

Data



We denote several types of data in the compute environment of the scientific systems:

- **Science generated:** New information generated by lab experiments or theoretical calculations form the basis for the scientific output of the university. This unique type of data should be stored on highly save media and kept available for at least 10 years (data management requirement).
- **Science data from elsewhere:** Often data is shared between institutes, where the above requirement should be fulfilled by the institute that generated the data. The institute that receives the data has no obligation wit regard to data management, but has an obligation to the users of the data.
- **Science data intermediate products:** While performing scientific analysis many intermediate products are created. These products can always (at certain costs) be recreated from the original data and associated software/scripts. This type of data can be denoted as scratch data for which low reliability hold.
- **Science publication data:** These final products are the scientifica output from the university and should be kept available for at least 10 years (data management requirement).
- **Office management information:** These data are personnel or personal data for which high security and high reliability are standard. The information is usually enclosed in database format. In addition software to access and manipulate this data shuld also be treated with high security and high reliability.
- **System management information:** This type of information is sensitive as it describes in detail the institute facilities for which many levels of security and reliability should be in place. In addition the access and manipulation tools should be highly secure.

Classification

Data type	Classification
Science generated	Restricted
Science data from elsewhere	Public
Science data intermediate products	Public
Science publication data	Public
Office management information	Private
System management information	Private

Current situation

Below a table denoting the types of data and their characteristics.

			capacity		latency	IOPS	throughput	granularity	reliability	energy use	accessibility	
			raw	density							random	sequential
Desktop	RAM		few GB	medium	nano sec	huge	high	very high	low	high	high	low
	OS/scratch		hundreds GB	high	micro sec	medium	high	high	medium	medium	medium	low
	permanent		several TB	high	milli sec	medium	high	medium	high	medium	medium	medium
KA	VM's server		several TB	high	micro sec	huge	huge	high	very high	high/medium	high	low
	home		several TB	high	micro sec	huge	huge	very high	very high	high/medium	very high	low
	application/group		tens TB	high	micro sec	large	medium	high	very high	medium	medium	medium
Compute servers	OS/scratch		hundreds GB	high	micro sec	huge	huge	high	medium	high/medium	medium	low
	permanent		tens TB	very high	milli sec	huge	huge	medium	very high	high/medium	low	high
Cluster	OS/scratch		100GB	high	micro sec	medium	medium	high	medium	medium	medium	low
	permanent		tens TB	very high	milli sec	huge	huge	medium	high	medium	medium	high
Nearline			hundreds TB	high	milli sec	medium	huge	medium	high	medium	medium	medium
Bulk	on-line		hundreds TB	huge	milli sec	medium	medium	low	medium	medium/low	low	high
	off-line		hundreds TB	huge	sec	small	small	low	high	extremely low		
Archive	on-line	meta-data	few TB	high	micro sec	high	medium	medium	medium	medium	medium	low
		data	tens TB	huge	sec	medium	small	low	very high	medium/low	low	high
	off-line	data	hundreds TB	huge	sec	small	tiny	low	extremely high	extremely low		

From:

<https://helpdesk.strw.leidenuniv.nl/wiki/> - Computer Documentation Wiki

Permanent link:

<https://helpdesk.strw.leidenuniv.nl/wiki/doku.php?id=policies:security:data&rev=1516034703>

Last update: 2018/01/15 16:45

