Introduction to Programming Discussion 3

Andrew Jianzhong Liu

ShanghaiTech University SIST

liujzh@shanghaitech.edu.cn

December 18, 2016

Overview

Searches We've Learned So Far...

Linear Search

Search through an array linearly. If items match then return True. If search exhausted and no match found, return False.

Binary Search

Search through a sorted array by removing half of the remaining items during each iteration. Better time complexity on sorted arrays.

Hash Function

Concept

A hash function is any function that can be used to map data of arbitrary size to data of fixed size.

Example

A function that takes in a student's name and returns the student ID.

Hashing Functions in Use

- Cyclic Redundancy Check (CRC), CRC16
- Checksums
- MD5
- SHA-1

Hash Tables

Concept

A Hash Table (hash map) is a data structure used to implement an associative array, a structure that can map keys to values. A hash table uses a hash function to compute an index into an array of buckets or slots, from which the desired value can be found.

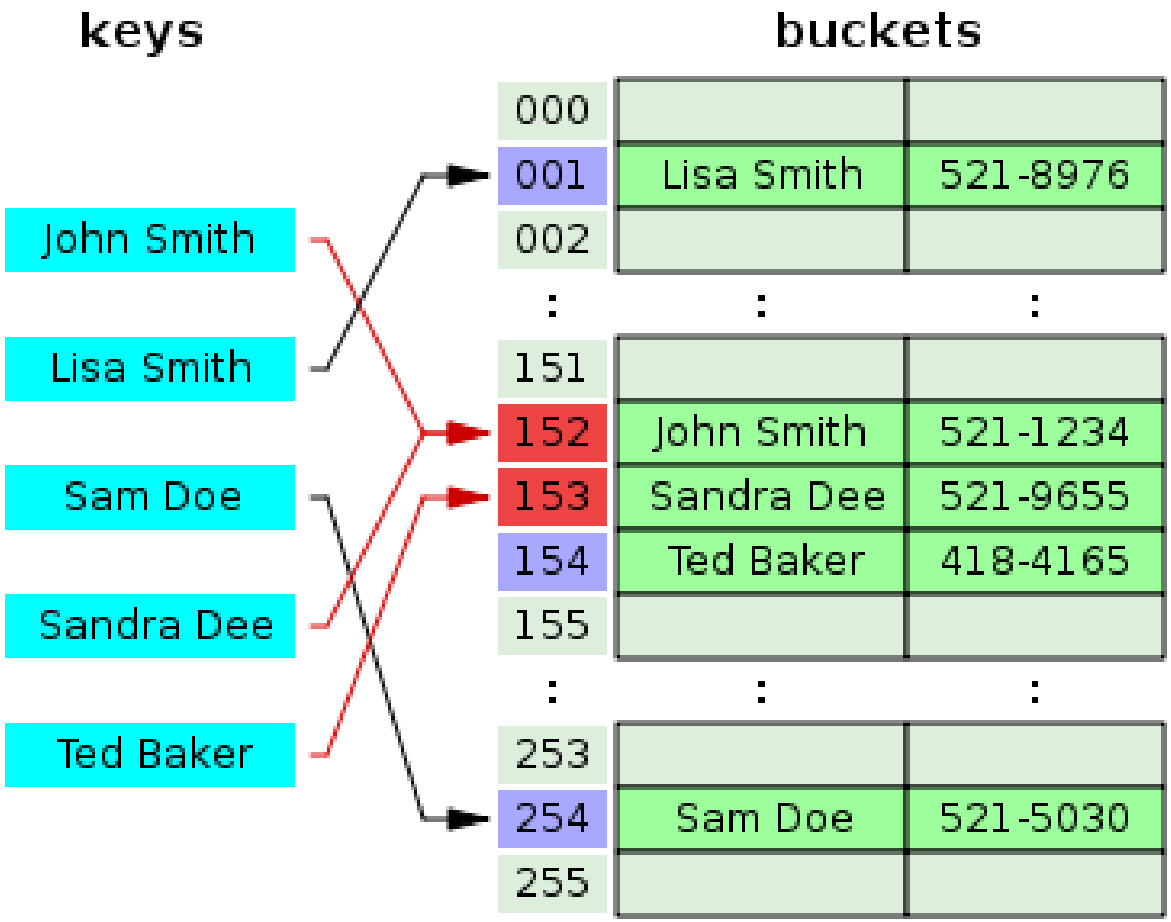
Applications

- Python Dictionary
- Databases
- Programming Language Objects

And a lot more...

