

# WHAT IS DATA ENGINEERING?

by Taras Kloba, Big Data Engineering Manager

**softserve**



## TARAS KLOBA

Big Data Engineering Manager

### QUICK FACTS

- Trainer, speaker, consultant
- More than 12 years of experience with databases
- Winner of Ukrainian IT Awards 2019 in Software Architecture
- Co-leader of PostgreSQL Ukraine and JOIN communities
- Certified Cloud Architect & Data Engineer on Google Cloud
- Certified Expert on Microsoft Azure
- SQL Saturday volunteer in Lviv and Krakow
- Ph.D. in Economics



x33M4y



M2HE6LPRJ6MV



Q62JCJRJGY77

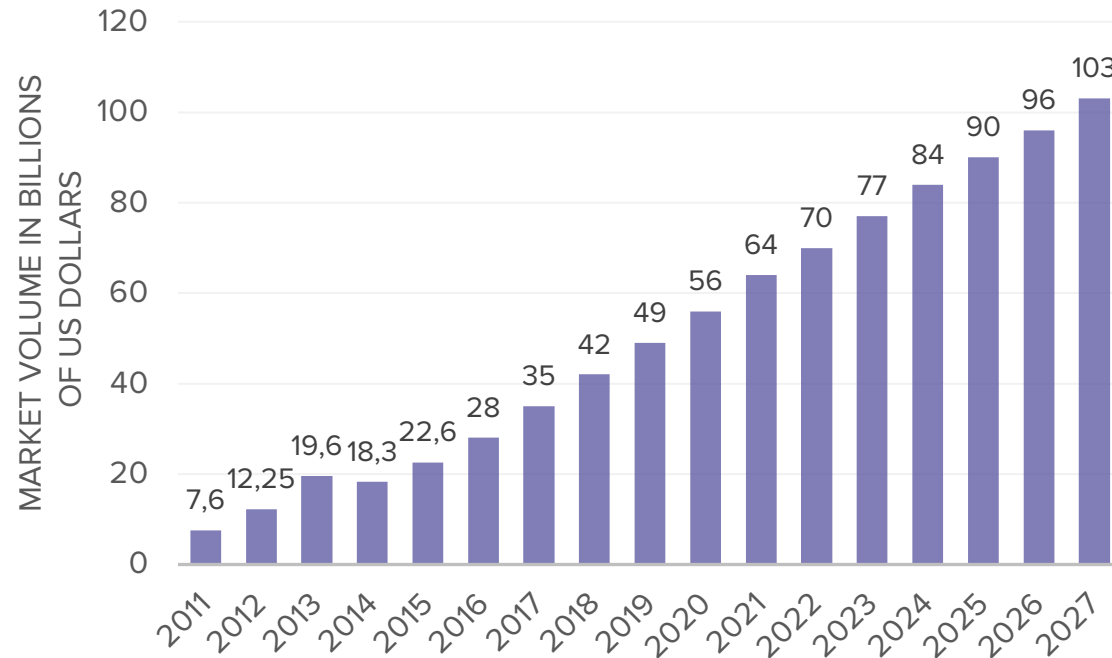
**1.7 MB/SECOND x 5'000'000'000**



softserve

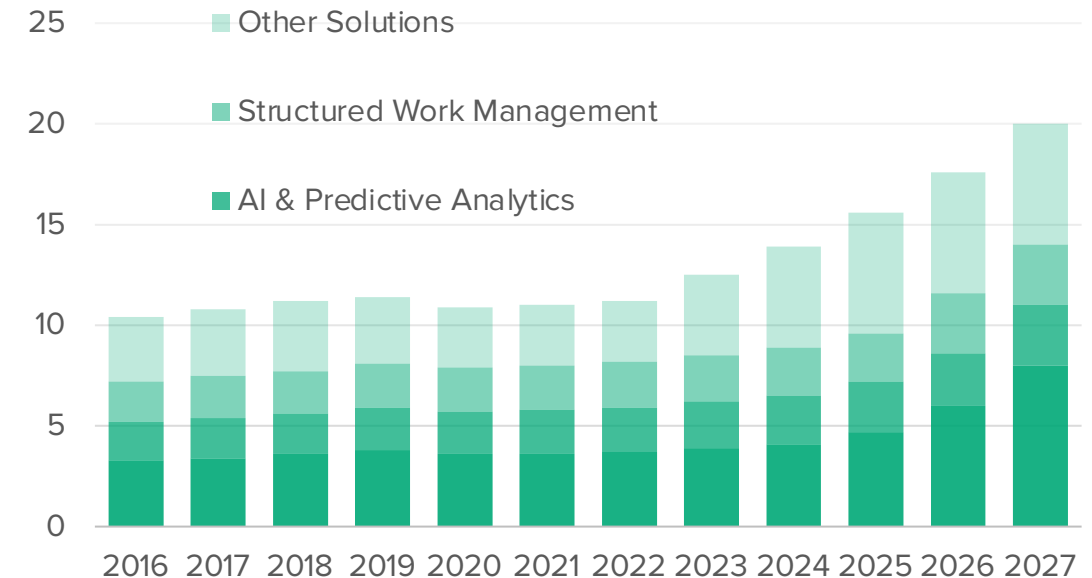
# MARKETS: BIG DATA VS SOFTWARE

The Big Data market size revenue forecast worldwide. During 2011–2027 (billions of US dollars)



54,36%

US productivity management software market size, by solution. During 2016–2027 (billions of US dollars)



41,38%

**softserve**



# HOW DOES A DISTRIBUTED SYSTEM WORK?

## 1 Machine:

- 4 I/O channels
- Each channel—100 MB/s
- Data = 1 Terabyte

Processing time:

00:45:00

HH:MM:SS



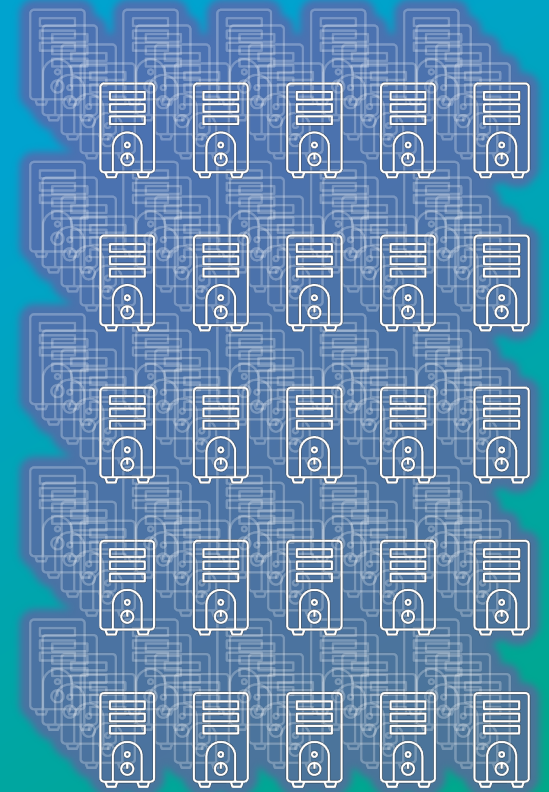
## 100 Machines:

- 4 I/O channels
- Each channel—100 MB/s
- Data = 1 Terabyte

Processing time:

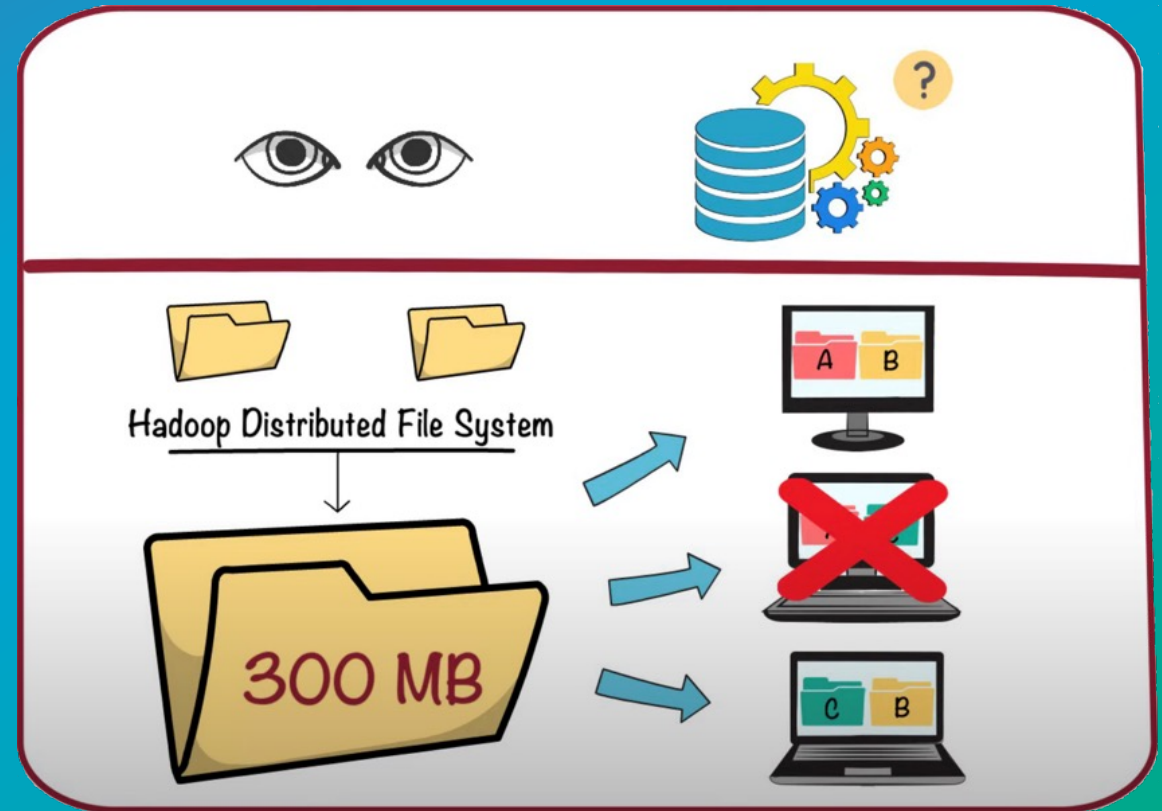
00:00:45

HH:MM:SS



softserve

# BIG DATA FRAMEWORKS



softserve

# 1000x DATAPROC CLUSTERS DAILY

[Multinational retail corporation] + SoftServe = <3

- Dataproc cluster pools—unique GCP feature
- Dataproc Omni—on GKE, on Azure
- Migration automation
- Profiler to optimize ephemeral clusters workloads
- Ranger-based centralized security
- Unified Monitoring and Alerting
- Persistent History Server solution

