Introduction to MongoDB

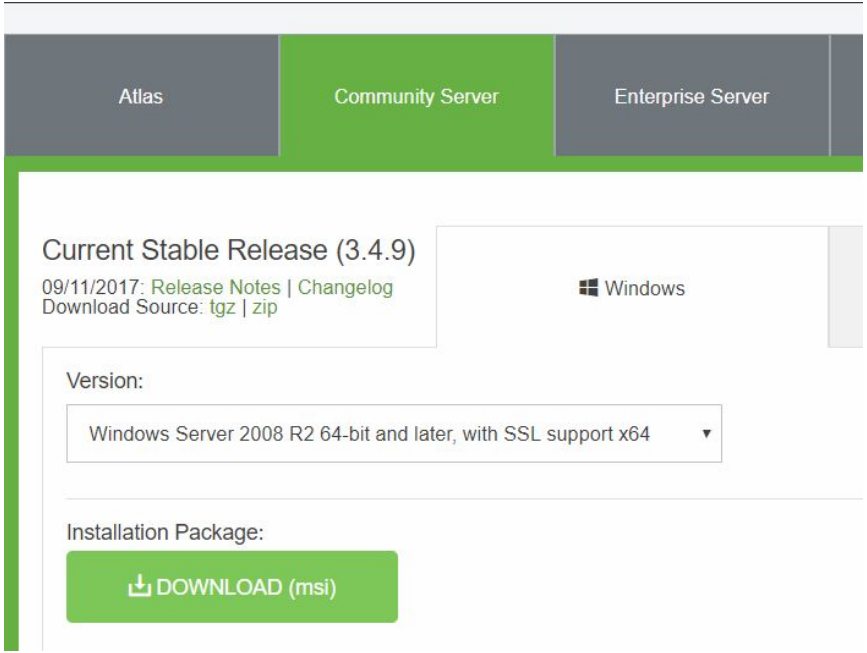
26/03/2018

What is MongoDB

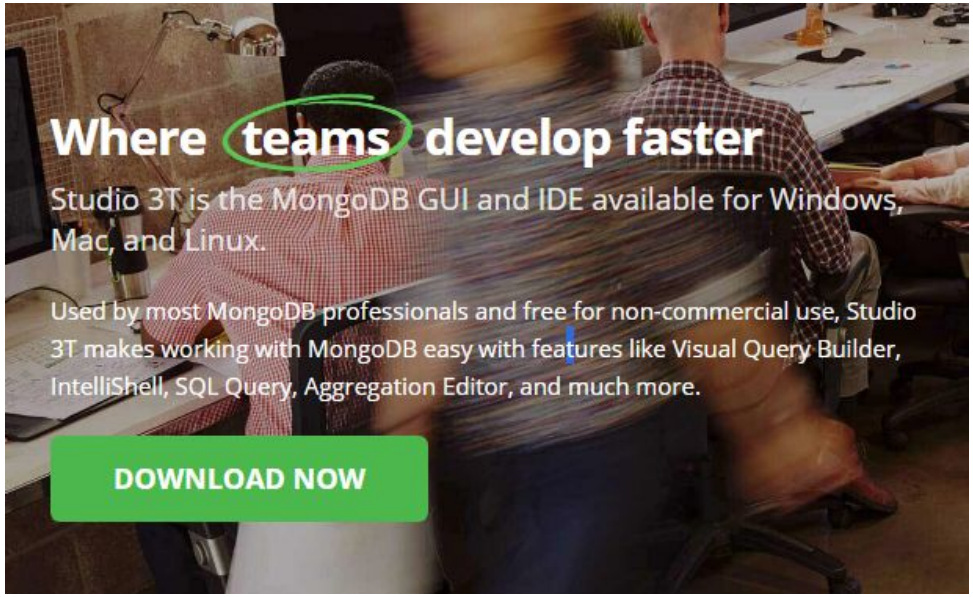
- It’s a free NoSQL document oriented database.
  - Not only SQL, since SQL queries are supported.
- It uses JSON-like documents to save data.
  - Flexible schemas - Allows missing fields.
- High performance, high availability, and automatic scaling.
  - Used by large companies like Amazon.

