

- A Parser is a program that takes a physical representation of some data and converts it into an in-memory form for the program as a whole to use. Parsers are used everywhere in software.
- An XML Parser is a parser that is designed to read XML and create a way for programs to use XML.

Explaining types of XML parsers

There are mainly three types of XML parsers:

- > SAX
- > DOM
- > Pull parser

SAX

SAX stands for 'Simple API for XML'. It does not create any internal structure. Clients do not know what methods to call. They just override the methods of the API and place his own code inside the method. It is an event-based parser, it works as an event handler in Java.

- Advantages
- Since it reads each unit of XML, it creates an event so that the calling program can use it.
- SAX uses what it likes to, by ignoring the bits which it doesn't care about.
- It is memory efficient.
- It's very fast and works for huge documents.
- Disadvantages
- The main disadvantage of SAX is that the Calling program must keep track of everything it might ever need.
- Since its Event-based, its API is less Intuitive.

DOM

DOM stands for 'Document Object Model'. A DOM Parser creates an internal structure in memory which is a DOM document object and the client applications get information of the original XML document by invoking methods on this document object. DOM Parser has a tree-based structure.

- Advantages
- It supports both Read and Write operations.
- It is preferred when there is random access to widely separated parts of the documents required.
- It builds the entire XML document representation in memory and then hands the calling program the whole chunk of memory.

- Disadvantages
- It consumes more memory since the whole XML document will be loaded into the memory.

Pull Parser

Pull parser waits for the application to come calling. That is, they ask for the next available event, and the application basically loops until it runs out of XML.

- Advantages
- It is designed to be used with large data sources.
- Pull parser chooses to skip the events (whole section of the document) which it is not interested in.