Hash Tables Cont.

Visualization



Hash Table Collisions

What are Collisions?

Collisions occur when two different keys are assigned the same value.

Example

Consider the hash function below:

```
int hash(int key, int len){ return key % len; }
```

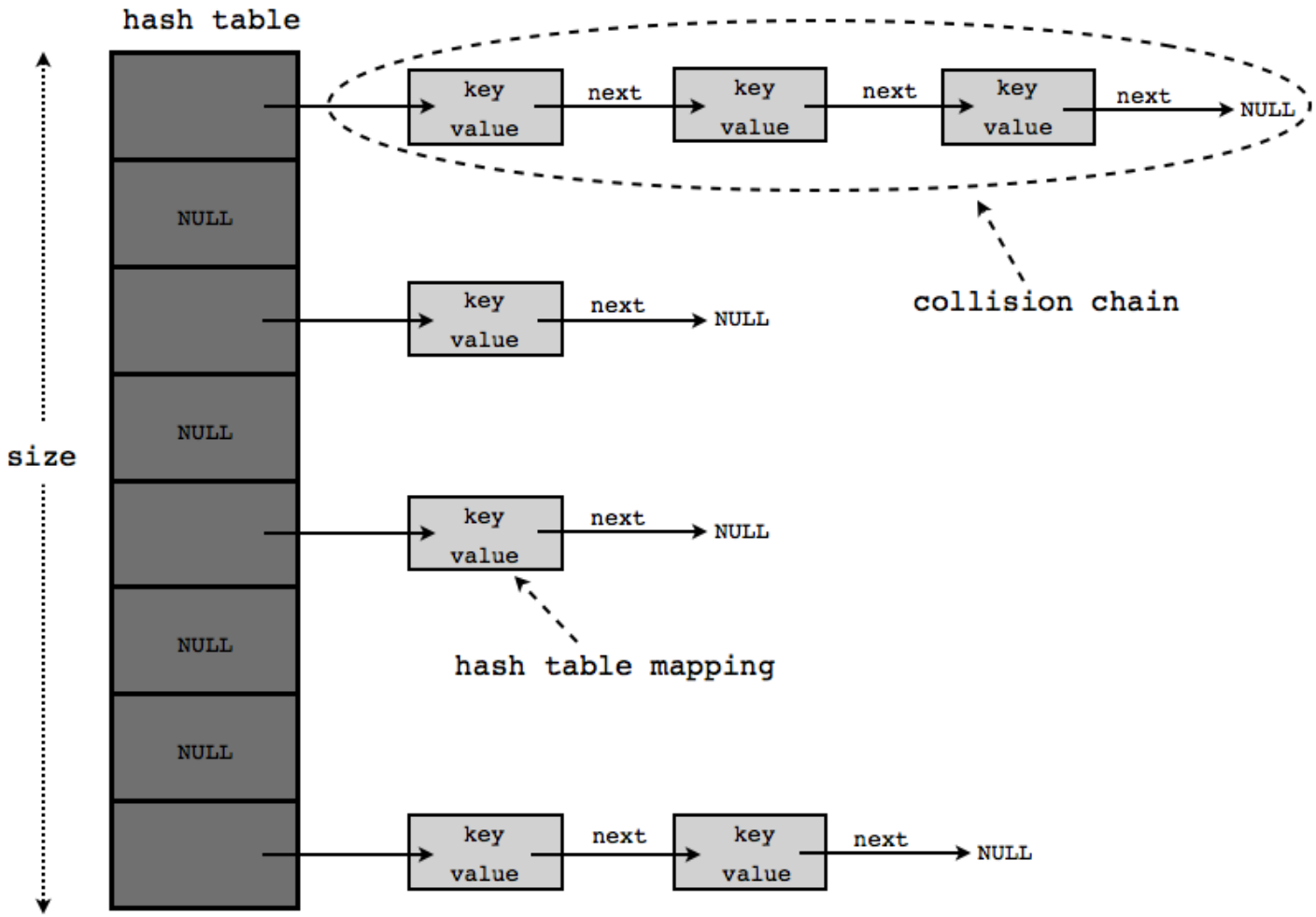
What if the length of the hash table was 5, and the inputs were 1 and 6?

