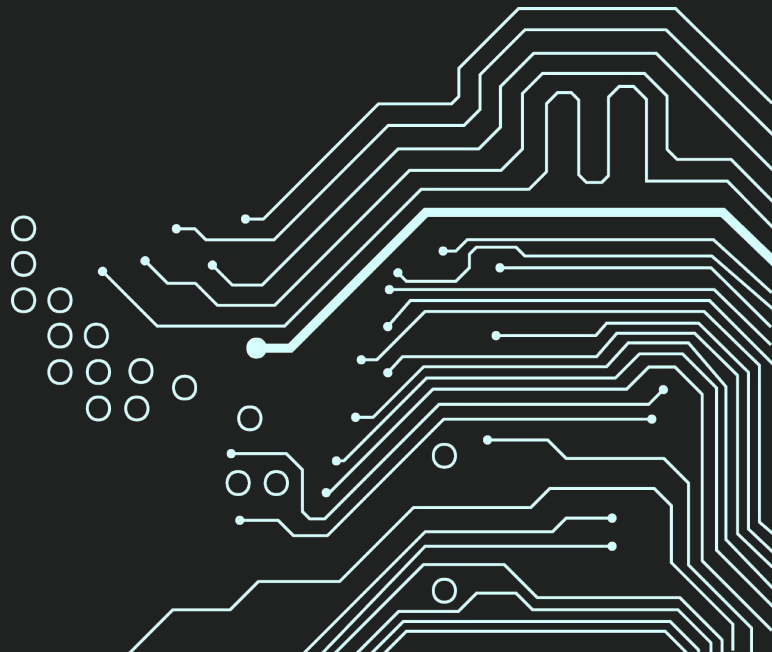
CSC 460: Team Cyber
Module 1 Presentation

Minsu Kim, Aila Choudhary, Joshua Jimenez,
Misael Perez, Kenny Zhu
Pin: 1289

Table of contents

01

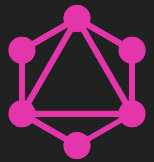
Module Overview



Pod 1: Data Science Programming Language + Data Extracting

Data Science Programming Language

Data Extracting



Overview

Data Science Programming Language

A programming language designed or widely used for data analysis, statistical modeling, machine learning, and data visualization.

- Enables efficient handling of large datasets.
- Provides libraries and frameworks for data manipulation and analysis.

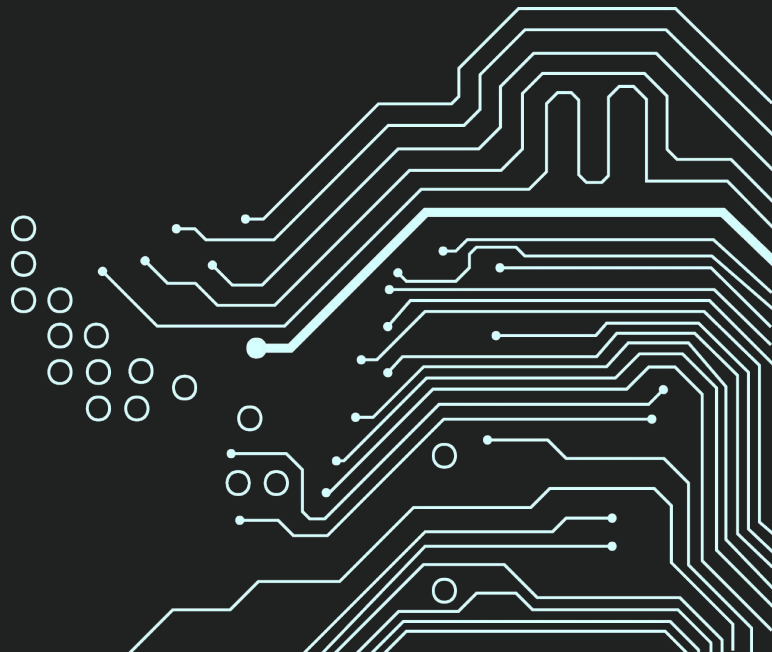
Data Extracting

The process of retrieving and collecting data from various sources, such as databases, APIs, web scraping, or files.

- Ensures access to relevant and high-quality data for analysis.
- Impacts the accuracy and reliability of insights derived from data.

