
Comprendre et modéliser l'empreinte environnementale du stockage de données : complexité des usages et limites des approches actuelles

■

Adrien Berthelot - Senior Research Associate in Sustainable ICT

■

Green Days 2026
10.03.2026 - Sophia Antipolis, France

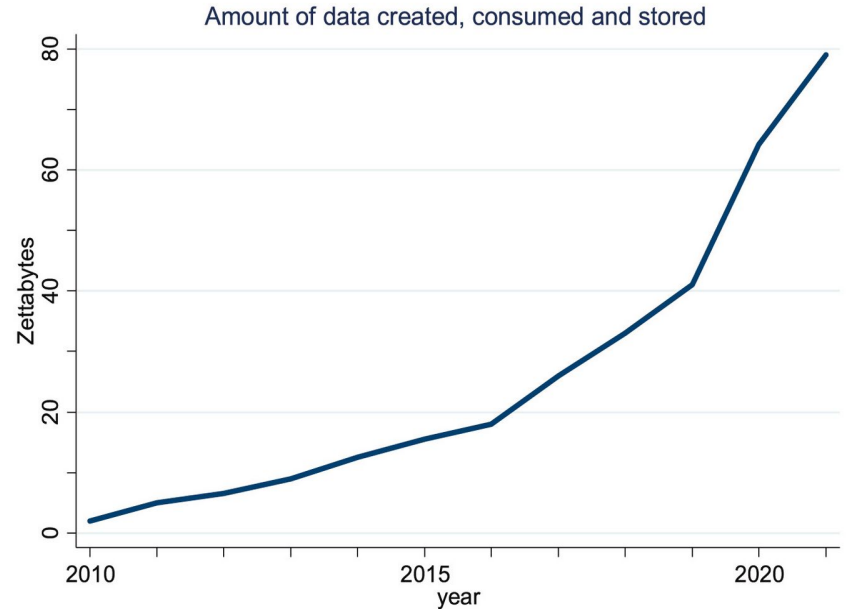


Introduction

- What is the cost of a Gigabyte ?
- Naive but important as data storage is one of the key features of ICT with computing and communicating

Introduction

- What is the cost of a Gigabyte ?
- Naive but important as data storage is one of the key features of ICT with computing and communicating
- According to IDC's report, the global size of the datasphere is or will be:
 - ~45 ZB in 2019
 - ... ~129 ZB in 2023
 - ... ~ 175 ZB in 2025
 - ... to ~393.9 zettabytes (ZB) in 2028



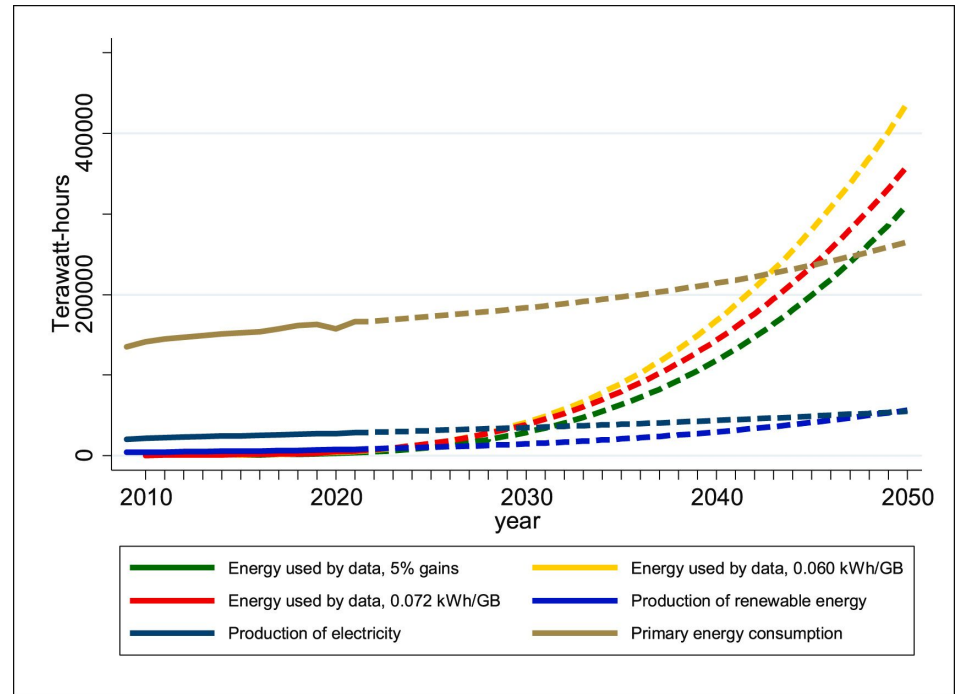
A "ZB" most commonly stands for a zettabyte, a unit of digital information equal to 10^{21} bytes or 100,000,000 GB

