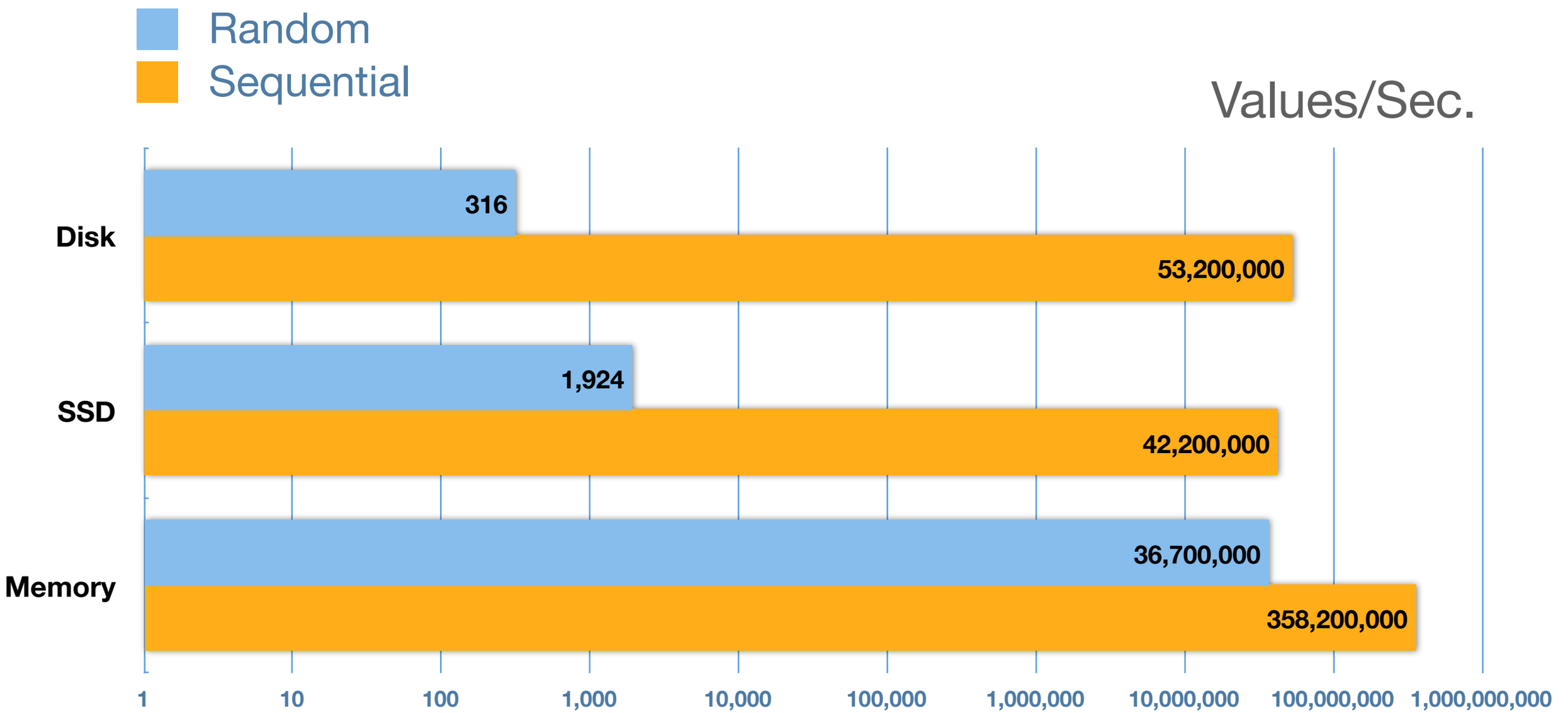
Datameer

Horizontal Virtualization On Commodity Hardware

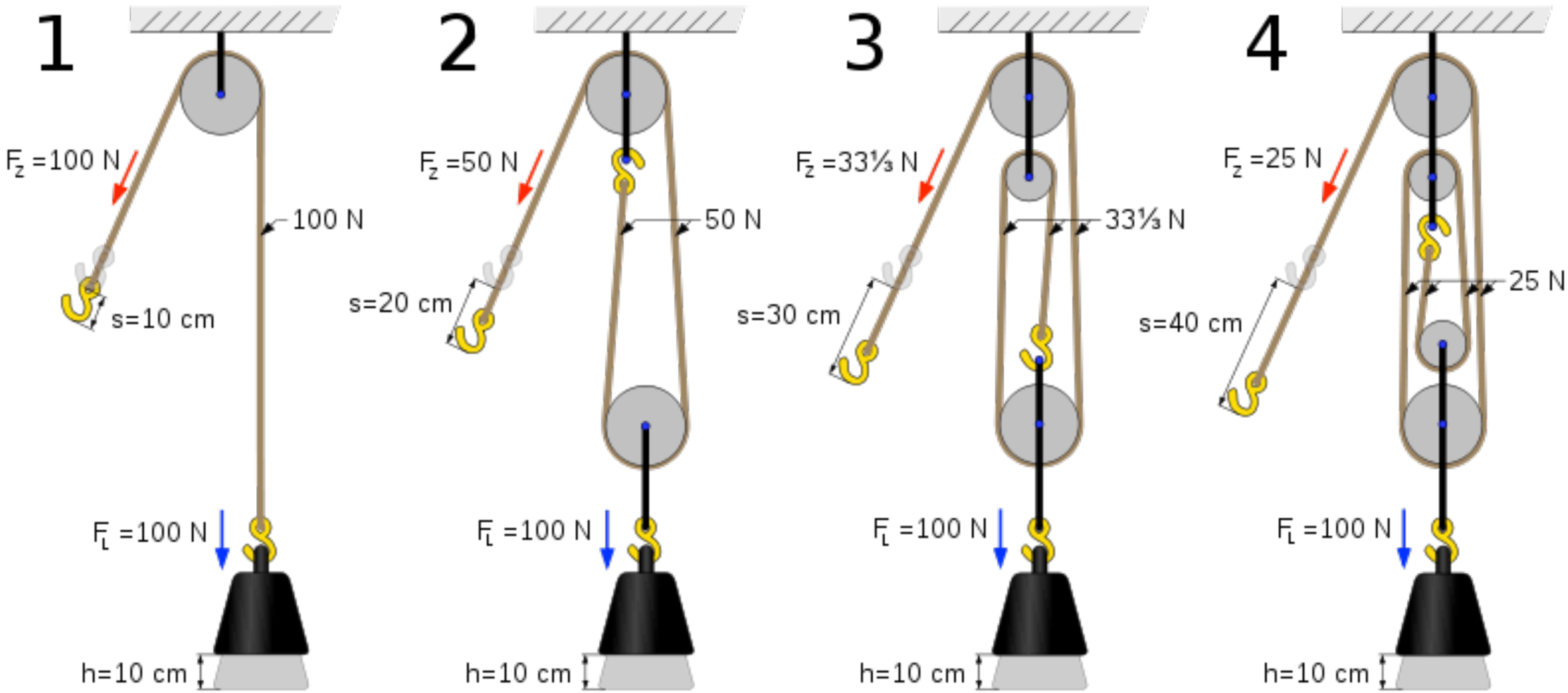
Stefan Groschupf
sg@datameer.com

Laws Of Physics



Adam Jacobs