Collision Resolution

Chaining using linked lists.

Hash Table Cont.

Visualization



Recursion

Concept

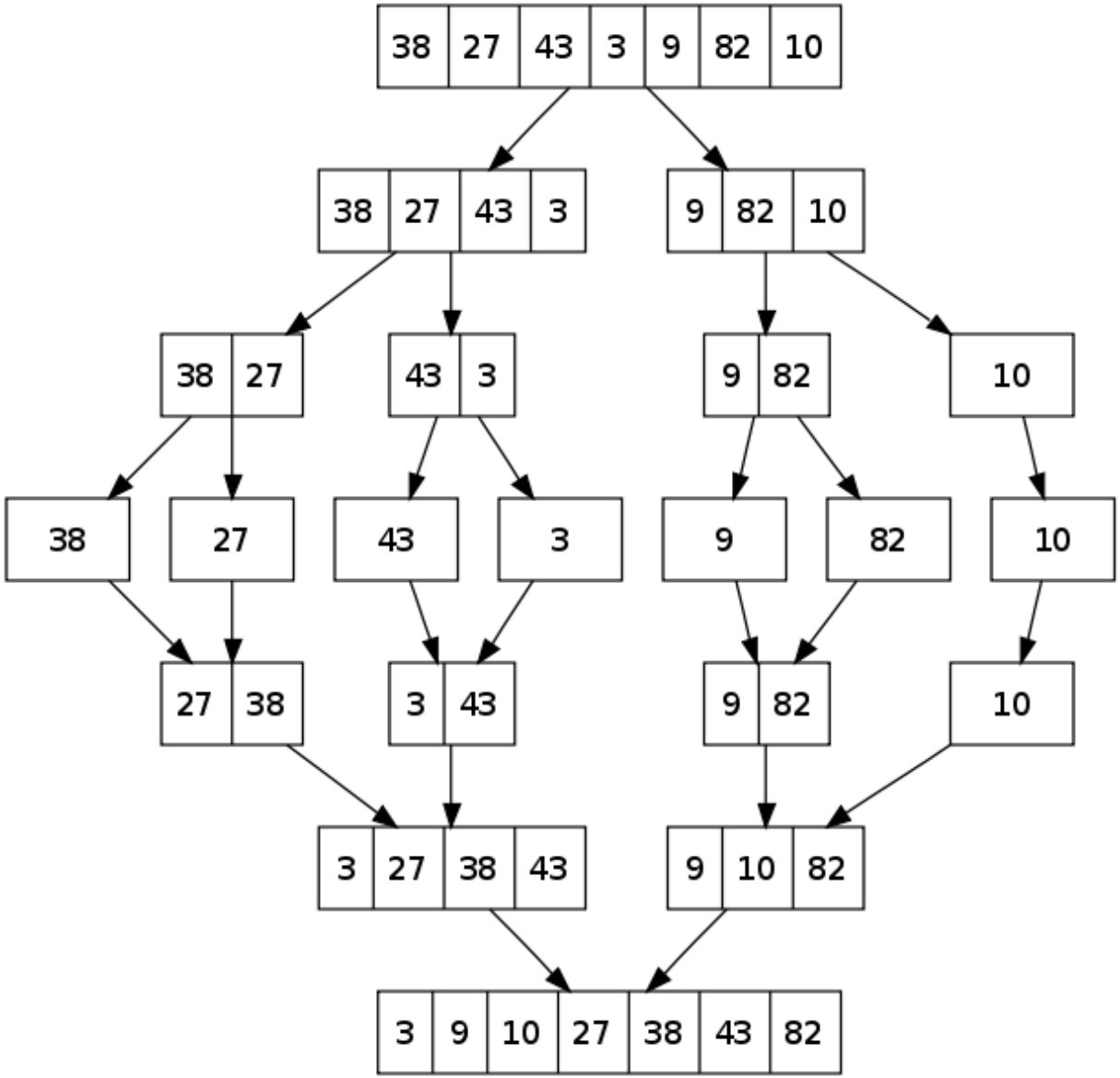
Recursion in computer science is a method where the solution to a problem depends on solutions to smaller instances of the same problem.

Comparison with Iteration

Performance Tail recursions tend to be on par with iterations, however specific comparison relies heavily on compilers.

Stack Size Stack size on most architectures is limited. Recursions tend to use more stack space than iterations.

