Why arrays (and dicts) are great, yet totally inadequate.

Jean-Claude Wippler
Equi 4 Software, NL

EuroTcl 2008, Strasbourg, FR

Collections

List: 1 2 3

Array:

1	X
2	Y
3	Z

Dict: 1 X 2 Y 3 Z

- Lists have 0 keys, arrays & dicts have 1 key

1	X	A	a
2	Y	B	b
3	Z	C	c

?

Iteration / traversal

foreach x 1 2 3 { … }

foreach x [array names] { … }

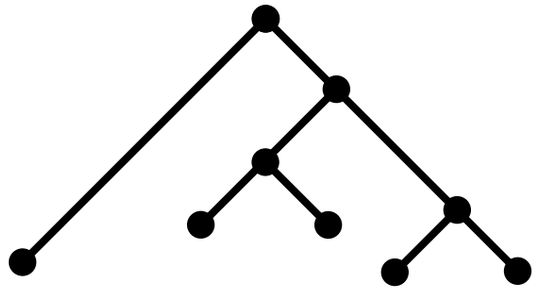
1	X
2	Y
3	Z

- **<u>Explicit</u> loops over <u>explicit</u> collections**

foreach x

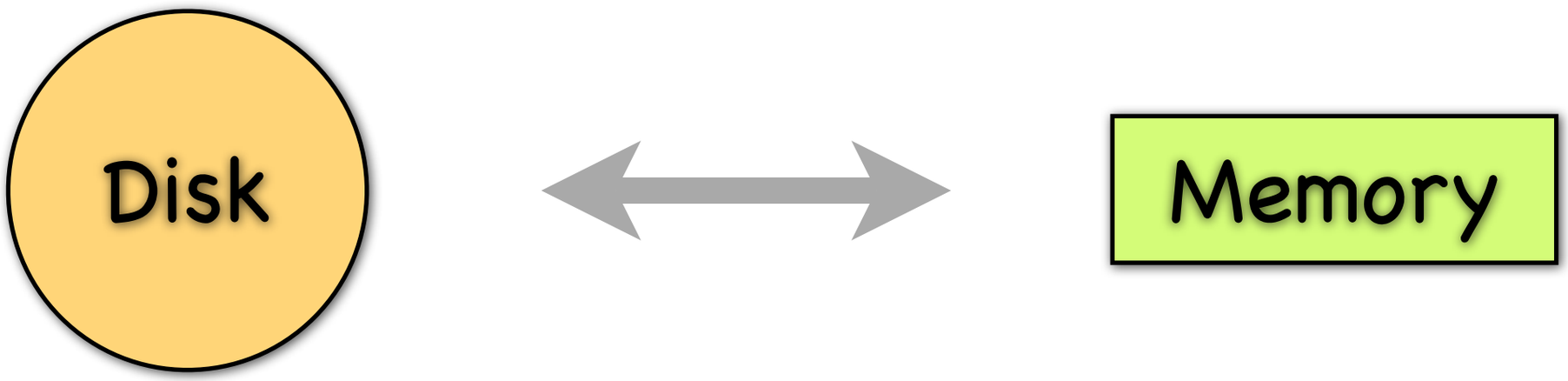
1	X	A	a
2	Y	B	b
3	Z	C	c

?

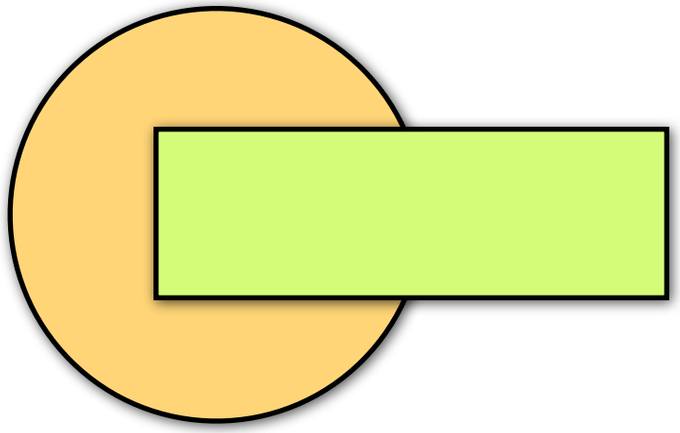


?

Two representations



- Data in both: Memory-mapped files



Generic operations

- Sorting: lists, dicts, arrays, db tables...
- Filtering/selection: loop & create new list
- Locating an entry by non-key properties
- Combine data: grouping / nesting / joining
- Set operations: shorthand for && and ||

Notation re-use

- Collections: lists vs. arrays/dicts

  "lset $a 1 ha" vs. "set a(1) ha"

- Stored data vs. data at run-time

  "SELECT * FROM x" vs. "foreach x [array ...]"

- Explicit loops vs. set-wise operations

Sample problems

- Change sort order of table shown on screen
- Chat app: up-to-date list of active users
- Backup all files changed since yesterday
- XML processing and transformations
- SW engineering: find all undocumented procs

One more big issue

- Procedural code: do this, then that, then ...
- No link back, <u>from</u> results <u>to</u> steps taken
  - think about drill-down, visual change
  - "apres-moi-le-déluge coding"
- We should make results in(tro)spectable
- Dataflow: remember spreadsheets?

A way out

- Rectangular data: zero or more columns
- For want of a better name, I call 'em Views
- Meta-data: column descriptions also a view
- Real view: data in memory or mem-mapped
- Derived view: transformation of other views
- Combined: wrapper around list, dict, db, ...

Views

- Properties of a view:
  - named columns
  - indexed rows
  - structure described by meta-view
  - v i r t u a l

Derived views

- Sort by last name:

	First	Last
0	Brent	Welch
1	Don	Libes
2	Jeff	Hobbs
3	John	Ousterhout

A

	First	Last
2	Jeff	Hobbs
1	Don	Libes
3	John	Ousterhout
0	Brent	Welch

B

- Constructed using a virtual row mapping:

  "remap" A with **2 1 3 0** produces view B

- It's all smoke and mirrors:
  result <u>looks</u> just like a sorted copy
  less mem use, can introspect to see **2 1 3 0**

Combined views

- Real view can be a:
  - list or dict, but not array (no row index)
  - Metakit view, including new Vlerq views
  - the result set of a SELECT in SQLite
- But why stop there:
  - directory listing, any VFS tree, XML
  - CSV or /etc/passwd (i.e. read and parse)
  - EXIM data associated with a JPEG image
  - any file format, exposed via "reader" code

In a nutshell

- By aiming for …
  - high-level general-purpose data structures
  - set-wise "wham" operators i.s.o. looping
  - "natively persistent" collections
- … we get
  - <u>simpler</u> and <u>faster</u> applications in Tcl