

([https://praxis.ac.in/data-science-program/?utm\\_source=Analytics%20India%20Magazine&utm\\_medium=Banner&utm\\_campaign=DS-Mar2020](https://praxis.ac.in/data-science-program/?utm_source=Analytics%20India%20Magazine&utm_medium=Banner&utm_campaign=DS-Mar2020)).

DEVELOPERS CORNER ([HTTPS://ANALYTICSINDIAMAG.COM/CATEGORY/LEARNING-CORNER/](https://analyticsindiamag.com/category/learning-corner/))

# Top Databases Used In Machine Learning Projects



BY AMBIKA CHOUDHURY ([HTTPS://ANALYTICSINDIAMAG.COM/AUTHOR/AMBIKA-CHOUDHURY/](https://analyticsindiamag.com/author/ambika-choudhury/))  
07/02/2020



One of the most critical components in machine learning projects is the database management system. With the help of this system, a large number of data can be sorted and one can gain meaningful insights from them. According to the Stack Overflow Survey [report](https://insights.stackoverflow.com/survey/2019#technology) (<https://insights.stackoverflow.com/survey/2019#technology>), 2019, Redis is the most loved database, whereas MongoDB is the most wanted database.

In this article, we list down 10 top databases used in machine learning projects.

*(The list is in alphabetical order)*



(<https://www.analytixlabs.co.in/>).

## 1| Apache Cassandra

Apache Cassandra (<https://analyticsindiamag.com/why-cassandra-and-hive-are-the-best-prospective-big-data-tools-for-ml/>) is an open-source and highly scalable NoSQL database management system that is designed to manage massive amounts of data in a faster manner. This popular database is being used by GitHub, Netflix (<https://www.slideshare.net/adrianco/migrating-netflix-from-oracle-to-global-cassandra>), Instagram, Reddit, among others. Cassandra has Hadoop integration, with MapReduce support.

Advantages:

- **Fault Tolerance:** In Cassandra, the data is automatically replicated to multiple nodes for fault-tolerance. Also, failed nodes can be replaced with [new nodes](https://analyticsindiamag.com/) (<https://analyticsindiamag.com/>).
- **Elastic Scalability:** Cassandra is designed with both read and write throughput, which increases linearly as new machines are added.

## 2| Couchbase

Couchbase Server (<https://www.couchbase.com/>) is an open-source, distributed, NoSQL document-oriented engagement database. It exposes a fast key-value store with managed cache for sub-millisecond data operations, purpose-built indexers for fast queries and a powerful query engine for executing SQL-like queries.

Advantages:

- **Unified Programming Interface:** The Couchbase Data Platform provides simple, uniform and powerful application development APIs across multiple programming languages, connectors, and tools that make building applications simple and accelerates time to market for applications.
- **Big data and SQL Integrations:** Couchbase Data platform includes built-in Big Data and SQL integration which allows a user to leverage tools, processing capacity, and data wherever it may reside.
- **Container and Cloud Deployments:** Couchbase supports all cloud platforms as well as a variety of container and virtualisation technologies.

## 3| DynamoDB

Amazon DynamoDB a fully managed, multi-region, durable database with built-in security, backup and restore, and in-memory caching for internet-scale applications. This accessible database has been using by Lyft, Airbnb, Toyota, Samsung, among others. DynamoDB offers encryption at rest which eliminates the operational burden and complexity involved in protecting sensitive data.

Advantages:

- **High Availability and Durability:** DynamoDB automatically spreads the data and traffic for the tables over a sufficient number of servers to handle the throughput and storage requirements while maintaining consistent as well as fast performance.
- **Performance at Scale:** DynamoDB provides consistent as well as single-digit millisecond response times at any scale. The DynamoDB global tables replicate the data across multiple AWS regions in order to provide fast and local access to data for globally distributed applications.

## 4| Elasticsearch

Elasticsearch (<https://www.elastic.co/>) is built on Apache Lucene and is a distributed, open-source search and analytics engine for all types of data including textual, numerical, geospatial, structured and unstructured data. Elasticsearch is the central component of the Elastic Stack which is a set of open-source tools for data ingestion, enrichment, storage, analysis, and visualisation.