Background

- Some confusion and contradiction
- A very sparkled research landscape:
 - From engineering new storage media ...
 - ... to discussing issue around “dark data”...
 - ... through energy efficiency of Cloud storage and database

Background

- Some confusion and contradiction
- A very sparkled research landscape:
 - From engineering new storage media ...
 - ... to discussing issue around “dark data”...
 - ... through energy efficiency of Cloud storage and database



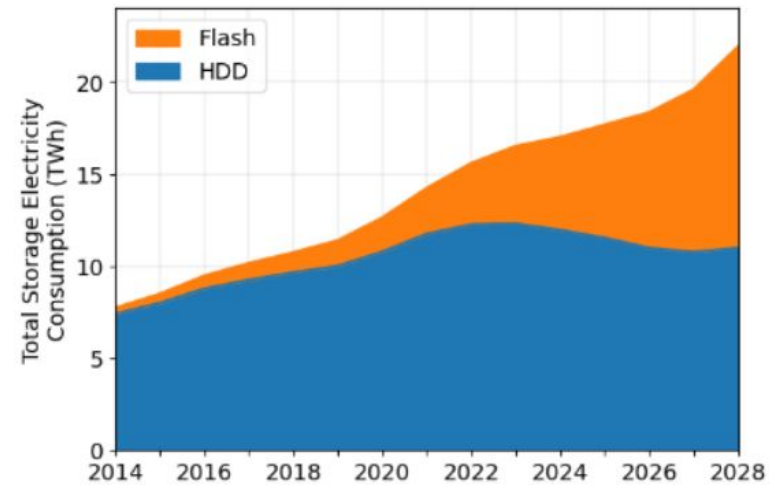
Castro, Vitor, et al. "Digital data demand and renewable energy limits: Forecasting the impacts on global electricity supply and sustainability." *Energy Policy* 195 (2024): 114404.

Background

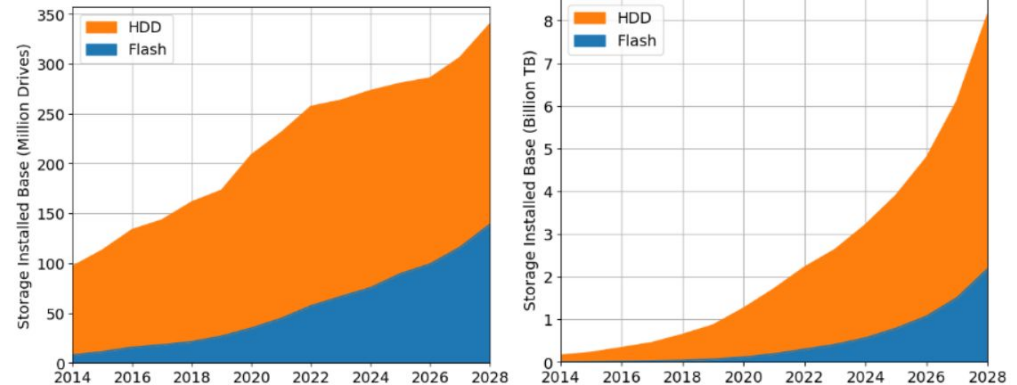
- From a data center perspective, reports are confirming a global trend
- Manufacturers of storage equipment are also confirming the trend

“Recent data from Azure suggests that storage-related emissions make up 33% of operational and 61% of embodied emissions. Storage racks alone account for 24% of operational and 45% of embodied emissions”

S. Mcallister et al., ‘A Call for Research on Storage Emissions’, *SIGENERGY Energy Inform. Rev.*, vol. 4, no. 5, pp. 67–75, 2025



Storage equipment annual electricity use by type. - Shehabi et al. 2024



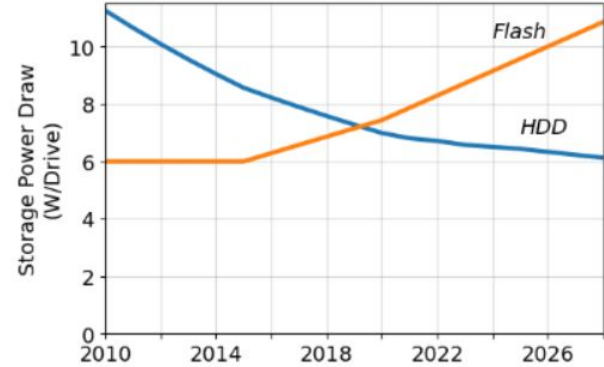
Installed base of storage devices in drive units (left) and TB capacity (right). - Shehabi et al. 2024

