Economics of Information Storage: The Value in Storing the Long Tail

James Hughes

MSST

2019 34th Symposium on Mass Storage Systems and Technologies

20-24 May 2019 • Santa Clara, California, USA

1975

History

- **Density has grown 36%/yr:**
  - 1956: 2 kb/in^2
  - 2005: 100 gb/in^2
- **Efficiency (B/$) grew 51%/yr:**
  - 1974: 200 MB disk drive price $450 k[1]
  - 2018: 10 TB Seagate disk price $300
- **Performance grew 2%/yr:**
  - 1974: 26 Op/s
  - 2018: 62 Op/s
- The market has consumed billions of these devices

[1] inflation adjusted

Questions

- **How was this possible?**
- **How did this happen?**
- **Will it continue?**
- **Will it happen for other classes of data?**

We show that the answers are

- **How was this possible?** The Long Tail
- **How did this happen?** Jevon's Paradox
- **Will it continue?** Yes
- **Will it happen for other classes of data?** Yes

Jevon's Paradox

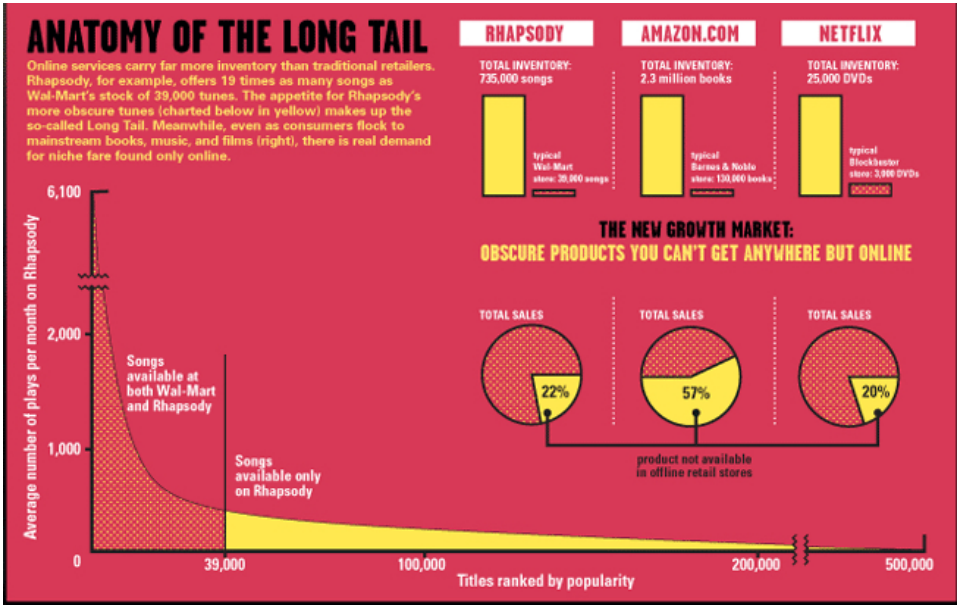
In economics, the Jevons paradox occurs when efficiency of a resource increases, but the rate of consumption of that resource rises.

- In 1865, he observed that technological improvements that *increased efficiency* of coal-use led to the *increased consumption* of coal in a wide range of industries.

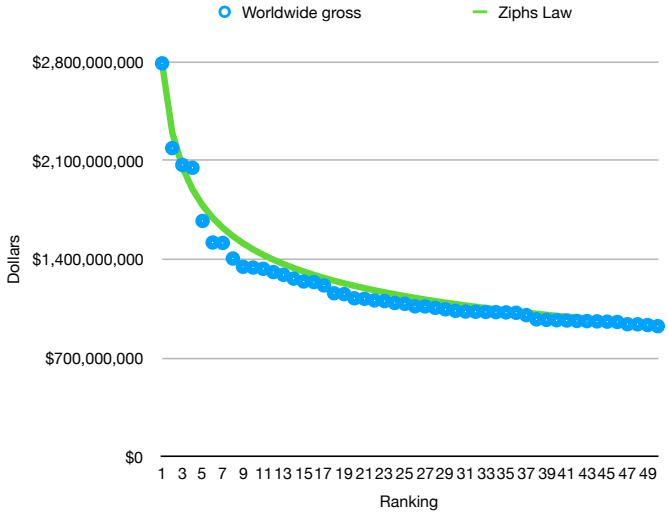
Table of contents

The Long Tail

The Long Tail



Ziph's Law vs. Movie revenue



Ziph's Law

The probability of the x entry being chosen.

