

Data and Files

Rupesh Nasre.
rupesh@cse.iitm.ac.in

Tabular Data

- Examples: Excel, googlesheet, csv, ...
- When all the entries are of the same type, we can use a 2D matrix.
- In general, different columns have different types, and we can use array of structures.
 - Bills, inventory, marks, ...
- Example
 - Data with Name, City
 - Any issues?

Name	City
Ananthram	Bangalore
Chanakya	Gottivalasa
Priya	Visakhapatnam
...	...

Name, City

- Update city of Chanakya to Chennai.
- Update name of the Bangalorean to Ananth.
- Issue: There is no guarantee of **uniqueness**.
 - multiple Chanakyas, multiple Bangaloreans
 - multiple Chanakyas in Gottivalasa
- Solution: Add a unique ID.
 - Roll number, Aadhar, PAN
 - How many do we need?

Name	City
Ananthram	Bangalore
Chanakya	Gottivalasa
Priya	Visakhapatnam
...	...

Roll Number, Name, City

- Update city of Chanakya to Chennai.
 - Update city of roll number 17 to Chennai.
- Update name of the Bangalorean to Ananth.
 - Update name of roll number 7 to Ananth.
- Find names of all the people from Dehradun with roll number ≥ 11 and roll number ≤ 20 .
- Find the smallest roll number with city as Kolkata and name as Babumoshay.

Roll Number	Name	City
7	Ananthram	Bangalore
17	Chanakya	Gottivalasa
34	Priya	Visakhapatnam
...	...	...

How to store this in a computer?

Storage in a Computer

- If N is provided at run-time?
 - `Student allStudents[N];` should still work.
- What if N is large?
 - Source: `ulimit2.c`
- What if most names are small?

```
typedef struct Student {  
    int rollnum;  
    char name[100];  
    char city[100];  
} Student;
```

```
Student allStudents[N];
```

Roll Number	Name	City
7	Ananthram	Bangalore
17	Chanakya	Gottivalasa
34	Priya	Visakhapatnam
...	...	...

Storage in a Computer

```
typedef struct Student {  
    int rollnum;  
    char *name;  
    char *city;  
} Student;  
  
Student *allStudents;  
  
...  
allStudents = (Student *)malloc(N * sizeof(Student));  
int rollnum; char name[100], city[100];  
  
for (int ii = 0; ii < N; ++ii) {  
    scanf("%d%s%s", &rollnum, name, city);  
    allStudents[ii].rollnum = rollnum;  
    allStudents[ii].name = name;  
    allStudents[ii].city = city;  
}  
// print all the student information.
```

For the input

7 ananthram bangalore
17 chanakya gottivalasa

What will be the output?

Storage in a Computer

```
typedef struct Student {  
    int rollnum;  
    char *name;  
    char *city;  
} Student;
```

Now write code to deallocate allStudents.

```
Student *allStudents;
```

For reading from a file, this changes to
fscanf(fileptr, "%d%s%s", ...);
Alternatively, we can use IO redirection.

```
...
```

```
allStudents = (Student *)malloc(N * sizeof(Student));
```

```
int rollnum; char name[100], city[100];
```

```
for (int ii = 0; ii < N; ++ii) {
```

```
    scanf("%d%s%s", &rollnum, name, city);
```

```
    allStudents[ii].rollnum = rollnum;
```

```
    allStudents[ii].name = (char *)malloc((1 + strlen(name)) *  
                                         sizeof(char));
```

```
    allStudents[ii].city = (char *)malloc((1 + strlen(city)) *  
                                           sizeof(char));
```

```
    strcpy(allStudents[ii].name, name);
```

```
    strcpy(allStudents[ii].city, city);
```

```
}
```

Why not sizeof(name)?

