

Structure of a URL

A valid value for the URL data type can be composed of the following parts.

Example URL:

```
http://www.app.example.co.uk/support
```

NOTE: IP addresses that include the protocol identifier (`http://1.2.3.4`) do not contain domain identifiers and need to be processed using a different set of methods. It might be easier to remove the protocol identifiers and change the data type to IP Address.

The hierarchy of domain names extends from right to left.

| Element Name | Examples | Wrangle Function | Notes |
|---------------------|---|---------------------------|--|
| Top-level domain | <ul style="list-style-type: none">co.ukcom, net, org | <i>SUFFIX Function</i> | Every valid URL must have at least one top-level domain. NOTE: When the <code>DOMAIN</code> function parses a multi-tiered top-level domain such as <code>co.uk</code> , the output is the first part of the domain value (e.g. <code>co</code>). |
| Second-level domain | example app.example | <i>DOMAIN Function</i> | This value can be extracted from a valid URL using the <code>DOMAIN</code> function. See <i>DOMAIN Function</i> . |
| Third-level domain | www | <i>SUBDOMAIN Function</i> | This value can be extracted from a valid URL using the <code>SUBDOMAIN</code> function. See <i>SUBDOMAIN Function</i> . |
| path | /support | | |
| protocol identifier | http:// https:// | | You can use pattern matching to locate these protocol identifiers. In your Wrangle transforms, use the following Trifacta® pattern: <pre>`http%?:/`</pre> For an example, see <i>IPOINT Function</i> . |
| hostname | http://www.app.example.com | | |