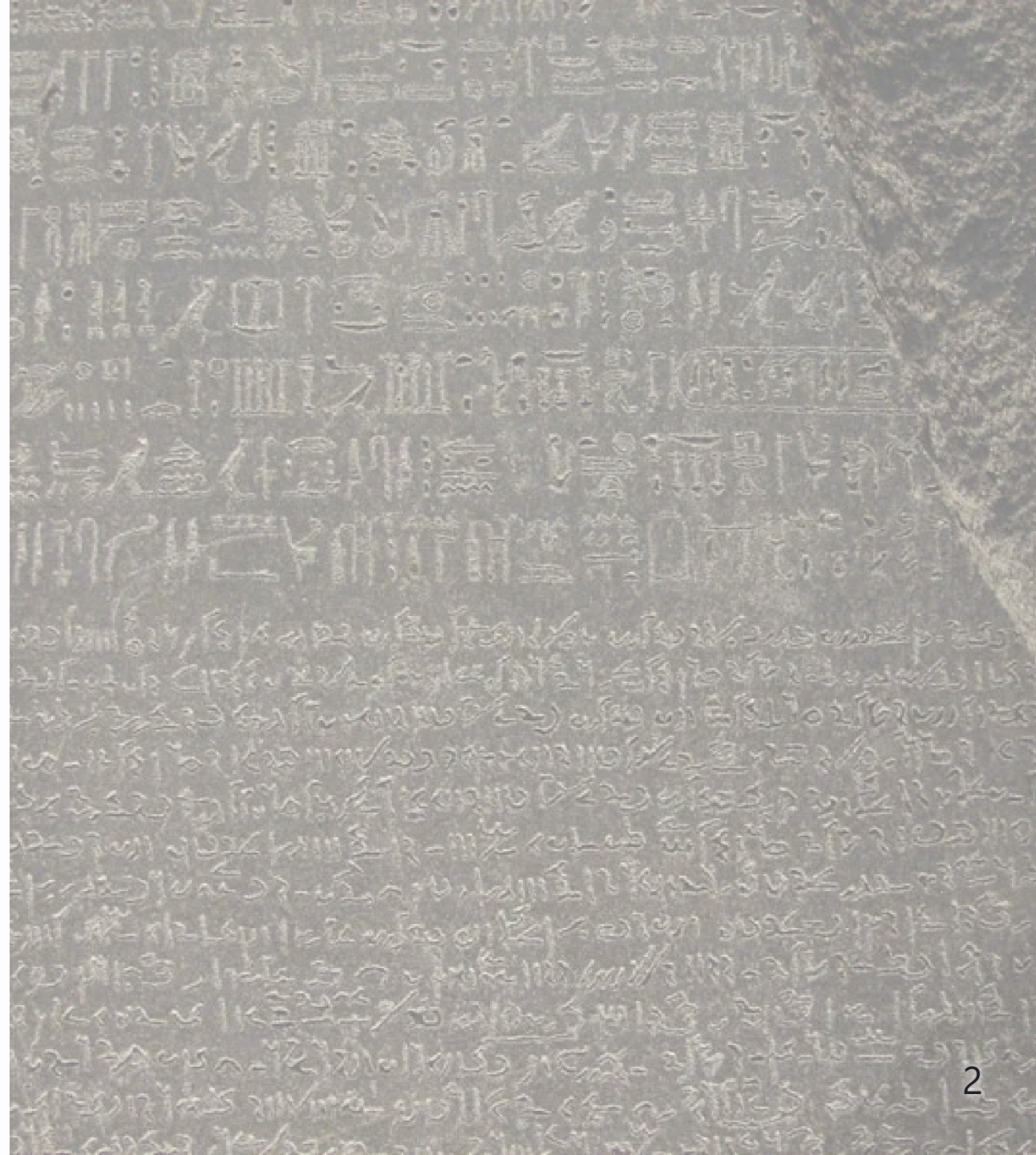# What a Character

`char` Data Type in Java, C# and JavaScript

Sualeh Fatehi

What is a char ?

What is a `char` **?**

- Unsigned primitive data type
- `char` is 16 bits in Java, C# and JavaScript
- Numeric value is a code point on the Basic Multilingual Plane (BMP)

Code Point Size

But, Unicode code points

- need **5 bits** for 17 code planes
- need **16 bits** for 65,536 in each code plane
- = **21 bits**

Solution: Surrogate Pairs

So,

- `char` is only for characters in the BMP
- Use **surrogate pairs** for characters outside the BMP
- Surrogate pairs can only be used in strings
- Language APIs allow for integer code points instead of surrogate pairs

Beware of Breakage

- String length gives number of `char` not number of code points
- Some string methods, such as substring may split surrogates
- String builder delete or remove methods may split surrogates

Code Examples

Slides and all code examples are on GitHub

https://github.com/**sualeh/What-a-Character**