Advantages:

- **Extensive Number of Features:** Besides speed, scalability and resiliency, Elasticsearch has several built-in features such as data rollups and index lifecycle management which makes efficient storing and searching data.
- **Faster in Manner:** Elasticsearch excels at full-text search and it is well-suited for time-sensitive use cases such as security analytics, infrastructure monitoring, etc.

## 5| MLDB

The Machine Learning Database (<https://mldb.ai/>) (MLDB) is an open-source system for solving big data machine learning problems, from data collection and storage through analysis and the training of machine learning models to the deployment of real-time prediction endpoints. In MLDB, machine learning models are applied using Functions, which are parameterised by the output of training Procedures, which run over Datasets containing training data.

Advantages:

- **Easy to Use:** MLDB provides a comprehensive implementation of the SQL SELECT statement, treating datasets as tables, with rows as relations [cm \(https://analyticsindiamag.com/\)](https://analyticsindiamag.com/) to learn and use for data analysts familiar with existing Relational Database Management Systems (RDBMS).



## 6| Microsoft SQL Server

Written in C and C++, [Microsoft SQL Server \(https://www.microsoft.com/en-in/sql-server/default.aspx\)](https://www.microsoft.com/en-in/sql-server/default.aspx) is a relational database management system (RDBMS). This database helps in gaining insights from all the data by querying across relational, non-relational, structured as well as unstructured data.

Advantages:

- **Flexible:** One can use the language and platform of choice with open source support.
- **Manage Big Data Environment:** With SQL Server, one can manage big data environment more easily with Big Data Clusters. It provides vital elements of a data lake such as Hadoop Distributed File System (HDFS), Apache Spark and analytics tools which are deeply integrated with SQL Server and fully supported by Microsoft

## 7| MySQL

Written in C and C++, [MySQL \(https://www.mysql.com/\)](https://www.mysql.com/) is one of the most popular open-source relational database management systems (RDBMS) powered by Oracle. It has been used by successful organisations such as Facebook, Twitter, YouTube, among others.

SEE ALSO



([https://an  
alyticsindi  
amag.com/  
top-10-  
python-  
open-  
source-  
projects-  
on-  
github-  
2019/](https://analyticsindiamag.com/top-10-python-open-source-projects-on-github-2019/))

[Top 10 Python Open Source Projects On GitHub: 2019.  
\(https://analyticsindiamag.com/top-10-python-open-source-projects-  
on-github-2019/\)](https://analyticsindiamag.com/top-10-python-open-source-projects-on-github-2019/)

Advantages:

- **Security and Scalability:** This database management system includes data security layers that protect sensitive data and it offers scalability to handle large amounts of data.
- **Backup Software:** mysqldump is a logical backup tool included with both community and enterprise editions of MySQL. It supports backing up from all storage engines.

## 8| MongoDB

[MongoDB \(https://www.mongodb.com/\)](https://www.mongodb.com/) is a general-purpose, document-based, distributed database which is built for advanced application developers. Since this is a document database, it mainly stores data in JSON-like documents. It provides support for aggregations and other modern use-cases such as geo-based search, graph search, and text search.

Advantages:

- **Data Store Flexibility:** MongoDB stores data in flexible, JSON-like documents which means fields can vary from document to document and data structure can be changed over time.
- **Distributed Database:** MongoDB is a distributed database at its core. Which is why high availability, horizontal scaling, and geographic distribution are built-in and easy to use.

## 9| PostgreSQL

AIM (https://analyticsindiamag.com/)



PostgreSQL (https://www.postgresql.org/) is a powerful, open-source object-relational database system which uses and extends the SQL language combined with many features that safely store and scale the most complicated data workloads. This database management system aims to help developers build applications, administrators to protect data integrity, build fault-tolerant environments and much more.

Advantages:

- **Security:** PostgreSQL has a robust access-control system as well as column and row-level security.
- **Extensibility:** This system has foreign data wrappers which connect to other databases or streams with a standard SQL interface.

## 10| Redis

Redis (https://redis.io/) is an open-source, in-memory data structure store which is used as a database, cache and message broker. It supports data structures such as strings, sorted sets with range queries, bitmaps, hyperloglogs, geospatial indexes, etc. The database has built-in replication, Lua scripting, LRU eviction, transactions and different levels of on-disk persistence.