Background

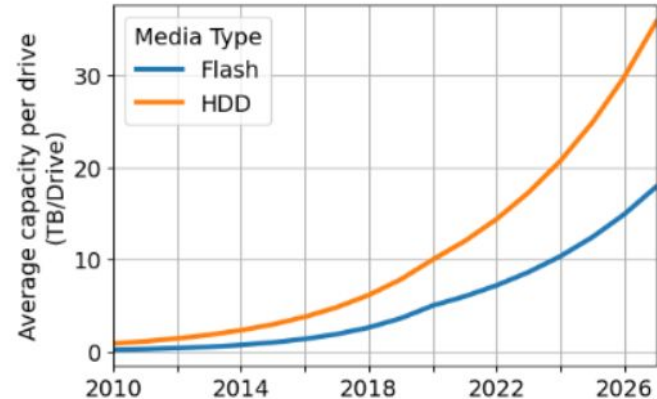
- From a data center perspective, reports are confirming a global trend
- Manufacturers of storage equipment are also confirming the trend
- Correlated with efficiency gains

“Storage often makes up 25%–45% of the total site power consumption. Where tape storage is provided, the system normally consumes another 10% of the site’s total power.”

The environmental impact, carbon emissions and sustainability of computing in the ATLAS experiment. The European Physical Journal C, 2025, vol. 85, no 12, p. 1397.



Average wattage of storage drives shipped in each year. - Shehabi et al. 2024



Assumed average drive capacity (TB) of new storage devices shipped in each year - Shehabi et al. 2024

Background

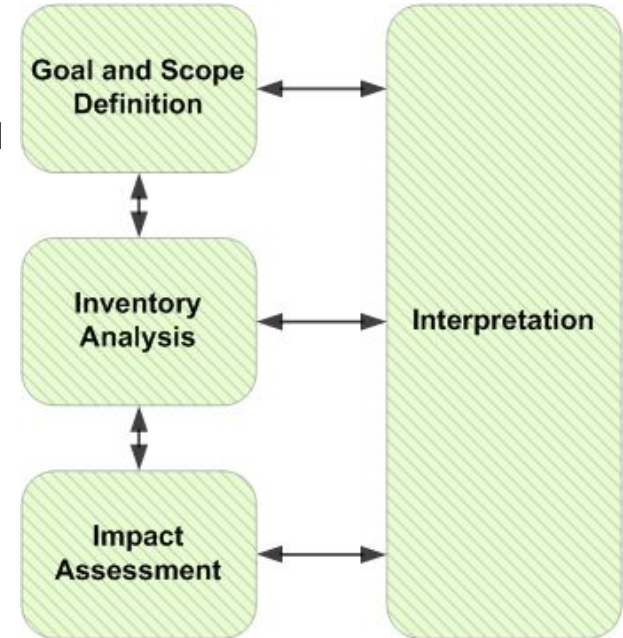
Various estimates for the cost of storing one TB for a year (in kWh per year):

- **6.132** - A. Shehabi et al., « 2024 United States Data Center Energy Usage Report », déc. 2024, doi: 10.71468/P1WC7Q.
- **87.6 to 700.8** - P. Banijamali et al. « Unveiling Key Performance Indicators for the Energy Efficiency of Cloud Data Storage » *Proceedings - IEEE 21st International Conference on Software Architecture Companion, ICSCA-C 2024*, 2024 doi: 10.1109/ICSCA-C63560.2024.00047.
- **11.38 (SSD) and 36.78(HDD)** - S. Tannu et P. J. Nair, « The Dirty Secret of SSDs: Embodied Carbon », *SIGENERGY Energy Inform. Rev.*, vol. 3, n° 3, p. 4-9, oct. 2023, doi: 10.1145/3630614.3630616.
- **26.28 to 61.32 (only idle)** - S. Raut et al. « Workload Management for Sustainability with Storage Systems in Hybrid Cloud Deployment » *WINTeCHCON 2023 Emerging Technologies from Silicon to Software for a Sustainable Future*, 2023. doi: 10.1109/WINTeCHCON58518.2023.10276943.
- **65700** - T. Jackson et al. « Is there a role for knowledge management in saving the planet from too much data? », *Knowledge Management Research & Practice*, vol. 21, n° 3, p. 427-435, mai 2023, doi: 10.1080/14778238.2023.2192580.
- **20.75** - Bristol University Research Data Storage Facility (Tape Storage excluded)