Recursion Example 1: Merge Sort



Recursion Example 2: Binary Tree Traversal

Pre-order Traversal

```
def preorder(root):
  print(root.key)
  if root.left != None:
    preorder(root.left)
  if root.right != None:
    preorder(root.right)
```

preorder.py

Mathematical Notations

Small-O Notation $f(x) = o(g(x)) \leftrightarrow \lim_{n\to\infty} \frac{f(x)}{g(x)} = 0$

i.e. Strict Upper Bound

Big-O Notation $f(x) = O(g(x)) \leftrightarrow \lim_{n\to\infty} \frac{f(x)}{g(x)} < \infty$

i.e. Loose Upper Bound

Big-Theta Notation $f(x) = \theta(g(x)) \leftrightarrow 0 < \lim_{n\to\infty} \frac{f(x)}{g(x)} < \infty$

i.e. Tight Bound

Big-Omega Notation $f(x) = \Omega(g(x)) \leftrightarrow \lim_{n\to\infty} \frac{f(x)}{g(x)} > 0$

i.e. Loose Lower Bound

Small-Omega Notation $f(x) = \omega(g(x)) \leftrightarrow \lim_{n\to\infty} \frac{f(x)}{g(x)} = \infty$

i.e. Strict Lower Bound

Complexity Analysis

```
def sort(arr):
    for j in range(len(arr)-1):
        minNum = j
        for i in range(j,len(arr)):
            if arr[minNum] > arr[i]:
                minNum = i
        if minNum != j:
            tmp = arr[minNum]
            arr[minNum] = arr[j]
            arr[j] = tmp
    return arr

a = [5,3,7,4,5,7,9,2]
a = sort(a)
print(a)
```

selectionsort.py

What if we need to ...

... pass in parameters with different types?

e.g. Shuttle Bus Dynamic Schedule
Parameters to pass in:

- Bus time
- Source
- Destination

... save information comprised of different types

e.g. Save information about students in the Programming course
Data to save:

- Name
- School
- Homework 0 grade
- etc...

Structs

Concept

- Complex data type
- Declares a list of variables
- Placed under one block of memory
- Accessible under one pointer
- Available in C, C++, C#, Rust ...

Sample Declaration

```
1 typedef struct node{
    int key;
3   struct node * left;
    struct node * right;
5 } Node;
```

struct.c

Classes

Concept

Classes are similar to structs in C++. Members are private by default instead of public.
Python has classes but not structs.

Usage in Python

```
class Heap
  def __init__(self):
    self.nodes = None
  def addNode(self,val):
    newNode = Node(val)
    if self.nodes = None:
      self.nodes = newNode
    else:
      pass
```

class.py

What is Object Oriented Programming?

Concept

Object Oriented Programming is a programming paradigm based on the concepts of objects. Objects may contain data called attributes or routines called methods. i.e. OOP is based on the values and behaviors of objects and their interaction.

Supported Languages

Java, C++, C#, Python, Perl, PHP etc.

Pros and Cons of OOP

CMU Eliminates OOP from freshmen curriculum

Object-oriented programming is eliminated entirely from the introductory curriculum, because it is both anti-modular and anti-parallel by its very nature, and hence unsuitable for a modern CS curriculum.

Pros and Cons

Pros

- Better development productivity and maintainability.
- Code reuse and encapsulation

Cons

- Slower speed and greater space usage
- Steep learning curve
- Problems in parallelization

Debugging, fun eh?

Types of Bugs ... not the flying ones!

Logic Bugs Bugs that result from logic errors in program code.

Syntax Bugs Bugs that originate from erroneous usage of syntax.

Arithmetic bugs, Resource bugs, Multithreading bugs...

How to Debug

1. Locate Error (Hardest!!!)
2. Design Error Repair
3. Patch Error
4. Retest Program

Process of Debugging

Find a Bug

Unexpected behavior of program may indicate a bug.

- Segmentation Fault, Bus Error, Exceptions, panic!();
- Wrong Answer for given input
- Unexpected halts, infinite loops, deadlocks etc.

Reproduce the Bug

- Easy to reproduce: Segmentation Faults, wrong answers.
- Hard to reproduce: Race conditions, deadlocks etc.

Find the origin and fix it.

Change what was wrong to what is right. Beware of generating new bugs! Then Retest.

