierul este protejat contra operațiilor de scriere (valoare 1) sau dacă scrierea este permisă (valoare 0), conform atributelor de fișier aferente.
  - b. Informațiile vor fi stocate într-un fișier din directorul curent, în format **JSON**, fișier numit *cache.json*
2. Când se primește în linia de comandă parametrul **query**:
  - a. Generează conținutul fișierului *cache.json*, în cazul în care acesta nu există deja, întocmai după cum se procedează la punctul 1.
  - b. Încarcă informațiile prezente în *cache.json* într-o structură de date echivalentă celei construite la punctul 1.a.
  - c. Afișează pe consolă următoarele informații despre fișierele descrise în *cache.json*:
    - i. Numărul total de fișiere pentru care există informații în *cache.json*, sub forma **Number of files: 422**
    - ii. Numele celor mai mari 5 fișiere (după dimensiune) în ordine crescătoare a dimensiunii lor, mesajul afișat fiind formatat precum **Top 5 largest files: a.txt, b.bin, c.c, arc.zip, 12.exe**
    - iii. Procentul fișierelor ascunse (raportat la numărul total de fișiere) după modelul **3% are hidden**
    - iv. Procentul fișierelor protejate contra scrierii, de exemplu **16% are read-only**
    - v. Numărul fișierelor care au suferit modificări în fiecare lună în care cel puțin un fișier a fost modificat, datele afișate pentru o astfel de lună fiind formate după modelul **2019/03: 4 modified files**, datele pentru mai multe luni fiind afișate pe linii separate (câte o lună per linie afișată).