Collecting Energy Consumption of Scientific Data

Energy Demands for Files During Their Life Cycle

Julian M. Kunkel,
Olga Mordvinova, Michael Kuhn, Thomas Ludwig

German Climate Computing Centre
Hamburg

Scientific Computing
Department of Informatics
University of Hamburg

16-09-2010

Motivation

Track energy consumption caused by I/O to

- quantify energy costs for storage and processing.
- pull off awareness for energy costs.
- identify expensive applications/workflows.
- identify chances to optimize the storage landscape.

- Do *you* know how much energy it costs to keep your files?

Energy Characteristics of Storage

Model	Capacity in GBytes	Power Consumption (transfer) in Watts	Power Consumption (idle) in Watts	IOPS	Transfer Rate in MBytes/s
WD Caviar Green 7,200	1,000	4.9	2.8	80	111
Seagate Cheetah 15K	600	16.4	11.7	184	204
Intel X25-M G2 Postville MLC	160	0.15	0.08	$\leq$ 35,000 (read) 8,600 (write)	250 (read) 100 (write)
Intel X25-E Extreme SLC	64	2.6	0.06	35,000 (read) 3,300 (write)	250 (read) 170 (write)
Fujitsu Siemens LTO1 (PRIMERGY)	100	18	N/A	0.013	16
Quantum LTO4	800	28.8	6.4	0.018	120
Tandberg LTO5	1,500	24	6.9	0.014	140

Metric

Total Energy for the File Life Cycle $=$ *TEFL*

The amount of energy consumed by file I/O and long-term storage.

Possibilities

- Measure real energy consumption.
- Provide an analytical model.

Difficulties with the Measurement

Accounting energy consumption per I/O requires to

- measure energy for all components of the parallel file system.
  - This requires measurement devices, complex implementation.
- divide energy consumption among concurrently executed I/Os.
  - How do we divide the energy consumption fairly among the I/Os?

Interpretation is hard

- How do we modify our access pattern/storage policy to reduce energy consumption?

Analytical Model

Required characteristics

- Divide energy consumption among files.
- Be invariant to concurrent operations and processing order.
- Ease interpretation of measured values to derive suggestions.

Analytical Model

Basic idea of the simple model

- Split transfer and idle costs depending on system capabilities:
  - E_{idle} = FileSizeOverTime / SystemCapacity · SystemIdleCosts
  - $E_{transfer}$ = TimeForIO · SystemActiveCosts
- Handle storage systems of the storage landscape individually.

Drawbacks

- Empty disk space is accounted to the computing facility.
- Assumes energy of the storage is invariant over its life time.

Collecting Energy Consumption

Design criteria

- Limited implementation effort
- Compatibility with existing systems
- Low performance overhead
- No background file scanning/update of values necessary

Metrics to Store

Alternatives

- Store accumulated energy consumption directly.
  - File system must know how to compute consumption.
- Account basic I/O information.
  - # reads, writes, blocks read/written and integrated file size.
  - $\Rightarrow$ Additional insight about file usage is provided.
  - Provide tools to compute energy consumption.

Storing of energy metrics

- Updates of values must be performed for each I/O.
- Similar to `atime`
- Allow caching – lazy persistency scheme
- Metrics could be stored in *Extended Attributes*.

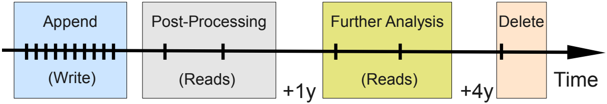
Example Metrics to Store For One Storage System

system.**iocount_read**	= 10
system.**iocount_write**	= 50
system.**blocks_read**	= 1000
system.**blocks_written**	= 1000
system.**file_size_over_time**	= 200000 *(Byte · Seconds)*
system.**last_file_size_update**	= 2010-05-12 17:00:00.01

1. Total Energy for the File Life Cycle
2. Analytical Model
3. Collecting Energy Consumption
4. **Example Scenario**
5. Summary

Example Scenario

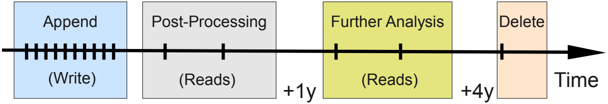
Application access pattern



Storage landscape configurations

- Tape & hard disks
- Tape & flash
- Hard disks or flash only

Access Pattern Details



- Iterative application:
  - In 10 minute intervals 10 GByte of data is appended to the file.
  - 100 iterations ⇒ 1 TByte of data
- I/O on disk is performed in a granularity of 10 MBytes.
- Post-processing and further analysis each read the whole data.

Storage Landscape Setup

Online storage

- 130 shelves a 16 disks, RAID-5, shelves are mirrored ⇒ RAID-51
- 2080 WD Cavier Green hard disks
- Total capacity about 1 Petabyte
- For Flash, Intel Extreme, same configuration ⇒ 64 TByte

Tape archive

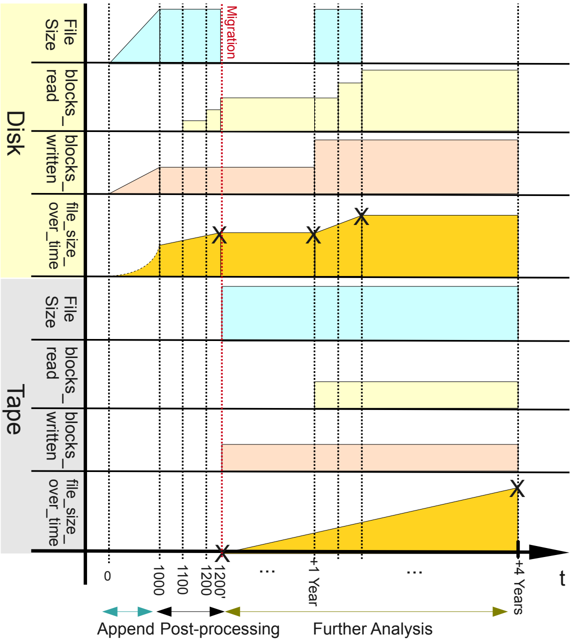
- StorageTek SL8500
- 10 Tape drive trays, capacity of 10,000 tapes a 1 TByte

