

# Introduction to awk

Matt Warner

---

## 1 What is awk?

- Created by:  
Aho, Weinberger and Kernighan.
- awk is a scripting language used for manipulating data and generating reports
- Versions of awk:  
awk, nawk, mawk, pgawk, ...
- GNU awk: gawk

## 2 What can you do with awk?

- awk operation:
  - reads a file line by line
  - splits each input line into fields
  - compares input line/fields to pattern performs action(s) on matched lines
- Useful for:
  - transforming data files
  - producing formatted reports
- Programming constructs
  - format output lines
  - arithmetic and string operations
  - conditionals and loops

## 3 Basic awk invocation

### *Syntax:*

```
awk 'script' file(s)
```

```
awk -f scriptfile file(s)
```

Common option: -F (to change the field separator)

## 4 Basic awk script

- consists of patterns & actions:

### *Syntax:*

```
pattern {action}
```

- If pattern is missing, action is applied to all lines
- if action is missing, the matched line is printed
- Must have either pattern or action

### Example:

```
# print all lines containing string "for" in testfile
```

```
awk '/for/ { print }' testfile
```

---

## 5 awk variables

awk reads input line into buffers *record* and *fields*

- field buffer:
  - One for each field in the current record
  - Variable names: \$1, \$2, ...
- record buffer:
  - \$0 holds the entire record

### 5.1 More variables

NR Number of the current record

NF Number of fields in the current record

*Also:*

FS Field separator (default=whitespace)

#### Example:

Say we had a file called emps:

```
~$ cat emps
```

```
Tom Jones      4424      5/12/66      543354
Mary Adams     5346      11/4/63      28765
Sally Chang    1654      7/22/54      6500000
Billy Black    1683      9/23/44      336500
```

```
$ awk '/Tom/ { print $0 }' emps
```

```
Tom Jones      4424      5/12/66      543354
```

```
$ awk '{ print NR, $0 }' emps
```

```
1 Tom Jones      4424      5/12/66      543354
2 Mary Adams     5346      11/4/63      28765
3 Sally Chang    1654      7/22/54      6500000
4 Billy Black    1683      9/23/44      336500
```

#### Example: Space as Field Separator

```
~$ cat emps
```

```
Tom Jones      4424      5/12/66      543354
Mary Adams     5346      11/4/63      28765
Sally Chang    1654      7/22/54      6500000
Billy Black    1683      9/23/44      336500
```

```
awk '{ print NR, $1, $2, $5 }' emps
```

```
1 Tom Jones      543354
2 Mary Adams     28765
3 Sally Chang    6500000
4 Billy Black    336500
```

---

### Example: Colon as Field Separator

```
~$ cat emps2
```

```
Tom Jones:4424:5/12/66:543354
```

```
Mary Adams:5346:11/4/63:28765
```

```
Sally Chang:1654:7/22/54:650000
```

```
Billy Black:1683:9/23/44:336500
```

```
awk -F: '/Jones/{ print $1, $2 }' emps2
```

```
Tom Jones 4424
```

```
awk -F: '/Jones/{print $1, $3}' emps2
```

```
Tom Jones 5/12/66
```

### Example: File Processing

*catinput*