

Public client-side API for DCC application

Main requirements specification, phase A, 2014-07-07

All client-side operations with DC service can be made using DCC application. All client-server interactions are based on the MOM event-based transport especially designed for HCE Python bindings. Networking transport layer is based on zmq library messaging. The DCC application has a command line user interface. Possible parameters are specified in the main DC service architecture document DC_application_architecture.docx.

All operations are requires the request data and produces response results data. Both request and response are formatted as json and represents serialized Python objects defined i the

`<DTS>/api/python/hce/dc/EventsObjects.py`

Complete list of Python objects dumps and correspondent json formatted serialized dumps represented in the **appendix B**. The “<DTS>” is a root of **Demo Tests Suit** directory, which after installation usually located in the directory:

`~/hce-node-tests/`

To use the DCC application the complete HCE DTS environment for Python language need to be installed with dependencies as described on main site install page in the “**Install Demo Test Suit Environment for Python language**” section. No another components of HCE required.

To perform some operation the operation name and request json file need to be specified. Common form is:

```
cd <DTS>/api/python/bin
./dc-client.py --config=./ini/dc-client.ini --command=<CMD> --file=<JSON_FILE> --fields=<JSON_FIELDS>
```

The “`dc-client.ini`” it is a configuration file contains most important parameters like host and port of DC service, request timeout and logging ini file path. Change the “`clientHost`” value to specify the DC’s service host IP address or domain name. Other configuration settings can be default.

The “<CMD>” is a name of requested command. Each command performs some action in the DC service and requires a time to execute. Depends on size of data, number of nodes in the distributed cluster and general load level of hardware host servers some operations can take a long time. If request returned timeout generally the “`timeout`” value in the “`dc-client.ini`” need to be increased. Operation request command names, input and output jsons are specified in the **appendix A**.

The “<JSON_FILE>” is a path and name of request json data for operation. All fields in the request json are mandatory. For operations like **SITE_UPDATE** and **URL_UPDATE** the fields of correspondent object that not required being updated need to be filled with “null” value.

Examples of all kind operation jsons files for functional tests on random generator site are provided in the directory:

`<DTS>/api/python/data/ftests/`

All jsons in this directory created especially for functional tests and need to be modified and used as templates in real applied system.

The “<JSON_FIELDS>” is a json string in format correspondent with the object structure for operation. For example, the **SITE_UPDATE** operation corresponds with the **SiteUpdate** object. The fields json can to have only fields that need to be set with values different from None (NULL) or defined in the `JSON_FILE`.

Any request can be executed with three forms:

- 1) Pre-created json file JSON_FILE with complete set of fields and de-serialization it to the target object.
- 2) Dynamically created target object with None all fields and updated values from the JSON_FIELDS. (With support of file reference as field value).
- 3) Mixed 1) and 2) i.e. update the values taken from the rpe-created JSON_FILE with values from JSON_FIELDS.

First way is useful for “NEW” operations type or in case of all fields values are static. Second way mostly useful for “UPDATE” operations when few fields updated but most of fields are not. And third type useful in more complex cases when pre-created jsons with complete set of fields can be used as templates. Second way also supports the file reference as field value. This means that if field value is string and starts from the “file://” prefix it treated not as a value, but tries to open file, read content and set it as a value of a field. For example:

```
--command=SITE_UPDATE --fields={'id':
```

```
"afe3b4bc8af61fa95d67f62122d55d04","updateType":2,"state":2,"properties":{"PROCESS_CTYPES":"text/html,text/xml","template":"file:///home/hce/template.txt"}}'
```

will update the record “template” of the “properties” field of the SiteUpdate object with value that is stored in the file “home/hce/template.txt”.

