

fibm11> echo "A pancia piena meglio si programma!"

Gruppo L: Informatica

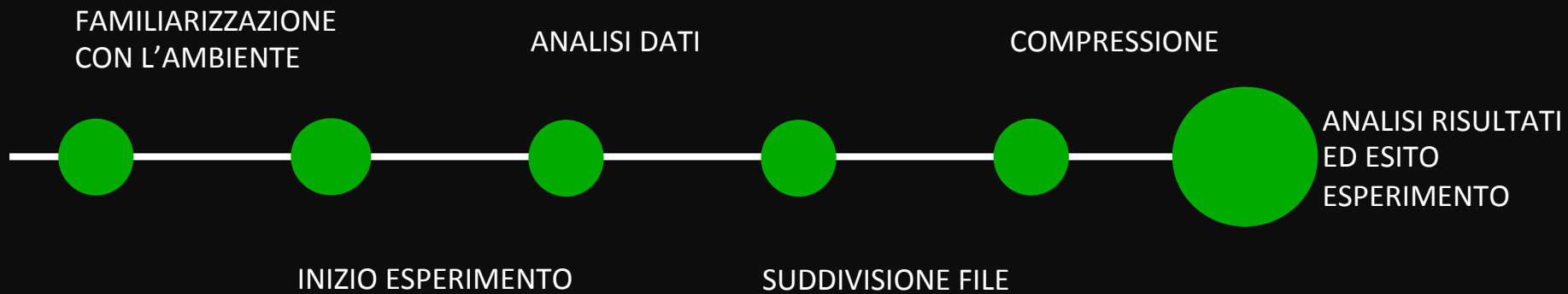
**PROGETTAZIONE E GESTIONE DI UN SISTEMA DATI PER UN
ESPERIMENTO CHE PRODUCE ALMENO 5400 TERABYTE.**

fibm11> echo "A pancia piena meglio si programma!"

SCOPO DEL PROGETTO:

Progettazione di un **algoritmo** dedicato all'**analisi** e alla **compressione** dei **dati** provenienti da KLOE.

fibm11> echo "A pancia piena meglio si programma!"



fibm11> echo "A pancia piena meglio si programma!"

AMBIENTE DI LAVORO:

The screenshot displays a terminal window titled 'dtterm' with a C program being edited. The program is a simple loop that prints the number of command-line arguments. Below the editor, a file system listing is visible, showing various files and directories with their permissions, sizes, and dates. The listing includes files like '68577792', '1149450267', and '68577792', as well as directories like '/home', '/admin', and '/proc'. The terminal prompt 'fibm11>' is visible at the bottom left.

```
#include<stdio.h>
#include<stdarg.h>

int main(int argc, char **argv)
{
    FILE *fin;
    int f,c;
    long int counter[256] = {0};
    if(argc < 2)
    {
        printf("Errore! numero parametri insufficiente\n");
        exit(1);
    }
    fin = fopen64(argv[1], "rb");
    while((c=getc(fin)) != EOF) counter[c]++;
    fclose(fin);

    int i;

    for(i = 0; i < 256; i++)
    {
        printf("\t%d\t->\t%d\n", i, counter[i]);
    }
}
```

```
total 33553792
-rw-r----- 1 stages1 usr 68577792 Jun 21 14:45 out3
-rw-r----- 1 stages1 usr 1149450267 Jun 21 12:15 out3.Z
-rw-r----- 1 stages1 usr 68577792 Jun 21 14:45 out4
-rw-r----- 1 stages1 usr 1350340481 Jun 21 12:15 out4.Z
-rw-r----- 1 stages1 usr 1073741824 Jun 11 10:49 raw038298N.000
-rw-r----- 1 stages1 usr 8883334592 Jun 11 10:51 raw077254N_ALL_r05_1_1_1.000
-rw-r----- 1 stages1 usr 274311160 Jun 11 10:51 raw077254N_L3BRA_r05_1_1_1.000
-rw-r----- 1 stages1 usr 2320701440 Jun 11 10:52 raw077254N_L3C0S_f05_1_1_1.000
-rw-r----- 1 stages1 usr 1148332032 Jun 11 10:52 raw077254N_L3DCC_f05_1_1_1.000
-rw-r----- 1 stages1 usr 6312 Jun 19 15:27 uscita
-rw-r----- 1 stages1 usr 4356 Jun 20 15:25 uscita2
```

```
% /home
1% /admin
/proc
/opt
% /scratch
% /var/adm/ras/livedump
28% /kloe/soft
4918 22% /calcolo/lnfa
84918 22% /calcolo/lnf
4% /runcond
1% /data/farm04
1% /data/farm08
2% /analysis
1% /ke00
1% /u
1% /var/loadl/spool
stdin
```