02

History & Evolution



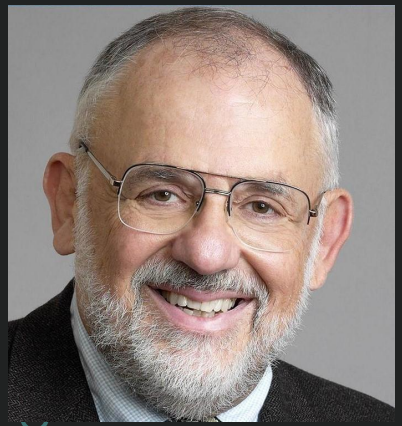
Python

Created in 1991 by Guido van Rossum.

Gained popularity due to its ease of use, extensive libraries, and versatility.

By the 2010s, Python became the dominant language in data science, thanks to libraries like NumPy (2006), Pandas (2008), and Scikit-learn (2007), enabling efficient data analysis and machine learning.

Matlab



Created in 1984 by Cleve Moler.

Gained popularity due to to its matrix-based computation, built-in functions, and ease of use for mathematical modeling.

Over time, MATLAB evolved with toolboxes for statistics, machine learning, and deep learning, making it relevant for data science.

R

Created in 1992 by Ross Ihaka and Robert Gentlemen at the University of Auckland in New Zealand

Gained popularity due to its strong statistical and data analytic capabilities, variety of high quality and customizable data visualization options, compatibility with other languages, and most of all its lack of cost

Today R is incorporated into the curriculum of many universities, and with the help of RStudio which provides a more user friendly interface, debugging tools and project management features, is now one of the most dominant languages in all fields involving statistics, research and data analysis

RESTful API

Created in 1984 by Roy T. Fielding. REST (Representational State Transfer)

In 2000s, REST gained popularity as major companies (Amazon, Facebook, Twitter and google) adopted it.

In 2010s, As Better security, documentation and efficiency needed, key standards were developed

WebSocket

Michael Carter led the creation of the first version of the protocol known as WebSocket in 2008.

Google Chrome was the first to implement full WebSocket support in 2009.

The WebSocket protocol was standardized in 2011.

A cornerstone of real-time web communication, widely used in connected devices, collaborative tools, and streaming services.

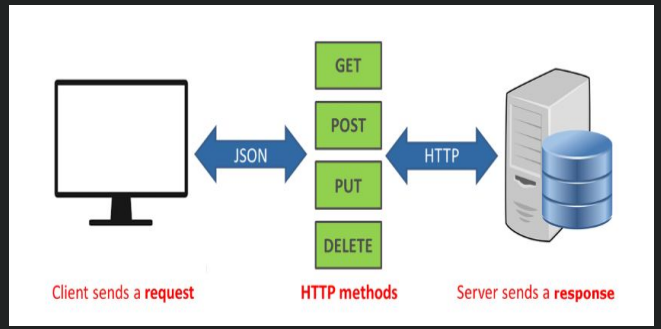
Evolution of GraphQL

Before GraphQL Transfering Data was:

- CSV (comma-separated values)
- HTML
- XML and JSON (structuring data)

Facebook wanted a better data transfer

APIs used Representational State Transfers (REST) - Use HTTP -> Access Domain URI

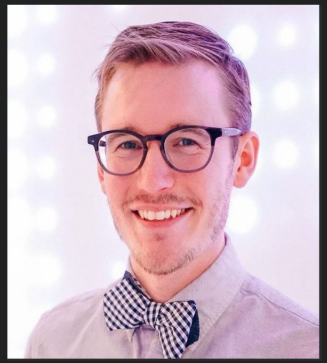


Evoluted To??

FQL (Facebook Query Language)