Installation

MongoDB community server

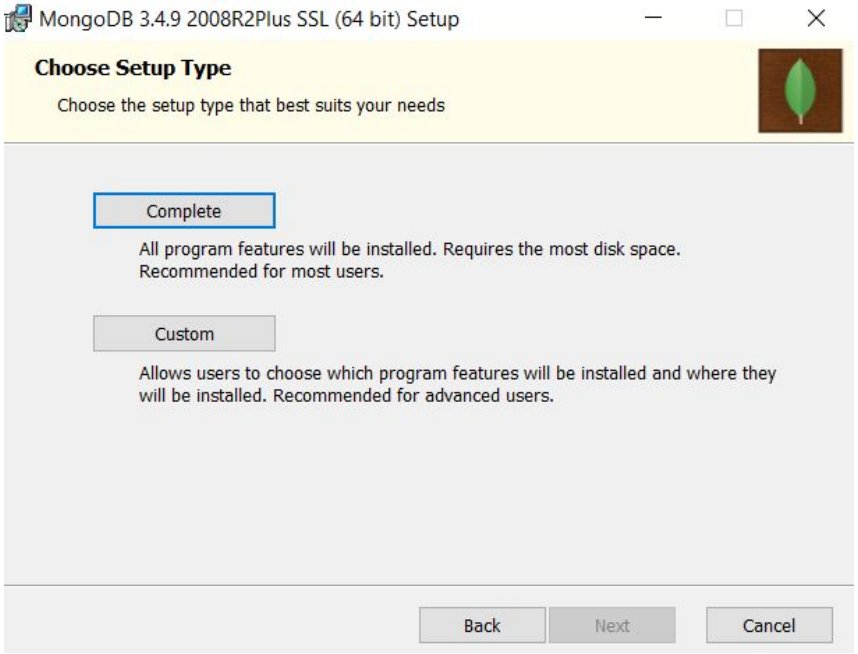


Studio 3T

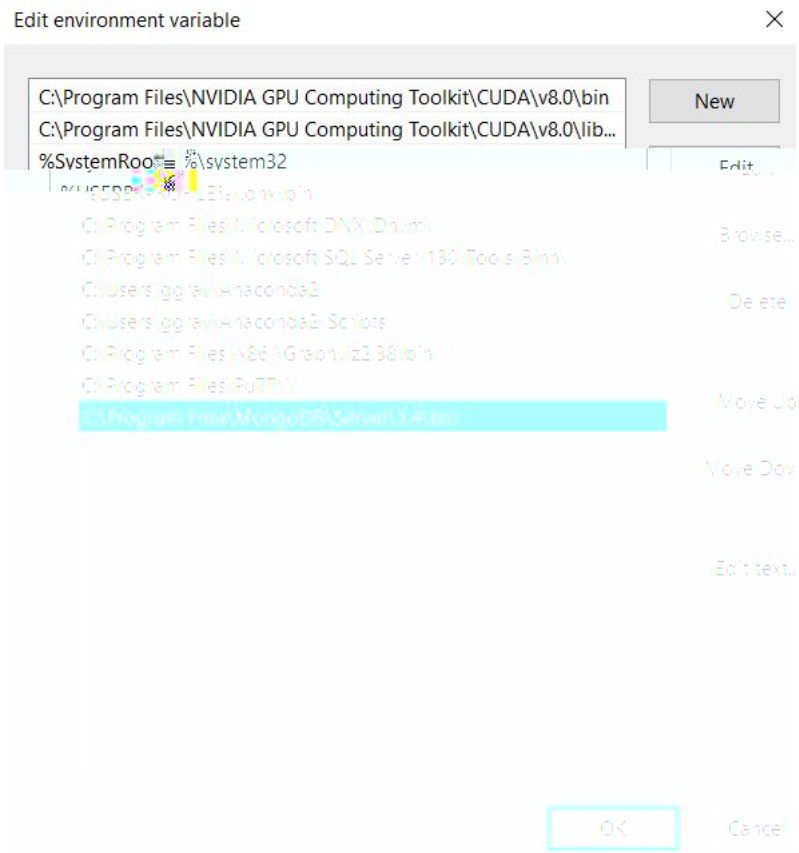


Installation

Step 1

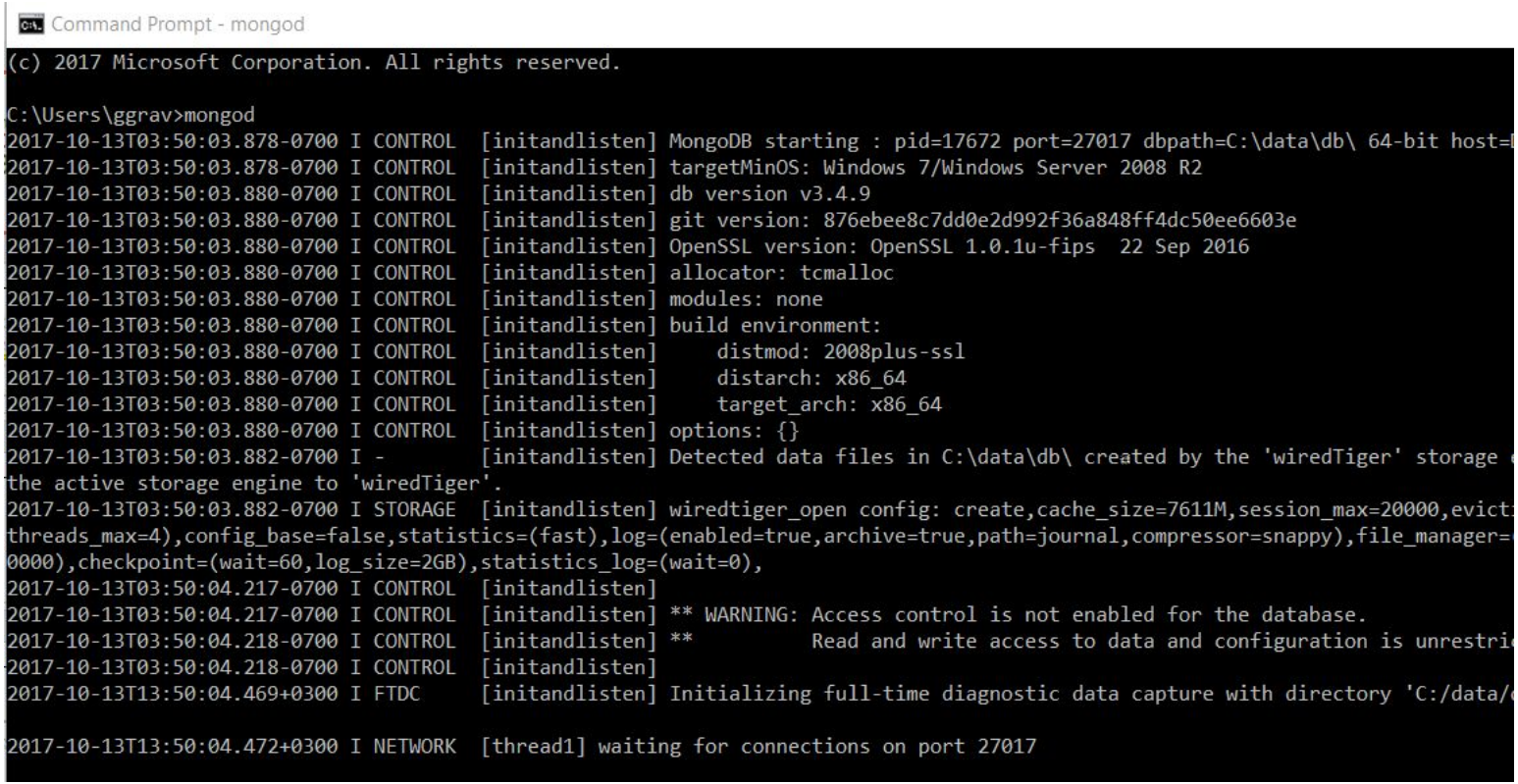


Step 2



Execution

Step 3 -> Create folder “C:\data\db\” -> Open cmd -> run “mongod”



What is a document

- A sum of key / value pairs

What is a collection

- A grouping of MongoDB **documents**.
- A **collection** is the equivalent of an RDBMS table.
- A **collection** exists within a single database.
- **Collections** do not enforce a schema.
- **Documents** within a collection can have different fields.
- **Collections** are grouped into **databases**.

Example: **db** = tweets, **collections** = greek_tweets, english_tweets, etc.

Operations

- Run / Connect to mongo
- Import documents
- Insert documents
- Create queries
- Quick data lookups

https://www.mongodb.com/what-is-mongodb

Connection

Studio 3T:

- Studio 3T -> Connect -> New connection -> (enter name) -> Save -> Connect
- Right click -> Add database -> (enter name) -> ok

Python:

```
1       import pymongo
2
3       client = pymongo.MongoClient('localhost', 27017)
4       db = client['yelp']
```

Import dataset

Studio 3T:

- Connect -> Select DB -> Import collections -> JSON -> Add sources -> Rename target collections -> Rename collections* -> Start import

* Rename collections to “restaurants”, “reviews” and “users”.

Insert documents

Python:

```
38          # Find the coordinates for each restaurant and