Life cycle assessment (LCA) - Quick introduction

- Developed in the 70s/80s as *cradle to grave* approach for the industry
- Standardised in the 90s as a common methods by the ISO and others
- Still less used in the digital sector as in other industrial sector

- Still a core concept and research topic in the field of environmental impact assessment
- Defined by 4 main characteristics
 - Full life-cycle perspective
 - Multi-criteria assessment
 - Quantitative results
 - Scientific approach



LCA of Digital services - Quick example

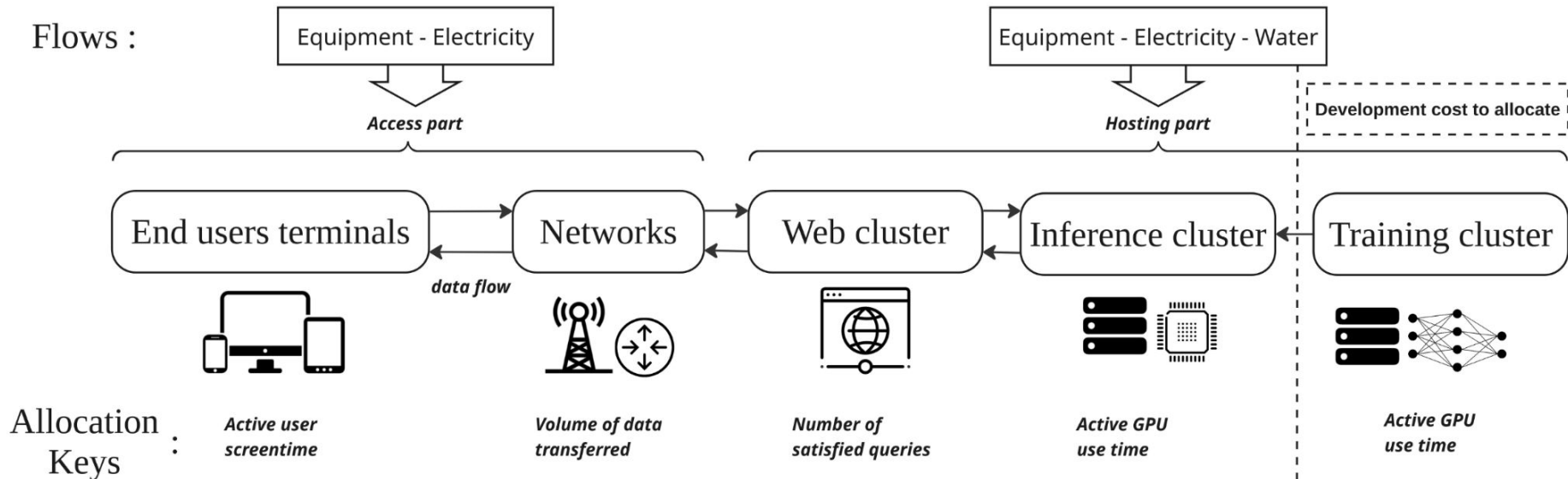
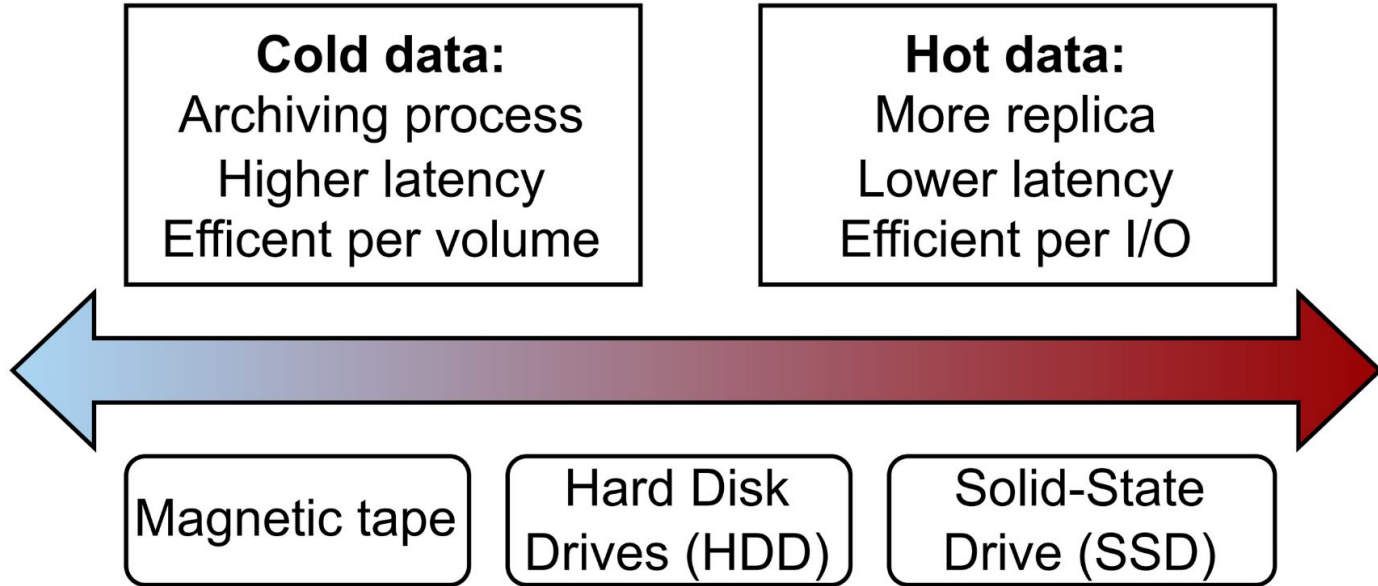