Calcolatore con **IBM AIX**

Editor testuale **Vi**

Linguaggio di programmazione **C**

SCREENSHOT DELL'AMBIENTE DI SVILUPPO

fibm11> echo "A pancia piena meglio si programma!"

KLOE produce ad ogni acquisizione 60 file da 8 GB l'uno con un flusso di 80 MB/s.

Le acquisizioni vengono salvate su nastri magnetici da 10 TB, che vengono riempiti ogni 2 giorni.

Dato l'elevato costo dei supporti di memorizzazione è necessario comprimere i dati prodotti per ridurre la spesa.

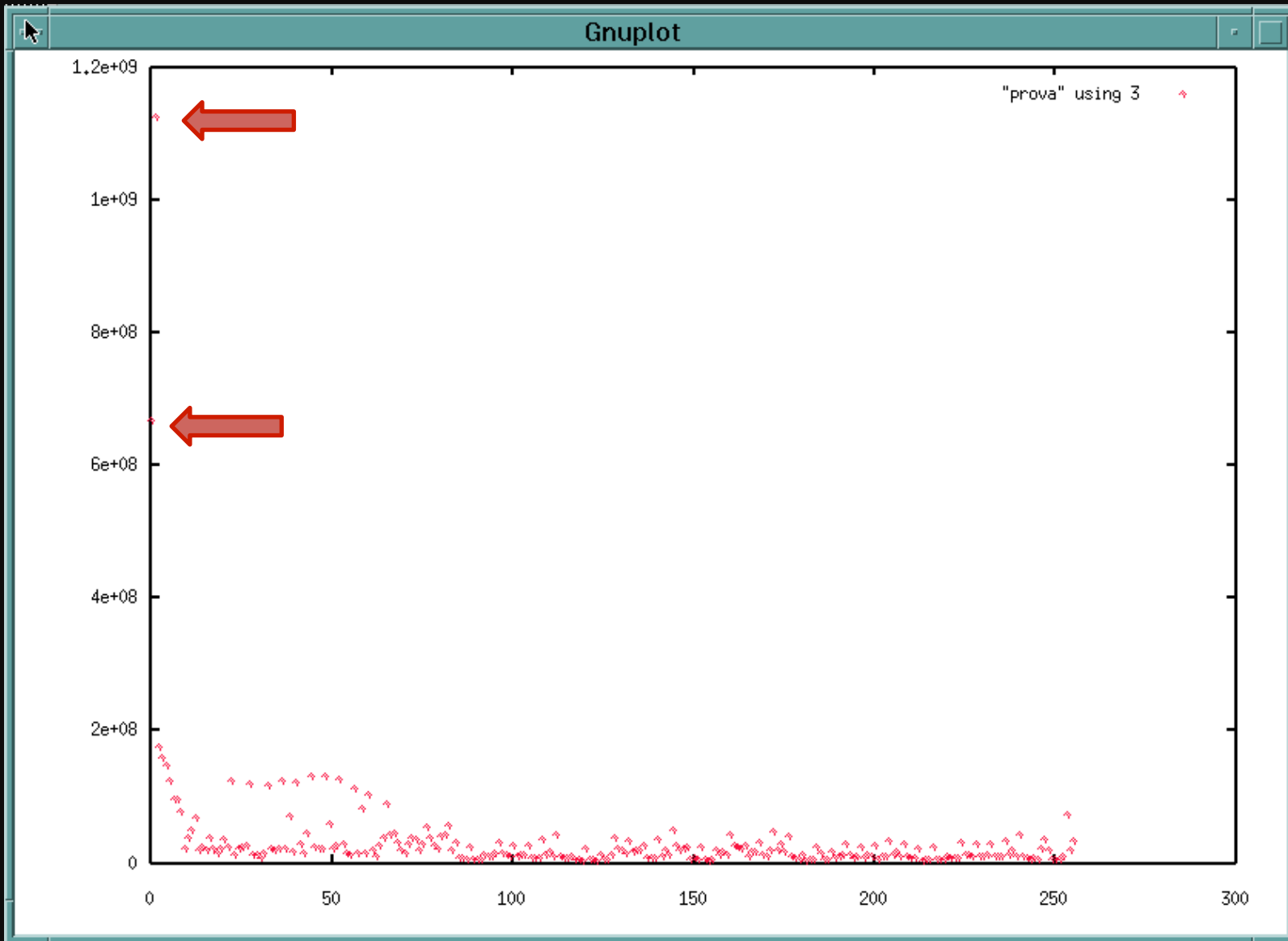
fibm11> echo "A pancia piena meglio si programma!"