Our contribution:



**softserve**



# [CANADIAN MINING COMPANY]

Leverage the vast amount of historical sensor and production data along with state-of-the-art Cloud and AI technologies to optimize haul road, truck and operator's performance

**39%**

Overall equipment effectiveness (OEE) improvement

**27%**

Maintenance cost decreased

**softserve**

# DEMO

softserve

# BQ VISUALISER

## BQ Visualiser

Select Job

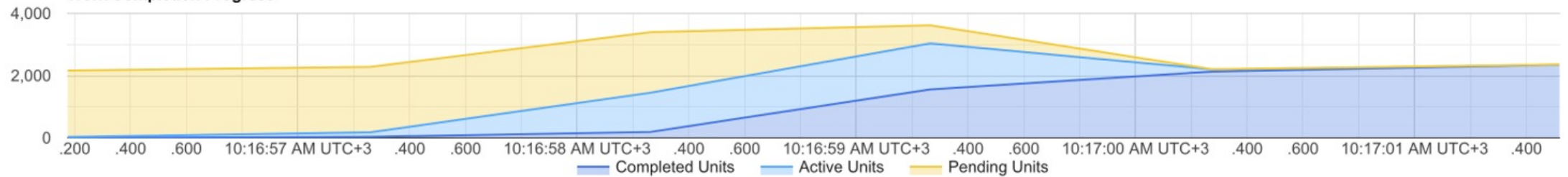
Tree

Timing

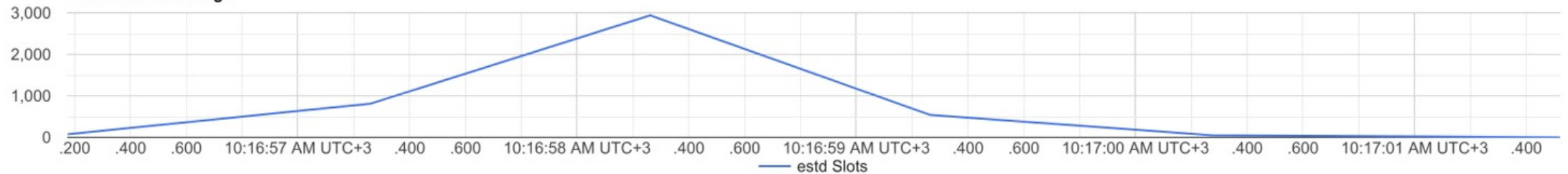
Progress

Debug Logs

### Work Completion Progress



### Estimated Slot Usage



Overview

Status

SQL

Timings

Statistics

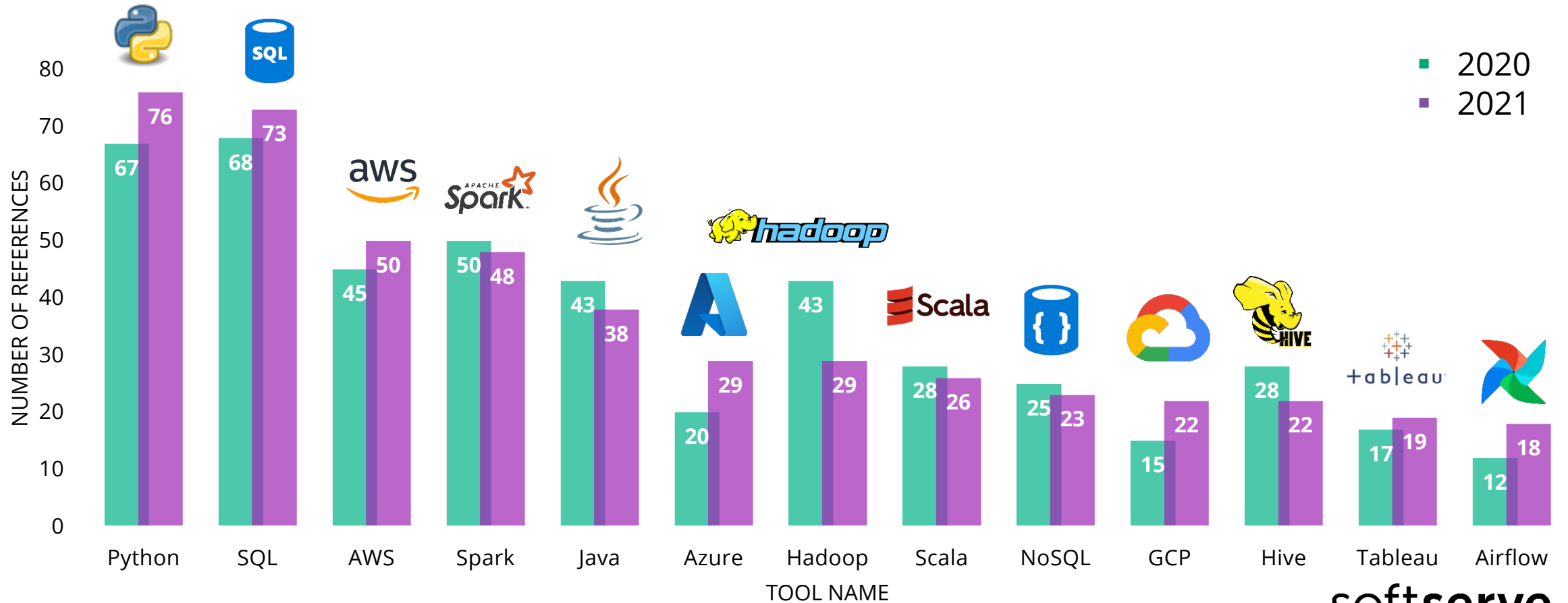
Settings

Display Options

softserve



# THE MOST COMMON TECHNOLOGIES IN JOB LISTING IN 2020 VS 2021



softserve



**BECOME  
PROFICIENT AT  
PROGRAMMING**

softserve

# SOFTWARE ENGINEERING (PROGRAMING LANGUAGES)



1991



Java™

1995



Scala

2004

2005 – Big Data

softserve

# RECOMMENDATIONS FOR BEGINNERS

- Python:
  - If you already know how to code, learn from the source itself:  
<https://wiki.python.org/moin/BeginnersGuide>
  - If you are new to programming:  
<https://wiki.python.org/moin/BeginnersGuide/NonProgrammer>
  - The wiki has plenty of learning resources, many of which are free:  
<https://www.reddit.com/r/learnpython/wiki/index>
- *Scala*
- *Java*



1 month



\$0.00



softserve

# TAKE VIDEO COURSES ON PYTHON

## Automate the Boring Stuff with Python Programming


A practical programming course for office workers, academics, and administrators who want to improve their productivity.

4.6 ★★★★★ (96,165 ratings) 1,012,727 students

Created by [Al Sweigart](#)

🕒 Last updated 4/2021 🌐 English 🗣️ English, French [Auto], 5 more

Share ⋮



Preview this course

Enroll now

This course includes:

- 9.5 hours on-demand video
- 95 downloadable resources
- Access on mobile
- Certificate of completion

### What you'll learn

- ✓ Automate tasks on their computer by writing simple Python programs.
- ✓ Programatically generate and update Excel spreadsheets.
- ✓ Crawl web sites and pull information from online sources.
- ✓ Use Python's debugging tools to quickly figure out bugs in your code.
- ✓ Write programs that can do text pattern recognition with "regular expressions".
- ✓ Parse PDFs and Word documents.
- ✓ Write programs that send out email notifications.
- ✓ Programatically control the mouse and keyboard to click and type for you.

<https://softserve.udemy.com/course/automate/>

\*For beginners



\$32.98



1 month

softserve

Programming Languages > Python

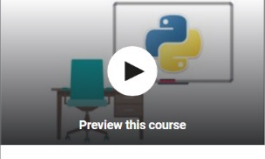
## Python for Data Structures, Algorithms, and Interviews!

Get a kick start on your career and ace your coding interviews!

4.5 ★★★★★ (10,014 ratings) 67,657 students

Created by [Jose Portilla](#)

🕒 Last updated 9/2019 🌐 English 🗣️ English, French [Auto], 8 more



Preview this course

\$9.99 ~~\$84.99~~ 88% off  
🕒 23 hours left at this price!

Add to cart

Buy now

30-Day Money-Back Guarantee

This course includes:






- 16.5 hours on-demand video
- 20 articles
- 43 downloadable resources
- Full lifetime access
- Access on mobile and TV
- Certificate of completion

[Apply Coupon](#) [Gift this course](#)

### What you'll learn

- ✓ Create a great resume
- ✓ Understand all major Data Structures and Algorithms
- ✓ Have a LinkedIn and Github profile for recruiters
- ✓ Ace coding interviews after preparing with the course's mock interviews

**Top companies offer this course to their employees**  
This course was selected for our collection of top-rated courses trusted by businesses worldwide. [Learn more](#)

<https://www.udemy.com/course/python-for-data-structures-algorithms-and-interviews/>

# CONTRIBUTE TO OPEN SOURCE

The screenshot shows the OPEN/TECH website. At the top is a navigation bar with links: HOME, PROJECTS, VOLUNTEERS, PARTNERS, BE A VOLUNTEER, ABOUT, UA, and LOG IN. The main header features the text "OPEN/TECH" in large white letters, followed by a description: "A charity crowdsourcing platform where small and individual contributions of time and skills combine to make significant changes in the world around us." Below this, a section titled "CURRENTLY WE HAVE" displays three statistics in a grid: "ACTIVE PROJECTS 13", "FINISHED PROJECTS 16", and "VOLUNTEERS 222". At the bottom of this grid are two buttons: "BROWSE ALL PROJECTS" and "SEE ALL VOLUNTEERS". The background of the page shows a blurred image of people working at computers.