Advantages:

- **Automatic Failover:** In Redis Sentinel, a failover process can be started where a replica is promoted to master and the other additional replicas can be reconfigured to use the new master.
- **Redis-ML:** Redis-ML is a Redis module which implements several machine learning models as built-in Redis data types. It is simple to load and deploy trained models from any platform (such as Apache Spark and scikit-learn) in a production environment.

Provide your comments below

0 comments

---

*If you loved this story, do join our **Telegram Community** (<https://t.me/joinchat/NJLxnhZB7GkX3CPvjs9QGQ>).*

Also, you can write for us and be one of the 500+ experts who have contributed stories at AIM. Share your **nominations here** (<https://analyticsindiamag.com/write-for-us/>).

---



AMBIKA CHOUDHURY ([HTTPS://ANALYTICSINDIAMAG.COM/AUTHOR/AMBIKA-CHOUDHURY/](https://analyticsindiamag.com/author/ambika-choudhury/))

A Technical Journalist who loves writing about Machine Learning and Artificial Intelligence. A lover of music, writing and learning something out of the box. Contact: [ambika.choudhury@analyticsindiamag.com](mailto:ambika.choudhury@analyticsindiamag.com)

f SHARE

(<https://www.facebook.com/sharer.php?u=https://analytic>

🐦 TWEET

(<https://twitter.com/share?text=Top%20Databases%20Used%20In%20Machine%20Learning%20Projects&via=Analyticsindiamag>  
[machine-learning-projects/](https://twitter.com/share?text=Top%20Databases%20Used%20In%20Machine%20Learning%20Projects&via=Analyticsindiamag))

📌 ([https://pinterest.com/pin/create/bookmarklet/?url=https://analyticsindiamag.com/top-databases-used-in-machine-learning-projects-content/uploads/2020/02/aim\\_database.jpg](https://pinterest.com/pin/create/bookmarklet/?url=https://analyticsindiamag.com/top-databases-used-in-machine-learning-projects-content/uploads/2020/02/aim_database.jpg))

in (<https://www.linkedin.com/cws/share?url=https://analyticsindiamag.com/top-databases-used-in-machine-learning-projects/>)

📧 (<https://wa.me/?text=Top%20Databases%20Used%20In%20Machine%20Learning%20Projects%20https://analyticsindiamag.com/top-databases-used-in-machine-learning-projects/>)  
(mailto:?)

Scroll To Top

subject=Top%20Databases%20Used%20In%20Machine%20Learning%20Projects&body=Top%20Databases%20Used%20In%20Mach  
databases-used-in-machine-learning-proje (https://analyticsindiamag.com/).

(https://reddit.com/submit?url=https://analyticsindiamag.com/top-databases-used-in-machine-learning-projects/).

(https://t.me/share?url?&text=Top%20Databases%20Used%20In%20Machine%20Learning%20Projects&url=https://analyticsindiam



(https://www.adasci.org/cds/).



(https://business.louisville.edu/learnmore/msba-india/?

utm\_source=analyticsIndia&utm\_medium=Banner&utm\_term=India&utm\_content=GetPaidIndia&utm\_campaign=MSBAIndia).

## OUR UPCOMING EVENTS

Data Engineering: Careers and Skills (https://register.gotowebinar.com/register/940328042302574863) | 12th Sep | Webinar

Deep Learning DevCon 2020 (https://dlcd.adasci.org/) | 29-30th Oct | Virtual



RELATED POSTS

## 10 Most Used Databases By Developers In 2020 (https://analyticsindiamag.com/10-most-used-databases-by-developers-in-2020/)

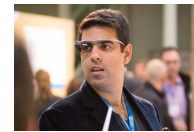
25/08/2020 · 3 MINS READ



(https://analyticsindiamag.com/10-most-used-databases-by-developers-in-2020/)

## From Engineer.AI To Builder.AI, Sachin Dev Duggal Reveals Why He Rebranded His Software Automation Startup (https://analyticsindiamag.com/sachin-duggal-speaks-on-why-he-rebranded-his-software-automation-startup-engineer-ai-to-builder-ai/)

01/05/2020 · 5 MINS READ



(https://analyticsindiamag.com/sachin-duggal-speaks-on-why-he-rebranded-his-software-automation-startup-engineer-ai-to-builder-ai/)