Figure 5: Flow diagram of ALCA

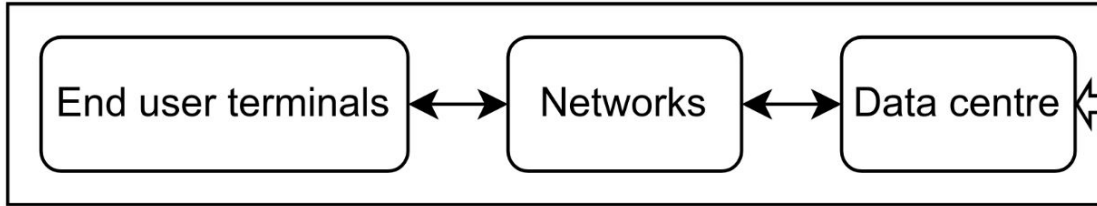
The complexity of HSM

- HSM : hierarchical storage management
- Even more complex and heterogeneous when you consider cloud storage
- This spectrum does apply also to:
 - interface, SATA or NVMe
 - protocol, RAID1 or RAID5

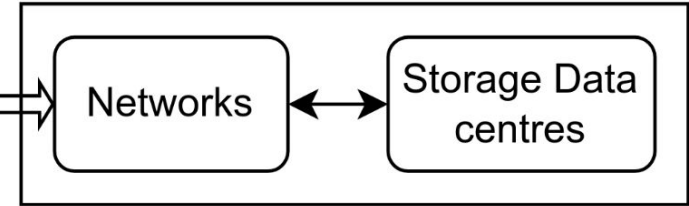


Challenges of assessing data storage

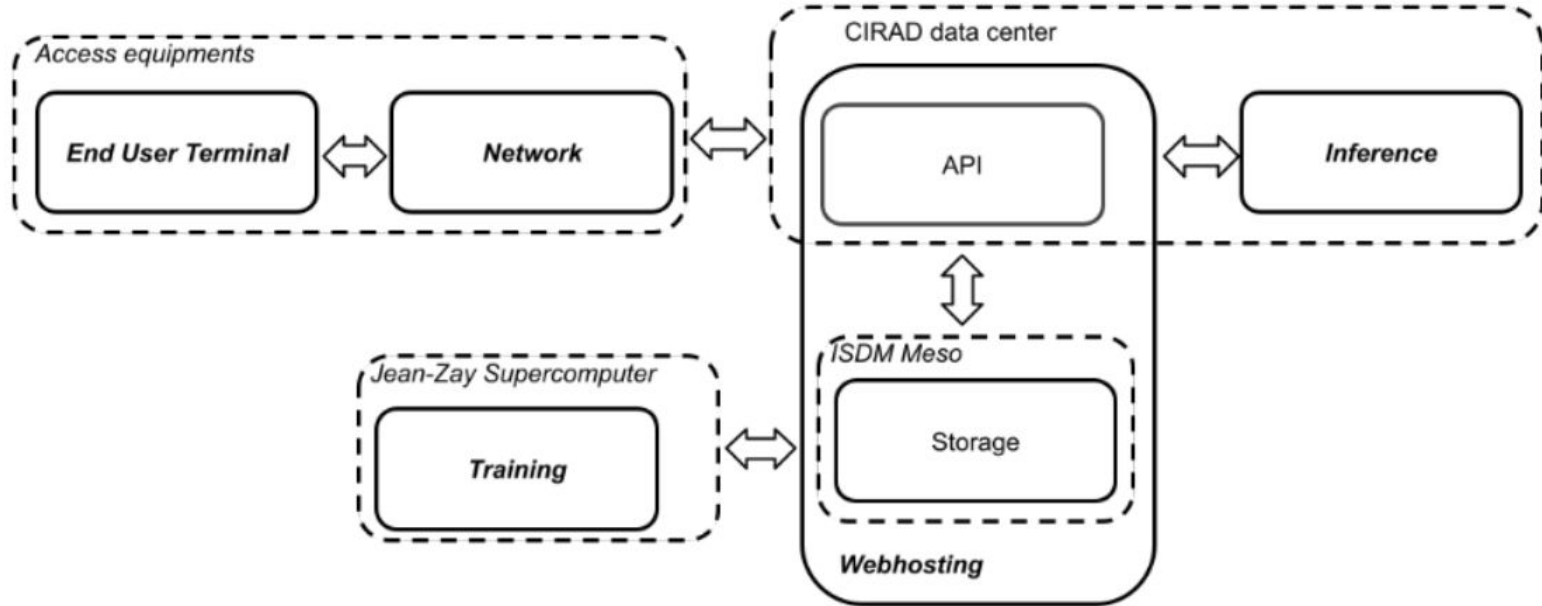
Standard 3 Tier of digital service



Storage as a digital service

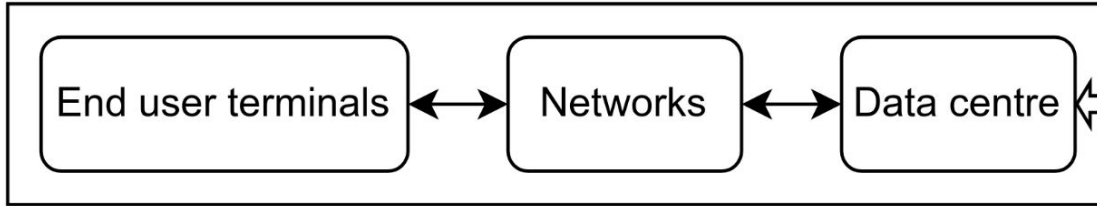


A consequence of specialisation

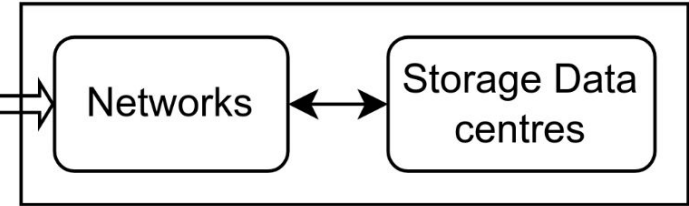


Challenges of assessing data storage

