Search Medium   Write

You have **2** free member-only stories left this month.

Sign up for Medium and get an extra one

Amine Kaabachi   Follow

Jan 1 · 4 min read · ✦ Member-only · ▶ Listen

2023 — Rockstar Data Engineer Roadmap

This article presents a roadmap for those who want to become Data Engineers in 2023. It also serves as a reference to learn and improve the understanding of the different skills required for this position.

For each skill category, I cite the skill points that are recommended with links to video/text courses and some reference books at the end. I have no association with the author, and this is not a referral article. The goal is to provide an easy way for you to navigate and improve your skills :)!

👏 84   💬 3

Photo by delfi de la Rua on Unsplash

To become a skilled data engineer, you need to know computer science, dev, networking, and database fundamentals.

CS Fundamentals

- *Data Structures and Algorithms.*

- *Basic Computer Architecture.*

- *APIs & REST.*

- *Linux (Terminal, Shell Scripting, Cronjobs, Processes).*

Reference Books

The Algorithm Design Manual

Computer Systems: A Programmer's Perspective

How Linux Works: What Every Superuser Should Know

Dev Fundamentals

- *Git.*

- *Python, Scala, Java, Go.*

- *Testing: Unit tests, Integration Tests, Functional Testing.*

- *VIM / VSCode.*

- *SCRUM / Agile Methodologies.*

- *SOLID / Refactoring.*

Reference Books

Operating Systems: Three Easy Pieces

Refactoring: Improving the Design of Existing Code

Networking Fundamentals

- *HTTP / HTTPs.*

- *TCP / IP / SSH / DNS.*

- *VPN, VPC.*

- *Firewalls.*

Reference Books

Computer Networking: A Top-Down Approach

Database Fundamentals

- *Relational Algebra / Normalisation.*

- *SQL.*

- *[ACID / CAP](#)*.

- *[OLTP vs OLAP.](#)*

- *[Datawarehouses and Datamarts.](#)*

- *[Datalake / Cloud Data Platform.](#)*

Reference Books

[Readings in Database Systems](#)

Learn Data Science in a Flash!? | Data Driven Investor

I was a trained classical pianist in my previous professional life. Remember those infomercials claiming that you could...

www.datadriveninvestor.com

Data Storage

- *Relational: MySql, PostgreSQL, MariaDB, Amazon Aurora.*

- *Document: MongoDB, Elasticsearch, Apache CouchDB.*

- *Wide column: Cassandra, Hbase, Google Bigtable.*

- *Graph: Neo4j, Amazon Neptune.*

- *Key-value: Redis, Memcached, Amazon DynamoDB.*

- *Messaging: RabbitMQ, Apache ActiveMQ.*

- *Data warehouses: Snowflake, Presto, Apache Hive, Amazon Redshift, Google BigQuery.*

- *Datalakes: S3, ADLS Gen2.*

Reference Books

Designing Data-Intensive Applications

Data Processing:

- *Cluster Computing: Hadoop, HDFS, MapReduce.*

- *Batch: Apache Pig, Apache Arrow, Apache Impala.*

- *Hybrid: Apache Spark, Apache Beam, Apache Flink.*

- *Streaming: Apache Kafka, Apache Storm, Apache Samza, Amazon Kinesis.*

- *Managed Solutions: Databricks, Amazon EMR, Google Dataproc, Azure HDInsight.*

Workflow Scheduling / ETC

- *Apache Airflow / Google Composer.*

- *Azure Data Factory.*

- *AWS Glue.*

- *Apache Oozie.*

DevOps & Security

- *Jenkins, Azure Devops, AWS CodePipeline, Github Actions.*

- *Docker, Kuberentes and Helm.*

- *Active Directory / Azure Active Directory.*

- *Encryption, Key Management.*

- *Data Governance, GDPR.*

Photo by davisuko on Unsplash

It's time to prove your skills. Here are some certifications that can help you get your dream job or just make you more confident.

Valuable Certifications

Databricks Certified Data Engineer Associate — Databricks

Key details about the certification exam are provided below.

databricks.com

Databricks Certified Data Engineer Professional — Databricks

Key details about the certification exam are provided below.

databricks.com

Databricks Certified Associate Developer for Apache Spark 3.0

The Databricks Certified Associate Developer for Apache Spark 3.0 certification exam assesses an understanding of the...

academy.databricks.com

Microsoft Certified: Azure Data Engineer Associate — Learn

This exam measures your ability to accomplish the following technical tasks: design Azure data storage solutions...

docs.microsoft.com

Certification AWS Certified Big Data — Specialty

L'examen AWS Certified Big Data — La certification Specialty est destinée aux personnes qui effectuent des analyses...

aws.amazon.com

Professional Data Engineer Certification | Certifications

A Professional Data Engineer enables data-driven decision making by collecting, transforming, and publishing data. A...

