

27 September 2023

DASAR SISTEM KOMPUTER



Nama dosen pengampu:

Dikerjakan oleh

Nama : Muhammad Rafi Rizaldi

NRP : 3123600001

Kelas : 1 D4 IT A

TUGAS 5

1. Jelaskan tentang format bilangan Float dengan format FP16, FP32, FP64 dan BF 16

Format bilangan floating-point (float) adalah cara untuk merepresentasikan angka desimal dengan presisi tertentu dalam komputasi. Ada beberapa format floating-point yang berbeda, termasuk FP16, FP32, FP64, dan BF16. Mari kita bahas masing-masing dari format tersebut:

1. FP16 (Floating-Point 16-bit):

- Juga dikenal sebagai "half-precision."
- Terdiri dari 16 bit.
- Terdiri dari 1 bit untuk tanda (positif/negatif), 5 bit untuk eksponen, dan 10 bit untuk mantissa (atau fraksi).
- Presisi rendah, cocok untuk aplikasi di mana penggunaan memori dan kecepatan lebih penting daripada akurasi numerik yang tinggi.

2. FP32 (Floating-Point 32-bit):

- Juga dikenal sebagai "single-precision."
- Terdiri dari 32 bit.
- Terdiri dari 1 bit untuk tanda, 8 bit untuk eksponen, dan 23 bit untuk mantissa.
- Digunakan secara luas dalam komputasi umum, termasuk sebagian besar perangkat keras dan perangkat lunak.

3. FP64 (Floating-Point 64-bit):

- Juga dikenal sebagai "double-precision."
- Terdiri dari 64 bit.
- Terdiri dari 1 bit untuk tanda, 11 bit untuk eksponen, dan 52 bit untuk mantissa.
- Memberikan presisi yang lebih tinggi dibandingkan FP32 dan sering digunakan dalam perhitungan ilmiah dan perhitungan yang memerlukan akurasi tinggi.

4. BF16 (BFloat16 atau Brain Floating-Point 16-bit):

- Format yang mirip dengan FP16 tetapi memiliki struktur yang sedikit berbeda.
- Terdiri dari 16 bit.
- Terdiri dari 1 bit untuk tanda, 8 bit untuk eksponen, dan 7 bit untuk mantissa.

- Dikembangkan oleh Google untuk digunakan dalam mesin cerdas dan kecerdasan buatan karena keseimbangan antara presisi numerik yang cukup dan kebutuhan komputasi yang lebih ringan.

Setiap format floating-point memiliki trade-off antara presisi numerik dan penggunaan memori/kecepatan komputasi. Pemilihan format tergantung pada aplikasi yang spesifik dan seberapa penting akurasi numerik dalam kasus tersebut. Semakin tinggi bit yang digunakan untuk mantissa dan eksponen, semakin tinggi pula presisi numeriknya, tetapi semakin besar pula penggunaan memori dan waktu yang dibutuhkan untuk perhitungan.

2. Build dan eksekusi program main.c yang ada di <https://github.com/ferryastika/floatConversion/blob/main/README.md> lalu amati untuk : a) ./main 3.141592654 b) ./main NaN c) ./main 0.0 d) ./main Inf

[illegible]