Generally, output json object is always the “**ClientResponse**”. Please, see appendix B for a structure representation. It represents client-level state. Fields “**errorCode**” and “**errorMessage**” reflects the client application operation state, like timeout or serialization to json error. The field “**itemsList**” is aimed to store array of “**ClientResponseItem**” items received from several nodes in distributed cluster. But, by default the DC service merges them in one. So, by default the array contains only one item and all fields exclude “**itemObject**” and “**id**” are textual and contain semicolon separated lists of sequentially merged values from all hosts. The “**id**” is unique request Id and the same for all nodes. The “**itemObject**” stores different serialized object depends on request. See appendix A to get relation between request and response objects. The “**itemObject**” is end user target data structure.

APPENDIX A.

Operation command name	Input json object	Output json object	Command description
Site object operations			
SITE_CLEANUP	SiteCleanup object	GeneralResponse object in "itemObject" field	Removes all crawled and processed data from all kind of storages, removes links history and crawling logs, initialize counter fields of site object representation.
SITE_DELETE	SiteDelete object	GeneralResponse object in "itemObject" field	Delete site object and all related data and structures from DC service.
SITE_FIND	SiteFind object	Array of Site objects in "itemObject" field	Find site object in the DC service by URL pattern and criterions to select.
SITE_NEW	Site object	GeneralResponse object in "itemObject" field	Creates new empty site object in the DC service, sets root URLs, initialize templates, patterns and another related structures. In case of conditions are succeed new site included in to the crawling right after creation.
SITE_STATUS	SiteStatus object	Site object in "itemObject" field	Returns serialized representation of the site object state inside DC service.
SITE_UPDATE	SiteUpdate object	GeneralResponse object in "itemObject" field	Updates fields of existing site object in the DC service.
URL object operations			
URL_CLEANUP	Array of URLCleanup objects	GeneralResponse object in "itemObject" field	The same as SITE_CLEANUP but operates only with selected URLs.
URL_CONTENT	Array of URLContentRequest objects	Array of URLContentResponse objects in "itemObject" field	Returns serialized data that was produced by crawling and/or processing of resource identified by link URL. Can include raw data of http response, headers and cookies.
URL_DELETE	Array of URLDelete objects	GeneralResponse object in "itemObject" field	Delete all resource data including link URL, crawled and processed contents. Removes resource from the DC service.
URL_FETCH	Array of URLFetch objects	Array of URL objects in "itemObject" field	Returns the array of url objects satisfies with criterions and selection conditions. Can be used to check state of resources, also, used internally by DC service to make crawling batches. Can also change state and other fields of selected URL objects atomic way right after selection.
URL_NEW	Array of URL objects	GeneralResponse object in "itemObject" field	Inserts the URL link in to the DC service and assign it for a proper site object.
URL_STATUS	Array of URLStatus objects	Array of URL objects in "itemObject" field	Similar with URL_FETCH but with exact definition of URL requested.
URL_UPDATE	Array of URLUpdate objects	GeneralResponse object in "itemObject" field	Updates fields of URL object in the DC service.

APPENDIX B.

