

Collection 'counters'

```
graph TD
    subgraph processes
        direction TB
        p1["*_id objectId"]
        p2["*_notification object"]
        p3["*_email string"]
        p4["*_webhook string"]
        p5["*_id objectId"]
        p6["*_ssePath string"]
        p7["*_currentStep object"]
        p8["*_seeds array[string]"]
        p9["*_maxPathLength int"]
        p10["*_maxPathProps int"]
        p11["*_whiteList array"]
        p12["*_blackList array"]
        p13["*_id objectId"]
        p14["*_steps array[object]"]
        p15["*_seeds array[string]"]
        p16["*_maxPathLength int"]
        p17["*_maxPathProps int"]
        p18["*_whiteList array"]
        p19["*_blackList array"]
        p20["*_id objectId"]
        p21["*_pathHeads object"]
        p22["*_pathHeads.http://derzis-val001.andrefs.com int"]
        p23["*_status string"]
        p24["*_createdAt date"]
        p25["*_updatedAt date"]
        p26["*_pid string"]
        p27["*___v int"]
    end

    subgraph resourceCaches
        direction TB
        rc1["*_id objectId"]
    end

    subgraph counters
        direction TB
        c1["*_id objectId"]
        c2["*_name string"]
        c3["*___v int"]
        c4["*_createdAt date"]
        c5["*_updatedAt date"]
        c6["*_value int"]
    end

    subgraph domains
        direction TB
        d1["*_id objectId"]
        d2["*_origin string"]
        d3["*___v int"]
        d4["*_crawl object"]
        d5["*_queued int"]
        d6["*_pathHeads int"]
        d7["*_delay int"]
        d8["*_nextAllowed date"]
        d9["*_ongoing int"]
        d10["*_success int"]
        d11["*_createdAt date"]
        d12["*_lastWarnings array"]
        d13["*_status string"]
        d14["*_updatedAt date"]
        d15["*_warnings object"]
        d16["*_E_ROBOTS_TIMEOUT int"]
        d17["*_E_RESOURCE_TIMEOUT int"]
        d18["*_E_DOMAIN_NOT_FOUND int"]
        d19["*_E_UNKNOWN int"]
        d20["*_id objectId"]
        d21["*_lastAccessed date"]
        d22["*_robots object"]
        d23["*_checked date"]
        d24["*_elapsedTime int"]
        d25["*_status string"]
        d26["*_text string"]
    end

    subgraph processResources
        direction TB
        pr1["*_id objectId"]
    end

    subgraph paths
        direction TB
        p1_1["*_id objectId"]
        p1_2["*_processId string"]
        p1_3["*_seed object"]
        p1_4["*_url string"]
        p1_5["*_id objectId"]
        p1_6["*_head object"]
        p1_7["*_url string"]
        p1_8["*_status string"]
        p1_9["*_domain object"]
        p1_10["*_status string"]
        p1_11["*_origin string"]
        p1_12["*_id objectId"]
        p1_13["*_id objectId"]
        p1_14["*_predicates object"]
        p1_15["*_count int"]
        p1_16["*_elems array[string]"]
        p1_17["*_id objectId"]
        p1_18["*_nodes object"]
        p1_19["*_count int"]
        p1_20["*_elems array[string]"]
        p1_21["*_id objectId"]
        p1_22["*_createdAt date"]
        p1_23["*_updatedAt date"]
        p1_24["*_lastPredicate string"]
        p1_25["*___v int"]
    end

    subgraph resources
        direction TB
        r1["*_id objectId"]
        r2["*_url string"]
        r3["*___v int"]
        r4["*_createdAt date"]
        r5["*_domain string"]
        r6["*_status string"]
        r7["*_triples array"]
        r8["*_updatedAt date"]
        r9["*_jobId int"]
        r10["*_crawlId object"]
        r11["*_domainTs date"]
        r12["*_counter int"]
        r13["*_id objectId"]
    end
```

Collection 'triples'

| triples | |
|---------------|---------------|
| * 🔑 _id | objectId |
| * 🔗 subject | string |
| * 🔗 object | string |
| * 🔗 predicate | string |
| * __v | int |
| * createdAt | date |
| * 🔗 nodes | array[string] |
| * sources | array[string] |
| * updatedAt | date |

| processTriples | |
|----------------|----------|
| * 🔑 _id | objectId |
| * 🔗 processId | string |
| * 🔗 triple | objectId |
| * __v | int |
| * createdAt | date |
| * updatedAt | date |

relationship

Main Diagram

| Collection counters | | |
|---------------------|-----------|-----------|
| Idx | Name | Data Type |
| * Pk | _id | objectId |
| * Idx | name | string |
| * | __v | int |
| * | createdAt | date |
| * | updatedAt | date |
| * | value | int |
| Indexes | | |
| Type | Name | On |
| Pk | _id_ | _id |
| | name_1 | name |

