



Cloud Computing and Big Data

Salvo Nicotra

Who Are We?



About me

- Italy/Catania
- Computer Science (proud Scienze dell'Informazione)
- Work together on data!

Something about you

- Where are you from?
- What is your first-level degree?
- What are your expectations for this course?

Note

Reach me on Teams  or by email 
salvatore.nicotra1@unict.it

Topics

Big Data

Literate Programming

Literate programming: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.

Donald Knuth (1984)



Source

Hadoop

Spark

Apache Spark™ is a unified analytics engine for large-scale data processing.

<https://spark.apache.org/>

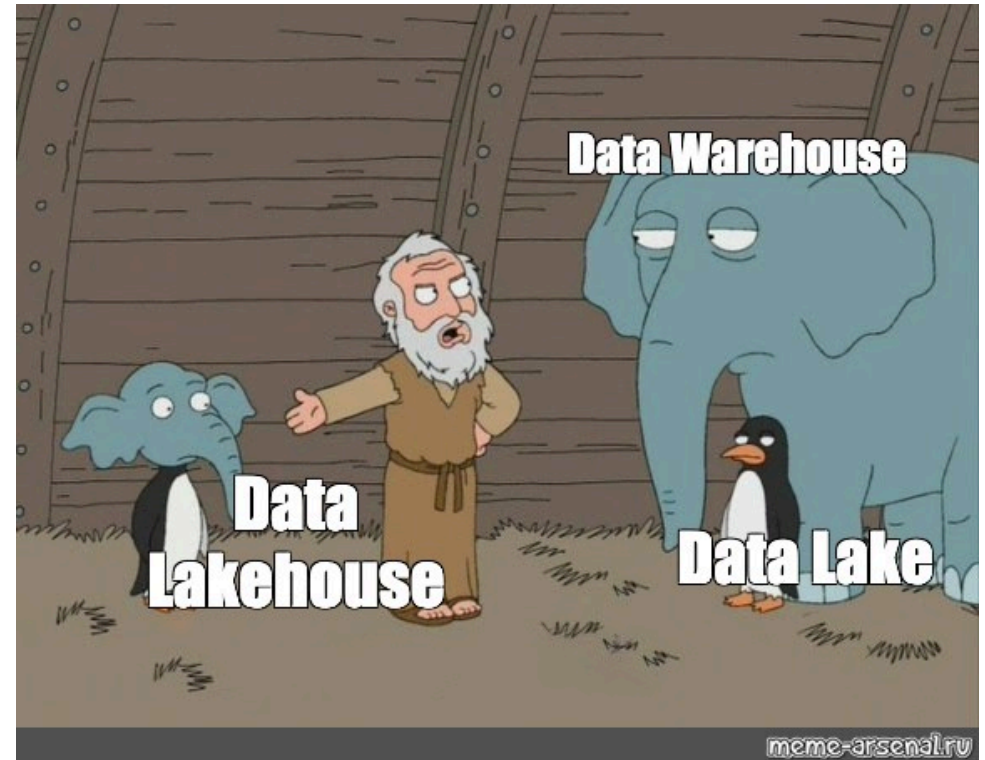


NicsMeme

Data (Base|Warehouses|Lakes)

A data lake is a system or repository of data stored in its natural/raw format, usually object blobs or files. A data lake is usually a single store of data including raw copies of source system data, sensor data, social data etc., and transformed data used for tasks such as reporting, visualization, advanced analytics and machine learning.

https://en.wikipedia.org/wiki/Data_lake



<https://meme-arsenal.com/create/meme/7205499>

Modern Data Platform (Data Catalogs)

A few years ago, data would primarily be consumed by the IT team in an organization. However, today data teams are more diverse than ever — data engineers, analysts, analytics engineers, data scientists, product managers, business analysts, citizen data scientists, and more. Each of these people have their own favorite and equally diverse data tools, everything from SQL, Looker, and Jupyter to Python, Tableau, dbt, and R.

