

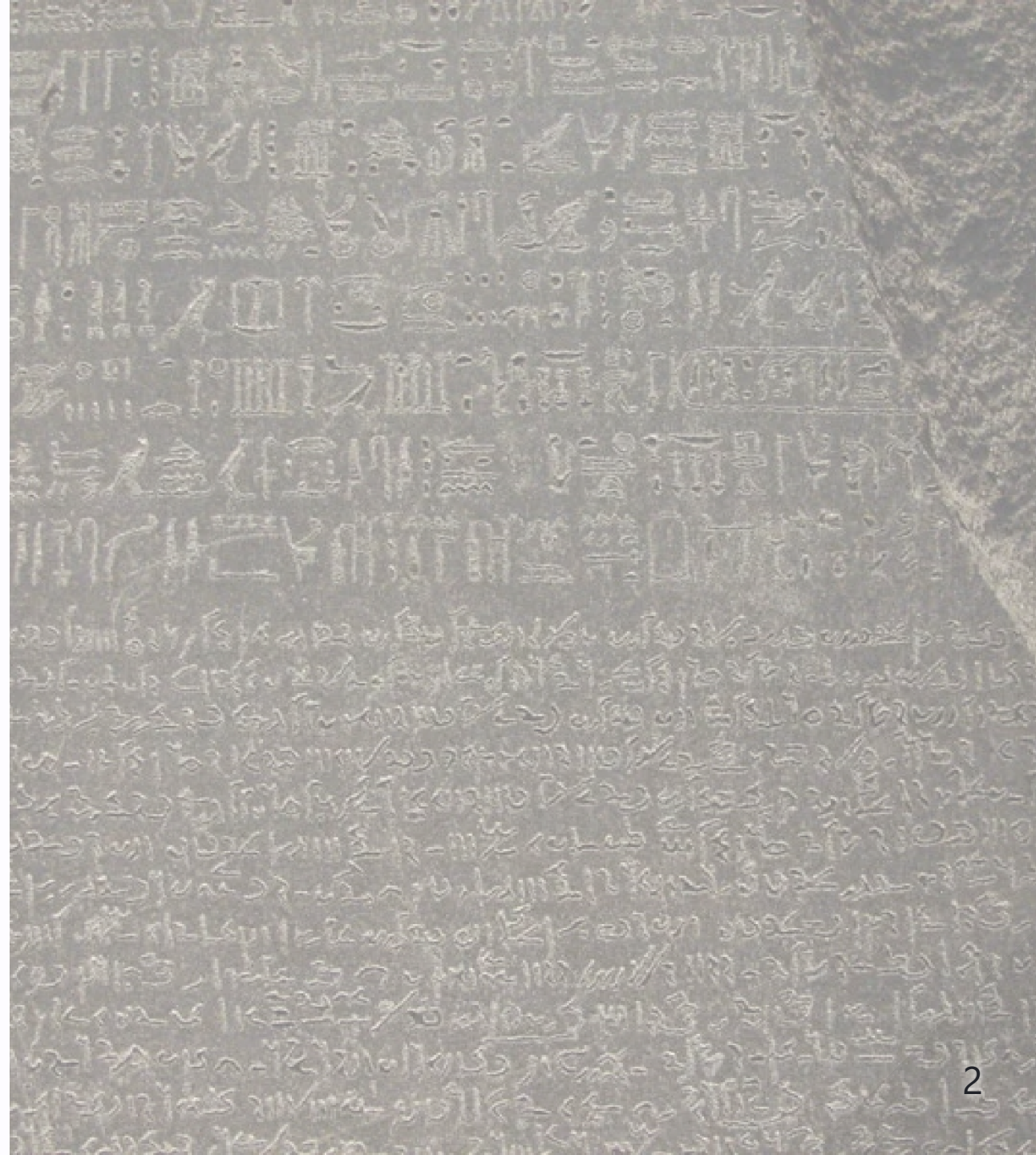
What a Character

Encoding in Java

Sualeh Fatehi

Encoding

See the [presentation on encoding concepts](#).



Compact Strings

- Before Java 9, strings were internally `char[]`
- Java now represents strings internally as `byte[]`
- External APIs are unchanged
- Two encodings are supported internally
 - UTF-16 uses one or two bytes per character
 - Latin-1 uses one byte per character

Charset

- Numerous character-encoding schemes supported via the `Charset` class
- Not all encoding schemes may be installed by default
- Many IBM EBCDIC code pages are supported
- `StandardCharsets` defines encodings that are always available
- Every encoding has a **canonical name** or MIME-preferred name

Code Examples

Slides and all code examples are on GitHub

<https://github.com/sualeh/What-a-Character>