Cons: Network Latency, Bad Recursion

Evolution of GraphQL



Lee Byron

Dan Schafer

Nick Schrock

Created GraphQL

September 14, 2015
Facebook

03

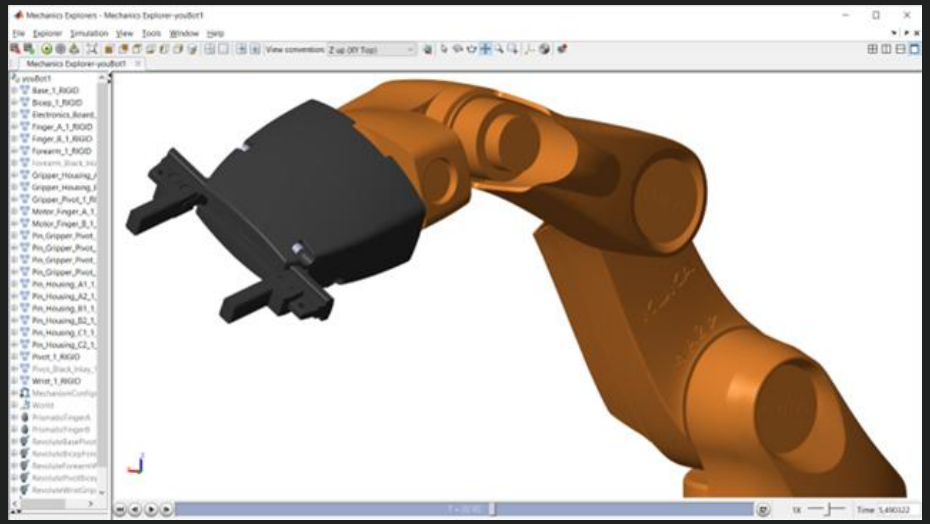
Popularity & Community Support

Python

Python has a vast and active community, and is widely adopted in academia, research, and industry, with strong support for statistics, data visualization, and AI.

Python's powerful libraries, including Pandas for data manipulation, NumPy for numerical computing, and TensorFlow for deep learning, make it a versatile tool for data science and AI.

Matlab



MATLAB has a smaller but strong community, primarily in engineering, physics, and applied sciences.It is widely used in academia and research for numerical computing, simulations, and signal processing. It offers toolboxes for statistics, machine learning, and AI.

R

Popularity

- Compatible with several languages such as Python, C/C++, SQL, Java etc
- Widely used in:
  - Universities & Research Institutions
  - Data Science & Analytics
  - Finance & Economics
  - Healthcare & Bioinformatics

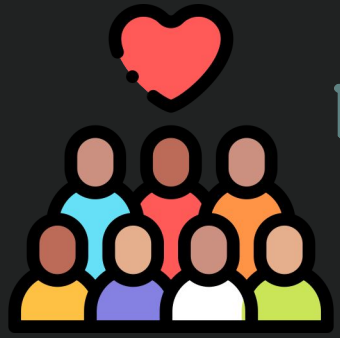
Community

- R has a huge community behind it including :
  - R - Bloggers
  - Stack Overflow
  - RStudio Community
  - DataCamp
  - Google, Facebook, Microsoft, etc

RESTful API

- Widely Adopted Across Industries
  - (Tech giants, cloud computing, mobile, IoT, AI)
- Strong Developer Community
  - (Open-source libraries, Stack Overflow, GitHub, tutorials)
- Long-Term Stability & Industry Trust
  - (Backward compatibility, security standards, API gateways)

WebSocket

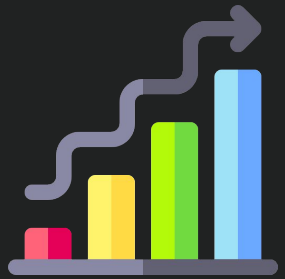


Popularity:

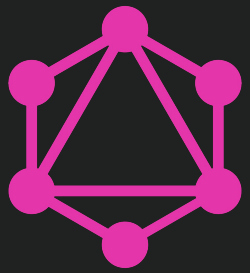
- Supported by popular libraries and frameworks.
- Widely used
  - in real-time applications.
  - for real-time data streaming.
  - in media for live video/audio streaming.

Community:

- Regular updates and improvements.
- Extensive documentation, tutorials, and resources available.
- Supported across all major browsers and platforms.



GraphQl



Popularity

GraphQl has been growing popular as companies have been switching to GraphQL

Community Support

The GraphQL Foundation is a community focus on collaboration and sharing new information about GraphQL.

04

Tabular Comparison

Criteria	Python	Matlab	R	RESTful API	WebSocket	GraphQL
Learning Curve	Easy	Moderate	Moderate	Moderate	Moderate	Moderate
Statistical Focus	Moderate	Low	High	Low	Low	Low
Machine Learning	Excellent	Good	Moderate	Low	Low	Low
Data Visualization	Good	Moderate	Excellent	Low	Low	Moderate
Real-Time Communication	Low	Low	Low	Excellent	Excellent	Good
API Flexibility	Moderate	Low	Low	Good	Good	Excellent
Scalability	High	Moderate	Moderate	High	High	High
Licensing	Free	Paid	Free	Free	Free	Free

05

Pros & Cons

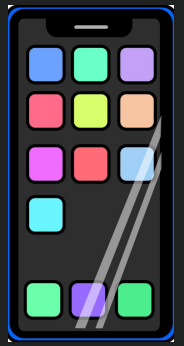
Python

Pros

- Easy to Learn
- Extensive Libraries
- Open-Source & Free
- Cross-Platform Compatibility
- Scalability & Versatility

