For this string:

Jéan   (é is code point 233)
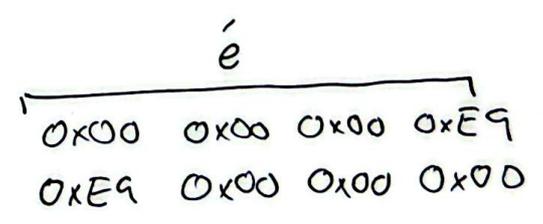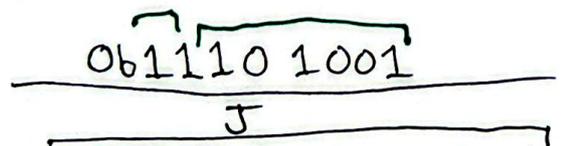
Q1: What is the UTF8 encoding?

Q2: What is the UTF32 encoding?

Q3: What is strlen of the UTF8 encoding? The UTF-32?

J	é	a	n
74	233	97	110

code points. 0x4A   0xE9   0x61   0x6E
in hex ~~(25000)~~

0b01001010

0b1110 1001
J

UTF32 encoding:

0x00 0x00 0x00 0x4A   0x00 0x00 0x00 0xE9
0x4A 0x00 0x00 0x00   0xE9 0x00 0x00 0x00

é

strlen returns 0 for top version
            1 for bottom version

é

UTF8: 0x4A 0xC3 0xA9 0x61 0x6E

strlen returns 5

| string.h |

int strlen (char* s)

void strcpy (char* dest, char* src)   copies bytes from
                                        src to dest

void strcat (char* dest, char* src)
                                        └ appends src to the
                                          end of dest

char str[]   vs.   char* str

In function arguments, these mean the same thing

char* : an address in memory where we can access char data

T* : an address in memory where we can access T data

)
T could be
int
double   Pointer Types
int16_t
char*   address = number representing a place in
                            memory (storage in our computer)

For a pointer T* p;

        p[index]   looks up the value offset by index
                        from p in memory

        p[index] = v   change the value offset by index
                            from p to the value v

x is a variable of type T
    &x   ⇒ evaluates to a T* that is
            the address where x is stored

Hashing

```
$ git commit -m "almost done"
[main 211935b] almost done
1 file changed ...
$ git log
commit: 211935b0cf003 ....  (40 hex char = 20 bytes)
```

```
$ ssh jpolitz@ieng6
ED25519 key fingerprint  SHA256: 8avDdt0 ...
```

                    just a byte array   32 bytes
                           ↓   ↓
SHA256 (char* input, int size, char* hash)

calculates a fixed-size "hash" of that data

   — deterministic
   — "one-way": hard to guess the input from output
   — unpredictably distributed: similar inputs produce different outputs