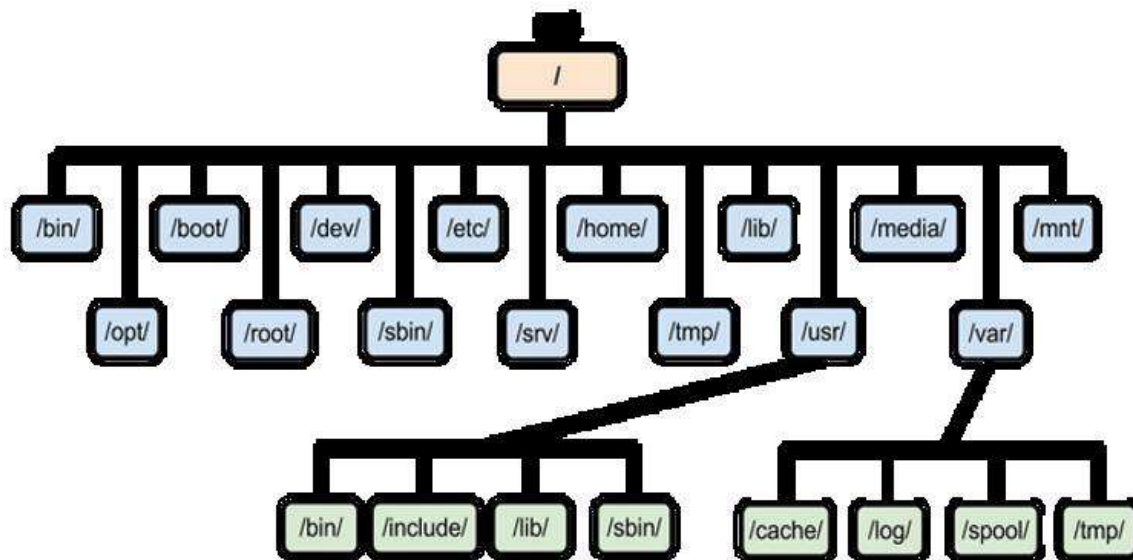


← May use strncpy.

malloc + strcpy is equivalent to **strdup**: `allStudents[ii].name = strdup(name);`

Files

- Almost everything in Linux is a file.
 - Directories, pipes, sockets, tty, USB, hard disk, devices, ...
- We deal with regular files (.c, .txt, a.out, chrome, ...)
 - `ls -l`



Support This Functionality.

- Find all Chanakya's cities.
- Find all Bangaloreans.
- Update city of Chanakya to Chennai.
- Update city of roll number 17 to Chennai.
- Update name of the Bangalorean to Ananth.
- Update name of roll number 7 to Ananth.

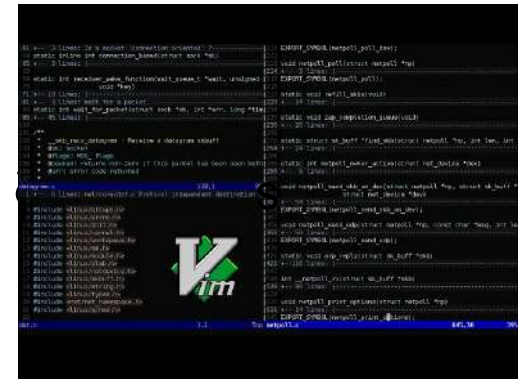
Let's also try this via Linux
command-line (grep, sed, awk).
File: id-name-city.txt

Roll Number	Name	City
7	Ananthram	Bangalore
17	Chanakya	Gottivalasa
34	Priya	Visakhapatnam
...	...	...

How to support these updates and queries? Any issues with the updates?

Try in Your Linux

- Editor: **Highlight syntax.**
- Editor: Split the window (read code problem statement together).
- Editor: Select code and auto-indent.
- Terminal: Play around with PS1 variable.
 - `export PS1="\[\e[1;42m\]\[\e[5;31m\]\w\\$ \[\e[0m\]"`
 - Understand what it means.



Homework: List faculty names.

- [This webpage](#) lists all CSE faculty members.
- From the terminal, list only their names .
- Use `grep`, `sed`, as needed.
 - Explore their options.

Practice Problems

- Use the output of `ls -l` to solve the following problems in C and shell.
 - List all directories.
 - List all files with size $\geq 1\text{MB}$.
 - List all .c files with size $\geq 1\text{MB}$.
 - Find the total size of all the files.
 - Sort the data based on file size.

In this course...

Application	Programming	Tools
Puzzle of grass, goat, lion	Pointers, structures	Editor, visualization
Data maintenance	File I/O	I/O redirection, make
Item grouping	Linked Lists	Debugging
Gene sequencing	Recursion, strings	Memory profiling, ulimit
Closest match	Matrices	Compute profiling, plotting
Keep the recent	Arrays	Version control
Cryptography	Strings, bitwise operators	Library (ncurses)
Unit testing	Command-line	Testing
Web-page creation, analysis	HTML, Strings	wget, shell scripting
Game	Graphics	Library (graphics)
Basics	Python	--

Our goal: *to have fun with automation*