Storage Landscape Configurations

Storage System	Power Consumption in Watts		IOPS		Transfer Rate in MBytes/s	
	(active estimate)	(idle)	(read)	(write)	(read)	(write)
Hard disk	10	5	80	80	125	125
FS with disks*	40,000	24,000	60,000	60,000	15,000	15,000
Flash disk	3	0.06	35,000	8,600	250	170
FS with flash*	24,600	13,132	10 *M*	2,5 *M*	30,000	20,000
LTO tape drive	53	35	0.01	0.01	210	210
Tape archive*	1,200	800	0.10	0.10	2,100	2,100

* Data of parallel file systems is inspired by vendor information.

Figure: File attributes for the migration use case.



Phases and Resulting Values for the I/O Metrics

	Processing Phase	Accessed Data (read) in Bytes	Accessed Data (write) in Bytes	I/O Operations (read)	I/O Operations (write)	File Size Over Time in Bytes · s
Online storage	Append	0	$1 \cdot 10^{12}$	0	$1 \cdot 10^{5}$	$30.3 \cdot 10^{15}$
	Post-processing	$3 \cdot 10^{12}$	$1 \cdot 10^{12}$	$3 \cdot 10^{5}$	$1 \cdot 10^{5}$	$42.3 \cdot 10^{15}$
	Further Analysis	$5 \cdot 10^{12}$	$2 \cdot 10^{12}$	$5 \cdot 10^{5}$	$2 \cdot 10^{5}$	$54.3 \cdot 10^{15}$
Tape	Append	0	0	0	0	0
	Post-processing	0	$1 \cdot 10^{12}$	0	1	0
	Further Analysis	$1 \cdot 10^{12}$	$1 \cdot 10^{12}$	1	1	$15.8 \cdot 10^{19}$

Table: Storage landscape with tape archive.

Processing Phase	Accessed Data (read) in Bytes	Accessed Data (write) in Bytes	I/O Operations (read)	I/O Operations (write)	File Size Over Time in Bytes · s
Append	0	$1 \cdot 10^{12}$	0	$1 \cdot 10^{5}$	$30.3 \cdot 10^{15}$
Post-processing	$2 \cdot 10^{12}$	$1 \cdot 10^{12}$	$2 \cdot 10^{5}$	$1 \cdot 10^{5}$	$42.3 \cdot 10^{15}$
Further Analysis	$4 \cdot 10^{12}$	$1 \cdot 10^{12}$	$4 \cdot 10^{5}$	$1 \cdot 10^{5}$	$15.8 \cdot 10^{19}$

Table: Only online storage available.

Phases and Resulting Energy Consumption

	Processing Phase	Energy Consumption in Joules			
		(disk & tape)		(flash & tape)	
		Idle	Busy	Idle	Busy
Online storage	Append	$0.7 \cdot 10^6$	$2.6 \cdot 10^6$	$6.2 \cdot 10^6$	$1.2 \cdot 10^6$
	Post-processing	$1.0 \cdot 10^6$	$10.4 \cdot 10^6$	$8.7 \cdot 10^6$	$3.5 \cdot 10^6$
	Further Analysis	$1.3 \cdot 10^6$	$18.3 \cdot 10^6$	$11.1 \cdot 10^6$	$6.3 \cdot 10^6$
Tape	Append	0	0	0	0
	Post-processing	0	$0.6 \cdot 10^6$	0	$0.6 \cdot 10^6$
	Further Analysis	$2.5 \cdot 10^6$	$1.1 \cdot 10^6$	$2.5 \cdot 10^6$	$1.1 \cdot 10^6$

Table: Storage landscape with tape archive.

	Energy Consumption in Joules			
	(disk)		(flash)	
	Idle	Busy	Idle	Busy
Append	$0.7 \cdot 10^6$	$2.6 \cdot 10^6$	$6.2 \cdot 10^6$	$1.2 \cdot 10^6$
Post-processing	$1.0 \cdot 10^6$	$7.8 \cdot 10^6$	$8.7 \cdot 10^6$	$2.7 \cdot 10^6$
Further Analysis	$3.8 \cdot \mathbf{10^9}$	$13 \cdot 10^6$	$32.4 \cdot \mathbf{10^9}$	$4.3 \cdot 10^6$

Table: Only online storage available.

Evaluation

Total energy spent

- Tape & disk: $23.3 \cdot 10^6$ Joule, 0.20 € per kWh $\Rightarrow$ 1.30 €
- Tape & flash: $21.1 \cdot 10^6$ Joule
- Online storage: disk $3.8 \cdot 10^9$ Joule $\Rightarrow$ 211 €, flash 9 times more

Observations for our scenarios

- Busy costs are negligible for online storage.
- Busy costs dominate for disks while idle costs dominate for the selected flash drive.

Summary & Conclusions

- Knowing energy consumption of files is valuable.
- I/O statistics can be treated similar as `atime`.

- The example scenarios show the importance of tape.
- Optimal migration reduced the TEFL to 0.5%.
- Capacity helps to leverage idle costs of hard disk drives.

- In the future we will embed statistics into (parallel) file systems.