Total DC's service Python objects dumps list with simple initialization:

```

Test ClientResponse object:
{'errorCode': 0, 'errorMessage': '', 'itemsList': [<dc.EventObjects.ClientResponseItem object at 0xb6e1c94c>]}
{
  "errorCode": 0,
  "errorMessage": "",
  "itemsList": [
    {
      "errorCode": 0,
      "errorMessage": "",
      "host": "",
      "id": 0,
      "itemObject": {
        "errorCode": 0,
        "errorMessage": "",
        "statuses": []
      },
      "node": "",
      "port": 0,
      "time": 0
    }
  ]
}
Test Site object:
{'urlType': 0, 'filters': [<dc.EventObjects.SiteFilter object at 0xb7255eec>], 'maxURLs': 0, 'userId': 0, 'iterations': 0,
'httpTimeout': 30000, 'id': 'c2ab041d4a4240eedb0df3b7e1d0f4c2', 'contents': 0, 'size': 0, 'errors': 0, 'maxResourceSize': 0,
'properties': {'PROCESS_CTYPES': 'text/html', 'HTTP_HEADERS': {}, 'STORE_HTTP_REQUEST': '1', 'HTTP_COOKIE': '',
'STORE_HTTP_HEADERS': '1'}, 'priority': 0, 'state': 1, 'processingDelay': 500, 'resources': 0, 'description': '', 'avgSpeed': 0,
'errorMask': 0, 'uDate': None, 'tcDate': None, 'maxErrors': 0, 'maxResources': 0, 'requestDelay': 500, 'cDate':
<app.Utils.SQLiteExpression object at 0xb7255dac>, 'urls': ['http://127.0.0.1/_site_random_generator.php?wrong=url'],
'avgSpeedCounter': 0}
{
  "avgSpeed": 0,
  "avgSpeedCounter": 0,
  "cDate": {
    "str": "NOW()"
  },
  "contents": 0,
  "description": "",
  "errorMask": 0,
  "errors": 0,
  "filters": [
    {
      "mode": 0,
      "pattern": "*",
      "siteId": "c2ab041d4a4240eedb0df3b7e1d0f4c2",
      "type": 1
    }
  ],
  "httpTimeout": 30000,
  "id": "c2ab041d4a4240eedb0df3b7e1d0f4c2",
  "iterations": 0,
  "maxErrors": 0,
  "maxResourceSize": 0,
  "maxResources": 0,
  "maxURLs": 0,
  "priority": 0,
  "processingDelay": 500,
  "properties": {
    "HTTP_COOKIE": "",
    "HTTP_HEADERS": {},
    "PROCESS_CTYPES": "text/html",
    "STORE_HTTP_HEADERS": "1",
    "STORE_HTTP_REQUEST": "1"
  },
  "requestDelay": 500,
  "resources": 0,
  "size": 0,
  "state": 1,
  "tcDate": null,
  "uDate": null,
  "urlType": 0,
  "urls": [
    "http://127.0.0.1/_site_random_generator.php?wrong=url"
  ],
  "userId": 0
}
Test SiteUpdate object:
{'urlType': 2, 'filters': [<dc.EventObjects.SiteFilter object at 0xb6f2c40c>], 'maxURLs': 2000, 'userId': None, 'iterations': 20,
'httpTimeout': 2000000000, 'id': 'b85ab149a528bd0a024fa0f43e80b5fc', 'contents': None, 'size': None, 'errors': 200000000000L,
'updateType': 1, 'maxResourceSize': 20000000, 'properties': {'PROCESS_CTYPES': 'text/plain', 'STORE_HTTP_REQUEST': '1',
'STORE_HTTP_HEADERS': '1'}, 'priority': 200, 'state': 2, 'processingDelay': None, 'resources': 2, 'description': 'Test update',
'avgSpeed': None, 'errorMask': 20000000000L, 'uDate': <app.Utils.SQLiteExpression object at 0xb6f2c36c>, 'tcDate':
'2012-12-12 12:12:12', 'urls': ['http://localhost/'], 'avgSpeedCounter': None}
{
  "avgSpeed": null,
  "avgSpeedCounter": null,
  "cDate": "2012-12-12 12:12:12",
  "contents": null,
  "description": "Test update",
  "errorMask": 200000000000,
  "errors": 200000000000,
  "filters": [
    {
      "mode": 0,
      "pattern": "http://localhost/*",
      "siteId": "b85ab149a528bd0a024fa0f43e80b5fc",
      "type": 1
    }
  ],
}
],

