

Continuous News Feeds in Near Real-time

Client: A media house with editorial strength

Challenge:

The client was a media company targeting a vast set of audience who like to read about everyday things ranging from politics, sports, celebrities, and the like. They were seeking a data acquisition engine that could not just collect all the relevant data they asked it to, but also did that within seconds of a news being published on one of their target sources on the web. An extremely powerful web crawler was the goal but building it required high-level expertise and meant some shift in focus from editorial.

Solution:

Keywords and the list of target sources provided by the client were fed into PromptCloud's low-latency component. The pipeline was set up and extracted data was indexed along with a markup indicating the category it belonged to. Only the data API layer was exposed to the client using which they downloaded data every time it appeared on it, and used that to build content for their own portal.

Benefits to the client:

- Complete data coverage and single point of lookout
- Automatic feeds arriving every time an article was published
- Zero data processing efforts at client's end
- Zero manpower required at client's end