Sono stati **analizzati** i file **grezzi** prodotti da **KLOE** per avere un'idea sul loro **contenuto**.

Tramite un **programma** sono stati analizzati i file **byte** per **byte**.

É stata notata una **ripetizione** dei valori **0** e **1**, circa **10** volte maggiore rispetto agli **altri**.

fibm11> echo "A pancia piena meglio si programma!"



FREQUENZA DI RIPETIZIONE DEI BYTE

```
fibm11> echo "A pancia piena meglio si programma!"
```

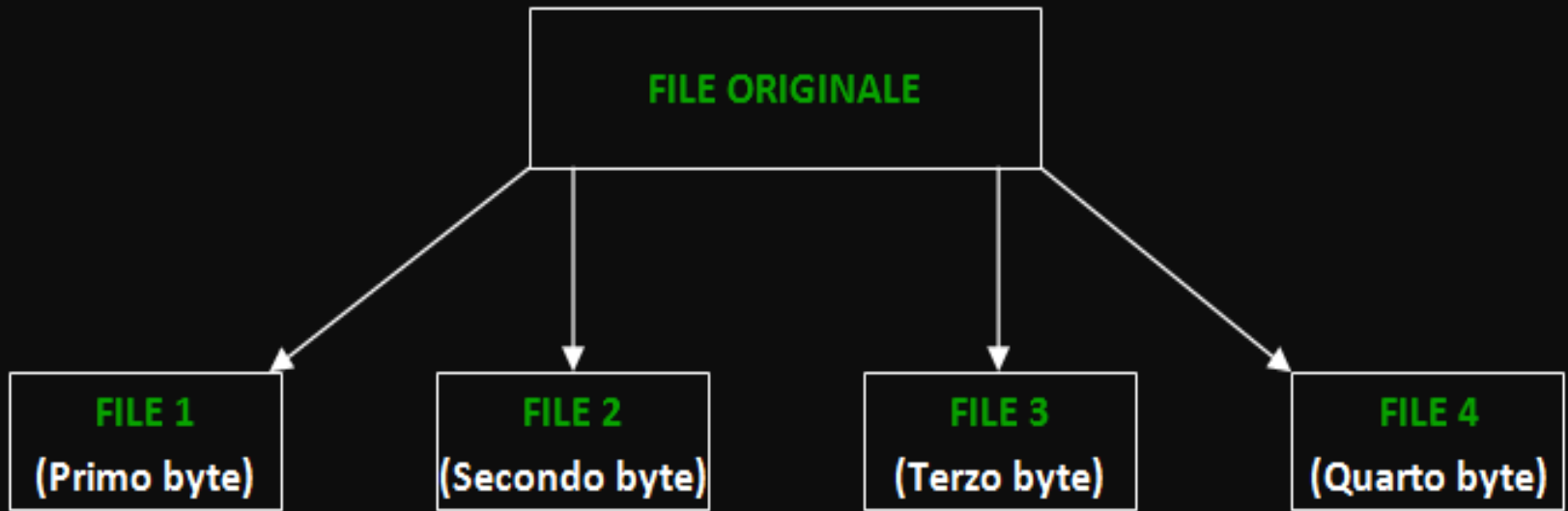
I numeri sul file erano scritti in floating point a 32 bit (4 byte).

Questi sono stati divisi in byte, e collocati in 4 file diversi in base alla loro posizione all'interno del numero.