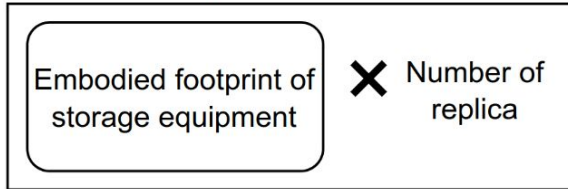
Standard 3 Tier of digital service



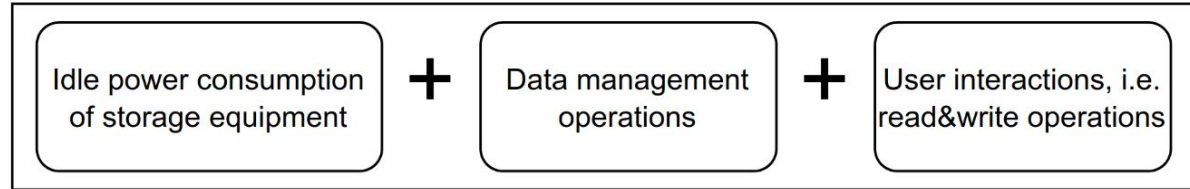
Storage as a digital service



Embodied footprint



Operational footprint



Limits and discussion

- Our project here, is mapping a digital use to an environmental cost, but at the end this cost will depend on our method's framework, i.e. attributional or consequential.
 - Attributional approach tends to favours efficiency
 - Consequential approach tends to favours *status quo*

Attributional LCA



What part of the global environmental burdens should be assigned to the product?