The Pathologies of Big Data

Laws Of Physics



Hadoop(mergesort)

DB(b-tree/index)

Thesis, Antithesis, Synthesis

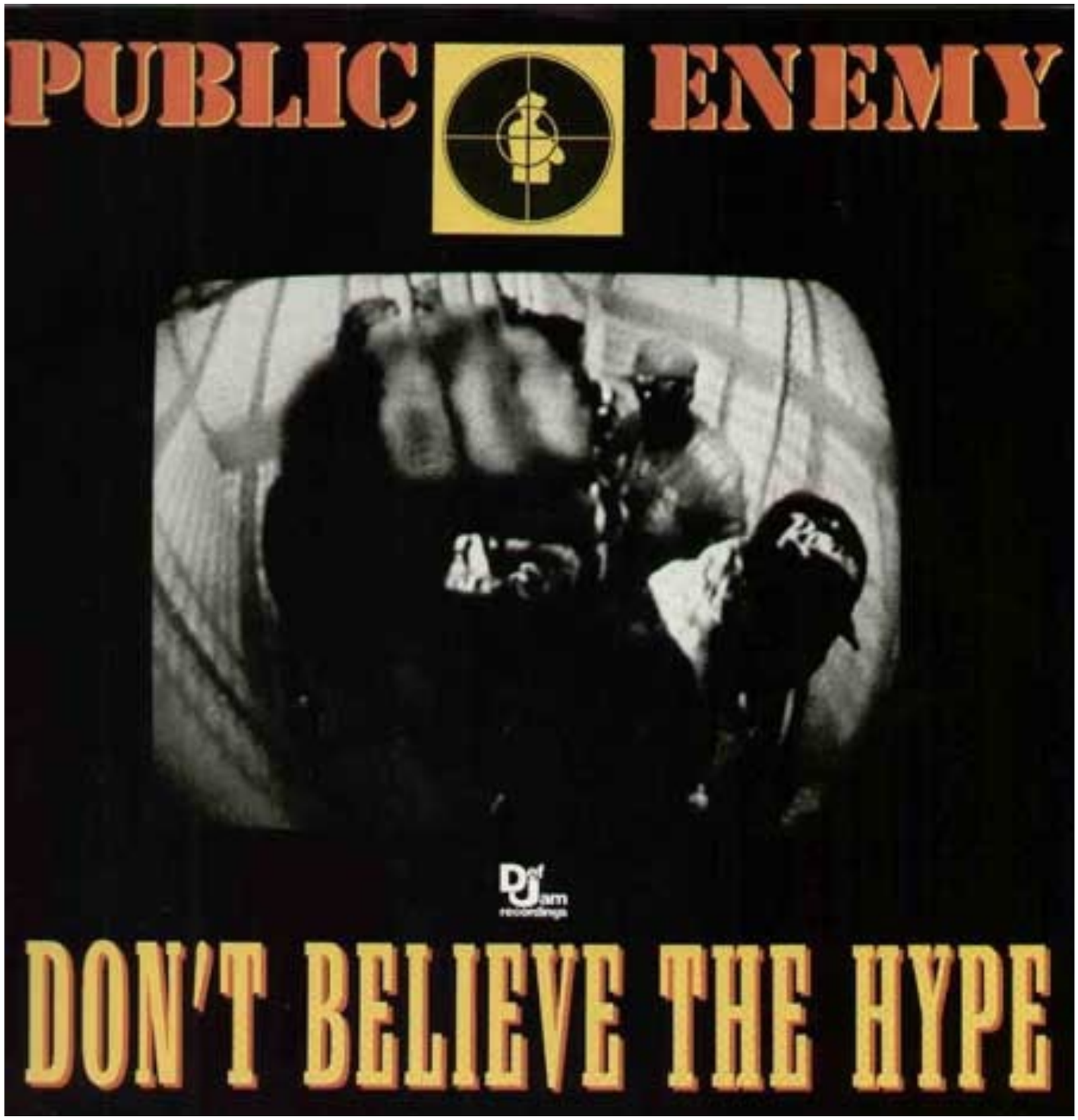
Thesis:
SQL/ACID solves all you data problems.