<https://towardsdatascience.com/data-catalog-3-0-modern-metadata-for-the-modern-data-stack-ec621f593dcf>



https://twitter.com/castordoc_data

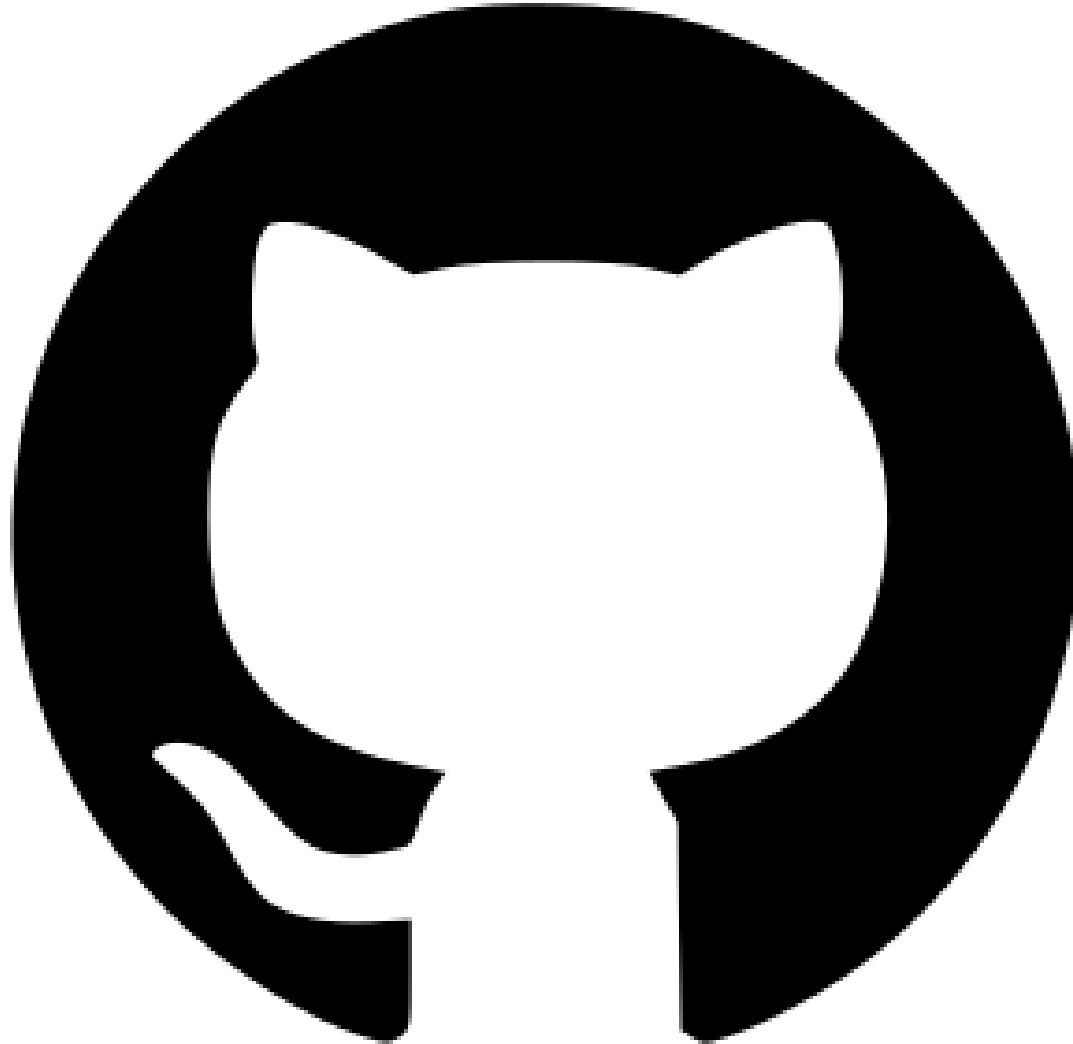
Tools

Teams



CCBD 2025

Github Organization



<https://github.com/unict-cloud-bigdata>

Cloud



Google Cloud

<https://cloud.google.com/>

aws



<https://aws.amazon.com/>



<https://azure.microsoft.com/>

Data Bricks



databricks

<https://www.databricks.com/>