Application Programming Interfaces (APIs)

Software Architecture

February 20, 2023

Brae Webb

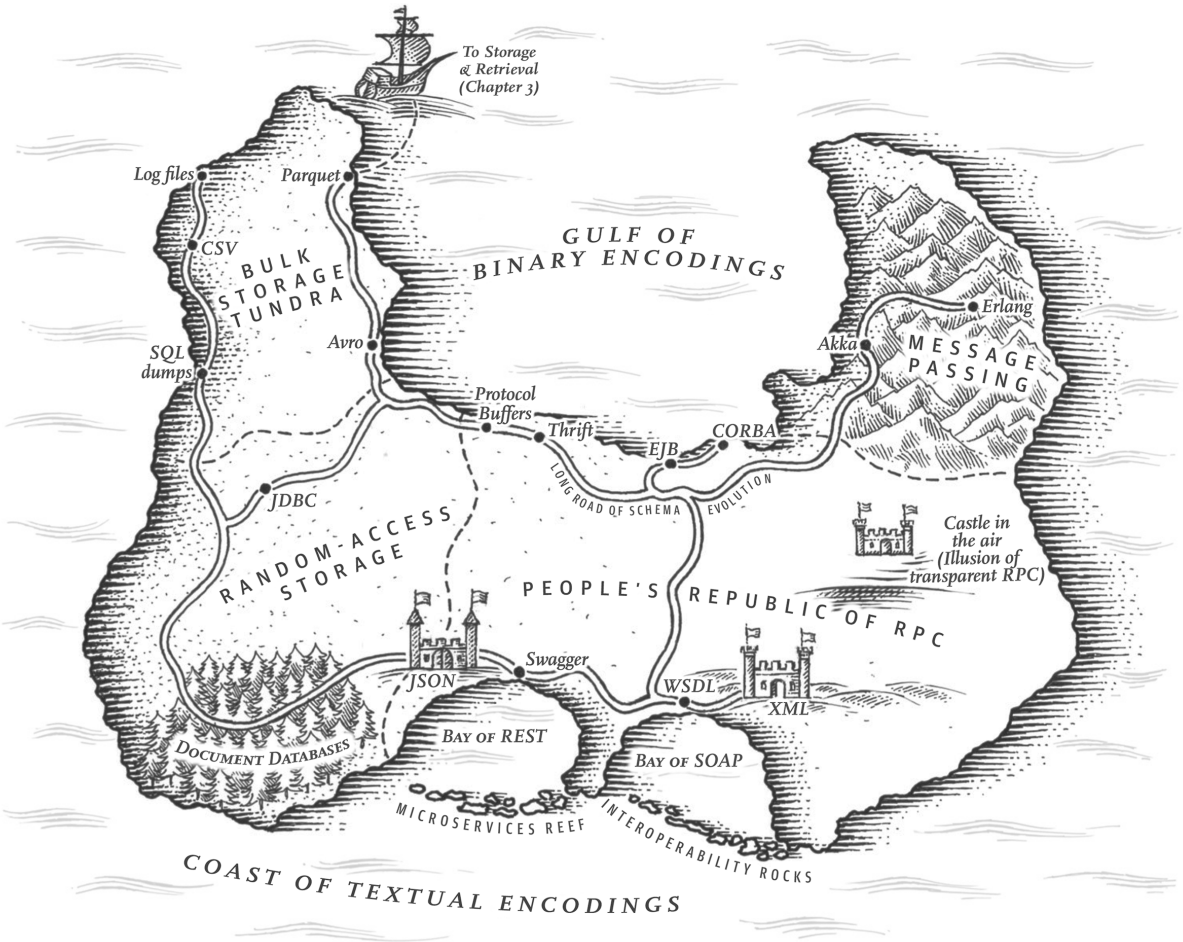

Figure 1: A map of communication techniques from Designing Data-Intensive Applications [1].

1 This Week

This week our goal is to:

- explore techniques used by developers use to communicate between components in distributed systems;
- initialize our GitHub repositories where we will be working on practical exercises; and
- build a minimal HTTP API of a todo app using the Flask framework.

2 Practicals

These practicals are designed to prepare you with the technical skills required for the cloud and team projects. We will normally spend the first section of the practicals gaining the relevant conceptual background for the practical. The second section will be a practical exercise where you will need to write and run some code.