ACTIVE PROJECTS	FINISHED PROJECTS	VOLUNTEERS
13	16	222

<https://opentech.softserveinc.com/en/projects>

The screenshot shows the CodeTriage website with the heading "Open source projects on GitHub that need your help." Below the heading is a filter bar set to "Python". A grid of project cards is displayed, each with a project name, a count of issues, and a brief description. The projects listed are: odoo (2467 issues), youtube-dl (2096 issues), youtube-dl (2085 issues), sympy (1955 issues), salt (1898 issues), pandas (1813 issues), pandas (1809 issues), scikit-learn (1435 issues), zulip (1386 issues), and nvda (1382 issues). Each card also includes a link to the project's GitHub repository.

Project	Issues
odoo	2467
youtube-dl	2096
youtube-dl	2085
sympy	1955
salt	1898
pandas	1813
pandas	1809
scikit-learn	1435
zulip	1386
nvda	1382

<https://www.codetriage.com/?language=Python>

 \$0.00

 1 month

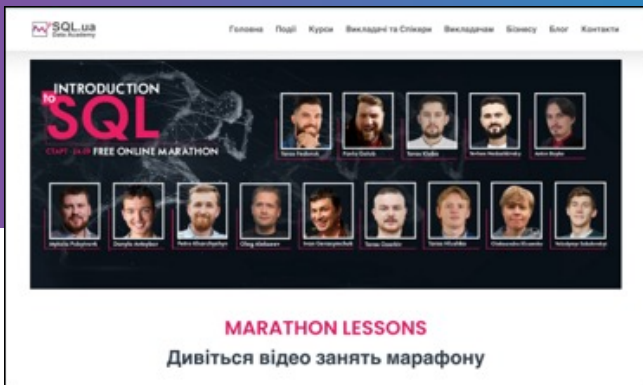
softserve





# UNDERSTAND YOUR DATABASES

softserve



<https://academy.sql.ua/intro-to-sql/>



<https://academy.sql.ua/course-postgresql-for-developers/>

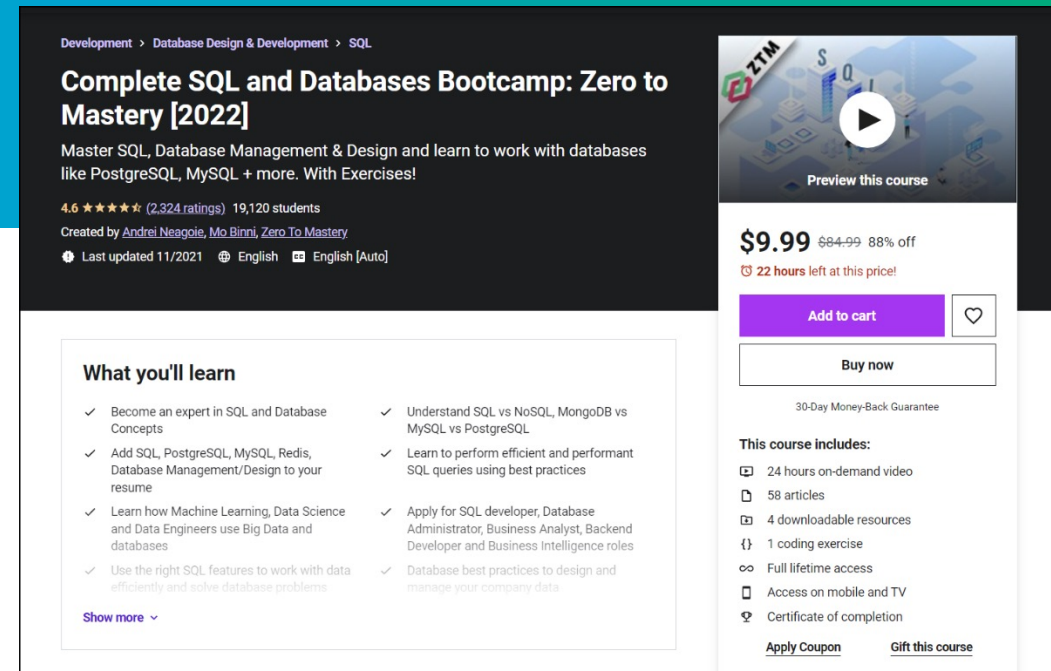


\$13.99



1 month

softserve



<https://www.udemy.com/course/complete-sql-databases-bootcamp-zero-to-mastery>

# PRACTICE SQL

SQL exercises

Login:

Password:

Enter

forgot password?

Register

Enter without login

» Ratings

» Statistics

» SQL interview

» Forums

» Certification

» Help

» Query optimization

» Personal page

» Developers & Thanks

» For employers

» Feedback

» References

» Support SQL-EX.RU

Language: English

December 01, 08:26 MSK

Today we have 171 visitors (28 new).  
9 exercises solved (2 on SELECT and 7 on DML),  
on learning stage - 353

Practical skills of SQL language

This site will help everyone to gain or improve skills in building SQL **Data Manipulation Language** statements. To train you will have to build yourself the SQL statements for retrieval or modification of specific data required in the exercises. When your query is incorrect, you will be able to see rows returned by the correct query along with that returned by your query. Furthermore, you may execute arbitrary **DML** statements on available databases by setting the "without checking" option. There are five levels of difficulty (from 1 to 5). You may see it in second column of exercises list. We propose the exercises on retrieving data (**SELECT** statement) and the exercises on modifying data (**INSERT**, **UPDATE**, **DELETE**, and **MERGE** statements). Your success in the solving the exercises are shown by a rating of participants. As this takes place, there are three stages: the first one (first 5 exercises) is performed without time control for an individual exercise, the second one (begin with the exercise 46) controls time for completion of each task. At the third stage which refers to **optimizing** and begins with exercise #139, it is required not only to solve an exercise correctly, but also time of execution of inquiry should be commensurate with time of execution of the author's solution.

Exercises of the first stage are available without registration and may be solved in any order you like. The solution of the rest of exercises requires registration. **REGISTRATION IS FREE** as this for all other services of the site. In the third column of exercises list you will be able to see "OK" notes with the numbers of done exercises, but that is available only to the registered users. In fact, that is the main reason for registration. If you would like to visit our web site again, you won't have to reselect which exercises you have done already and which you haven't. If you don't want to register, you may enter as a guest, but in that case your results won't be traced by the system. Registered users also may discuss the solutions to exercises in our forum.

NOTE: The query stated incorrectly may return the "correct" data on a current state of database. For this reason you should not be surprised if the results of incorrect query are coincide with the results of right one with your query is estimated as incorrect by the WebFing system.

NOTE: Your browser should support Cookies and Javascript to provide correct usage of this site. If you use content filter, it should allow opening child windows to explore help pages.

Certification

Successful testing of yours can be confirmed by the certificate "SQL Data Manipulation Language Specialist". We support quality of our certificate by periodically replacing exercises and changing the certificate requirements.


How about certification. What does the certificate give to you?

SQL syntax used

User's queries are executed by SQL server that brings some limitations to the syntax of SQL statements. Now we use **Microsoft SQL Server 2010 (11.0)** on the rating stages, and **MariaDB 10.4 (compatible with MySQL 4.1)**, **PostgreSQL 12.2**, and **Oracle Database 11g** on the learn stage in addition. That is why you should follow the syntax of these realizations when building queries. It should be noted that SQL syntax implemented in Microsoft SQL Server is close to that of **SQL-92** standard. But there are some distinctions, among them in absence of NATURAL JOIN of tables. Supplied help on SQL Data Manipulation Language is held in accordance with SQL-92 standard and contains information necessary for solving the exercises. In the same place it is possible to find features of used realization (SQL Server 2005).

Top 10 results

Person	Scores	Days	Days_2	Days_3	Scores_3
Chernov D.V. (Kunskit)	587	1850	1600.401	1710.840	254
Kostomarov A.V. (a129)	649	3279	4336.841	2473.867	251
Doshchenko V.N. (mcsn)	842	3108	2042.968	524.573	251
Krasovskij E.A. (papiropik)	869	3441	216.518	14.367	248
Kraslavskij O.H. (Arcan)	678	5376	284.501	225.486	247
Boiko D. (Angellora)	654	5138	5433.872	3825.743	237
Bezhaev A.Yu. (Baser)	638	4101	201.482	178.821	234
Dubinskij A.V. (L_vella)	652	4129	2797.568	3371.376	227
Karayova N.V. (vikam)	655	5635	292.052	242.855	225
Kurochkin P.A. (qerqer)	638	4139	257.852	164.392	211

 **SQLBolt**  
Learn SQL with simple, interactive exercises.

Interactive Tutorial

More Topics

Introduction to SQL

Welcome to SQLBolt, a series of interactive lessons and exercises designed to help you quickly learn SQL right in your browser.

What is SQL?

SQL, or Structured Query Language, is a language designed to allow both technical and non-technical users query, manipulate, and transform data from a relational database. And due to its simplicity, SQL databases provide safe and scalable storage for millions of websites and mobile applications.

Did you know?

There are many popular SQL databases including SQLite, MySQL, Postgres, Oracle and Microsoft SQL Server. All of them support the common SQL language standard, which is what this site will be teaching, but each implementation can differ in the additional features and storage types it supports.


Relational databases


Before learning the SQL syntax, it's important to have a model for what a relational database actually is. A relational database represents a collection of related (two-dimensional) tables. Each of the tables are similar to an Excel spreadsheet, with a fixed number of named columns (the attributes or properties of the table) and any number of rows of data.

For example, if the Department of Motor Vehicles had a database, you might find a table containing all the known vehicles that people in the state are driving. This table might need to store the model name, type, number of wheels, and number of doors of each vehicle for example.