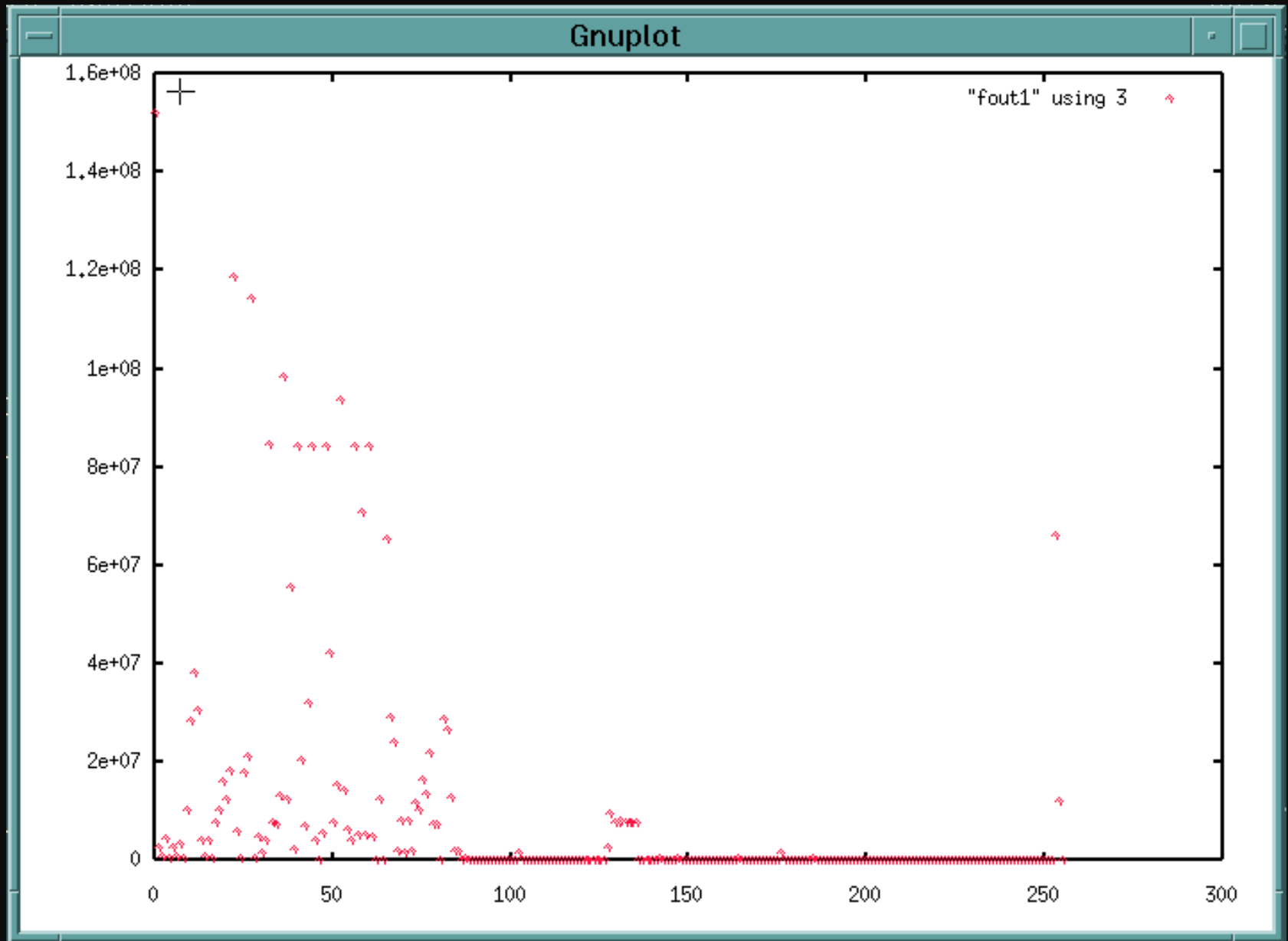
L'oscillazione è presente maggiormente nell'ultimo byte, diminuendo progressivamente fino al primo.

fibm11> echo "A pancia piena meglio si programma!"