cloud.google.com

CCA Spark and Hadoop Developer Certification — Cloudera

This may have been caused by one of the following: Your request timed out A plugin/browser extension blocked the...
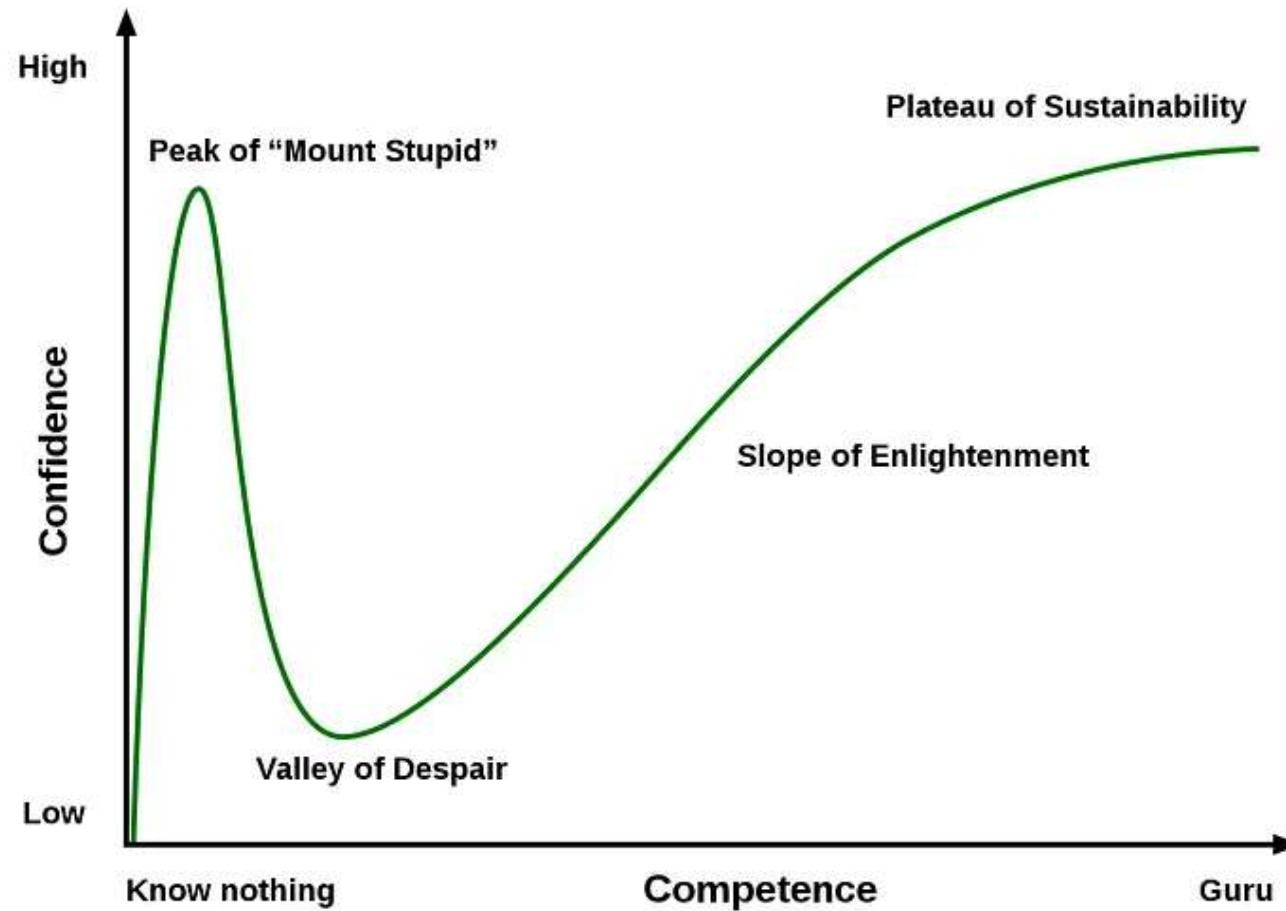
www.cloudera.com

Certified Kubernetes Application Developer (CKAD) | Cloud Native Computing Foundation

As one of the highest velocity projects in the history of open source, Kubernetes use is exploding. The Cloud Native...

www.cncf.io

Conclusion

Keep this story bookmarked! It can help you learn efficiently and become a better data engineer. You are not born a guru, It is continuous trial and error that improves your skills.

Dunning–Kruger Effect

[Dunning–Kruger Effect](#) — Wikimedia

Data Engineering   Data Science   Data   Programming   Certification

Get an email whenever Amine Kaabachi publishes.

Your email

✉⁺ Subscribe

By signing up, you will create a Medium account if you don't already have one. Review our Privacy Policy for more information about our privacy practices.

About   Help   Terms   Privacy