Chapter 01 - Creating APIs That Data Scientists Will Love

API for Data Scientists

- Data Scientist need different type of APIs
  - API in this Book is = software programs that you call using HTTP to retrieve data or execute commands
    - Through what we use it? Over Internet or Internal network
    - How we use it? API consumer send a request to an API, the API receives this request and sends back a response = API Communication

Data Scientists Activities

- Preparing or Cleansing Data (38%)
  - During Exploratory Data Analysis (EDA) Process on new data set, Analyzing:
    - Contents
    - Formats
    - Patterns
  - Scheduled Data Pipeline (Data Engineer Job):
    Sequence of software tasks that pull from multiple data sources and reformat or remove errors from the data so that it can be used for downstream tasks such as:
    - Visualizations
    - Reports
    - Model
- Creating Reports, Presentations, Data Visualization (29%)
  - Extract Insight from data $\xrightarrow{\text{result into}}$ enable organization to monitor its operations + make better decisions
  - Analytics (Data Analytics Job)
  - Data Scientist make APIs call to provide data for variety analytics products (Meta Base, Power BI, etc.)
- Selecting, Training, Deploying Models (27%)
  - Use Machine Learning, Mathematical Model $\xrightarrow{\text{so}}$ make prediction / cluster data into groups / perform natural language processing / etc.
  - 2 types of API call in this part:
    - API consumers
      - API as an input source for ML model

- API Producers (ML engineers)
    - Deploy their models as APIs for others to use
        - Internal consumer $\xrightarrow{\text{so}}$ Host API in their network
        - External consumer $\xrightarrow{\text{so}}$ Host API in over internet

Data Scientists Tools

Programming Languages:

- Python
- SQL
- R
  Development Environment:
- Command Line
- IDEs
    - VS Code
    - PyCharm
    - RStudio
- Notebook Environment
  * Interactive programming environment that allow Markdown descriptions to be interlaced with program code and output of commands
  Maintenance:
- Venv
- Conda
- Docker: Packaging environments + deploying application
    - Dev container
        - Github Codespace

API Design for Data Scientists

- API should always return data in JSON format
    - JSON instead of XML
    - Why?
        - Python has an strong ecosystem around JSON
        - JSON is:
            - Lightweight data format
            - Support hierarchies

- Human readable
- Can be convert into List & Dictionaries
  -
    - Most web APIs returns JSON
- Provide an SDK (Software Development Kit) to consume the API
  - It can happen directly, but you can make the life easier by publishing python library
    -
      - enforce good coding practices
- Provide external identifiers in data
  - Make the combination of multiple resources easier if your provide industry-standard identifier
- Data should conform to data type definitions
  - Little things that won't show themselves in web, can be so much critical in data science field
    - Data returned from APIs must conform to its definitions
      - [OpenAPI Specification (OAS)](#) file
- Provide a method for bulk downloads
  - Prevent from:
    - Big use of data scientist local resource + API provider resource
    - Timeout
    - Memory overflows
  - Good for initial full load of new data pipeline
  - Good formats are:
    - Comma-separated values (CSV)
    - Apache Parquet
- Should support querying by last changed data
  - Provide last changed date query parameter $\xrightarrow{\text{result into}}$ data pipeline to retrieve any new records + update records

What is an Good API

an API with a good story = fulfill the needs of real consumers

Project 1:

- Chapter 1: Understanding your users and selecting the right API

- Chapter 2: Selecting your API architecture and setting up your development environment
- Chapter 3: Creating your database
- Chapter 4: Developing the FastAPI code
- Chapter 5: Documenting your API
- Chapter 6: Deploying your API to the cloud
- Chapter 7: Creating an SDK for your API

Signs That You Need an API

- Extend your reach
- Need to provide access to your applications or system for your partner
- People access the website through web scrapping or reverse engineering website APIs $\xrightarrow{\text{which shows}}$ people want the data
- Valuable data, analytics or metrics to provide to the public or partners
- You created statistical or machine learning models to share
- You developed a generative AI models to share with application builders

Choose and Create Your First API

Check This Boxes

Base on:

- User desirability:
    - Your users want the produce.
- Technical feasibility:
    - Your technical environment and team can create it.
- Economic viability:
    - You expect it to be worth the investment.

Create User Stories

- User stories help you understand what you are building
- User story = informal descriptions of a feature or product that are written from the end user's perspective
- User Story template

- As a (user type)
- I want to (goal or intent)
- So that (motivation or benefits)

Additional Resources

Document Your Idea of an API:

- [Design Thinking](#)
- [The Lean Startup](#)
- [APIOps Cycles](#)
  API management:
- [APIs: A strategy guide](#)
- [Continuous API Management](#)