This semester we will be working on the creation of a scalable and fault-tolerant todo application. You should aim to keep up with the practicals or you will not be able to complete the projects.

3 Communication

The world relies (heavily) on distributed systems, single machines are no longer sufficient to handle the demands of modern applications. There is no machine in the world powerful enough to process the requests Google receives every second[1]. But you don't need to be the size of Google to require distributed systems, even relatively small applications require multiple machines to handle the load. This inter-machine teamwork requires each machine to know how to talk to the others.

In this practical, we'll be exploring the different techniques used to communicate between components in distributed systems. We will then build a simple HTTP API using the Flask framework which allows other applications to interact with our application.

4 Data Formats

Communication requires an exchange of information. On single computer systems information is often stored at runtime in memory as primitive data which your programming language can interpret; bytes, integers, strings, etc. In object-oriented languages, primitive data is wrapped up into useful packages: objects. If we want this information to escape the confines of our programming language runtime, we need to package it up in a language-independent format. To be language-independent we need many languages to implement an encoding and decoding mechanism for the format. We have several language-independent formats available but a few defacto standards.

4.1 XML

Extensible Markup Language (XML) is one of the most widely used language-independent formats. The use cases of XML are extensive, it is the foundation for many popular utilities[2], such as SVG file formats, SAML authentication, RSS feeds, and ePub books.

```
<root>
  <item>
    <key>Course Code</key>
    <value>CSSE6400</value>
  </item>
  <item>
    <key>Course Title</key>
    <value>Software Architecture</value>
  </item>
</root>
```

[1] Current estimates are that Google requires over a million servers

[2] https://en.wikipedia.org/wiki/List_of_XML_markup_languages

XML is designed as a markup language, similar to HTML, it is not designed as a data exchange format. Developers have come to point out that the verbosity and complexity of XML, compared to alternatives such as JSON, are deal breakers. While XML can be used as a data exchange format it is not designed for it, and as a result APIs built around XML as a data format are becoming less common.

4.2 JSON

JavaScript Object Notation (JSON) is quickly replacing XML as the data format used in APIs. As you will note, it is more succinct and communicates the important points to a human reader better. The popularity of JSON is largely due to its compatibility with JavaScript which has taken over as the defacto web development language. JSON is the map-esque data type in JavaScript. Detractors of JSON claim that its main disadvantage compared to XML is that it lacks a schema. However, schemas are possible in JSON[3], they are optional, just as in XML, but are used much less than in XML.

```
» cat csse6400.json

{
    "Course Code": "CSSE6400",
    "Course Title": "Software Architecture"
}
```

4.3 MessagePack

It should not be a surprise that the JSON and XML formats are not resource efficient. Nowadays, we are less concerned with squeezing data into a tiny amount of data on the hard drive as storage is cheap. However, we are often concerned with how much data is being transmitted via network communication as bandwidth can be expensive.

In the example JSON snippet above, we use 78 bytes to encode the message. MessagePack[4] is a standard for encoding and decoding JSON. When encoding our original JSON snippet with MessagePack we shrink to just 57 bytes. At our scale, a negligible difference, but at the scale of terrabytes or petabytes, a significant factor.

```
» cat csse6400.msgpack

82 ab 43 6f 75 72 73 65 20 43 6f 64 65 a8 43 53 53 45 36 34 30 30 ac 43 6f 75 72 73
   65 20 54 69 74 6c 65 b5 53 6f 66 74 77 61 72 65 20 41 72 63 68 69 74 65 63 74 75
   72 65
```

Info

For those interested, 0x82 specifies a map type (0x80) with two fields (0x02). Followed by a string type (0xa0) of size eleven (0x0b). The rest is left as an exercise: https://github.com/msgpack/msgpack/blob/master/spec.md.