Cons

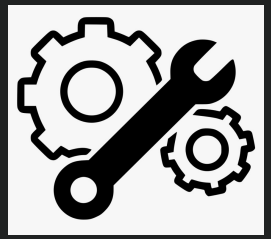
- Slower Execution Speed
- High Memory Consumption
- Weak in Mobile Development



Matlab

Pros

- Industry & Academic Use
- Built-in Toolboxes
- Excellent Visualization & Plotting



Cons

- Expensive Licensing
- Weaker Community Support
- Less Scalable for Large Data Science Applications

R

Pros

- Very effective in Statistical Analysis & Data Science
- Advanced Data Visualization
- Over 19,000 packages to adapt to almost any field
- Completely free to use
- Easily integrated with other languages

Cons

- Slower than other languages
- Uses a large amount of memory
- Difficult to learn for beginners
- Inconsistent package quality

RESTful API

Pros

- Simplicity & Easy of Use
- Scalability & Performance
- Widespread Adoption & Compatibility
- Flexibility & Data Format Support
- Separation of Client & Server

Cons

- Over-fetching & Under-fetching of Data
- Lack of Real-Time Communication
- Statelessness Can Lead to Redundant Data Transfer
-
- Complexity in API Versioning & Management

WebSocket

Pros

- Real-Time Data Streaming.
- Bidirectional Communication.
- Scalable for Large Data Streams.

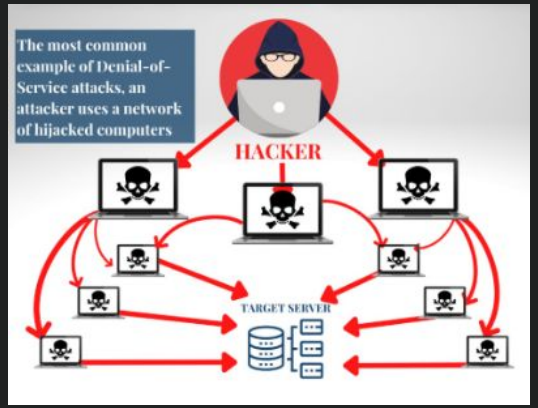
Cons

- Complex Implementation.
- Stateful Connections.
- Resource Intensive.

GraphQL

Pros

- SIngle Endpoints between frontend and backend
- No Over-fetching data
- Requesting only fields

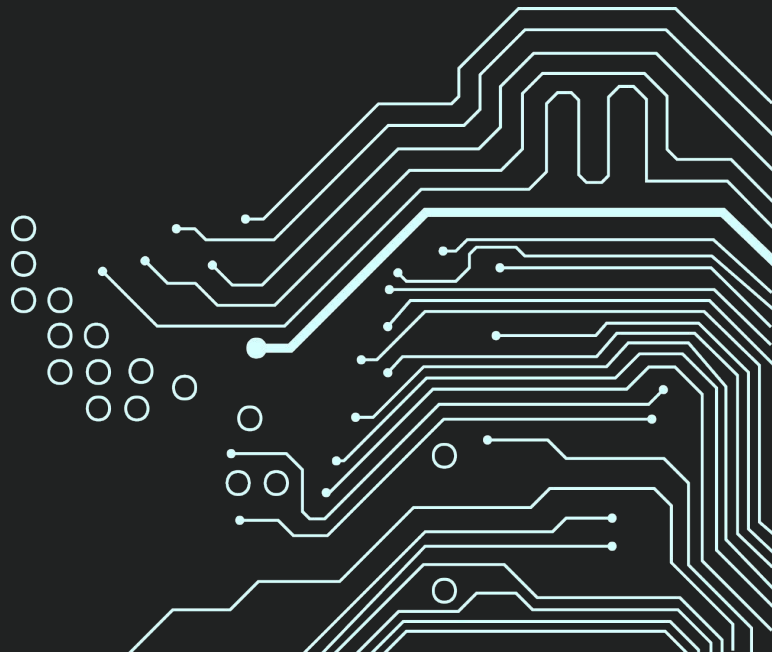


Cons

- Take time learning the language
- Can cause issues for complex queries
- Risk of Denial of Service attacks

06

Industry & Academic Relevance



Python

Industry

- Software development


- Finance


- Artificial Intelligence

Academia

- Computer Science
- Data Science
- AI Research

Why Companies or Colleges Choose Python

- Open-source & free
- Extensive libraries for data analysis, AI, and web development.
- Scalable & cross-platform

Matlab

Industry

- Aerospace & Defense

- Automotive

Academia

- Engineering
- Applied Mathematics
- Robotics
- Physics

Why Companies or Colleges Choose Python

- Powerful for mathematical modeling, simulations, and numerical analysis.

