

-----oOo-----

Phần 1: Sử dụng awk để làm việc với file text

AWK là một ngôn ngữ lập trình hướng dữ liệu, thường được dùng cho việc xử lý dữ liệu text dựa trên việc tìm kiếm mẫu dữ liệu. Dữ liệu đầu vào được chia thành các bản ghi (dòng), mỗi bản ghi được chia thành các trường (cột). AWK thường được dùng để lọc và chuẩn hóa dữ liệu đầu ra từ dữ liệu đầu vào ban đầu.

Cú pháp:

```
awk 'BEGIN {câu_lệnh_awk} {thân_chương_trình} END {câu_lệnh_awk}'
```

Chúng ta thấy câu lệnh AWK chia thành các khối

Khối BEGIN

Cú pháp

```
BEGIN { awk_commands }
```

Khối này chỉ được chạy duy nhất một lần lúc bắt đầu, trước khi awk thực thi khối body cho tất cả các dòng trong file input

- Khối này không bắt buộc
- Từ khóa BEGIN bắt buộc phải viết hoa
- Có thể cho nhiều lệnh vào trong khối BEGIN
- Khối BEGIN này có thể hữu dụng cho việc in các report headers, khởi tạo các biến

Khối thân chương trình

Cú pháp

```
/pattern/ {action}
```

Khối lệnh trong body được thực thi mỗi lần duyệt một dòng trong input file. Ví dụ nếu trong input file có 10 record thì body sẽ được thực thi 10 lần. Không có từ khóa nào đánh dấu cho khối body này.

Khối END

Cú pháp

END { awk_commands }

Khối END chỉ được thực thi một lần ngay sau khi khối body xử lý xong toàn bộ file input. Một số đặc tính của khối END

- Không bắt buộc
- Từ khóa END phải được viết hoa
- Có thể có nhiều lệnh trong khối END
- Khối này hữu ích cho việc in report footer và làm các thao tác dọn dẹp

Một số biến định nghĩa sẵn trong AWK

\$1: Cột đầu tiên của file

\$2: Cột thứ 2 của file

\$n: Cột thứ n của file

NR: thứ tự hiện tại của bản ghi so với khởi điểm của đầu vào

FILENAME: tên của file đầu vào hiện tại

Ví dụ : File test.txt có nội dung

1	Bill_Gates	k60a2
2	Barack_Obama	k60a3
3	Lionel_Messik	k60a3
4	Kim_Jong_Un	k60a2

VD1: Câu lệnh in ra cột đầu tiên của tất cả các dòng

```
awk '{print $1}' test.txt
```

Kết quả:

```
1
2
3
4
```

VD2: Câu lệnh để in ra cột thứ 2 của file test.txt như sau:

```
awk 'BEGIN {print "Ho va ten"} {print $2} END {print "Het"}' test.txt
```

Kết quả:

Ho va ten

Bill_Gates

Barack_Obama

Lionel_Messi

Kim_Jong_Un

Het

VD3: Câu lệnh in ra tên file của file test.txt như sau:

```
awk 'BEGIN {print "Noi dung"} {print $2} END {print FILENAME; print "Het"}'
test.txt
```

Kết quả:

Noi dung

Bill_Gates

Barack_Obama

Lionel_Messi

Kim_Jong_Un

test.txt

Het

Cấu trúc điều khiển trong AWK

Awk cũng có các cấu trúc điều khiển như if...else, for, while... giống các ngôn ngữ lập trình khác

VD4: Câu lệnh để in ra tên những học sinh thuộc lớp k60a2 như sau:

```
awk 'BEGIN {print "Ho va ten"} { if ($3 == "k60a2") print $2} END {print "Het"}'
test.txt
```

Kết quả:

Ho va ten
Bill_Gates
Kim_Jong_Un
Het

VD5: Câu lệnh để in ra mỗi dòng 4 lần

```
awk 'BEGIN {print "Ket qua"} {for (i=1;i<3;i++) print "In dong ",$1,"lan" ,i} END {print "Het"}}' test.txt
```

Hoặc

```
awk 'BEGIN {i=1; print "Ket qua"} {while (i<3){print "In dong ",$1,"lan" ,i; i++}} END {print "Het"}}' test.txt
```

Kết quả:

Ket qua
In dong 1 lần thứ 1
In dong 1 lần thứ 2
In dong 2 lần thứ 1
In dong 2 lần thứ 2
In dong 3 lần thứ 1
In dong 3 lần thứ 2
In dong 4 lần thứ 1
In dong 4 lần thứ 2
Het

Các hàm cơ bản trong Awk

Hàm lấy độ dài của chuỗi (length)

VD6: Lấy ra độ dài của tên các học sinh

```
awk 'BEGIN {print "Ket qua"} {print length($2)} END {print "Het"}}' test.txt
```

Kết quả:

10
12
12
11

Hàm viết hoa hoặc viết thường cả chuỗi (toupper/tolower)

VD7:

```
awk 'BEGIN {print "Ket qua"} {print toupper($2)} END {print "Het"}' test.txt
```

Kết quả:

Ket qua

BILL_GATE

BARACK_OBAMA

LIONEL_MESSI

KIM_JONG_UN

Het

```
awk 'BEGIN {print "Ket qua"} {print toupper($2)} END {print "Het"}' test.txt
```

Kết quả:

Ket qua

bill_gates

barack_obama

lionel_messi

kim_jong_un

Het

Hàm cắt chuỗi **substr**

Cú pháp: *substr(chuỗi_đầu_vào, cắt_từ_vị_trí, cắt_đến_vị_trí)*

VD: Cắt từ đầu đến ký tự thứ 4 của tên học sinh

```
awk 'BEGIN {print "Ket qua"} {print substr($2,0,4)} END {print "Het"}' test.txt
```

Kết quả:

Ket qua

Bill

Bara

Lion

Kim_

Het

Hàm lấy ra số thứ tự của ký tự muốn tìm đầu tiên trong chuỗi (INDEX)

Cú pháp: *index(chuỗi_đầu_vào, ký_tự_muốn_tìm_vị_trí)*

VD: Tìm vị trí của ký tự l trong chuỗi họ tên sinh viên:

```
awk 'BEGIN {print "Ket qua"} {print index($2,"l")} END {print "Het"}' test.txt
```

Kết quả:

Ket qua

3

0

1

0

Het

Phần 2: Bài tập thực hành

Bài 1: Cho file diemso.txt như sau:

1 *Nguyen_Van_A* 5 6 7

2 *Pham_Thi_B* 6 5 4

3 *Nguyen_Van_C* 9 6 8

3 cột cuối là điểm của 3 môn Toán, Lý, Hóa. Hãy dùng awk để hiển thị điểm trung bình của các học sinh như sau:

Diem trung binh

Nguyen_Van_A 6

Pham_Thi_B 5

Nguyen_Van_C 7.6666

Ket thuc

Bài 2:Hiển thị họ của tất cả các học sinh trong lớp.

Bài 3: Cho file sinhvien.txt có nội dung như sau:

Sau day la thong tin cac sinh vien trong lop K60A2

1 *Nguyen_Van_A* *Thaibinh*

2 *Pham_Thi_B* *Namdinh*

3 *Nguyen_Van_C* *Thanhhoa*

Hãy hiển thị họ và tên các sinh viên trong lớp K60A2

Phần 3: Liên lạc

STT	Họ và tên	Email	ĐT
1	Nguyễn Minh Hải	nguyenminhhai06@gmail.com	
2	Nguyễn Thị Tâm	nguyenthitam.hus@gmail.com	
3	Trần Thị Hương	tranthihuong.hus@gmail.com	