Antithesis:
Nobody needs SQL/ACID, let's throw it all out.

Synthesis:
By carefully considering data integrity constraints we can find a more optimal data management solution for a particular problem.

voidpointer:
http://news.ycombinator.com/item?id=1163516

PUBLIC ENEMY



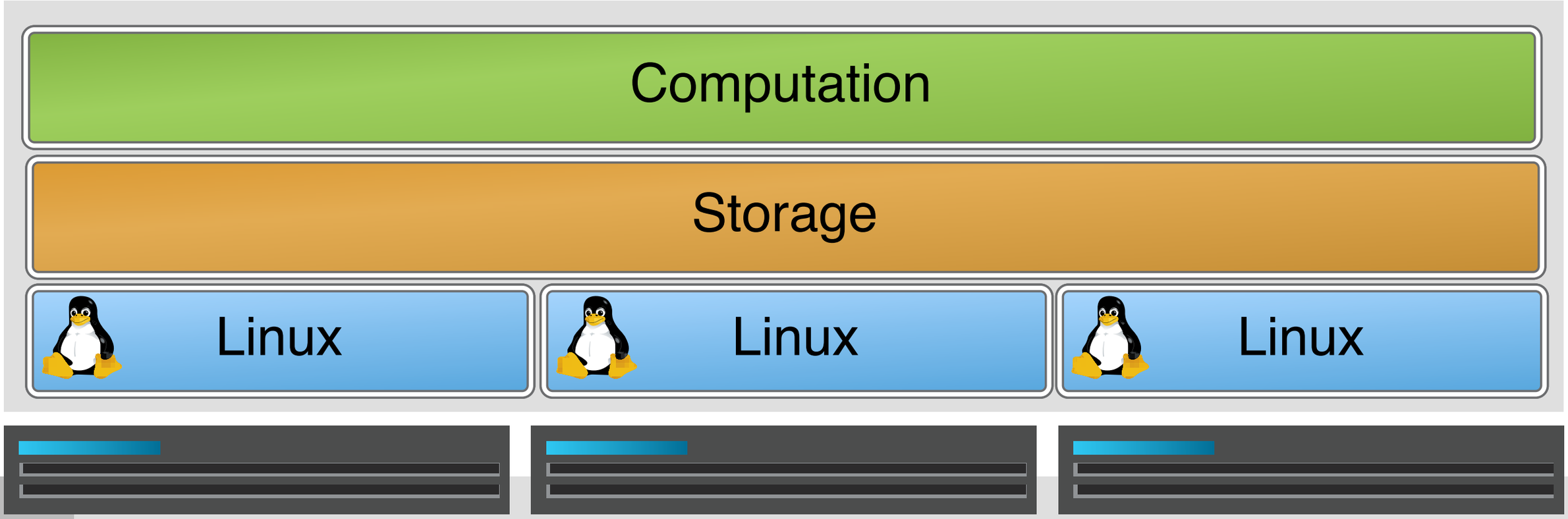
DON'T BELIEVE THE HYPE

Considerations

- Keep it simple!
  => Simple to distribute.
- What is really the problem to solve?
  - Serve Data or Analyze Data?
- Sequential Write/Read performs best.
  - Data aggregation challenges.
- Index only if really necessary.
  - For data serving.

Distribute on low cost HW

- Distributed Storage, Computation.
- Distributed Index.
- Horizontal Virtual Operating System.
- Open Source Platform.



Meet the ...

www.datameer.com

Free for .edu!

Street Cred

- Long time open source contributor

nutch hadoop katta Bixo

http://github.com/sgroschupf/zkclient

http://github.com/sgroschupf/aws-tasks

NOKIA Connecting People EMI