<https://sql-ex.ru/>

<https://sqlbolt.com/>

 \$0.00

 2 weeks

softserve



# **MASTER DATA PROCESSING TECHNIQUES**

softserve

Development > Database Design & Development > Apache Kafka

## Apache Kafka Series - Learn Apache Kafka for Beginners v2

START HERE: Learn Apache Kafka 2.0 Ecosystem, Core Concepts, Real World Java Producers/Consumers & Big Data Architecture

**Created by** **Stephane Maarek** | AWS Certified Cloud Practitioner, Solutions Architect, Developer, Conductor Inc

4.7 ★★★★★ (27,597 ratings) 126,932 students

Created by Stephane Maarek | AWS Certified Cloud Practitioner, Solutions Architect, Developer, Conductor Inc

● Last updated 11/2021 ● English 🗣 English [Auto], Z more

**What you'll learn**

- ✓ Understand Apache Kafka Ecosystem, Architecture, Core Concepts and Operations
- ✓ Start a personal Kafka development environment
- ✓ Create your Producers and Consumers in Java to interact with Kafka
- ✓ Extended APIs Overview (Kafka Connect, Kafka Streams), Case Studies and Big Data Architecture
- ✓ Master Concepts such as Topics, Partitions, Brokers, Producers, Consumers
- ✓ Learn major CLIs: kafka-topics, kafka-console-producer, kafka-console-consumer, kafka-consumer-groups, kafka-configs
- ✓ Program a Real World Twitter Producer & ElasticSearch Consumer
- ✓ Practice and Understand Log Compaction

**\$12.99** ~~\$199.99~~ 88% off

🕒 15 hours left at this price!

**Add to cart** **Buy now**

30-Day Money-Back Guarantee

**This course includes:**

- 📺 7.5 hours on-demand video
- 📄 18 articles
- 🔑 Full lifetime access
- 📱 Access on mobile and TV
- 📜 Certificate of completion

**Apply Coupon** **Gift this course**

Data Science > Apache Beam

## Batch Processing with Apache Beam in Python

Easy to follow, hands-on introduction to batch data processing in Python

**Highest Rated** 4.7 ★★★★★ (23 ratings) 108 students

Created by Alexandra Abbas

● Last updated 9/2020 ● English 🗣 English [Auto]

**What you'll learn**

- ✓ Core concepts of the Apache Beam framework
- ✓ How to install Apache Beam locally
- ✓ How to read and write CSV data from Apache Beam
- ✓ How to deploy your pipeline to Cloud Dataflow on Google Cloud
- ✓ How to design a pipeline in Apache Beam
- ✓ How to build a real-world ETL pipeline in Apache Beam
- ✓ How to apply built-in and custom transformations on a dataset

**\$9.99** ~~\$69.99~~ 83% off

🕒 15 hours left at this price!

**Add to cart** **Buy now**

30-Day Money-Back Guarantee

**This course includes:**

- 📺 1 hour on-demand video
- 📄 1 downloadable resource
- 🔑 Full lifetime access
- 📱 Access on mobile and TV
- 📜 Certificate of completion

**Apply Coupon** **Gift this course**

<https://www.udemy.com/course/apache-kafka/>

<https://www.udemy.com/course/apache-beam-python/>

YouTube

Введите запрос

Главная  
Навигатор  
Подписки

Библиотека  
История  
Входящие видео  
Смотреть позже  
Понравившиеся

ПОДПИСКИ

**Distributed Systems lecture series**

23 видео • 211 102 просмотра • Обновлено 13 окт. 2021 г.

Accompanying lecture notes:  
<https://www.cl.cam.ac.uk/teaching/212...>  
These videos form an 8-lecture series on distributed systems, given as part of the undergraduate computer science course at the University of Cambridge. It is preceded by an 8-lecture course on concurrent systems for which videos are not publicly available, but slides can be found on the course web page:  
<https://www.cl.cam.ac.uk/teaching/212...>

Martin Kleppmann **ПОДПИСАТЬСЯ**

- 1 Distributed Systems 1.1: Introduction Martin Kleppmann 14:36
- 2 Distributed Systems 1.2: Computer networking Martin Kleppmann 13:57
- 3 Distributed Systems 1.3: RPC (Remote Procedure Call) Martin Kleppmann 15:43
- 4 Distributed Systems 2.1: The two generals problem Martin Kleppmann 11:33
- 5 Distributed Systems 2.2: The Byzantine generals problem Martin Kleppmann 10:43
- 6 Distributed Systems 2.3: System models Martin Kleppmann 10:43
- 7 Distributed Systems 2.4: Fault tolerance Martin Kleppmann 8:19

<https://www.datacamp.com/courses/big-data-fundamentals-with-pyspark>

[Video course on YouTube](#)



\$31.41



1 month

datacamp **WE'RE HIRING** Products ▾ For Business ▾ Pricing

## Big Data Fundamentals with PySpark

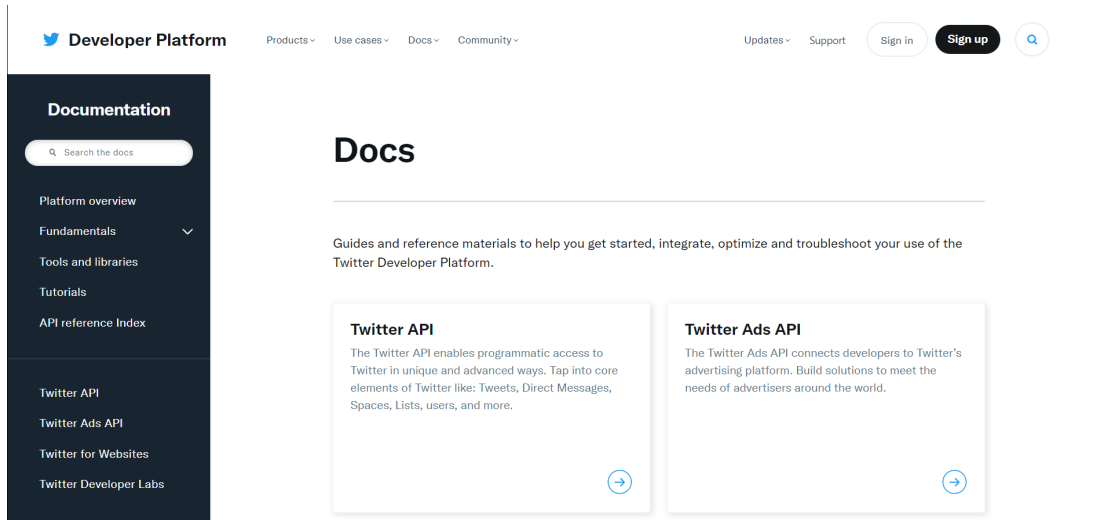
Learn the fundamentals of working with big data with PySpark.

**Start Course For Free**

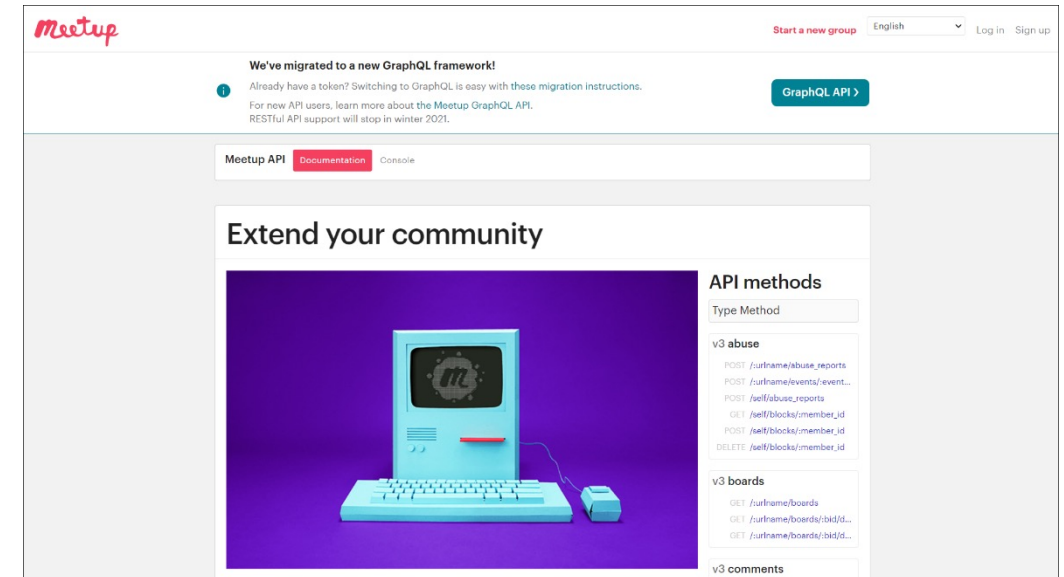
🕒 4 Hours 📺 16 Videos 📄 55 Exercises 👤 26,415 Learners

📊 4800 XP

# PRACTICE WITH AVAILABLE STREAMING DATA SOURCES



<https://developer.twitter.com/en/docs>



[https://www.meetup.com/meetup\\_api/docs/](https://www.meetup.com/meetup_api/docs/)

 \$0.00

 2 weeks

softserve





# **SCHEDULE YOUR WORKFLOWS**

softserve

Development > Software Engineering > Apache Airflow

## The Complete Hands-On Introduction to Apache Airflow

Learn to author, schedule and monitor data pipelines through practical examples using Apache Airflow


**Bestseller** 4.6 ★★★★★ (4,393 ratings) 22,454 students

Created by **Marc Lambert**

Last updated 10/2021 English English [Auto], French [Auto], 1 more

### What you'll learn

- ✓ Create plugins to add functionalities to Apache Airflow.
- ✓ Master core functionalities such as DAGs, Operators, Tasks, Workflows, etc
- ✓ The difference between Sequential, Local and Celery Executors, how do they work and how can you use them.
- ✓ Install and configure Apache Airflow
- ✓ Using Docker with Airflow and different executors
- ✓ Understand and apply advanced concepts of Apache Airflow such as XCOMs, Branching and SubDAGs.
- ✓ Use Apache Airflow in a Big Data ecosystem with Hive, PostgreSQL, Elasticsearch etc.
- ✓ Think, answer and implement solutions using Airflow to real data processing problems



**\$9.99** ~~\$49.99~~ 80% off

14 hours left at this price!