```

```

    "httpTimeout": 200000000,
    "id": "b85ab149a528bd0a024fa0f43e80b5fc",
    "iterations": 20,
    "maxErrors": 20000,
    "maxResourceSize": 20000000,
    "maxResources": 2000,
    "maxURLs": 2000,
    "priority": 200,
    "processingDelay": null,
    "properties": {
      "PROCESS_CTYPES": "text/plain",
      "STORE_HTTP_HEADERS": "1",
      "STORE_HTTP_REQUEST": "1"
    },
    "requestDelay": 200000000,
    "resources": 2,
    "size": null,
    "state": 2,
    "tcDate": {
      "str": "NOW()"
    },
    "uDate": {
      "str": "NOW()"
    },
    "updateType": 1,
    "urlType": 2,
    "urls": [
      "http://localhost/"
    ],
    "userId": null
  }
}
Test SiteUpdate object:
{
  "urlType": 2, "filters": [<dc.EventObjects.SiteFilter object at 0xb6f2c40c>], "maxURLs": 2000, "userId": None, "iterations": 20,
  "httpTimeout": 200000000, "id": '699fcf4591fc23e79b839d8819904293', "contents": None, "size": None, "errors": 20000000000L,
  "updateType": 0, "maxResourceSize": 20000000, "properties": {'PROCESS_CTYPES': 'text/plain', 'STORE_HTTP_HEADERS': '1',
  'STORE_HTTP_HEADERS': '1'}, "priority": 200, "state": 2, "processingDelay": None, "resources": 2, "description": 'Test update',
  "avgSpeed": None, "errorMask": 20000000000L, "uDate": <app.Utils.SQLEXPRESS object at 0xb747006c>, "tcDate":
  <app.Utils.SQLEXPRESS object at 0xb6f2c36c>, "maxErrors": 20000, "maxResources": 2000, "requestDelay": 200000000, "cdate":
  '2012-12-12 12:12:12', "urls": ['http://localhost/'], "avgSpeedCounter": None
}
{
  "avgSpeed": null,
  "avgSpeedCounter": null,
  "cdate": "2012-12-12 12:12:12",
  "contents": null,
  "description": "Test update",
  "errorMask": 20000000000,
  "errors": 200000000000,
  "filters": [
    {
      "mode": 0,
      "pattern": "http://localhost/*",
      "siteId": "699fcf4591fc23e79b839d8819904293",
      "type": 1
    }
  ]
}
,
  "httpTimeout": 200000000,
  "id": "699fcf4591fc23e79b839d8819904293",
  "iterations": 20,
  "maxErrors": 20000,
  "maxResourceSize": 20000000,
  "maxResources": 2000,
  "maxURLs": 2000,
  "priority": 200,
  "processingDelay": null,
  "properties": {
    "PROCESS_CTYPES": "text/plain",
    "STORE_HTTP_HEADERS": "1",
    "STORE_HTTP_REQUEST": "1"
  },
  "requestDelay": 200000000,
  "resources": 2,
  "size": null,
  "state": 2,
  "tcDate": {
    "str": "NOW()"
  },
  "uDate": {
    "str": "NOW()"
  },
  "updateType": 0,
  "urlType": 2,
  "urls": [
    "http://localhost/"
  ],
  "userId": null
}
}
Test SiteStatus object:
{'deleteTaskId': None, 'id': '699fcf4591fc23e79b839d8819904293'}
{
  "deleteTaskId": null,
  "id": "699fcf4591fc23e79b839d8819904293"
}
}
Test SiteDelete object:
{'id': '699fcf4591fc23e79b839d8819904293', 'taskType': 1}
{
  "id": "699fcf4591fc23e79b839d8819904293",
  "taskType": 1
}
}
Test SiteCleanup object:
{'id': 'b85ab149a528bd0a024fa0f43e80b5fc', 'taskType': 1}
{
  "id": "b85ab149a528bd0a024fa0f43e80b5fc",
  "taskType": 1
}
}
Test SiteFilter object:
{'pattern': '*', 'siteId': '235325634634263', 'mode': 0, 'type': 1}
Test URL object:

```

```
{'totalTime': 0, 'linksI': 0, 'linksE': 0, 'processingTime': 0, 'Cdate': None, 'mRateCounter': 0, 'httpTimeout': 30000, 'size': 0, 'requestDelay': 0, 'mRate': 0.0, 'charset': '', 'statusCode': 0, 'state': 0, 'urlMd5': '701ccc5c1c589041d31d13dae8dce90d', 'crawlingTime': 0, 'type': 0, 'parentMd5': '', 'status': 1, 'siteSelect': 0, 'contentType': '', 'processingDelay': 0, 'errorMask': 0, 'httpMethod': '', 'eTag': '', 'siteId': 'b85ab149a528bd0a024fa0f43e80b5fc', 'freq': 0, 'rawContentMd5': '', 'url': 'http://127.0.0.1/_site_random_generator.php?a=1', 'crawled': 0, 'uDate': None, 'batchId': 0, 'depth': 0, 'processed': 0, 'lastModified': ''}
{
  "Cdate": null,
  "UDate": null,
  "batchId": 0,
  "charset": "",
  "contentType": "",
  "crawled": 0,
  "crawlingTime": 0,
  "depth": 0,
  "eTag": "",
  "errorMask": 0,
  "freq": 0,
  "statusCode": 0,
  "httpMethod": "",
  "httpTimeout": 30000,
  "lastModified": "",
  "linksE": 0,
  "linksI": 0,
  "mRate": 0.0,
  "mRateCounter": 0,
  "parentMd5": "",
  "processed": 0,
  "processingDelay": 0,
  "processingTime": 0,
  "rawContentMd5": "",
  "requestDelay": 0,
  "siteId": "b85ab149a528bd0a024fa0f43e80b5fc",
  "siteSelect": 0,
  "size": 0,
  "state": 0,
  "status": 1,
  "totalTime": 0,
  "type": 0,
  "url": "http://127.0.0.1/_site_random_generator.php?a=1",
  "urlMd5": "701ccc5c1c589041d31d13dae8dce90d"
}
Test URLStatus object:
{'url': '701ccc5c1c589041d31d13dae8dce90d', 'urlType': 1, 'siteId': 'b85ab149a528bd0a024fa0f43e80b5fc'}
{
  "siteId": "b85ab149a528bd0a024fa0f43e80b5fc",
  "url": "701ccc5c1c589041d31d13dae8dce90d",
  "urlType": 1
}
}
Test BatchItem object:
{'urlId': '235325634634234', 'siteId': '235325634634263', 'urlObj': <dc.EventObjects.URL object at 0xb6f2c5ac>}
{
  "siteId": "235325634634263",
  "urlId": "235325634634234",
  "urlObj": {
    "Cdate": null,
    "UDate": null,
    "batchId": 0,
    "charset": "",
    "contentType": "",
    "crawled": 0,
    "crawlingTime": 0,
    "depth": 0,
    "eTag": "",
    "errorMask": 0,
    "freq": 0,
    "statusCode": 0,
    "httpMethod": "",
    "httpTimeout": 30000,
    "lastModified": "",
    "linksE": 0,
    "linksI": 0,
    "mRate": 0.0,
    "mRateCounter": 0,
    "parentMd5": "",
    "processed": 0,
    "processingDelay": 0,
    "processingTime": 0,
    "rawContentMd5": "",
    "requestDelay": 0,
    "siteId": "235325634634263",
    "siteSelect": 0,
    "size": 0,
    "state": 0,
    "status": 1,
    "totalTime": 0,
    "type": 0,
    "url": "http://127.0.0.1/",
    "urlMd5": "699fcf4591fc23e79b839d8819904293"
  }
}
}
Test Batch object:
{'items': [], 'crawlerType': 1, 'id': [<dc.EventObjects.BatchItem object at 0xb6f2c62c>, <dc.EventObjects.BatchItem object at 0xb6f2c6cc>]}
{
  "crawlerType": 1,
  "id": [
    {
      "siteId": "235325634634263",
      "urlId": "235325634634234",
      "urlObj": {
        "Cdate": null,
        "UDate": null,
        "batchId": 0,
        "charset": "",
        "contentType": "",
        "crawled": 0,
        "crawlingTime": 0,

