

Quora uses cookies to improve your experience. [Read More](#)

Database Systems Machine Learning Computer Science

## Which database is best for machine learning?

Ad by DataCamp

•••

### How can I become a data scientist?

After being dubbed "sexiest job of the 21st Century" by Harvard Business Review, data science has stirred the interest of the general public. Many people are intrigued by the job...[\(Continue reading\)](#)

9 Answers



**Vlado Boza**, Phd. Computer Science, Comenius University in Bratislava (2017)

Answered January 15, 2018 · Author has 239 answers and 353.3K answer views

Originally Answered: Which database is best for machine learning ?

Short answer: PostgreSQL is second best database for anything ([Is PostgreSQL good enough?](#)). This also applies here.

Long answer: Depends on what is in your data (images, text, time series, marketing data, ...) and how big it is (mostly whether it fits in one machine or not). Also depends what is the source of your data and what do you want to do with them.

For example, if you only want to download ImageNet dataset and train some network on top of it, there is zero point in using any database. You just store your data on disk.

Sometimes, you want to also do some exploratory analytics on data. Then it really depends on type of data. Is it a lot of text? Use Elasticsearch. Is it a lot of columns with categorical/numerical features? Maybe use some SQL database. Is it super big and unstructured/semistructured? Look for some NoSQL database like HBase/Cassandra.

And sometimes your data is already in some database (like data about e-shop purchases). It might make sense to directly use that DB as a source instead building and maintaining another database. But be careful and do not overwrite anything there.

And finally, you usually do not want to store your model in database, but somewhere else, because it is usually just one big file, which you load in your production application/microservice.

24K views · View 19 Upvoters

### Related Questions

[More Answers Below](#)

[Is PostgreSQL the best choice of database for artificial intelligence using machine learning?](#)

[Which is the best way to build a database for machine learning?](#)

[Why don't Mongo DB and machine learning work well together?](#)

[Which database should I use for data science, MongoDB or MySQL?](#)

[What database should one use when doing machine learning operations?](#)



**Eugene Sokolov**, Computer engineer, world traveller and athlete with many hobbies

Answered March 25, 2016 · Upvoted by Vlado Boza, Phd. Computer Science, Comenius University in Bratislava (2017)

Originally Answered: Which database is best for machine learning ?

This question is a bit vague and there is no explanation or background to it.

I will assume the question is in regards to picking a database management system (DBMS) for some sort of machine learning project. One would not pick a "database" (you probably meant DBMS) for a machine learning project, but rather pick data (found on the Internet or bought or made or web scraped) to which machine learning algorithms are applied to.

In general, there are two types of DBMS: SQL (Structured Query Language) and NoSQL (Non SQL). One should take a Databases course to fully understand how databases work but the Wikipedia articles should do well for a start:

The main difference is how the data is stored, aka. how the data is structured. A few good references on the the SQL vs NoSQL discussion:

[SQL vs NoSQL Database Differences Explained with few Example DB](#)

[SQL vs. NoSQL- What You Need to Know](#)

[NoSQL Databases Explained](#)

[SQL \(MySQL\) vs NoSQL \(CouchDB\)](#)

Basically, you choose SQL or NoSQL based on the type of data you have. It is a way to structure your data. Once you have decided on either SQL or NoSQL, you pick a database management system. Within the structure type, all of the systems are extremely similar. A few examples of SQL type systems include: MySQL, PostgreSQL, SQL Server, and SQL lite. Some NoSQL type systems include MongoDB, Neo4j, Firebase, Microsoft Azure Tables, and CouchDB. The system is usually picked based on user preference, or sometimes performance or integration within an existing system. Syntactically, most systems are very similar with possible slight variations.

If you had meant which data is best for machine learning, then you my friend are lost.

32.3K views · View 36 Upvoters

Sponsored by DatadogHQ.com

#### Code-free, self-maintaining browser tests w/ Datadog Synthetics.

Easily build browser tests without code in minutes using a Web Recorder. Give it a try for free.

[Free Trial](#)



**Chris Richardson**, BS Biology, Computer Science, Psychology, Interesting Trivia & Computer Science, Florida Atlantic University ...

Answered March 6, 2018

Originally Answered: Which database is best for machine learning ?

I don't know what other databases the others described, but for me, it is MUMPS. It describes a hypothetical architecture that does not define how large a character is. There is only one data type, strings. C. J. Date, one of the designers of relational databases said that if he was to start again, he would have only one data type, character strings. We take the concept to another dimension, sparse matrices and no preallocation of anything as well as being multi-user on all of the platforms it runs on. If you create the first element in an array and then the millionth element in the array, and you have created only two elements. Also MUMPS makes little distinction between data and code. You can store code in a database and pull it out and execute it at run-time. It is beautiful for AI and large data storage. It can make a laptop look like a huge system. I have the Knight's tour problem written in MUMPS and have generated over 311 million solutions on a laptop at this time and the routine is surprisingly small and is re-entrant as well as being recursive. Best of all the interpreter is free for download and will run under Linux. I am running it currently on Ubuntu Mate. Greystone Technologies MUMPS supports UTF-8, the universal character set. There is also Vista, the system that has run the VA hospitals for 40 years and it is free for download. It is FOIA and as is, 180 aspects of the hospital. It is source code that runs the system in the MUMPS interpreter. Go to [Welcome to the WorldVista homepage](#) and see what we are all about.

