Monitor Flight Schedules

Client: A popular airline company from Spain

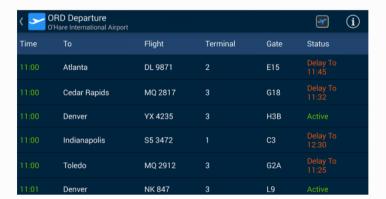**Solution:** Site-specific crawl and extraction

Domain: Travel

Challenge:

The client was on the lookout for flight schedule data for different aircrafts (extraction based on model numbers) from a flight tracking website. They wanted to incorporate this data into their analytics system and derive insights that would help them optimize their internal flight schedules. Since the data presented by the flight monitoring website was unstructured, they couldn't access it programmatically so as to perform analyses. The data was to be extracted at a frequency of 3 days and delivered in CSV format.

The Solution:

The client provided us with the specifics of the requirement such as source website, URLs to be crawled and the data points to be extracted. The frequency of crawls was set to 3 days. After establishing the feasibility of the crawls, our team set up the crawlers to extract the required data fields from the target site. Since this is a custom use case, it comes under our site specific crawls offering where we build crawler set ups from the scratch for the target sites. We completed the initial setup in a few days and the first set of data containing about 300 k records was delivered to the client.

Time	To	Flight	Terminal	Gate	Status
11:00	Atlanta	DL 9871	2	E15	Delay To 11:45
11:00	Cedar Rapids	MQ 2817	3	G18	Delay To 11:32
11:00	Denver	YX 4235	3	H3B	Active
11:00	Indianapolis	S5 3472	1	C3	Delay To 12:30
11:00	Toledo	MQ 2912	3	G2A	Delay To 11:25
11:01	Denver	NK 847	3	L9	Active

ORD Departure — O'Hare International Airport

Benefits to the client:

- All the complex aspects of web scraping were taken care of by us

- The initial setup was completed in a matter of few days and data flow was consistent thereafter

- We set up monitoring systems for the sources to ensure a smooth data flow

- Huge amounts of data was handled effortlessly by our extensive infrastructure

- The client could start consuming the data immediately after delivery because of the clean and ready-to-use format