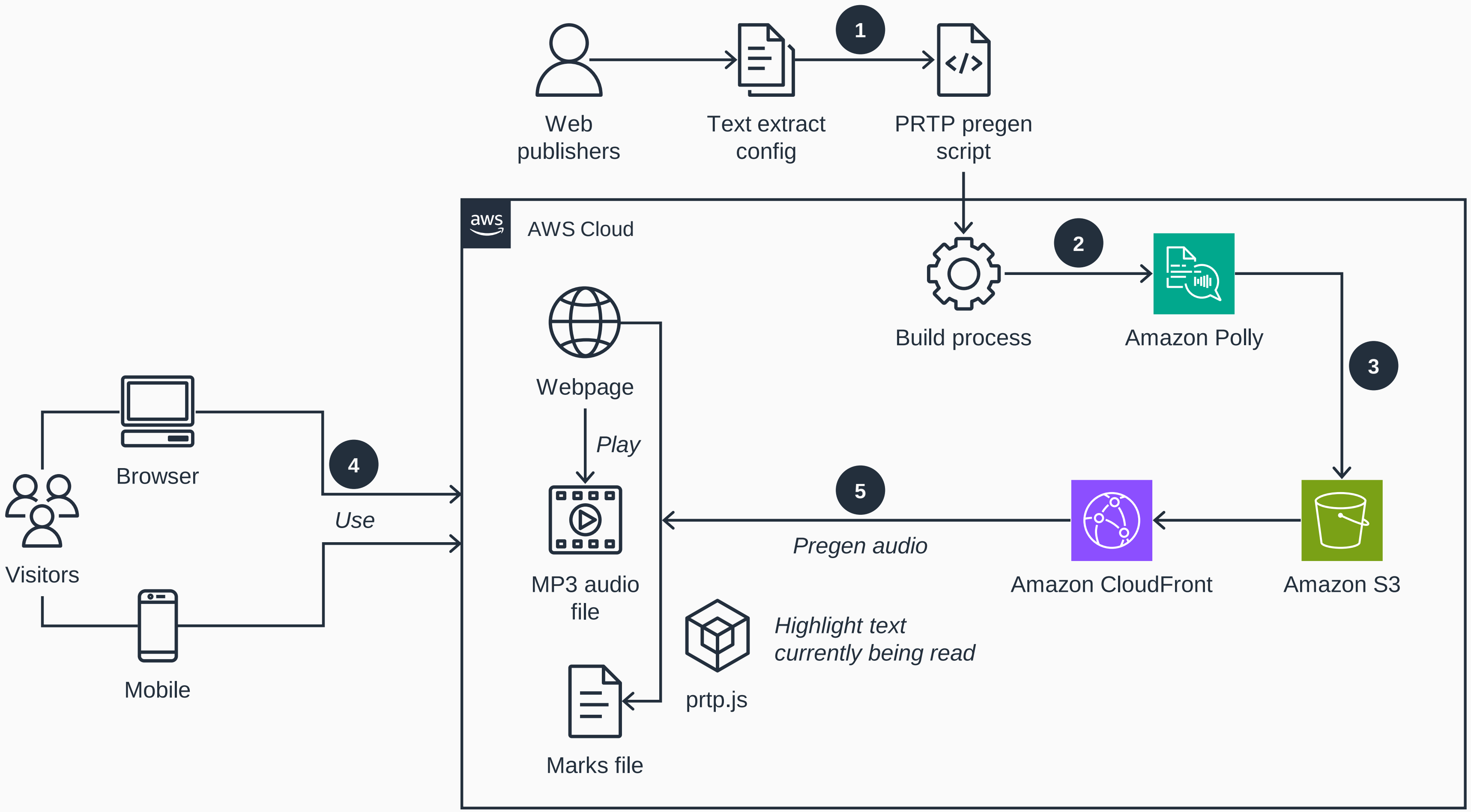
Guidance for Reading & Highlighting Web Content with Amazon Polly

Static Webpages

This architecture diagram shows how to use Amazon Polly to read and highlight content on static webpages. Amazon Polly outputs the files in a storage bucket. When you play the audio, the browser downloads the MP3.



1. Web publishers pre-generate the audio files at build time by defining a text extraction configuration (extract config) file in order to identify the content to be converted from the HTML page.

2. A PollyReadsThePage (PRTP) pre-generation process invokes **Amazon Polly** to generate an audio version of the specified text. This process takes the HTML page and the extract config file as inputs.

3. **Amazon Polly** stores the generated audio file in an **Amazon Simple Storage Service (Amazon S3)** bucket.

4. When visitors play the audio using a browser, the audio is downloaded from the **Amazon S3** bucket through **Amazon CloudFront**.

5. As visitors listen to the audio version of the text, prtp.js uses the marks file to highlight the text currently being read.

