

## Introducción a la Arquitectura de Sistemas – Apunte Representación de Números en Coma Flotante

Uno de los mayores inconvenientes que presentan las **representaciones de coma fija** es la imposibilidad de representar cantidades con diferencias de varios órdenes de magnitud, debido a que el error absoluto está fijo.

Por ejemplo, si se desea medir la distancia de la tierra a la luna se puede tolerar un error de algunas decenas de kilómetros. Si se mide la distancia entre dos ciudades ese mismo error no sería aceptable. Si se mide la distancia entre dos lugares dentro de una ciudad pequeña, el error invalidaría la medida.

Por esto, elegimos los sistemas de coma flotante que tienen como ventaja representar en un mismo sistema, números muy grandes y números muy pequeños donde el **error relativo** (Avance/ValorMax) se mantiene constante y el **error absoluto** (Avance) varíe en función del orden de magnitud elegido.

### ➤ Sistema **IBM 360**

Posee un ancho de palabra de 32 bits, los cuales están distribuidos de la siguiente forma:

<b>S</b>	<b>exponente CD(2,7)</b>	<b>mantisa SVA(16,6)</b>
<b>31 30</b>	<b>24 23</b>	<b>0</b>

- Signo: ocupando el bit mas significativo, (0 positivo, 1 negativo).
- Exponente: expresado en CD(2,7) con frontera equilibrada.

$$f = \frac{b^d}{2} = 1000000$$

- Mantisa: expresada en SVA(16,6) con normalización 0,X.  
La representación en función de la mantisa y el exponente es

$$r = 0, m \times 10_h^e$$

Ejemplo, convertir a IBM 360 el siguiente número:

$$125,42 \times 10^{25} = 125,42 \times \frac{10^{25}}{16^y} 16^y$$

$$\frac{10^{25}}{16^y} = 1 \rightarrow 10^{25} = 16^y$$

$$\log 10^{25} = \log 16^y$$

$$25 = y \cdot \log 16 \rightarrow y = 20,76$$

Entonces:

$$125,42 \times 10^{25} = 125,42 \times 16^{20,76}$$

$$125,42 \times 10^{25} = 125,42 \times 16^{0,76} \times 16^{20}$$

$$125,42 \times 10^{25} = 1031,57 \times 16^{20} = 407,91_h \times 16^{20}$$

$$125,42 \times 10^{25} = 0,40791_h \times 16^{23}$$

signo = 0

exponente:  $23 = 10111_b$  en  $CD(2,7) = (1\ 000\ 000_b + 10111_b) = 1\ 010\ 111_b$

mantisa:  $40791_h$  SVA(16,6) normalización 0,X

IBM 360 = 0 1010111<sub>b</sub> 407910<sub>h</sub>

empaquetado = 57407910<sub>h</sub>

➤ Sistema **PDP/11**

Este sistema ofrece diferencias significativas con respecto al IBM 360, también posee un ancho de 32 bits, y su distribución es la siguiente:

<b>S</b>	<b>exponente CD(2,8)</b>	<b>mantisa SVA(2,24)</b>
<b>31 30</b>	<b>23 22</b>	<b>0</b>

- Exponente: expresado en CD(2,8) con frontera equilibrada

$$f = \frac{b^d}{2} = 10000000$$

- Mantisa: con normalización 0,1X donde el 1 es un bit oculto o implícito, es decir, no es almacenado en la representación, permitiendo así una ganancia de precisión.

Ejemplo, convertir a PDP/11 el siguiente número:

$$-128 \times 10^{-20} = -128 \times \frac{10^{-20}}{2^y} 2^y$$

$$\frac{10^{-20}}{2^y} = 1 \rightarrow 10^{-20} = 2^y$$

$$\log 10^{-20} = \log 2^y$$

$$-20 = y \cdot \log 2 \rightarrow y = -66,43$$

Entonces:

$$-128 \times 10^{-20} = -128 \times 2^{-66,43}$$

$$-128 \times 10^{-20} = -128 \times 2^{-0,43} \times 2^{-66}$$

$$-128 \times 10^{-20} = -95,01 \times 2^{-66} = -1011111,0001_b \times 2^{-66}$$

$$-128 \times 10^{-20} = -0,10111110001_b \times 2^{-59}$$

signo = 1

exponente = -59 = -111011<sub>b</sub>      en CD(2,8) = (10 000 000<sub>b</sub> + -111011<sub>b</sub>) = 01000101<sub>b</sub>

mantisa = 011110001<sub>b</sub>      SVA(2,24) normalización 0,1X

PDP/11 = 1 01000101 011110001000000000000000<sub>b</sub>

empaquetado = A2BC4000<sub>h</sub>

➤ Sistema **IEEE 754**

Este sistema, a diferencia de los anteriores, permiten representar valores especiales.

El estándar define representaciones para números de coma flotante, con precisión simple y doble utilizando anchos de palabra de 32 y 64 bits respectivamente, los cuales están distribuidos de la siguiente forma:

IEEE 754 corto (32 bits):

<b>S</b>	<b>exponente CD(2,8)</b>	<b>mantisa SVA(2,24)</b>
<b>31 30</b>	<b>23 22</b>	<b>0</b>

- Exponente: expresado en CD(2,8) con frontera no equilibrada.

$$f = \frac{b^d}{2} - 1 = 01111111$$

- Mantisa: con normalización 1,X

La representación en función de la mantisa y el exponente es:

$$r = 1, m \times 10_b^e$$

Ejemplo, convertir a IEEE 754 corto el siguiente número:

$$2,5 \times 10^{-39} = 2,5 \times \frac{10^{-39}}{2^y} \times 2^y$$

$$\frac{10^{-39}}{2^y} = 1 \rightarrow 10^{-39} = 2^y$$

$$\log 10^{-39} = \log 2^y$$

$$-39 = y \cdot \log 2 \rightarrow y = -129,55$$

Entonces:

$$2,5 \times 10^{-39} = 2,5 \times 2^{-129,55}$$

$$2,5 \times 10^{-39} = 2,5 \times 2^{-0,55} \times 2^{-129}$$

$$2,5 \times 10^{-39} = 1,7 \times 2^{-129} = 1,1011_b \times 2^{-129}$$

$$2,5 \times 10^{-39} = 0,011011_b \times 2^{-127}$$

signo = 0

exponente = -127 = -01111111<sub>b</sub> en CD(2,8) = (01111111<sub>b</sub> + 01111111<sub>b</sub>) = 00000000<sub>b</sub>

mantisa = 011011<sub>b</sub> SVA(2,24) normalización 0,1X

IEEE 754 corto = 0 00000000 011011000000000000000000<sub>b</sub>

empaquetado = 00360000<sub>h</sub> Es un número Subnormal.

IEEE 754 largo (64 bits):

<b>S</b>	<b>exponente CD(2,11)</b>	<b>mantisa SVA(2,53)</b>
<b>63 62</b>	<b>52 51</b>	<b>0</b>

- Exponente: expresado en CD(2,11) con frontera no equilibrada.

$$f = \frac{b^d}{2} - 1 = 11111111$$

- Mantisa: con normalización 1,X  
La representación en función de la mantisa y el exponente es:

$$r = 1, m \times 10_b^e$$

Existen, además de los sistemas de precisión simple (32 bits) y precisión doble (64 bits) otros sistemas, como el de precisión extendida ( $\geq 80$  bits) y de precisión cuádruple (128 bits)