$$P(x) = Cx^{-\alpha}$$

Where α is the decay rate and C is a value to make PDF sum to 1.
We calculate the revenue to be the probability of use $P(x)$ times the price v.

$$v_x = vCx^{-\alpha}$$

$\alpha = -0.278$ and $vC = \$2.8\text{B}$

Curating physical artifacts

"Select, organize, and look after the items in (a collection or exhibition)"

- Museums, Libraries. 3000yrs of history
  - Select
  - Preserve
  - Present
- Value from
  - The collection
  - The presentation

$$V = \sum_{i}^{n} v_i$$

Select/Ingest

Acquire the stuff

- Physical Aritifacts
  - "Things", books, art
- Digital Artifacts
  - objects, BLOBs, Collisions from LHC

The value of the items effect how fast the value of the collection grows, not the value of the already collected stuff.

Preserve

Ensure the stuff stays safe

- Physical Aritifacts
  - Warehouse, heat, lighting, people, maintenance, security
  - Linear to the warehouse size.
- Digital Artifacts
  - Datacenter, power, cooling, people, maintenance, security
  - Linear to the storage system size (point in time)

Cost of preserving the artifacts is linear to the storage space it holds and keeping the stuff safe.

Present

- Physical Artifacts
  - Create an exhibition, let public pay to see
  - Sell items
- Digital Artifacts
  - Present the data to the paying customer
  - Presenting faster can allow more revenue to be achieved on the same content value.

Acquiring and preserving are costs. Presenting is where value is realized.

Information

- Amount of information to store
- Value of information
- Value of a collection of information
- Value of a storage system as storage efficiency increases

Amount of information

- Eddington number, N_{edd} argues that there are 10^{80} protons in the universe.
- Philosophers argue we could indeed be living in a simulation and there could be an infinite number of simulations.

Value of information

- Objective value: What has been paid
- Subjective value: What might it is worth to a person

Objective value

General agreed upon method of

- Physical Artifacts
  - Assessment of a house
  - base price for an auction.
- Digital Artifacts
  - Movies streamed
  - Files accessed

An agreed upon value that other assessors would agree with.

Subjective value

Personal worth or "bet" of future value

- Physical Artifacts
  - Houses near family members
  - Value of Marvel Comic collectiion
  - Bidding value up at auction
- Digital Artifacts
  - Family photos
  - Backup of hard drive
- Value above the objective value for personal reasons > 0
- Objective value is lower bounds of value

Value of a collection of information

$$
\begin{aligned}
V &= \sum_{x=1}^{n} vP(x) \\
&= \sum_{x=1}^{n} vCx^{-\alpha} \\
&= vC \sum_{x=1}^{n} x^{-\alpha} \\
&= vCH_n \\
&\approx vC \log(n)
\end{aligned} \tag{1}
$$

Objective value of a storage system as storage efficiency increases

New storage system value $V \rightarrow V'$ if the storage devices can store 50% more objects for the same price, from $n = 1 \times 10^9 \rightarrow n' = 1.5 \times 10^9$

$$\begin{aligned} \frac{V'}{V} &= \frac{vC \log n'}{vC \log n} \\ &= \frac{\log n'}{\log n} \\ &\approxeq 1.0196 \end{aligned} \tag{2}$$

Doubling the efficiency adds to the long tail

- 2% to the value to the storage system.
- 2% to the access rate to the storage system.

History

- 50% CAGR effeciency increase
- 2% CAGR performance increase

What about other media?

- Nothing in this analysis was predicated on media type.
- Efficiency MB/$ is the key criteria
- Efficiency dominates until there are two classes with the same MB/$.
  - Has happened with 2.5" disks.
  - Could happen with Flash and Persistent RAM.

Conclusion

Reality is more complex, but the rules:

The increase in value and utilization of a storage system as the capacity increases is the ratio of the logs of the stored objects.

There will always be more lower value data to store

Stored information will continue to grow as device efficiency continues to grow

Questions?

Baskin Engineering
UC SANTA CRUZ

ssrc
Storage Systems Research Center

Center for Research in Storage Systems

Thank You

Baskin Engineering
UC SANTA CRUZ

ssrc
Storage Systems Research Center

Center for Research in Storage Systems