```

```

        "depth": 0,
        "eTag": "",
        "errorMask": 0,
        "freq": 0,
        "httpCode": 0,
        "httpMethod": "",
        "httpTimeout": 30000,
        "lastModified": "",
        "linksE": 0,
        "linksI": 0,
        "mRate": 0.0,
        "mRateCounter": 0,
        "parentMd5": "",
        "processed": 0,
        "processingDelay": 0,
        "processingTime": 0,
        "rawContentMd5": "",
        "requestDelay": 0,
        "siteId": "235325634634263",
        "siteSelect": 0,
        "size": 0,
        "state": 0,
        "status": 1,
        "totalTime": 0,
        "type": 0,
        "url": "http://127.0.0.1/",
        "urlMd5": "699fcf4591fc23e79b839d8819904293"
    }
},
{
    "siteId": "335325634634264",
    "urlId": "335325634634235",
    "urlObj": {
        "CDate": null,
        "UDate": null,
        "batchId": 0,
        "charset": "",
        "contentType": "",
        "crawled": 0,
        "crawlingTime": 0,
        "depth": 0,
        "eTag": "",
        "errorMask": 0,
        "freq": 0,
        "httpCode": 0,
        "httpMethod": "",
        "httpTimeout": 30000,
        "lastModified": "",
        "linksE": 0,
        "linksI": 0,
        "mRate": 0.0,
        "mRateCounter": 0,
        "parentMd5": "",
        "processed": 0,
        "processingDelay": 0,
        "processingTime": 0,
        "rawContentMd5": "",
        "requestDelay": 0,
        "siteId": "235325634634263",
        "siteSelect": 0,
        "size": 0,
        "state": 0,
        "status": 1,
        "totalTime": 0,
        "type": 0,
        "url": "http://127.0.0.1/",
        "urlMd5": "699fcf4591fc23e79b839d8819904293"
    }
}
]
}
}
items": []
}
Test URLFetch object:
{'sitesCriteriaions': {'ORDER BY': 'Priority DESC, TcDate ASC'}, 'algorithm': 0, 'urlsCriteriaions': {'LIMIT': 20, 'ORDER BY': 'CDate ASC'}, 'urlUpdate': None, 'maxURLs': 20, 'sitesList': []}
{
    "algorithm": 0,
    "maxURLs": 20,
    "sitesCriteriaions": {
        "ORDER BY": "Priority DESC, TcDate ASC"
    },
    "sitesList": [],
    "urlUpdate": null,
    "urlsCriteriaions": {
        "LIMIT": 20,
        "ORDER BY": "CDate ASC"
    }
}
}
Test URLFetch object:
{'sitesCriteriaions': {'ORDER BY': 'Priority DESC, TcDate ASC'}, 'algorithm': 0, 'urlsCriteriaions': {'LIMIT': 20, 'ORDER BY': 'CDate ASC'}, 'urlUpdate': None, 'maxURLs': 20, 'sitesList': ['235325634634263', '235325634634234']}
{
    "algorithm": 0,
    "maxURLs": 20,
    "sitesCriteriaions": {
        "ORDER BY": "Priority DESC, TcDate ASC"
    },
    "sitesList": [
        "235325634634263",
        "235325634634234"
    ],
    "urlUpdate": null,
    "urlsCriteriaions": {
        "LIMIT": 20,
        "ORDER BY": "CDate ASC"
    }
}
}
Test URLFetch object:

```

```
{'sitesCriteriaions': {'ORDER BY': 'Priority DESC, TcDate ASC'}, 'algorithm': 0, 'urlsCriteriaions': {'WHERE': "Status=7 AND
Crawled>0 AND Processed>0 AND CDate BETWEEN '2014-06-28 00:00:01' AND '2014-06-28 23:59:59'", 'LIMIT': 20, 'ORDER BY': 'CDate
ASC'}, 'urlUpdate': None, 'maxURLs': 20, 'sitesList': []}
{
  "algorithm": 0,
  "maxURLs": 20,
  "sitesCriteriaions": {
    "ORDER BY": "Priority DESC, TcDate ASC"
  },
  "sitesList": [],
  "urlUpdate": null,
  "urlsCriteriaions": {
    "LIMIT": 20,
    "ORDER BY": "CDate ASC",
    "WHERE": "Status=7 AND Crawled>0 AND Processed>0 AND CDate BETWEEN '2014-06-28 00:00:01' AND '2014-06-28 23:59:59'"
  }
}
}
Test URLUpdate object:
{'totalTime': None, 'linksI': None, 'linksE': None, 'criteriaions': {'LIMIT': 1}, 'processingTime': None, 'CDate': None,
'mRateCounter': None, 'httpTimeout': None, 'size': None, 'requestDelay': None, 'mRate': None, 'charset': None, 'httpCode': None,
'state': None, 'urlMd5': '699fcf4591fc23e79b839d8819904293', 'crawlingTime': None, 'type': None, 'parentMd5': None, 'status': 1,
'siteSelect': None, 'contentType': None, 'processingDelay': None, 'errorMask': None, 'httpMethod': None, 'eTag': None, 'siteId':
'525326523434525', 'freq': None, 'rawContentMd5': None, 'url': 'http://127.0.0.1/', 'crawled': None, 'UDate':
<app.Utils.SQLExpression object at 0xb6f2c92c>, 'batchId': None, 'depth': None, 'processed': None, 'lastModified': None}
{
  "CDate": null,
  "UDate": {
    "str": "NOW()"
  },
  "batchId": null,
  "charset": null,
  "contentType": null,
  "crawled": null,
  "crawlingTime": null,
  "criteriaions": {
    "LIMIT": 1
  },
  "depth": null,
  "eTag": null,
  "errorMask": null,
  "freq": null,
  "httpCode": null,
  "httpMethod": null,
  "httpTimeout": null,
  "lastModified": null,
  "linksE": null,
  "linksI": null,
  "mRate": null,
  "mRateCounter": null,
  "parentMd5": null,
  "processed": null,
  "processingDelay": null,
  "processingTime": null,
  "rawContentMd5": null,
  "requestDelay": null,
  "siteId": "525326523434525",
  "siteSelect": null,
  "size": null,
  "state": null,
  "status": 1,
  "totalTime": null,
  "type": null,
  "url": "http://127.0.0.1/",
  "urlMd5": "699fcf4591fc23e79b839d8819904293"
}
}
Test URLUpdate object:
{'totalTime': None, 'linksI': None, 'linksE': None, 'criteriaions': {'LIMIT': 1}, 'processingTime': None, 'CDate': None,
'mRateCounter': None, 'httpTimeout': None, 'size': None, 'requestDelay': None, 'mRate': None, 'charset': None, 'httpCode': None,
'state': 0, 'urlMd5': '701ccc5c1c589041d31d13dae8dce90d', 'crawlingTime': None, 'type': None, 'parentMd5': None, 'status': 1,
'siteSelect': None, 'contentType': None, 'processingDelay': None, 'errorMask': None, 'httpMethod': None, 'eTag': None, 'siteId':
'b85ab149a528bd0a024fa0f43e80b5fc', 'freq': None, 'rawContentMd5': None, 'url': None, 'crawled': None, 'UDate':
<app.Utils.SQLExpression object at 0xb6f2ca4c>, 'batchId': None, 'depth': None, 'processed': None, 'lastModified': None}
{
  "CDate": null,
  "UDate": {
    "str": "NOW()"
  },
  "batchId": null,
  "charset": null,
  "contentType": null,
  "crawled": null,
  "crawlingTime": null,
  "criteriaions": {
    "LIMIT": 1
  },
  "depth": null,
  "eTag": null,
  "errorMask": null,
  "freq": null,
  "httpCode": null,
  "httpMethod": null,
  "httpTimeout": null,
  "lastModified": null,
  "linksE": null,
  "linksI": null,
  "mRate": null,
  "mRateCounter": null,
  "parentMd5": null,
  "processed": null,
  "processingDelay": null,
  "processingTime": null,
  "rawContentMd5": null,
  "requestDelay": null,
  "siteId": "b85ab149a528bd0a024fa0f43e80b5fc",
  "siteSelect": null,
  "size": null,
  "state": 0,
  "status": 1,