## This Is What Google TensorFlow Is Giving Away For Free Now (https://analyticsindiamag.com/this-is-what-google-tensorflow-is-giving-away-for-free-now/)

01/05/2020 · 3 MINS READ




(https://analyticsindiamag.com/this-is-what-google-tensorflow-is-giving-away-for-free-now/)

## Top Free 9 Resources To Learn Python For Machine Learning (https://analyticsindiamag.com/top-free-9-resources-to-learn-python-for-machine-learning/)

30/04/2020 · 3 MINS READ



(https://analyticsindiamag.com/top-free-9-resources-to-learn-python-for-machine-learning/)

 [\(https://analyticsindiamag.com/\)](https://analyticsindiamag.com/)

# Micron Launches Open-Source Storage Engine Designed For SSDs & Storage Class Memory.

[\(https://analyticsindiamag.com/micron-launches-open-source-storage-engine-designed-for-ssds-storage-class-memory/\)](https://analyticsindiamag.com/micron-launches-open-source-storage-engine-designed-for-ssds-storage-class-memory/)



[\(https://analyticsindiamag.com/micron-launches-open-source-storage-engine-designed-for-ssds-storage-class-memory/\)](https://analyticsindiamag.com/micron-launches-open-source-storage-engine-designed-for-ssds-storage-class-memory/)

28/04/2020 · 3 MINS READ

# 11 Best Firms In India For Data Scientists To Work For – 2020

[\(https://analyticsindiamag.com/11-best-firms-in-india-for-data-scientists-to-work-for-2020/\)](https://analyticsindiamag.com/11-best-firms-in-india-for-data-scientists-to-work-for-2020/)



[\(https://analyticsindiamag.com/11-best-firms-in-india-for-data-scientists-to-work-for-2020/\)](https://analyticsindiamag.com/11-best-firms-in-india-for-data-scientists-to-work-for-2020/)

27/04/2020 · 12 MINS READ

More than 1,00,000 people  
are subscribed to our  
newsletter

Subscribe now to receive in-depth stories on AI & Machine Learning.

[ABOUT US\(HTTPS://ANALYTICSINDIAMAG.COM/ABOUT/\)](https://analyticsindiamag.com/about/)

[ADVERTISE\(HTTPS://ANALYTICSINDIAMAG.COM/ADVERTISE-WITH-US/\)](https://analyticsindiamag.com/advertise-with-us/)

[WRITE FOR US\(HTTPS://ANALYTICSINDIAMAG.COM/WRITE-FOR-US/\)](https://analyticsindiamag.com/write-for-us/) [\(https://analyticsindiamag.com/\)](https://analyticsindiamag.com/)

[COPYRIGHT\(HTTPS://ANALYTICSINDIAMAG.COM/COPYRIGHT-TRADEMARKS/\)](https://analyticsindiamag.com/copyright-trade-marks/)

[PRIVACY\(HTTPS://ANALYTICSINDIAMAG.COM/PRIVACY-POLICY/\)](https://analyticsindiamag.com/privacy-policy/)

[TERMS OF USE\(HTTPS://ANALYTICSINDIAMAG.COM/TERMS-USE/\)](https://analyticsindiamag.com/terms-use/)

 [\(HTTPS://FACEBOOK.COM/ANALYTICSINDIAMAGAZINE\)](https://facebook.com/analyticsindiamagazine/)

 [\(HTTPS://TWITTER.COM/ANALYTICSINDIAMAG\)](https://twitter.com/analyticsindiamag/)

 [\(HTTPS://INSTAGRAM.COM/ANALYTICSINDIAMAGAZINE\)](https://instagram.com/analyticsindiamagazine/)

 [\(HTTPS://PINTEREST.COM/ANALYTICSINDIAMAG\)](https://pinterest.com/analyticsindiamag/)

  [\(HTTPS://YOUTUBE.COM/CHANNEL/UCALWRSGEJAVG1VW9QSH9DMA\)](https://youtube.com/channel/UCALWRSGEJAVG1VW9QSH9DMA)

 [\(HTTPS://MEDIUM.COM/ANALYTICS-INDIA-MAGAZINE\)](https://medium.com/analytics-india-magazine/)