6.7K views · View 2 Upvoters



**Galvin Widjaja**, CEO / Founder at Lauretta.io (2017-present)

Answered November 13, 2017 · Author has 718 answers and 2.2M answer views

Originally Answered: Which database is best for machine learning ?

The most likely answer is Spark with Hadoop HDFS.

The most common areas where machine learning will peel away from traditional statistical analytics is with large amounts of unstructured data.

The three main requirements of a database that will allow you to care less about it as you experiment with Big Data are quantity, data structure and database performance. Unless you are doing streaming analytics, you probably won't have to worry about velocity.

Spark + HDFS works well for this. The Spark architecture allows you to analyse in-memory and then store the data in big, low cost infrastructure.



Search for questions, people, and topics

Sign In

would already be duplicated 6 times.

26.7K views · View 18 Upvoters

---

Related Questions

More Answers Below

[Is learning MongoDB a good skill for machine learning or data scientist?](#)[Which database do you use for your online machine learning models?](#)[How do I make MySQL use machine learning to learn which is the correct rows to select from a database?](#)[How important is knowledge of databases in Machine Learning?](#)[For machine learning, do I need to learn a database management system?](#)**Sreejith S**, Software Developer

Answered November 13, 2017 · Upvoted by Vlado Boza, Phd. Computer Science, Comenius University in Bratislava (2017)

Originally Answered: Which database is best for machine learning ?

Well there isn't really any **Best Database** specifically for machine learning as such. The main goal of machine learning is basically to get enough quality data so that you algorithm/model can trained on it to learn the correlations/patterns present withing the data set to solve specific problems.

This data can come from a large databases like Oracle, MySQL, Redis, MongoDB, etc. or even from a small excel spreadsheet.

10.2K views · View 4 Upvoters · Answer requested by Prashant Pattanshetti

---

Sponsored by JetBrains**A must-have set of tools for multiplatform development.**

Get JetBrains Toolbox with its 15+ code editors for all languages and technologies included in one app.

 [Download](#)**Talvinder Singh**, loves ML, experience in RNN.

Answered March 10, 2017 · Author has 241 answers and 160.1K answer views

Originally Answered: Which database system is used for machine learning?

It depends on the requirement of your application. For example Hadoop architecture for data processing is different from traditional RDBMS. In my opinion, for simple applications traditional databases can be used (SQL or NoSQL). ML applications that requires heavy data processing can be processed using Spark. For research analysis, MATLAB can be used, also it depends in the capacity of your processing machine,

12.1K views · View 1 Upvoter

**Swapnil Powar**, Software Engineer

Answered March 26, 2018

Originally Answered: Which database is best for machine learning ?

**MLDB : The Machine Learning Database**

MLDB is an opensource database designed for machine learning. Send it commands over a RESTful API to store data, explore it using SQL, then train machine learning models and expose them as APIs.

6.9K views · View 7 Upvoters

**Miguel Paraz**, Bachelor's Computer Science, Ateneo de Manila University (1995)

Answered September 6, 2019 · Author has 7.5K answers and 11.2M answer views

Here's a product I came across. It's marketed as a database, but is arguably an API service - [Aito.ai](#)

But, I think more databases and similar services will be built for easier machine learning, and, mainstream databases will have machine learning built in.

4.2K views · View 6 Upvoters

 Search for questions, people, and topics

Sign In

Originally Answered: Which database is best for machine learning ?

Let try with Python

<http://Practical Machine Learning with Python>

Mariadb

[Database Design Example Phase 1: Analysis ↗](#)

7.3K views

#### Related Questions

[Is PostgreSQL the best choice of database for artificial intelligence using machine learning?](#)

[Which is the best way to build a database for machine learning?](#)

[Why don't Mongo DB and machine learning work well together?](#)

[Which database should I use for data science, MongoDB or MySQL?](#)

[What database should one use when doing machine learning operations?](#)

[Is learning MongoDB a good skill for machine learning or data scientist?](#)

[Which database do you use for your online machine learning models?](#)

[How do I make MySQL use machine learning to learn which is the correct rows to select from a database?](#)

[How important is knowledge of databases in Machine Learning?](#)

[For machine learning, do I need to learn a database management system?](#)

[What Should I use for creating a large database in order to use machine learning?](#)

[What is meant by a database, and how is it useful to machine learning?](#)

[Is there an open-source database to practice machine learning?](#)

[When would you use an SQL database for machine learning?](#)

[Which database is good for a recommender system?](#)

#### Related Questions

[Is PostgreSQL the best choice of database for artificial intelligence using machine learning?](#)

[Which is the best way to build a database for machine learning?](#)

[Why don't Mongo DB and machine learning work well together?](#)

[Which database should I use for data science, MongoDB or MySQL?](#)

[What database should one use when doing machine learning operations?](#)

[Is learning MongoDB a good skill for machine learning or data scientist?](#)