It's Not Where Your Data Is, It's How It Got There

Gala Yadgar
Roman Shor
Eitan Yaakobi
Assaf Schuster

Computer Science
Technion
In The Past 30 Years...

• Computers looked like this:

• Storage looked like this:
  ➢ Sequential = good
  ➢ Random = bad
Where Is The Data?

- Where we wrote it
In The Last 10 Years...

• Computers look like this:

• Storage looks like this:
  - Read = fast
  - Write = pretty fast (usually)
  - Erase = really slow (when?)
Where Is The Data?

• Could be anywhere...
Why Do We Care?

We care **how** data moves (not **where**)

- Performance
- Durability

For example:

- **Write Amplification (WA)** = \( \frac{\text{user writes} + \text{internal writes}}{\text{user writes}} \)
- Larger overprovisioning \(\rightarrow\) lower WA \(\rightarrow\) less erasures
Example

- Write amplification is an overall measure
- How can we tell what caused it?
SSDPlayer

One video is worth a thousand histograms
SSDPlayer Design

• General
  ➢ No timing

• Simple
  ➢ Appearance
  ➢ Input
  ➢ Features

• Flexible

• Java open source project
Now In Color

Skewed workload with temperature tags
Now For Real

Visualization mode: your own data and platform
What Else?

• More flash
  - Parity and striping
  - Flash based caches
  - Bit error rate
  - ...

• More layout
  - Content based optimizations: deduplication, compression...
  - File systems and databases: snapshots, versions, clones...
  - Anything that moves: shingled magnetic recording, log structured file systems...
Now You

- Download 🔄
- Play 🎬
- Tell us what you learned

http://www.cs.technion.ac.il/~gala/SSDPlayer/