Exercise 1: Merge Sort using Recursion

Take Home Exercise 2: Implement Dictionaries using OOP and Hash Tables

Why VCS?

Ever ran into these problems?

- Rewrote some parts of the program and failed, need to revert immediately.
- Need to track modifications easily.
- More than 2 people collaborating on the same project. All writing code.

VCS Use in Computing and Programming

- Save previous versions of a file
- Able to revert and compare differences between versions
- Allows collaborations in programming and other tasks.

Popular VCSs

Git Written by Linus Torvalds and collaborators, written in C, Shell, Perl, Tcl, Python, distributed version control system.

Subversion(SVN) Distributed under the Apache License, aims to compatible with CVS, widely used by the free software community. Examples include FreeBSD, GCC.

Concurrent Versions System (CVS) Client-server version control system, influential but no releases since 2008.

Concepts in Version Control Systems

Branch A set of files under version control may be branched or forked at a point in time so that, from that time forward, two copies of those files may develop at different speeds or in different ways independently of each other.

Commit To commit (check in, ci or, more rarely, install, submit or record) is to write or merge the changes made in the working copy back to the repository. The terms 'commit' and 'checkin' can also be used as nouns to describe the new revision that is created as a result of committing.

Pull, Push Copy revisions from one repository into another. Pull is initiated by the receiving repository, while push is initiated by the source. Fetch is sometimes used as a synonym for pull, or to mean a pull followed by an update.

Concepts in Version Control Systems Cont.

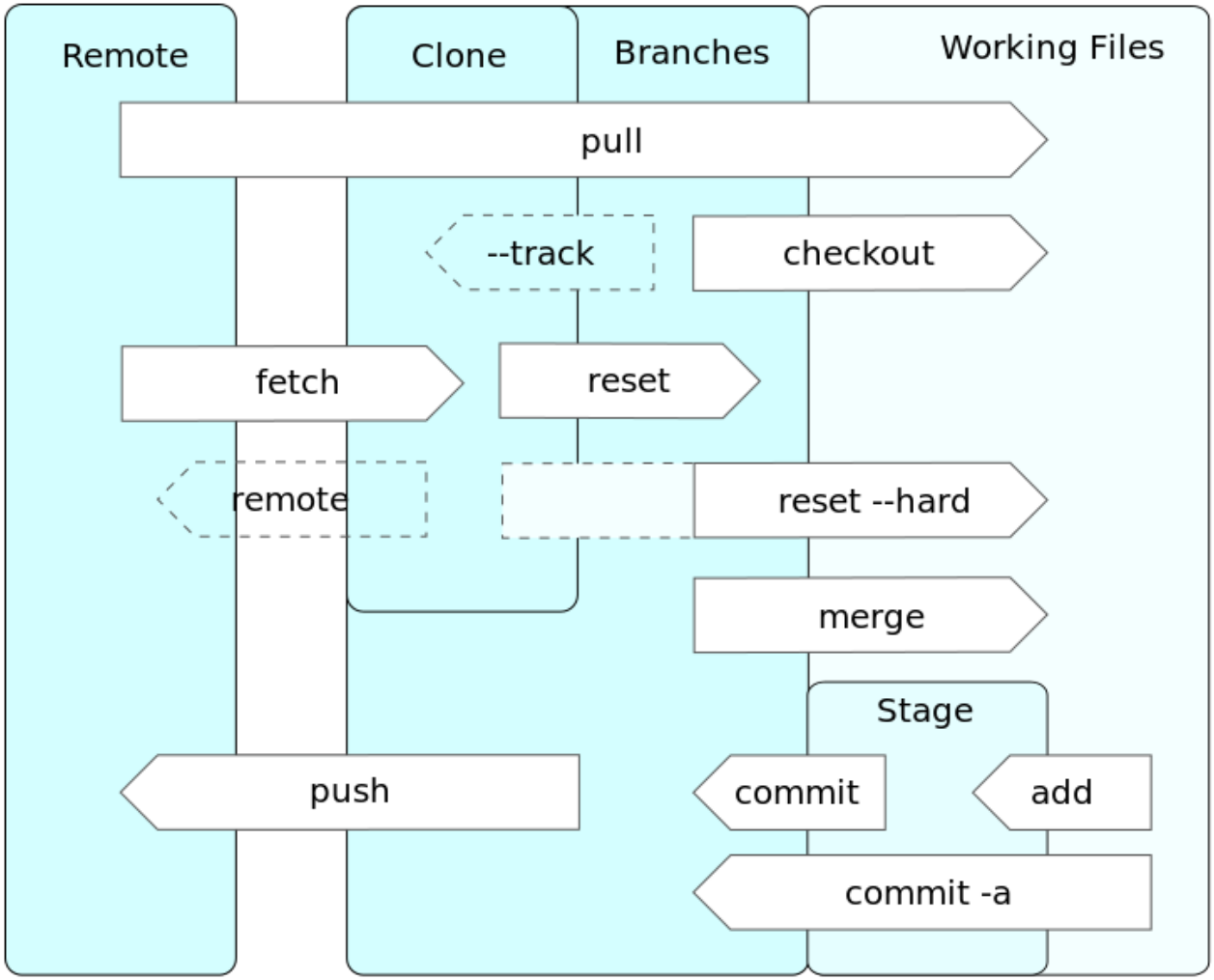
Repository The repository is where files' current and historical data are stored, often on a server. Sometimes also called a depot.