[3] https://json-schema.org/
[4] https://msgpack.org/

4.4 Protobuf

Protocol Buffers (protobuf) is another type of binary encoding. However, unlike MessagePack, the format was designed from scratch, allowing a more compact and better designed format. Protobufs require all data to be defined by a schema. For example:

```
   » cat csse6400.proto

message Course {
    required string code = 1;
    required string name = 2;
}
```

Protobufs differ from XML, JSON, and MessagePack via their method of integration. In the previous examples, your language would have a library to encode and decode the data format into and out of your language's type system. With protobuf, an external tool, *protoc*, takes the schema and generates a model of the schema in your target language. This gives every language a native interface for interacting with the data format, often allowing developers to not be aware of the underlying encoding.

```
$ protoc --java_out=. csse6400.proto
```

```
   » cat csse6400.java

Course softarch = Course.newBuilder()
    .setCode("CSSE6400")
    .setName("Software Architecture")
    .build();
output = new FileOutputStream(args[0]);
softarch.writeTo(output);
```

5 API Protocols

It is worth being aware of a few of the common protocols and conventions used in APIs. Fortunately, we don't need to understand the low-level communication protocols, such as TCP/IP, HTTP, and HTTPS, as these are handled by the underlying libraries. We will focus on the higher-level protocols and conventions.

We outline a few but note that in this course we will primarily focus on REST APIs. This is only an indication that they are better understood by course staff, not their superiority.

5.1 XML-RPC

XML-RPC was one of the first API standards. It is a Remote Procedure Call (RPC) based API which communicates via XML. It is not a commonly used standard anymore.

5.2 SOAP

After XML-RPC came SOAP. SOAP was impactful for service-oriented (now microservices) architectures but has largely fallen out of favour in-place of REST APIs. We do not cover SOAP in any meaningful depth due to its increasing irrelevance and complexity.

5.3 REST

REST is not a standard like SOAP once was, instead REST is an architectural style guided by a set of architectural constraints that allows us to build flexible APIs.

In this course we do not dive too deep into the architectural style and instead opt for a more surface level understanding. It is a common mistake for people to refer to REST as a HTTP based web service API, they are different. In this course we chose to embrace this mistake and often refer to a HTTP based web service API when saying REST.
An example of this type of API might be:

GET /api/v1/todo List all tasks todo

POST /api/v1/todo Create a task todo

GET /api/v1/todo/id List all details about a certain task

PUT /api/v1/todo/id Update the fields of an existing task

DELETE /api/v1/todo/id Delete a specific task

5.4 JSON-RPC

A JSON RPC is another RPC based API based around communication via JSON. When deciding whether to use an RPC-based API or a REST/SOAP based API, the distinction to make is that RPC APIs should be action based, i.e. "I want to do X". Whereas REST/SOAP APIs are Create Read Update Delete (CRUD) based for providing a interface to mutable data.

5.5 GraphQL

GraphQL is a query language which allows more dynamic querying of data than REST based APIs. You can create nested queries and specify the fields you want to receive in response. GraphQL APIs are well-suited for building APIs designed for developers to consume but when dealing with rigid inter-service based requests, REST APIs are generally preferable.

5.6 gRPC

Another RPC framework, gRPC has started gaining a lot of attention since its first release in 2016. gRPC is based on the protobuf format. In addition to a more type-safe system, gRPC based APIs provide authentication and streaming mechanisms.

6 GitHub

We will use GitHub to host our practical work. This is strongly encouraged as it will help you to get experience with the assessment submission process. Additionally, committing your work is a good habit to get into and will be useful for your future career.

6.1 Creating a GitHub Account

If you do not already have a GitHub account, you will need to create one. You can do this by visiting https://github.com/join.

6.2 Joining the Course Organization

Once you have created an account, you will need to join the course organization. If you have not yet filled out the Google Form, you will need to do so before you can join the organization. The link to the Google Form can be found on Blackboard.

Once you have filled out the form, tell your tutor your GitHub username and they will add you to the organization.

6.3 Joining the GitHub Classroom

Once you have joined the organization, you will need to join the GitHub Classroom.

TODO: Add link to GitHub Classroom

6.4 Creating a Practical Repository

7 TODO

7.1 The API design

TODO: Add description of all the API endpoints

7.2 Implementation with Flask

TODO: getting our github repo going

TODO: Intro to getting python up and running

TODO: Intro to flask

TODO: returning json with flask

TODO: Calling your api locally

References

[1] M. Kleppmann, *Designing Data-Intensive Applications: The big ideas behind reliable, scalable, and maintainable systems*. O'Reilly Media, Inc., March 2017.