R

Industry

- Sports Analytics
- Marketing
- Healthcare
- Finance & Banking

Academia

- Statistics
- Data Science
- Economics
- Engineering

Why Choose R?

- Open Source
- Great for Statistical Analysis & Data Science
- Large collection of packages

RESTful API

Industry

- Widely Used in Tech & Business
- Cloud & SaaS Integration
- Mobile & Web Development
- IoT & AI Applications
- Industry Leaders Use it

Academia

- Taught in Computer Science & Engineering
- Research & Data Science Applications
- Standard for Educational APIs
- Foundation for Emerging Technologies

WebSocket

Industry

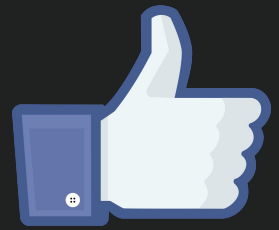
- Financial Services
- Social media
- Gaming
- Collaboration Tools
- Media Streaming

Academia

- Research on real-time systems.
- Prototyping real-time applications.
- Teaching web development and real-time communication technologies.

Why Choose WebSocket?

- Real-Time Capabilities.
- Efficiency.
- Scalability.
- Versatility.



GraphQL

Companies


Industries

- Social Media Platforms
- Entertainment Media
- Telecommunications
- E-commerce
- Retail

Academia

- Taught in Educational Institutions
- Online Courses

Why Choose GraphQL

- Efficient
- Better Performance
- Flexibility

07

Recommendations & Conclusion

Data Science Programming Languages

Python

Recommended For:

- Students and faculty working on machine learning, deep learning, or big data projects.
- Those who want a single language for end-to-end data science workflows.
- Beginners due to its readability and ease of learning.

MATLAB

Recommended For:

- Faculty and students in engineering, physics, or applied mathematics.
- Projects requiring heavy numerical computation or simulations.

R

Recommended For:

- Statistical Analysis
- Data Visualization
- Machine Learning

Not Recommended For:

- Real Time / High Performance Applications
- Web / App development
- Big Data Processing

Data Extraction

RESTful API

- **Industry Impact:**
  - REST APIs power cloud services, SaaS, mobile apps, and IoT.
- **Standardization & Security:**
  - Advances like OAuth 2.0, OpenAPI, and API gateways improved API management.
- **Modern API Evolution:**
  - While GraphQL, gRPC, and WebSockets offer alternatives, REST remains dominant due to its flexibility, compatibility, and ease of use.
- **Industry & Academic Relevance:**
  - REST is widely taught in universities and used by top companies, making it a critical skill for developers.

Data Extraction

WebSocket

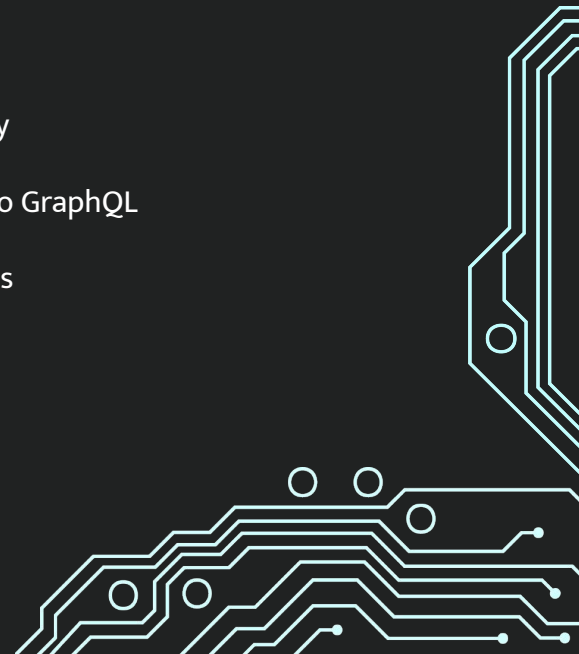
Recommended For:

- Real-time applications like live notifications, collaborative tools, or streaming data.
- Projects where low latency and continuous data flow are critical.

GraphQL

Recommended For:

- Growing Popularity
- Growing Community
- Companies Switch to GraphQL
- Faster than Rest APIs



08

Q&A