**Add to cart**

**Buy now**

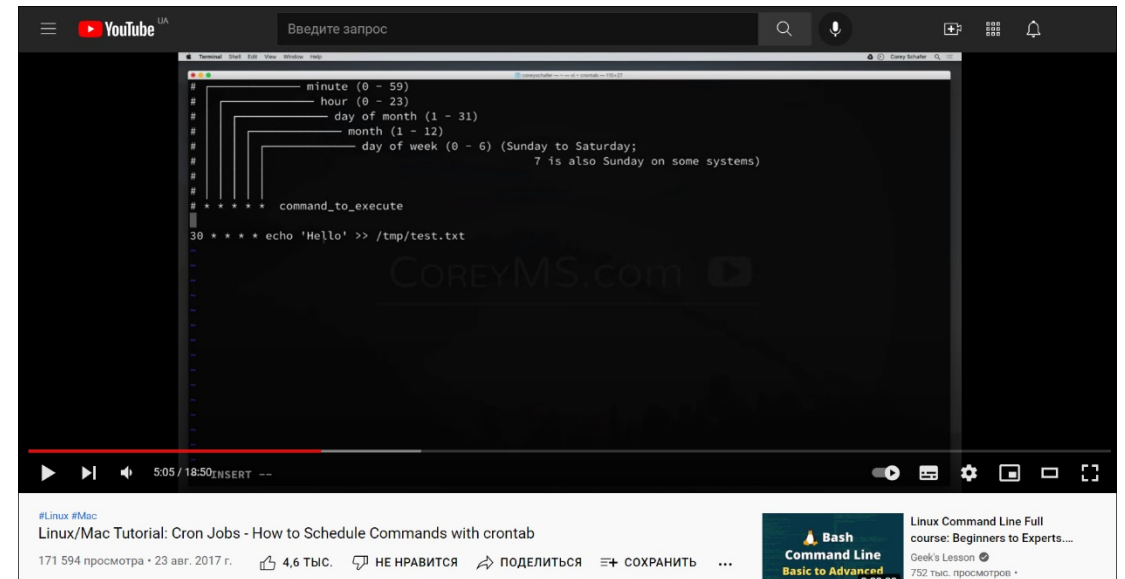
30-Day Money-Back Guarantee

**This course includes:**

- 6 hours on-demand video
- 16 articles
- 5 downloadable resources
- Full lifetime access
- Access on mobile and TV
- Certificate of completion

[Apply Coupon](#) [Gift this course](#)

<https://www.udemy.com/course/the-complete-hands-on-course-to-master-apache-airflow/>



[Linux/Mac Tutorial: Cron Jobs - How to Schedule Commands with crontab](#)

 \$0.00

 2 weeks

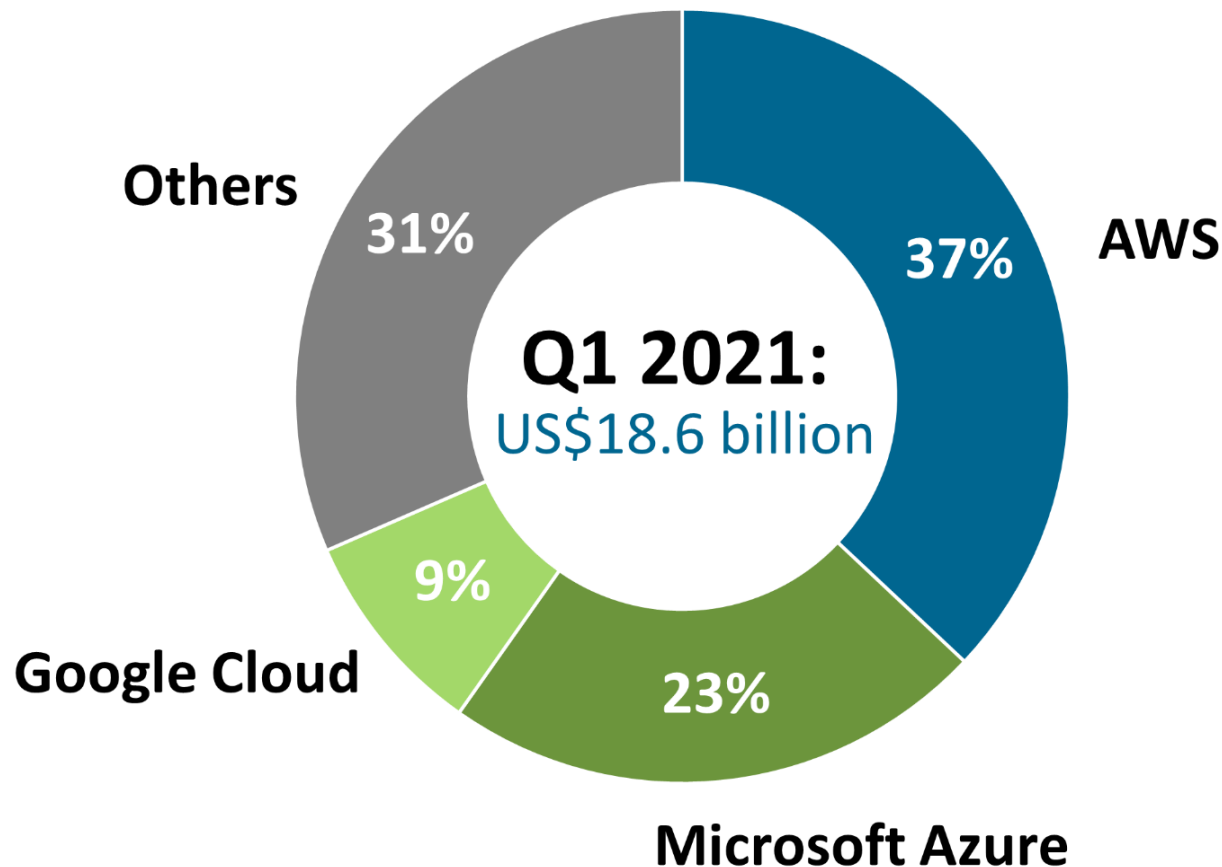
**softserve**



**UNDERGO  
PROFESSIONAL  
CERTIFICATION**

softserve

# US CLOUD INFRA SERVICES SPEND



Q1 2021 cloud infrastructure spend in the United States grew **29%** to US\$18.6 billion, driven by a focus on digital transformation

Source: Canalys estimates, May 2021

**softserve**

# GET PROFESSIONAL CERTIFICATION



<https://aws.amazon.com/certification/certified-data-analytics-specialty/>

**\$300**



<https://docs.microsoft.com/en-us/learn/certifications/exams/ms-203>

**\$165**

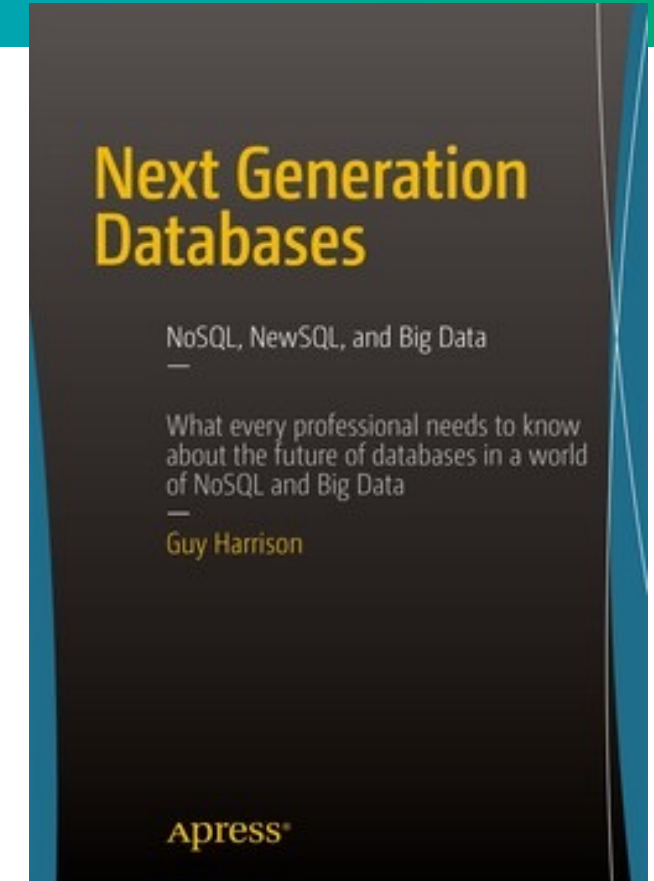
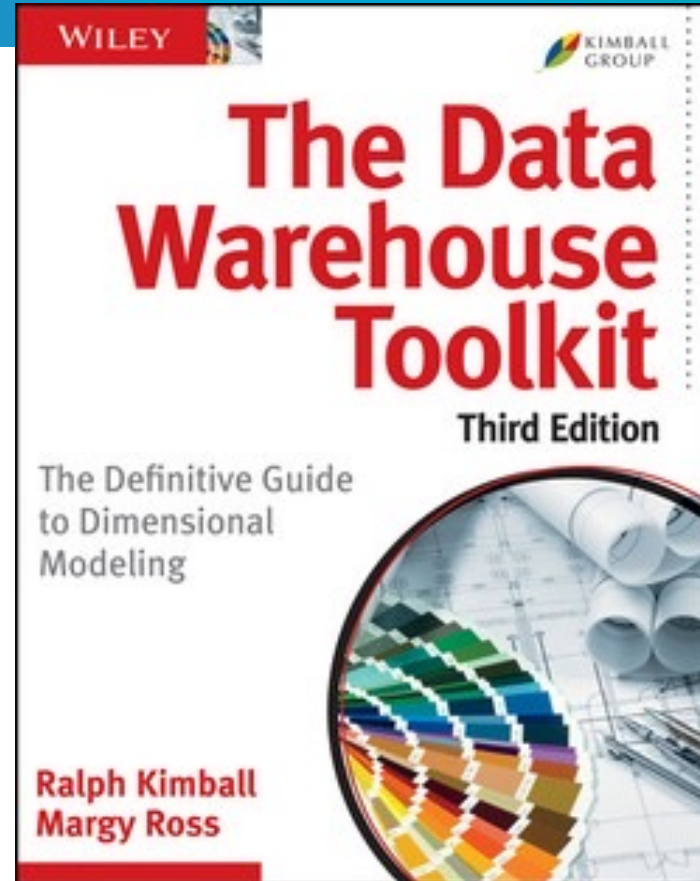