XX kg CO₂-equ.
etc.

Consequential LCA



What is the impact of the product on the global environmental burdens?

ZZ kg CO₂-equ.
etc.

Two types of LCA
To respond to
different questions

Limits and discussion

- **Attributional LCA**
 - Allocation of infrastructure according to “use” share, likely storage space
 - Good for comparing storage systems, reporting, and assess overall efficiency
- **Short-term Consequential LCA**
 - Only looking at energy consumption scenarios, excluding infrastructure
 - Assessing energy leverage, effectiveness of carbon aware strategies
- **Long-term Consequential LCA**
 - Assessing effect of decisions on the infrastructure and how this influence trend of growth or degrowth of the storage system
 - Suited for discussing how data storage system can or cannot fit into more global emission reduction scenarios

Limits and discussion

- Too much focus on individual use could be counterproductive
- Creating at worse, wrong incentives such as inciting user too spend too much time deleting unused data
- It is difficult to have precise information about the breakdown of the global datasphere between usage (private, work, etc.)
 - But it is likely that most of data are not personal data but rather professional one

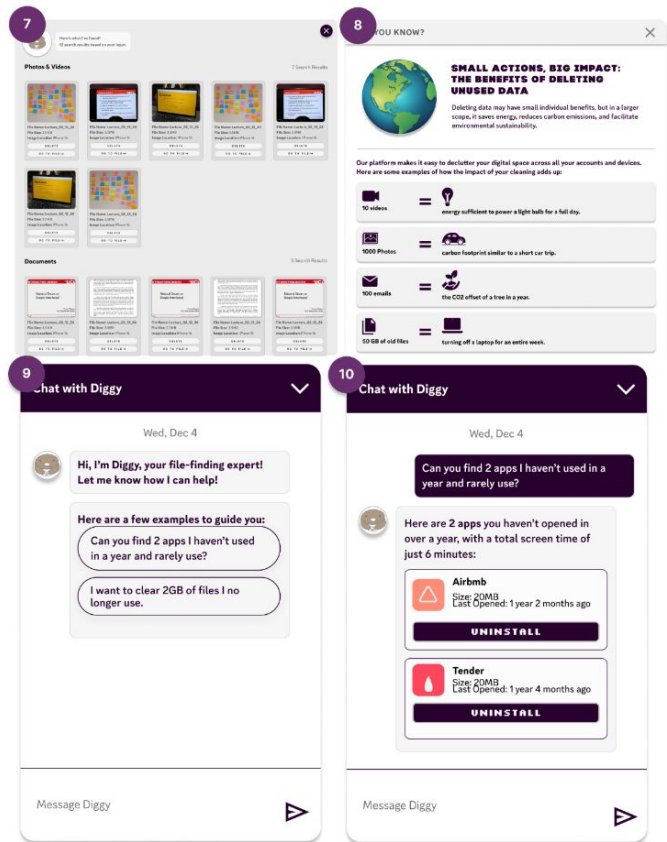
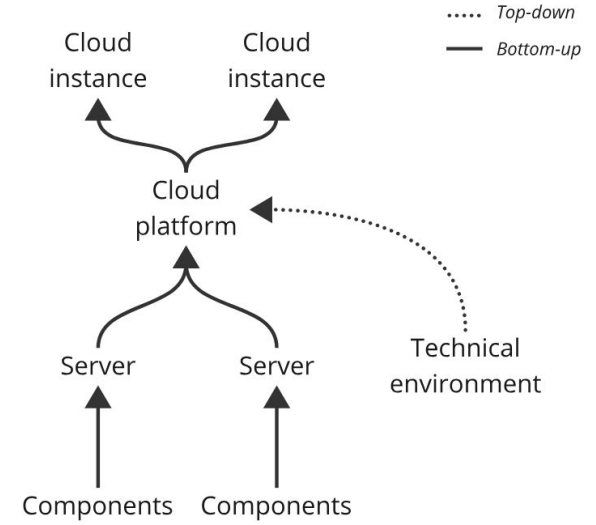
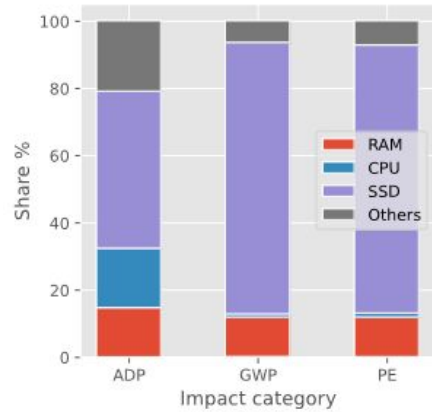


Figure 9: Digibyte Final Prototype Interfaces. (7) Smart Search (8) Awareness (9-10) Diggy the chatbot.