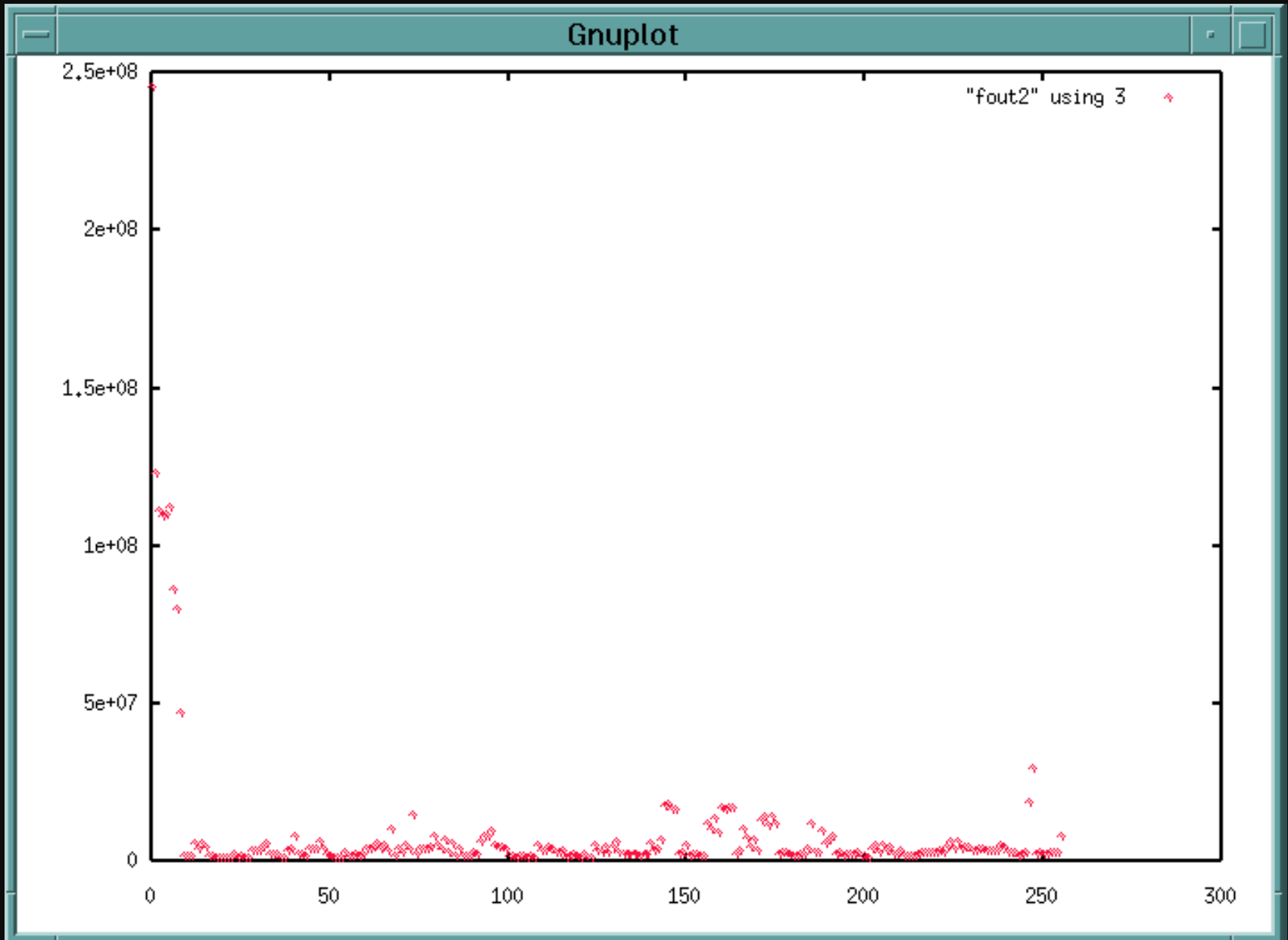
SCOMPOSIZIONE DI UN FLOATING POINT:



