Protobuf

Sandra Persson

TIAMATICA

What is Protobuf?

- *Protocol Buffers*
- *A format for serializing structured data developed by Google*
- *Designed to be small, fast and simple*
- *Language-neutral, platform-neutral, open-source*
- *Where is it used? gRPC*

Protobuf {JSON} <XML> CSV Avro...

Protocol buffers

Protobuf vs. JSON

```
{
  "name": "Cajal”,
  "id": 2,
  "email": "ramonycajal@gmail.com"
}
```

Fieldnr type

0 0 0 0 1 0 1 0

Protocol Buffers
Bytes sequence:

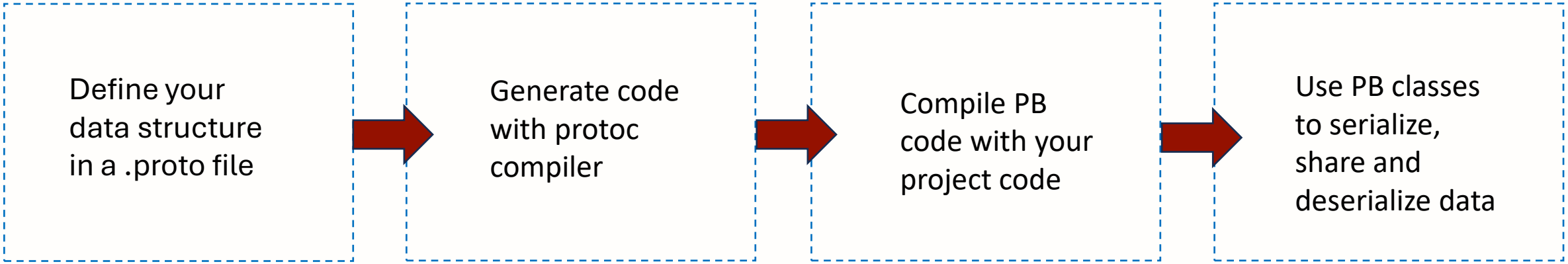
10 5 67 97 106 97 108 16 2 26 21 114 97 109 111 110 121
99 97 106 97 108 64 103 109 97 105 108 46 99 111 109

Breakdown:

tag	length	C	a	j	a	l
10	5	67	97	106	97	108

tag	2
16	2

tag	length	ramonycajal@gmail.com
26	21	114 97 109 111 110 121 99 97 106 97 108 64 103 109 97 105 108 46 99 111 109



.proto file:

```
message Person {
  optional string name = 1;
  optional int32 id = 2;
  repeated string email = 3;
}
```

```python
# -*- coding: utf-8 -*-
# Generated by the protocol buffer compiler.  DO NOT EDIT!
# NO CHECKED-IN PROTOBUF GENCODE
# source: SearchRequest.proto
# Protobuf Python Version: 5.27.0
"""Generated protocol buffer code."""
from google.protobuf import descriptor as _descriptor
from google.protobuf import descriptor_pool as _descriptor_pool
from google.protobuf import runtime_version as _runtime_version
from google.protobuf import symbol_database as _symbol_database
from google.protobuf.internal import builder as _builder
_runtime_version.ValidateProtobufRuntimeVersion(
    _runtime_version.Domain.PUBLIC,
    5,
    27,
    0,
    '',
    'SearchRequest.proto'
)
# @@protoc_insertion_point(imports)

_sym_db = _symbol_database.Default()



DESCRIPTOR = _descriptor_pool.Default().AddSerializedFile(b'\n\x13

_globals = globals()
_builder.BuildMessageAndEnumDescriptors(DESCRIPTOR, _globals)
```

10 5 67 97 106 97 108 16 2 26 21 114 97
109 111 110 121 99 97 106 97 108 64 1
03 109 97 105 108 46 99 111 109

Advantages and disadvantages

- Compact = Small and fast
- Schema as documentation
- Forward- and backward compatibility

- Not readable by humans
- Better for statically typed languages
- How much faster?

Supported languages

Languages supported directly by Google:

C++, C#, Java, Kotlin, Objective-C, PHP, Python, Ruby, Dart, Go

Third party:

C, Haskell, Javascript, Julia, Matlab, Perl, PHP, R, Rust, Scala, Swift, Typescript…

APL?

Can it be implemented in APL?

- As a Proof of concept
- As a project to learn APL
- Obstacles = lots of learning

Double
Float
Int64
Uint64
Int32
Fixed64
Fixed32
Bool
String
Message

Bytes
Uint32
Enum
Sfixed32
Sfixed64
Sint32
Sint64

Demo

- Serializing from Python
- Parsing in APL
- And then serializing again

```
syntax = "proto3";

package AddressBook;

message Person {
  optional string name = 1;
  optional int32 id = 2;
  optional string email = 3;

   enum PhoneType {
    PHONE_TYPE_UNSPECIFIED = 0;
    PHONE_TYPE_MOBILE = 1;
    PHONE_TYPE_HOME = 2;
    PHONE_TYPE_WORK = 3;
  }

   message PhoneNumber {
    optional string number = 1;
    optional PhoneType type = 2;
  }

  repeated PhoneNumber phones = 4;
}

message AddressBook {
  repeated Person people = 1;
}
```

Thanks for listening

TIAMATICA