

Locality of storage access

Why is access locality important?

- Access locality refers to the ability of software to make good use of the cache. (details on Cache in a following video)
- Memory is broken up into pages.
- Software that uses the same or neighboring pages repeatedly has good access locality.
- Hardware is designed to speed up such software.

Temporal Locality

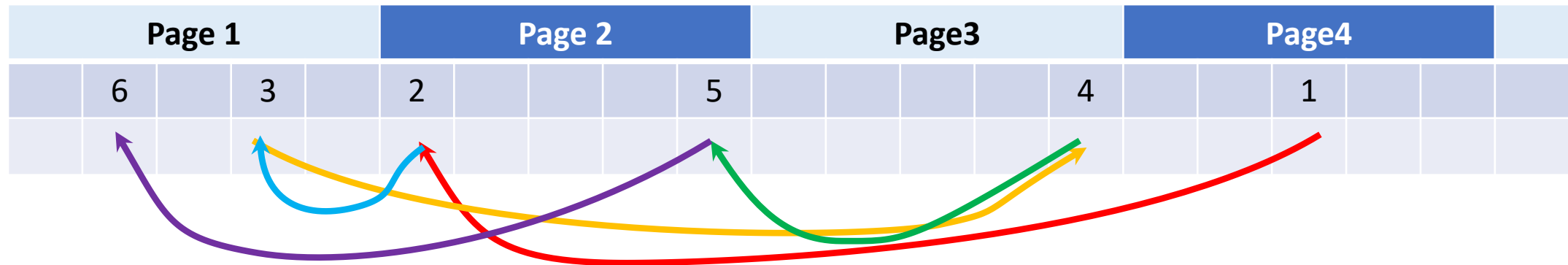
- **Task:** compute the function $f_{\theta}(x)$ on a long sequence x_1, x_2, \dots, x_n
- θ is a parameter vector – example: the weights in a neural network.
- The parameters θ are needed for each computation.
- If θ fits in the cache – access is fast
- If θ does **not** fit in the cache – each x_i causes at least one cache miss – program will be much slower.
- **Temporal Locality:** repeated access to the same memory location

Spatial locality

- **Task:** compute the function $\sum_{i=1}^{n-1} (x_i - x_{i+1})^2$ on x_1, x_2, \dots, x_n
- Contrast two ways to store x_1, x_2, \dots, x_n :
- Linked list (poor locality)
- Indexed array (good locality)

Linked List

Let x_1, x_2, \dots, x_n be 1,2,3,4,5,6



Traversal of 6 elements touches 4 pages

array

Let x_1, x_2, \dots, x_n be 1,2,3,4,5,6

Page 1					Page 2					Page3					Page4					
							1	2	3	4	5	6								

Traversal of 6 elements touches 2 pages

Summary

- Improved memory locality reduces run-time
- Why? Because computer memory is organized in pages.
- Next: Two notebooks demonstrating the effect of locality.