Annexes

Open source database



Simon Thibault, David Ekchajzer, Adrien Berthelot, Eric Fourboul, Samuel Rince, et Romain Rouvoy. 2024. « BoaviztAPI: a bottom-up model to assess the environmental impacts of cloud services ». In HotCarbon'24. Workshop on Sustainable Computer Systems. Santa Cruz, United States.

Consequential LCA of digital services

Attributional LCA



What part of the global environmental burdens should be assigned to the product?

XX kg CO₂-equ.
etc.

Consequential LCA



What is the impact of the product on the global environmental burdens?

ZZ kg CO₂-equ.
etc.

Two types of LCA
To respond to
different questions

Besoin de mesurer l'empreinte à d'autres échelles

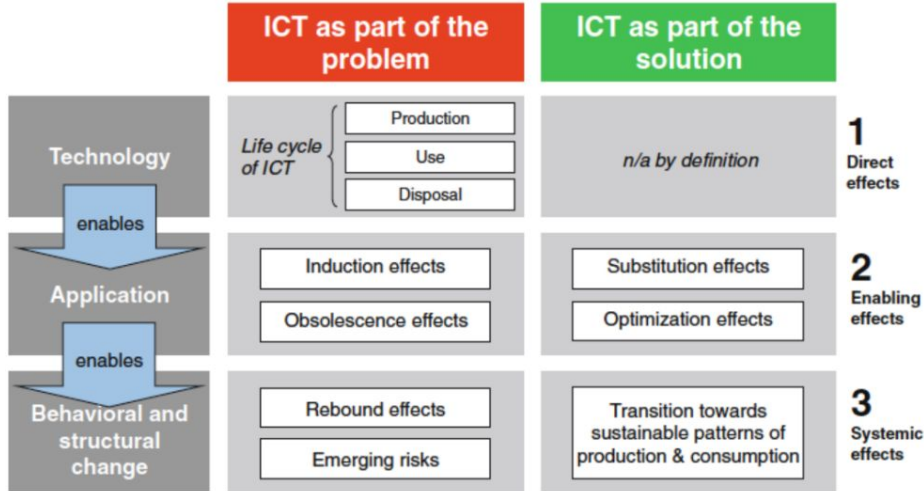


Figure 2: Matrix of ICT effects, based on [13] and taken from [14]

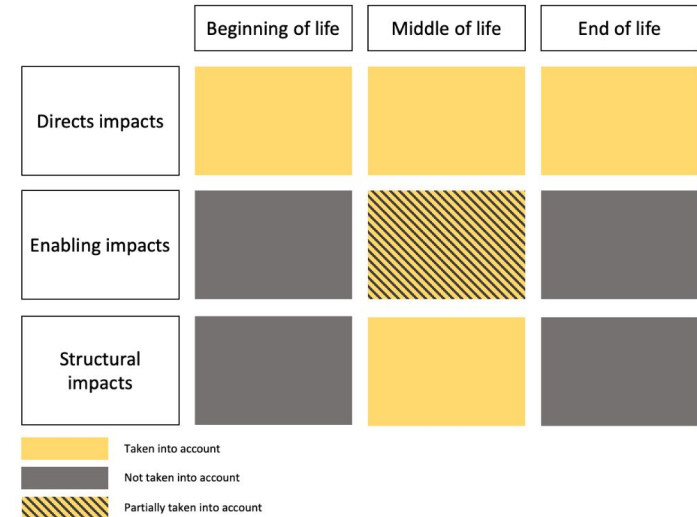


Figure 3: Considered impacts related to the ICT life phases in Hilty's proposal

Grimal, Lou, Ines Di Loreto, Nicolas Burger, et Nadège Troussier. 2021. « Design of an interdisciplinary evaluation method for multi-scaled sustainability of computer-based projects. A workbased on the Sustainable Computing Evaluation Framework (SCEF) ». LIMITS Workshop on Computing within Limits.