| Collection domains | | |
|--------------------|---|--|
| Idx | Name | Data Type |
| * Pk | _id | objectId |
| * Unq | origin | string |
| * | __v | int |
| * | crawl | object |
| * | crawl.queued | int |
| Idx | crawl.pathHeads | int |
| | crawl.delay | int |
| Idx | crawl.nextAllowed | date |
| | crawl.ongoing | int |
| | crawl.success | int |
| * | createdAt | date |
| * | lastWarnings | array |
| * Idx | status | string |
| * | updatedAt | date |
| * | warnings | object |
| * | warnings.E_ROBOTS_TIMEOUT | int |
| * | warnings.E_RESOURCE_TIMEOUT | int |
| * | warnings.E_DOMAIN_NOT_FOUND | int |
| * | warnings.E_UNKNOWN | int |
| * | warnings._id | objectId |
| | lastAccessed | date |
| | robots | object |
| * | robots.checked | date |
| * | robots.elapsedTime | int |
| * Idx | robots.status | string |
| * | robots.text | string |
| Indexes | | |
| Type | Name | On |
| Pk | _id_ | _id |
| Unq | origin_1 | origin |
| | robots.status_1 | robots.status |
| | crawl.nextAllowed_-1 | crawl.nextAllowed |
| | status_1_crawl.pathHeads_1_crawl.nextAllowed_-1 | status, crawl.pathHeads, crawl.nextAllowed |

| Collection paths | | |
|------------------|------|-----------|
| Idx | Name | Data Type |

| Collection paths | | |
|------------------|--|--------------------------------------|
| * Pk | _id | objectId |
| * Idx | processId | string |
| * | seed | object |
| * Idx | seed.url | string |
| * | seed._id | objectId |
| * | head | object |
| * Idx | head.url | string |
| * | head.status | string |
| * | head.domain | object |
| * | head.domain.status | string |
| * | head.domain.origin | string |
| * | head.domain._id | objectId |
| * | head._id | objectId |
| * | predicates | object |
| * Idx | predicates.count | int |
| * | predicates.elems | array[string] |
| * | predicates._id | objectId |
| * | nodes | object |
| * Idx | nodes.count | int |
| * | nodes.elems | array[string] |
| * | nodes._id | objectId |
| * | createdAt | date |
| * | updatedAt | date |
| * | lastPredicate | string |
| * | __v | int |
| Indexes | | |
| Type | Name | On |
| Pk | _id_ | _id |
| | head.url_1_nodes.count_1 | head.url, nodes.count |
| | seed.url_1_head.url_1_predicates.count_1 | seed.url, head.url, predicates.count |
| | processId_1 | processId |

| Collection processResources | | |
|-----------------------------|------|-----------|
| Idx | Name | Data Type |
| * Pk | _id | objectId |
| Indexes | | |
| Type | Name | On |
| Pk | _id_ | _id |

| Collection processTriples | | |
|---------------------------|----------------------|-------------------|
| Idx | Name | Data Type |
| * Pk | _id | objectId |
| * Unq | processId | string |
| * Unq | triple | objectId |
| * | __v | int |
| * | createdAt | date |
| * | updatedAt | date |
| Indexes | | |
| Type | Name | On |
| Pk | _id_ | _id |
| Unq | processId_1_triple_1 | processId, triple |

Collection processTriples

| Relationships | | |
|---------------|---|----|
| Type | Name | On |
| Vir | Relationship (triple) ref triples (_id) | |

Collection processes

| Idx | Name | Data Type |
|-------|--|---------------|
| * Pk | _id | objectId |
| * | notification | object |
| * | notification.email | string |
| * | notification.webhook | string |
| * | notification._id | objectId |
| * | notification.ssePath | string |
| * | currentStep | object |
| * | currentStep.seeds | array[string] |
| * | currentStep.maxPathLength | int |
| * | currentStep.maxPathProps | int |
| * | currentStep.whiteList | array |
| * | currentStep.blackList | array |
| * | currentStep._id | objectId |
| * | steps | array[object] |
| * | steps.seeds | array[string] |
| * | steps.maxPathLength | int |
| * | steps.maxPathProps | int |
| * | steps.whiteList | array |
| * | steps.blackList | array |
| * | steps._id | objectId |
| * | pathHeads | object |
| * | pathHeads.http://derzis-val001.andrefs.com | int |
| * Idx | status | string |
| * Idx | createdAt | date |
| * | updatedAt | date |
| * Unq | pid | string |
| * | __v | int |

| Indexes | | |
|---------|-------------|-----------|
| Type | Name | On |
| Pk | _id_ | _id |
| Unq | pid_1 | pid |
| | createdAt_1 | createdAt |
| | status_1 | status |

Collection resourceCaches

| Idx | Name | Data Type |
|------|------|-----------|
| * Pk | _id | objectId |

| Indexes | | |
|---------|------|-----|
| Type | Name | On |
| Pk | _id_ | _id |

Collection resources

| Idx | Name | Data Type |
|-------|------|-----------|
| * Pk | _id | objectId |
| * Idx | url | string |

Collection resources

| | | |
|-------|------------------|----------|
| * | __v | int |
| * | createdAt | date |
| * Idx | domain | string |
| * Idx | status | string |
| * | triples | array |
| * | updatedAt | date |
| | jobId | int |
| | crawlId | object |
| * | crawlId.domainTs | date |
| * | crawlId.counter | int |
| * | crawlId._id | objectId |

Indexes

| Type | Name | On |
|------|-------------------|----------------|
| Pk | _id_ | _id |
| | domain_1_status_1 | domain, status |
| | url_1_status_1 | url, status |

Collection triples

| Idx | Name | Data Type |
|-------|-----------|---------------|
| * Pk | _id | objectId |
| * Unq | subject | string |
| * Unq | object | string |
| * Unq | predicate | string |
| * | __v | int |
| * | createdAt | date |
| * Idx | nodes | array[string] |
| * | sources | array[string] |
| * | updatedAt | date |

Indexes

| Type | Name | On |
|------|--------------------------------|----------------------------|
| Pk | _id_ | _id |
| Unq | subject_1_predicate_1_object_1 | subject, predicate, object |
| | nodes_1 | nodes |