Checkout To check out (or co) is to create a local working copy from the repository. A user may specify a specific revision or obtain the latest. The term 'checkout' can also be used as a noun to describe the working copy.

Clone Cloning means creating a repository containing the revisions from another repository. This is equivalent to pushing or pulling into an empty (newly initialized) repository. As a noun, two repositories can be said to be clones if they are kept synchronized, and contain the same revisions.

Working Copy The working copy is the local copy of files from a repository, at a specific time or revision. All work done to the files in a repository is initially done on a working copy, hence the name. Conceptually, it is a sandbox.

Git VCS Model



Examples of Git in Use

- git add Cargo.toml src/lib.rs
- git commit
- git push origin master
- git clone metastasis@gradebot.org:/user/username/1/3
- git remote add origin metastasis@gradebot.org:user/username/1/5
- git status
- git diff
- git reset --hard
- git checkout <hash>

A Taste of Using CLI

Let's get a taste of what commands are like.

Open a terminal. You can use a tty or a terminal emulator.

For Ubuntu, use Ctrl+Alt+T to open a terminal window.

A Taste of Using CLI Cont.

Let's create a directory for the purpose of this demonstration.

```
cd ~
mkdir clidemo
cd clidemo
```

Let's look at some system information

```
uname -a
 lsb_release  -a
```

Let's look at some of the running processes

```
top
ps -el
```

A Taste of Using CLI Cont.

Okay now let's manipulate some files

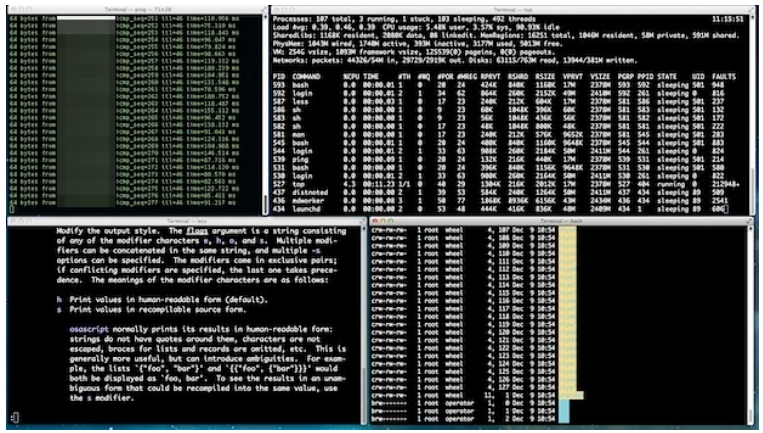
```
echo "I am a programmer" > file1
cat file1
cp file1 file2
dd if=file1 of=file3
diff -y file2 file3
mv file2 file4
rm *
```

Now for some funky stuff

```
ls /usr/bin -al | grep 'gcc'
ls /dev -al | grep 'sd'
```

WTF, don't we have GUIs already?

Why We Use CLIs



- Complete control over program and system
- Faster than using GUIs in many tasks
- Use less memory
- Offers scripting functions

A CLI-only day

Tools

- Internet Browsing:
  - links
  - w3m
- Email:
  - Mutt
  - Alpine
- Instant Messaging
  - Weechat (IRC)
- Audio Players
  - cmus
  - VLC
- Text Editors
  - vim
  - nano

Any Questions?

Thanks for coming!