<https://cloud.google.com/certification/data-engineer>

**\$200**

**softserve**

# ADDITIONAL RECOMMENDED BOOKS



softserve

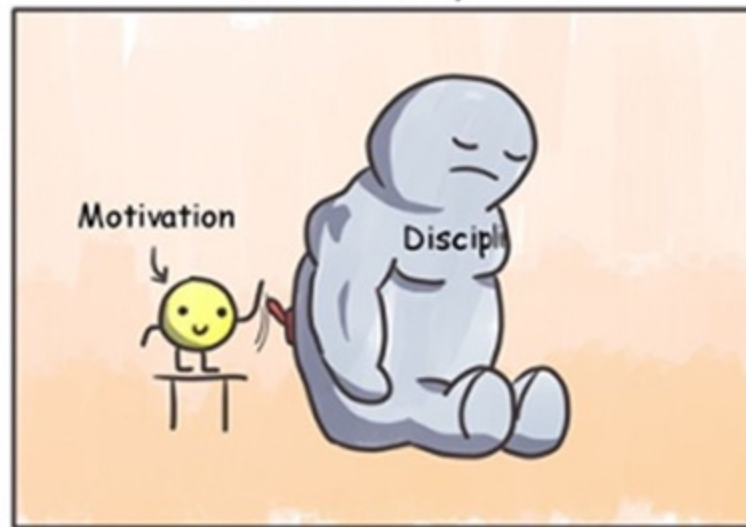


# MOTIVATION AND DISCIPLINE

How people think motivation works



How it actually works



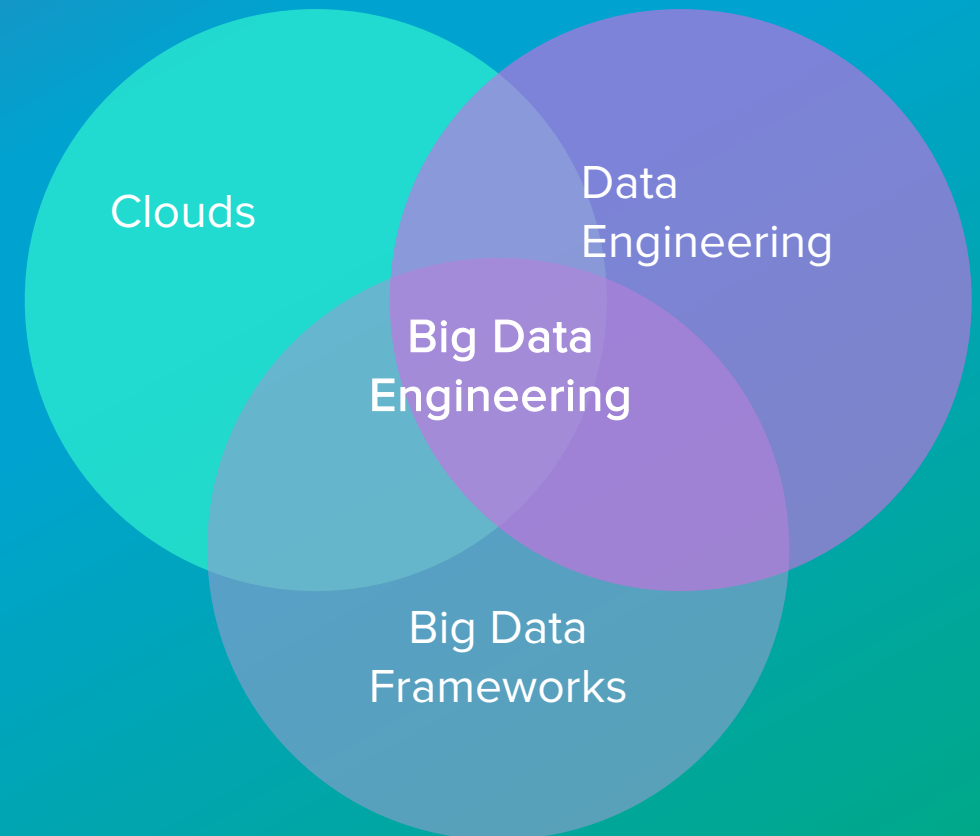
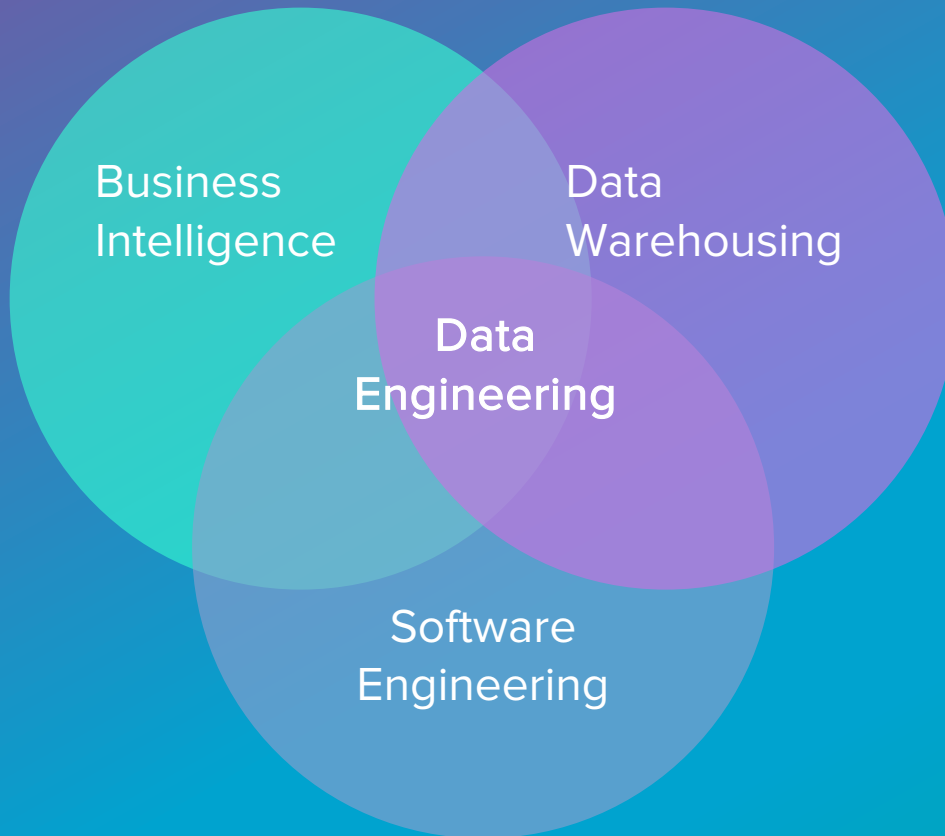
Total: \$257.37-\$392.37



