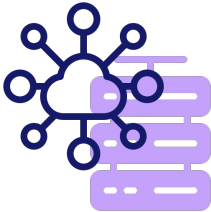




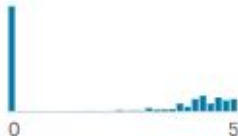

UNIVERSITÀ TELEMATICA  
INTERNAZIONALE UNINETTUNO

# **Progetto:** ***Piattaforme per i Big Data***

**Alessio Cimino - a.a. 2024**



# Google PlayStore Android App Data (2.3 Million+ App Data)

▲ App Name	▲ App Id	▲ Category	# Rating	▲ Installs	✓ Free	▲ Size		
Name of the app	Package name	App category	Average rating	Approximate install count	Whether app is Free or Paid	Size of application package		
2177946 unique values	2312944 unique values	Education	10%		100+ 19%		Varies with device	
		Music & Audio	7%					1,000+ 17%
		Other (1916948)	83%					Other (1471377) 64%
Gakondo	com.ishakwe.gakondo	Adventure	8.8	18+	True	18M		
Ampere Battery Info	com.webserveis.batteryinfo	Tools	4.4	5,000+	True	2.9M		
Vibook	com.doantiepvien.crm	Productivity	8.8	58+	True	3.7M		

**App Name**

**Rating Count**

**Free**

**Minimum Android**

**Released**

**Ad Supported**

**App Id**

**Installs**

**Price**

**Developer Id**

**Privacy Policy**

**InAppPurchases**

**Category**

**Minimum Installs**

**Currency**

**Developer Website**

**Last Updated**

**Editor Choice**

**Rating**

**Maximum Installs**

**Size**

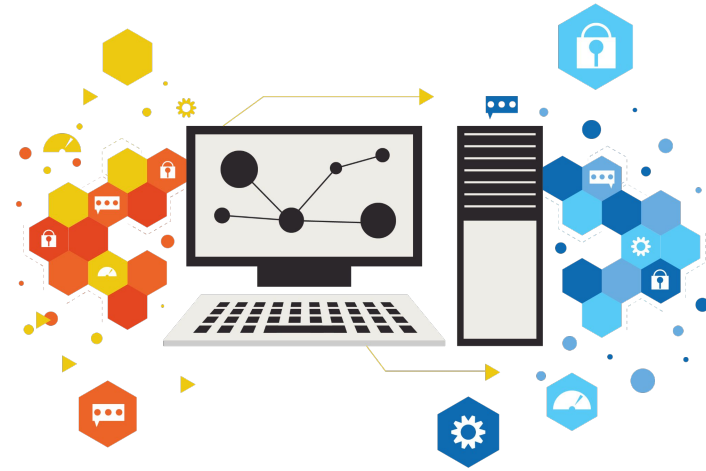
**Developer Email**

**Content Rating**



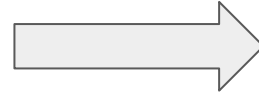
# Funzionalità

- *INFO App*
- *INFO Content Rating and Price*
- *INFO Developer x App*
- *INFO App x developer*
- *TOP Download*
- *TOP Rating*
- *TOP Paid x Download*
- *TOP Paid x Price*
- *TOP Release Update*
- *TOP Pegi 18*
- *TOP Best Developers*



# Funzionalità

- *INFO App*
- *INFO Content Rating and Price*
- *INFO Developer x App*
- *INFO App x developer*
- *TOP Download*
- *TOP Rating*
- *TOP Paid x Download*
- *TOP Paid x Price*
- *TOP Release Update*
- *TOP Pegi 18*
- *TOP Best Developers*



OUTPUT on  
monitor



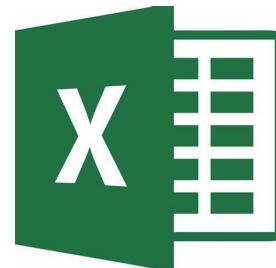
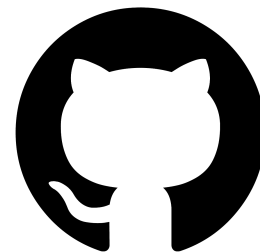
OUTPUT  
.CSV





## INFRASTRUTTURA

- *Github*: dove si trova depositato il codice dell'applicazione per poter lavorare in remoto;
- *PySpark*: usato come modello di programmazione per ottenere dati dei .csv e processare i risultati;
- *Google Cloud*: per poter comparare i risultati rispetto all'esecuzione in locale;
- *Microsoft Excel*: per poter mostrare i grafici verificando l'esattezza dei risultati ottenuti.