39          # save them to an external collection
40          all_restaurants = find_all_restaurants()
41          for restaurant in all_restaurants:
42              json_obj = {
43                  'name': restaurant['name'],
44                  'business_id': restaurant['business_id'],
45                  'longitude': restaurant['longitude'],
46                  'latitude': restaurant['latitude']
47              }
48              insert_to_db(json_obj, 'restaurants_coordinates')
```

Querying

Studio 3T:

- Right click collection -> Open Intellishell
- Examples:
  - db.restaurants.find({})
  - db.restaurants.find({"neighborhood": "Downtown"})
  - db.restaurants.find({ **$and**: [{"neighborhood": "Southeast"}, {"city": "Las Vegas"}]})

Python:

```
24          db['restaurants'].find({'neighborhood': neighborhood})
25          db['restaurants'].find_one({'business_id': restaurant_id})
26          db['reviews'].find({"$and": [{"business_id": restaurant_id}, {"stars": 5}]})
```

Data lookups - For quick data checking

Studio 3T:

- Right click collection -> Open Intellishell
  - Ex: db.collection_name.find({ "Search_Field": "value" }, { "Field_to_display": 1 })
  - db.restaurants.find({ "neighborhood": "Downtown" }, { "name": 1 })
  - db.restaurants.find({ "name": /.*pollos.*/ }, { "text": 1 })
    - /.* is the regex equivalent for: “any single character, 0 or more times”

Python:

```
27      def find_reviews_that_contain_a_word(word):
28          return db['reviews'].find({'text': {'$regex': '.*' + word + '.*'}})
29
```