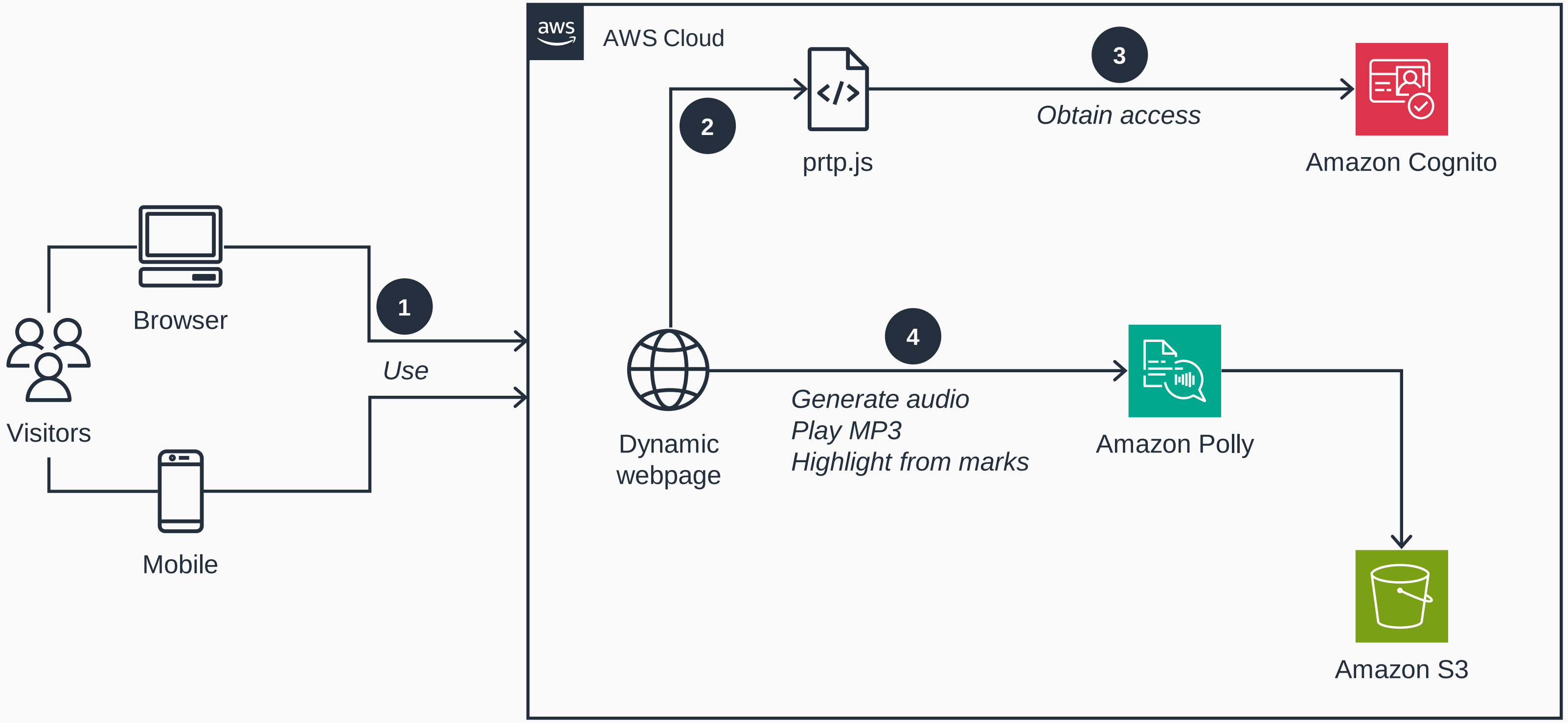
Reviewed for technical accuracy June 20, 2024

AWS Reference Architecture

Guidance for Reading & Highlighting Web Content with Amazon Polly

Dynamic Webpages

This architecture diagram shows how to use Amazon Polly to read and highlight content on dynamic webpages. When you play the audio, the page uses prtp.js to generate the audio in Amazon Polly, which highlights the synthesized audio using the same approach as for static pages.



1. The content of a dynamic page changes in response to visitor interactions, so audio can't be pre-generated. Instead, it must be synthesized dynamically using prtp.js when visitors play the audio.
2. First, the page uses prtp.js to generate the audio in **Amazon Polly**. Then, prtp.js highlights the synthesized audio using the same approach as it uses for static pages.
3. To access AWS services from the browser, the script uses an AWS identity provided by an **Amazon Cognito** identity pool. This identity pool is scoped to allow the visitor just enough access to **Amazon Polly** and the **Amazon S3** bucket to render the audio.
4. Generating both MP3 audio and speech marks requires **Amazon Polly** to synthesize the same input twice and return the response.

Reviewed for technical accuracy June 20, 2024

AWS Reference Architecture