```



```

    "totalTime": null,
    "type": null,
    "url": null,
    "urlMd5": "701ccc5c1c589041d31d13dae8dce90d"
  }
  Test URLDelete object:
  {'url': 'http://127.0.0.1/', 'urlType': 0, 'siteId': '233243243242423'}
  {
    "siteId": "233243243242423",
    "url": "http://127.0.0.1/",
    "urlType": 0
  }
  Test URLDelete object:
  {'url': '701ccc5c1c589041d31d13dae8dce90d', 'urlType': 1, 'siteId': 'b85ab149a528bd0a024fa0f43e80b5fc'}
  {
    "siteId": "b85ab149a528bd0a024fa0f43e80b5fc",
    "url": "701ccc5c1c589041d31d13dae8dce90d",
    "urlType": 1
  }
  Test URLCleanup object:
  {'url': 'http://127.0.0.1/', 'urlType': 0, 'state': None, 'status': 1, 'siteId': '346436436436346'}
  {
    "siteId": "346436436436346",
    "state": null,
    "status": 1,
    "url": "http://127.0.0.1/",
    "urlType": 0
  }
  Test URLCleanup object:
  {'url': '235325634634263', 'urlType': 1, 'state': 1, 'status': None, 'siteId': '3463463463463463'}
  {
    "siteId": "3463463463463463",
    "state": 1,
    "status": null,
    "url": "235325634634263",
    "urlType": 1
  }
  Test URLContentRequest object:
  {'url': 'http://127.0.0.1/', 'contentTypeMask': 3, 'urlMd5': '699fcf4591fc23e79b839d8819904293', 'siteId': '325632523424234234', 'urlFetch': None}
  {
    "contentTypeMask": 3,
    "siteId": "325632523424234234",
    "url": "http://127.0.0.1/",
    "urlFetch": null,
    "urlMd5": "699fcf4591fc23e79b839d8819904293"
  }
  Test URLContentRequest object:
  {'url': 'http://127.0.0.1/', 'contentTypeMask': 9, 'urlMd5': '699fcf4591fc23e79b839d8819904293', 'siteId': '3464364364363634', 'urlFetch': None}
  {
    "contentTypeMask": 9,
    "siteId": "3464364364363634",
    "url": "http://127.0.0.1/",
    "urlFetch": null,
    "urlMd5": "699fcf4591fc23e79b839d8819904293"
  }
  Test URLContentRequest object:
  {'url': '', 'contentTypeMask': 3, 'urlMd5': 'd41d8cd98f00b204e9800998ecf8427e', 'siteId': '', 'urlFetch': <dc.EventObjects.URLFetch object at 0xb6f2ca0c>}
  {
    "contentTypeMask": 3,
    "siteId": "",
    "url": "",
    "urlFetch": {
      "algorithm": 1,
      "maxURLs": 10,
      "sitesCriteriaions": {
        "ORDER BY": "Priority DESC, TcDate ASC",
        "WHERE": "State IN (1,2,3)"
      },
      "sitesList": [],
      "urlUpdate": null,
      "urlsCriteriaions": {
        "LIMIT": 20,
        "ORDER BY": "Cdate ASC",
        "WHERE": "Status=7 AND Crawled>0 AND Processed>0 AND CDate BETWEEN '2014-06-28 00:00:01' AND '2014-06-28 23:59:59'"
      }
    },
    "urlMd5": "d41d8cd98f00b204e9800998ecf8427e"
  }
  Test URLContentResponse object:
  {'url': 'http://127.0.0.1/', 'status': 0, 'processedContents': <dc.EventObjects.Content object at 0xb6f2c4cc>, 'urlMd5': None, 'rawContents': <dc.EventObjects.Content object at 0xb6f2c74c>}
  {
    "processedContents": {
      "buffer": [
        "test content 1"
      ],
      "cDate": 0
    },
    "rawContents": {
      "buffer": [
        "<html>test content 1</html>"
      ],
      "cDate": 124124354
    },
    "status": 0,
    "url": "http://127.0.0.1/",
    "urlMd5": null
  }
}

```