

Nombre: Diogo Donadoni Santos

1. Explicar en qué contexto se ha recolectado la información.

Explique por qué el sitio web elegido proporciona dicha información.

La información fue recolectada de el sitio web (<https://www.top500.org/list/2019/06/?page={}>) para listar los 500 computadores mas potentes.

2. Definir un título para el dataset. Elegir un título que sea descriptivo.

Top500Computers

3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).

Extrae el listado de los 500 computadores mas potentes (<https://www.top500.org/list/2019/06/?page={}>)

4. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	DOE/SC/Oak Ridge National LaboratoryUnited States	Summit- IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail 2,282,544	122,300.0	187,659.3	8,806	
2	National Supercomputing Center in WuxiChina	Sunway TaihuLight- Sunway MPP, Sunway SW26010 260C 1.45GHz, SunwayNRCPC	10,649,600	93,014.6	125,435.9	
3	DOE/NNSA/LLNLUnited States	Sierra- IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail N 1,572,480	71,610.0	119,193.6		
4	National Super Computer Center in GuangzhouChina	Tianhe-2A- TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000N4,981,760	61,444.5	100,678.7	18,482	
5	National Institute of Advanced Industrial Science and Technology (AIST)Japan	AI Bridging Cloud Infrastructure (ABCI)- PRIMERGY CX2590 M4, Xeon Gold 6148 20C 2.4GHz, N 391,680	19,880.0	32,576.6	1,649	
6	Swiss National Supercomputing Centre (CSCS)Switzerland	Piz Daint- Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100Cray 361,760	19,590.0	25,326.3	2,272	
7	DOE/SC/Oak Ridge National LaboratoryUnited States	Titan- Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20xCray Inc. 560,640	17,590.0	27,112.5	8,209	
8	DOE/NNSA/LLNLUnited States	Sequoia- BlueGene/Q, Power BQC 16C 1.60 GHz, CustomIBM	1,572,864	17,173.2	20,132.7	
9	DOE/NNSA/LANL/SLNLUnited States	Trinity- Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnectCray Inc.	979,968	14,137.3	43,902.6	
					3,844	

5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

Rank - Posición en el rank de los 500 computadores mas potentes

System - Información del sistema

Cores - Cantidad de cores en el sistema

Rmax (TFlop/s) - La maxima performance alcanzada del sistema

Rpeak (TFlop/s) - La maxima performance teorica del sistema

Power (kW) - Consumo de energia

6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).

<https://www.top500.org> (Por fornecer el conjunto de datos y dejar el sitio abierto para web scrapping.)

<https://www.top500.org/robots.txt>

User-agent: *

Allow: /

Host: www.top500.org

7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.

Interes por conocer los computadores mas potentes y donde estan ubicados

8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:

Released Under CC0: Public Domain License

Estoy seleccionando esta licencia porque la información es publica y puede ser usada libremente.

9. Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

<https://github.com/diogodonadoni/data/blob/master/Top500Computers.py>

10. Dataset. Presentar el dataset en formato CSV

<https://github.com/diogodonadoni/data/blob/master/Top500Computers.csv>

Contribuciones	Firma
Investigación previa	Diogo Donadoni Santos
Redacción de las respuestas	Diogo Donadoni Santos
Desarrollo código	Diogo Donadoni Santos