3-6 month

**softserve**

# DATA VS BIG DATA

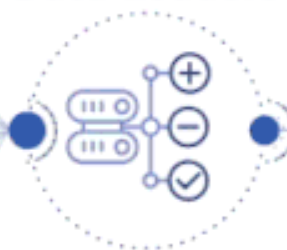


softserve

## Datasources



## Data Loaders



## Data Lake



Raw Data



Trusted Data



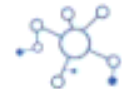
Service Data



Catalog



Discovery



Lineage

## Data Sharing



## Data Visualization



## M.L Development



## Add-ons



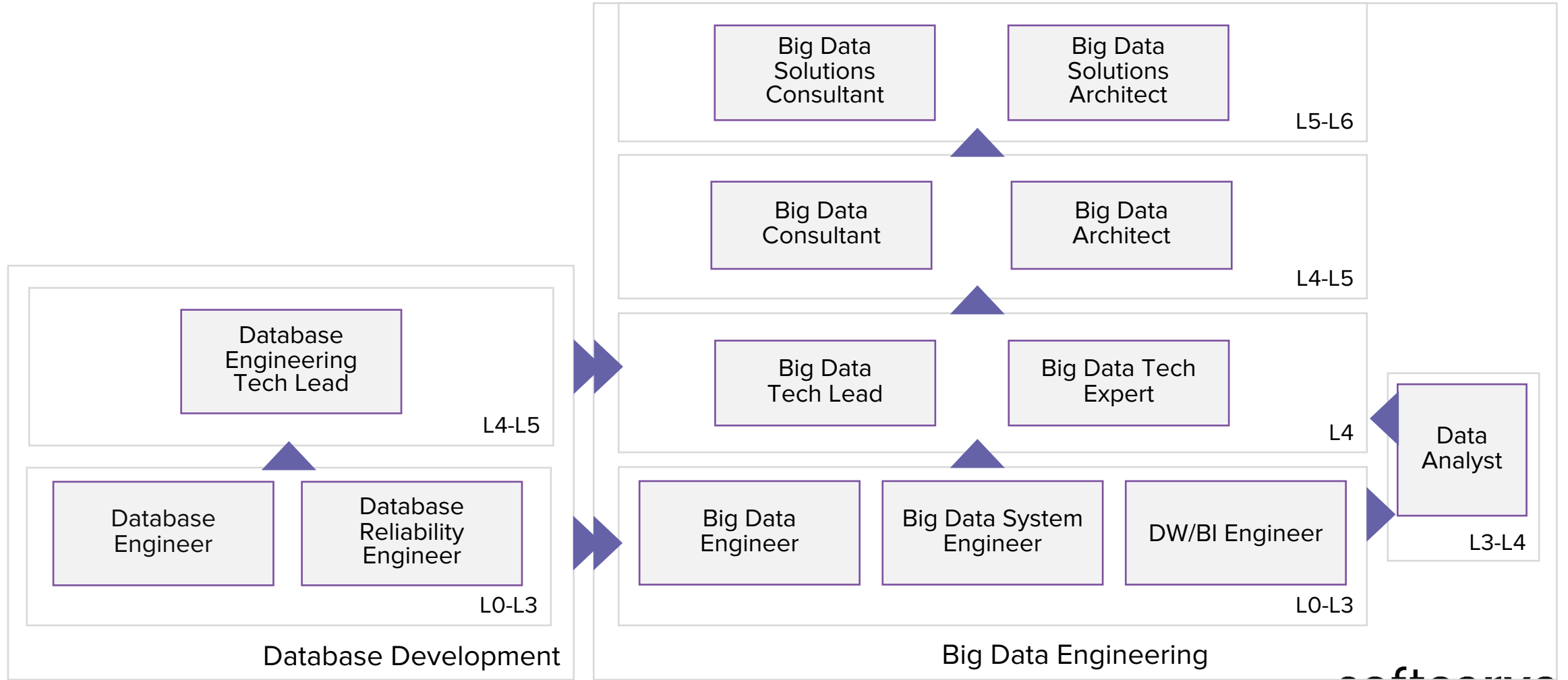
Insights Stores



A.I.Store

softserve

# DATA ENGINEERING PROFILES



# DATA PLATFORM AND ROLES



**softserve**

# DATA TOOLS

## Data Governance

Informatica, SailPoint, collibra, dremio, OKERA

## Data Analysis & Visualization

Looker, Amazon QuickSight, Sisense, QlikQ, MicroStrategy, Tableau, PowerBI, GCP Data Studio

## Data Exploration

Jupyterhub, binder, colab, GCP Datalab, Azure Notebooks, Amazon SageMaker

## Data Ingestion/Transformation

Informatica, MuleSoft, Fivetran, talend, Pentaho, Trifacta, Amazon Glue, Amazon Athena, Azure Data Factory, StreamSets, dbt

## Data Processing

AWS Kinesis, GCP DataFlow, Databricks, Confluent, Apache Kafka, Azure Functions, AWS Lambda, GCP Cloud Functions

## Orchestration Systems

Apache Airflow, talend, oozie

## NoSQL

Bigtable, DynamoDB, CosmosDB, MongoDB, Cassandra, redis

## Data Storages

AWS S3, Azure Storage, GCP Cloud Storage, AWS Redshift, GCP Big Query, Azure Synapse

## Relational Databases

Microsoft SQL Server, Oracle, PostgreSQL, MySQL

## Data Infrastructure & Big Data Frameworks

Hadoop, Cloudera, AWS EMR, GCP Dataproc, Azure HDInsight, Pivotal, SAP Cloud Platform, IBM InfoSphere, Spark, beam, Flink

## Database Administration

Microsoft SQL Server, Oracle, PostgreSQL, MySQL

Data Science

Business Analysts

DevOps

Project Managers

Product Managers

QAs

...

# RETRAINING PROGRAMS FOR EVERYONE\*

For BI Engineers:

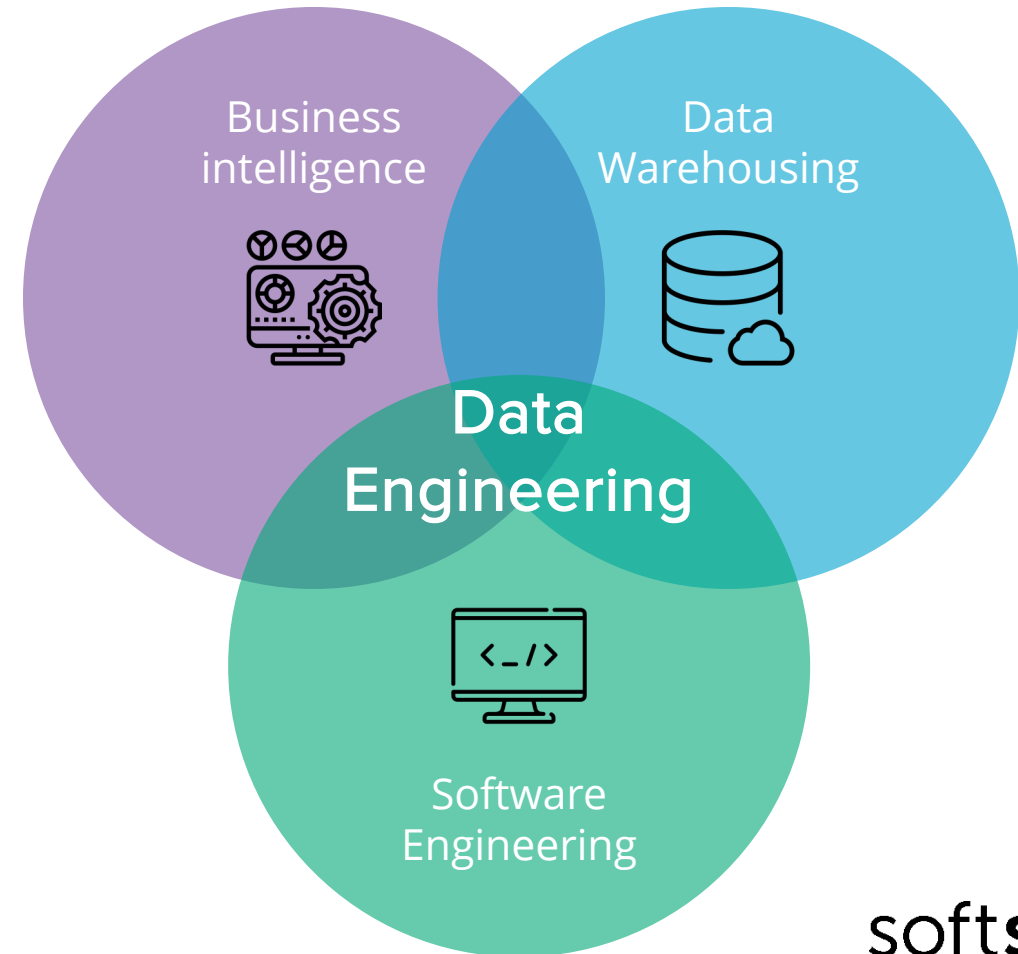
- Python Big Data retraining program
- Scala Big Data retraining program

For Software Engineers:

- AWS for Big Data retraining program
- Azure for Big Data retraining program
- GCP for Big Data retraining program

For Data Warehouse Engineers:

- BI retraining program



**softserve**



# REQUERMENTS FOR CANDIDATES

## Requirements:

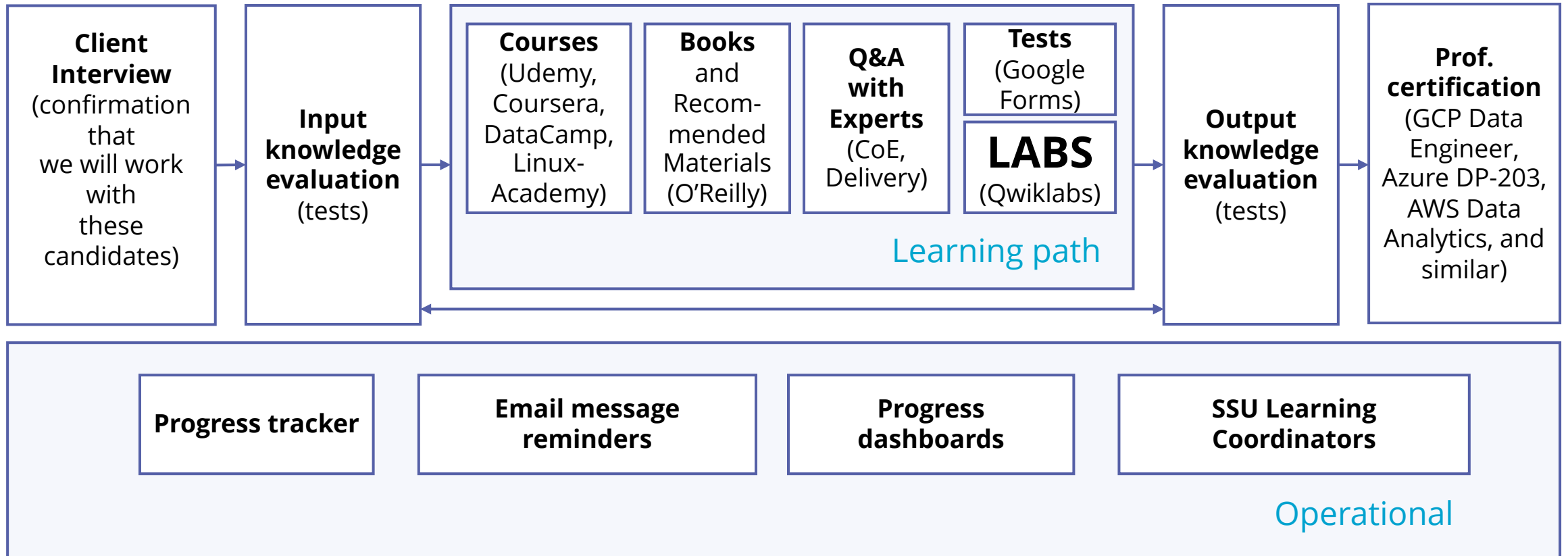
- Python developer with **two or more years of experience** (back-end development)
- Skilled in **SQL**
- Showing understanding of **data engineering key principles** and concepts

## Nice to have:

- Showing exposure to **cloud services**
- Demonstrating **Upper-Intermediate English** level

softserve

# RETRAINING PROGRAM STAGES



\* Duration of the retraining program is 8 weeks

**softserve**

# GAPS ANALYSIS

Greetings!

You made it! I believe you have all chances to pass AWS Data Analytics Certification (if you have not made it yet). It will be a pleasure to work with you!