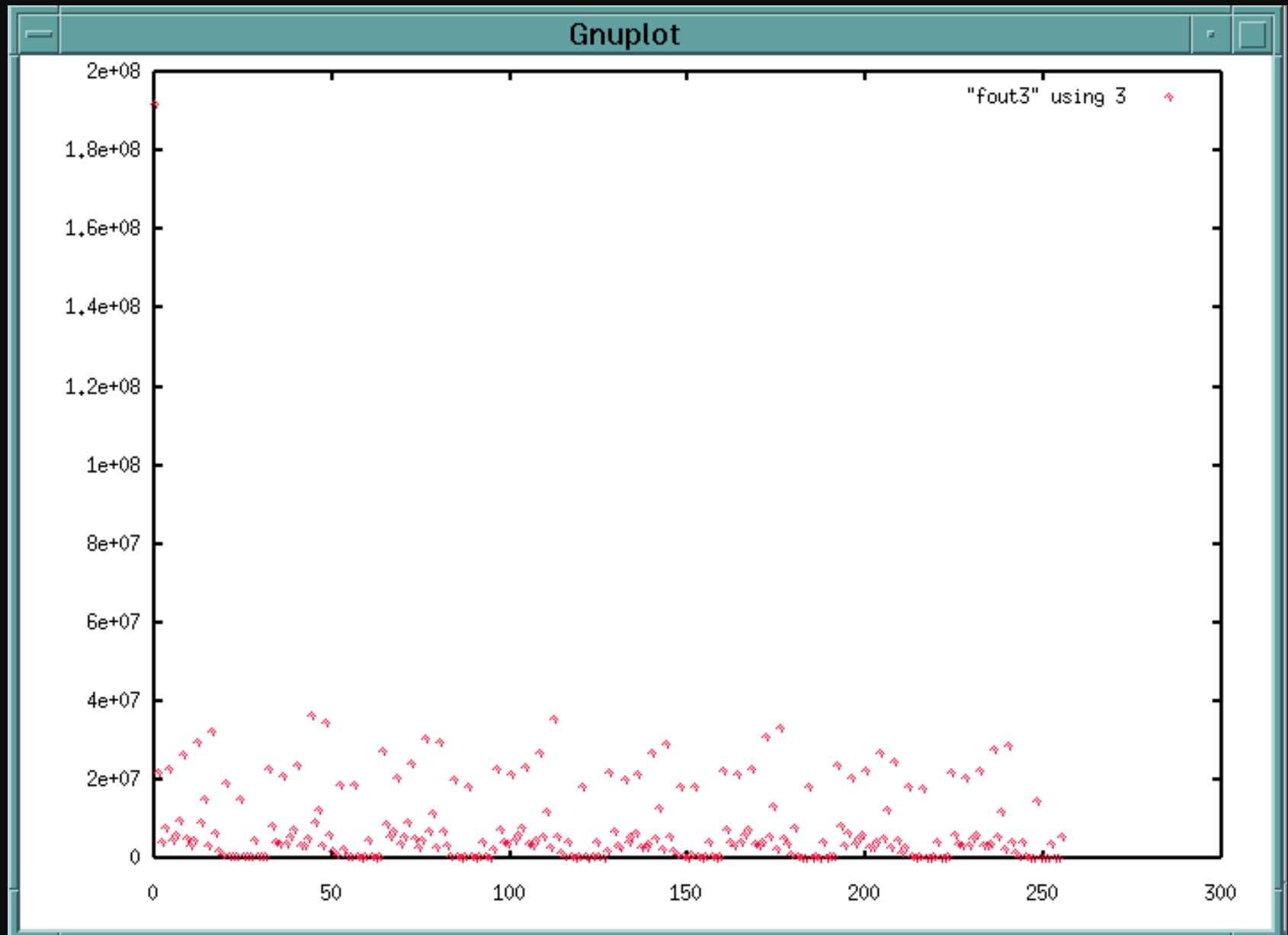
fibm11> echo "A pancia piena meglio si programma!"