# Fasi del codice utilizzato



## Scelta della categoria:

```
[?] Choose the Categories:
[ ] SelectAll
[X] Action
[ ] Adventure
> [ ] Arcade
[ ] Art
[X] Audio
[X] Auto
[ ] Beauty
[ ] Board
[ ] Books
[ ] Business
[ ] Card
[ ] Casino

How many occurrences do you want printed? (max 1000)
→ 100
```

# RENDIMENTO

Tested on:

**LOCAL:** Oracle Virtual Box (free version) on Laptop:

*CPU AMD 3020e (2C / 2T, 1.2 / 2.6GHz, 1MB L2 / 4MB L3)*

*Memoria RAM: 8GB SO-DIMM DDR4-2400*

*SSD de 256GB M.2 2242 PCIe NVMe 3.0x2*

*Scheda grafica integrata AMD Radeon Graphics*

*Sistema operativo: Windows 10*

*Kali ver. 6.0.34 Debian 64-bit*

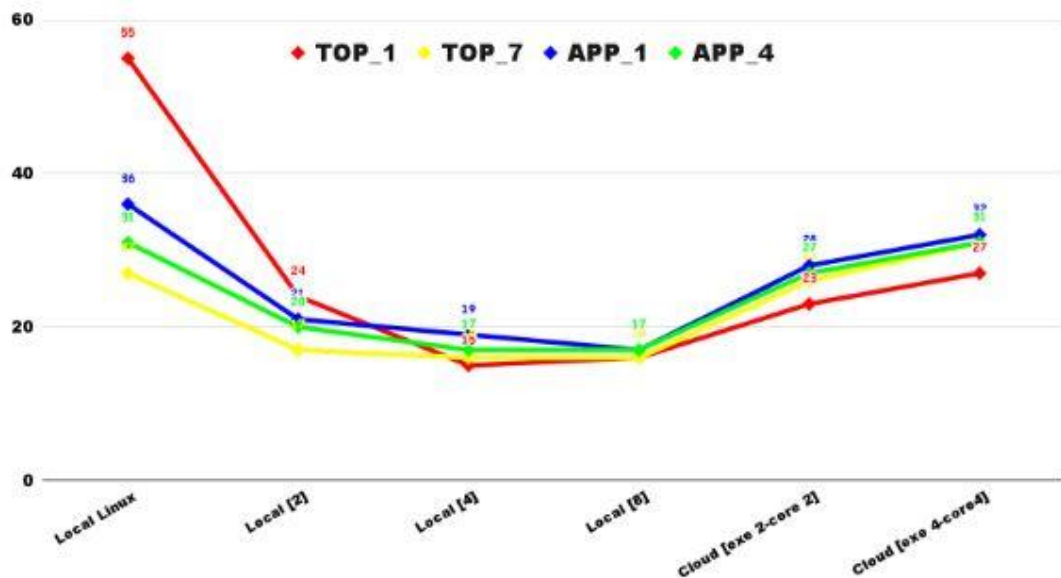
**CLOUD:** Google Cloud

num executors 2/executive cores 4

num executors 4/executive cores 4



## SpeedTest



# CONCLUSIONI

Il progetto rappresenta una solida base per lo sviluppo di future versioni puntando ad un miglioramento della grafica e dell'interfaccia utente, risultando utile sia per gli utenti interessati alle tendenze di mercato che per gli sviluppatori che cercano di comprendere l'andamento e l'analisi del settore delle applicazioni