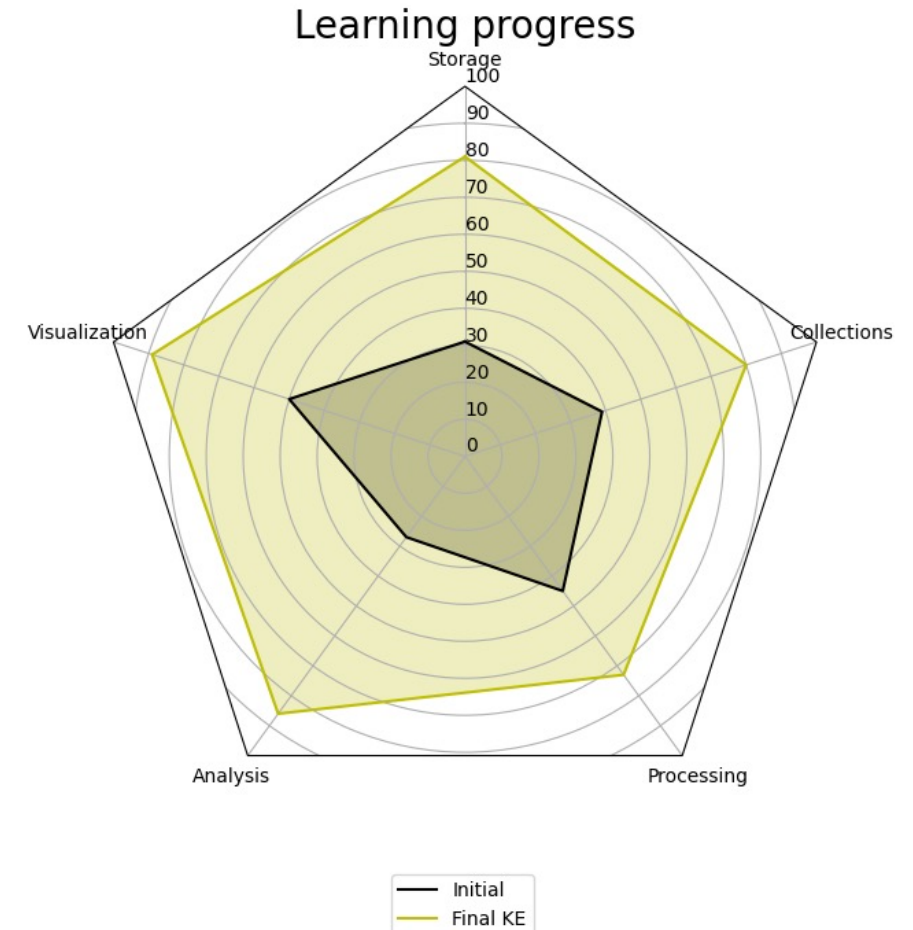
Your total score is 85 out of 92. This test contains questions related to different knowledge areas. Here are your scores for each knowledge area:

- Storage – 32 out of 33.
- Visualisation – 3 out of 3.
- Analysis – 14 out of 16.
- Processing – 17 out of 20.
- Collections – 18 out of 20.

Thank you for the enthusiasm!

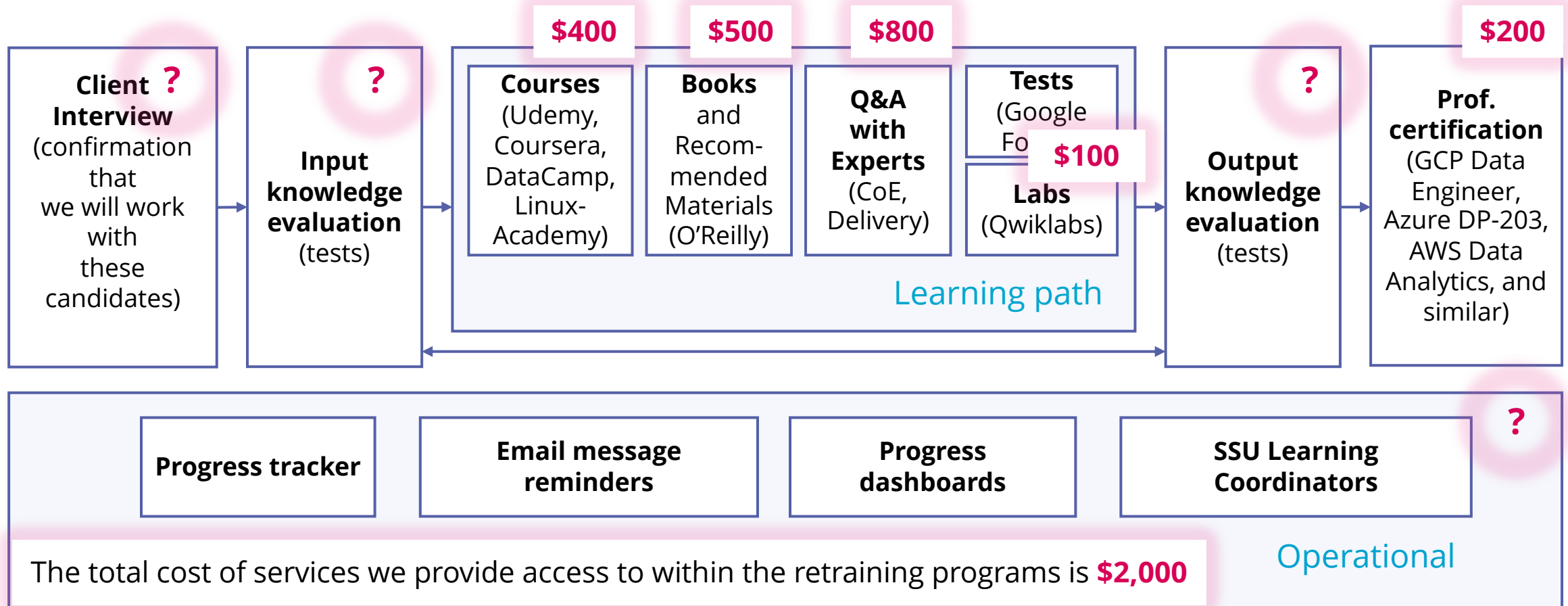
Best regards,

Big Data & Analytics Competence Development Team



**softserve**

# COST TO PARTICIPATE IN SUCH PROGRAM WITHOUT SOFTSERVE



# KEEP GOING

- <https://www.dataengineeringpodcast.com/> - Data Engineering Podcast
- <https://www.reddit.com/r/dataengineering/> - Data Engineering Subreddit
- <https://maximebeauchemin.medium.com/> - M. Beauchemin's blog
- <https://dataawesome.substack.com/> - Monthly Newsletter of awesome data curated content
- <https://news.ycombinator.com/> - Hacker News



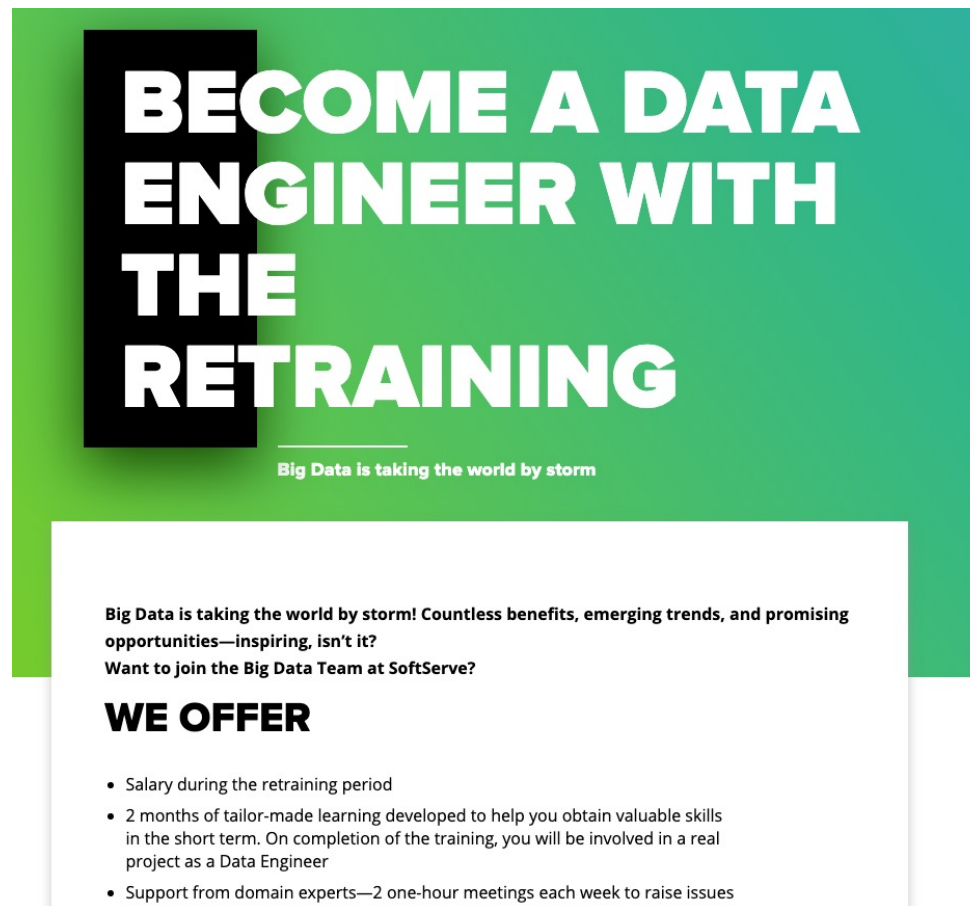
<https://hubs.ly/H0PVvGj0>

**SoftServe's Big Data Digest**

**softserve**

# MORE DETAILS AND REGISTRATION

<https://career.softserveinc.com/en-us/technology/course/big-data>



**BECOME A DATA ENGINEER WITH THE RETRAINING**

Big Data is taking the world by storm

Big Data is taking the world by storm! Countless benefits, emerging trends, and promising opportunities—inspiring, isn't it?  
Want to join the Big Data Team at SoftServe?

**WE OFFER**

- Salary during the retraining period
- 2 months of tailor-made learning developed to help you obtain valuable skills in the short term. On completion of the training, you will be involved in a real project as a Data Engineer
- Support from domain experts—2 one-hour meetings each week to raise issues



**softserve**



**TIME FOR  
YOUR  
QUESTIONS!**

softserve