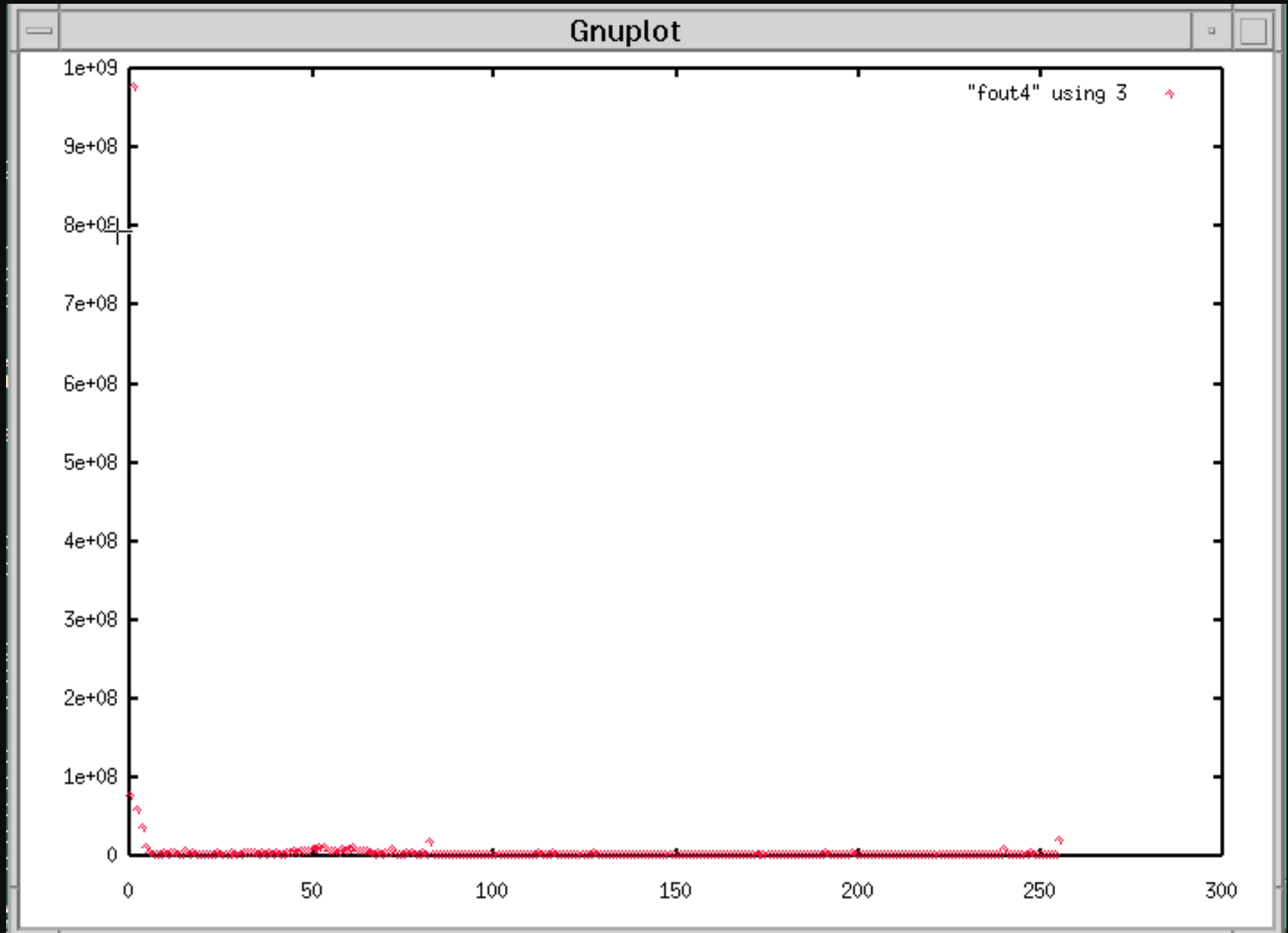
fibm11> echo "A pancia piena meglio si programma!"



fibm11> echo "A pancia piena meglio si programma!"



fibm11> echo "A pancia piena meglio si programma!"



fibm11> echo "A pancia piena meglio si programma!"

Grazie a questa **soluzione**, è possibile comprimere i **4 file** ottenendo una compressione quasi del **60%**.

TIPO DI FILE	COMPRESSIONE
ALL	59,29%
BAH	54,5%
COS	57,75%
DCC	58,25%
MEDIA	57,45%

fibm11> echo "A pancia piena meglio si programma!"

In **conclusione** tramite questi passaggi è stato possibile **comprimere** i file **grezzi** provenienti da **KLOE** riducendone il volume quasi del **60%**, **riducendo** così **significativamente** lo spazio necessario alla loro **memorizzazione** e di conseguenza la **spesa** per i **supporti**.