Language Guide

This guide describes how to use the protocol buffer language to structure your protocol buffer data, including `.proto` file syntax and how to generate data access classes from your `.proto` files. It covers the **proto2** version of the protocol buffers language: for information on the newer **proto3** syntax, see the Proto3 Language Guide (https://developers.google.com/protocol-buffers/docs/proto3?hl=zh-cn).

Proto3 While **`proto2`** will continue to be supported, we encourage new code to use **`proto3`** instead, whic to use and supports more languages.

This is a reference guide – for a step by step example that uses many of the features described in this document, see the tutorial (https://developers.google.com/protocol-buffers/docs/tutorials?hl=zh-cn) for your chosen language.

Defining A Message Type

First let's look at a very simple example. Let's say you want to define a search request message format, where each search request has a query string, the particular page of results you are interested in, and a number of results per page. Here's the `.proto` file you use to define the message type.

```
ige SearchRequest {
quired string query = 1;
:ional int32 page_number = 2;
:ional int32 result_per_page = 3;
```

The `SearchRequest` message definition specifies three fields (name/value pairs), one for each piece of data that you want to include in this type of message. Each field has a name and a type.

Specifying Field Types

In the above example, all the fields are scalar types (#scalar): two integers (`page_number` and `result_per_page`) and a string (`query`). However, you can also specify composite types for your fields, including enumerations (#enum) and other message types.

Assigning Field Numbers

As you can see, each field in the message definition has a **unique number**. These numbers are used to identify your fields in the message binary format (https://developers.google.com/protocol-buffers/docs/encoding?hl=zh-cn), and should not be changed once your message type is in use. Note that field numbers in the range 1 through 15 take one byte to encode, including the field number and the field's type (you can find out more about this in Protocol Buffer Encoding (https://developers.google.com/protocol-buffers/docs/encoding.html?hl=zh-cn#structure)). Field numbers in the range 16 through 2047 take two bytes. So you should reserve the field numbers 1 through 15 for very frequently occurring message elements. Remember to leave some room for frequently occurring elements that might be added in the future.

The smallest field number you can specify is 1, and the largest is 2^{29} - 1, or 536,870,911. You also cannot use the numbers 19000 through 19999 (`FieldDescriptor::kFirstReservedNumber` through

`FieldDescriptor::kLastReservedNumber`), as they are reserved for the Protocol Buffers implementation - the protocol buffer compiler will complain if you use one of these reserved numbers in your `.proto`. Similarly, you cannot use any previously reserved (#reserved) field numbers.

Specifying Field Rules

You specify that message fields are one of the following:

- `required`: a well-formed message must have exactly one of this field.
- `optional`: a well-formed message can have zero or one of this field (but not more than one).
- `repeated`: this field can be repeated any number of times (including zero) in a well-formed message. The order of the repeated values will be preserved.

For historical reasons, `repeated` fields of scalar numeric types aren't encoded as efficiently as they could be. New code should use the special option `[packed=true]` to get a more efficient encoding. For example:

```
ıted int32 samples = 4 [packed=true];
```

You can find out more about `packed` encoding in Protocol Buffer Encoding (https://developers.google.com/protocol-buffers/docs/encoding.html?hl=zh-cn#packed).

ˈed Is Forever You should be very careful about marking fields as **`required`**. If at some point you wish to s
ˌ or sending a required field, it will be problematic to change the field to an optional field – old readers will
ler messages without this field to be incomplete and may reject or drop them unintentionally. You should
ler writing application-specific custom validation routines for your buffers instead. Some engineers at Goc
ome to the conclusion that using **`required`** does more harm than good; they prefer to use only **`optional`**
`ted`. However, this view is not universal.

Adding More Message Types

Multiple message types can be defined in a single `.proto` file. This is useful if you are defining multiple related messages – so, for example, if you wanted to define the reply message format that corresponds to your `SearchResponse` message type, you could add it to the same `.proto`:

```
ıge SearchRequest {
quired string query = 1;
tional int32 page_number = 2;
tional int32 result_per_page = 3;


ıge SearchResponse {
```

ning Messages leads to bloat While multiple message types (such as message, enum, and service) can b d in a single `.proto` file, it can also lead to dependency bloat when large numbers of messages with varyi dencies are defined in a single file. It's recommended to include as few message types per `.proto` file as ple.

Adding Comments

To add comments to your `.proto` files, use C/C++-style `//` and `/* ... */` syntax.

```
earchRequest represents a search query, with pagination options to
ndicate which results to include in the response. */

ıge SearchRequest {
quired string query = 1;
tional int32 page_number = 2;  // Which page number do we want?
tional int32 result_per_page = 3;  // Number of results to return per page.
```

Reserved Fields

If you update (#updating) a message type by entirely removing a field, or commenting it out, future users can reuse the field number when making their own updates to the type. This can cause severe issues if they later load old versions of the same `.proto`, including data corruption, privacy bugs, and so on. One way to make sure this doesn't happen is to specify that the field numbers (and/or names, which can also cause issues for JSON serialization) of your deleted fields are `reserved`. The protocol buffer compiler will complain if any future users try to use these field identifiers.

```
ıge Foo {
served 2, 15, 9 to 11;
served "foo", "bar";
```

Note that you can't mix field names and field numbers in the same `reserved` statement.

What's Generated From Your `.proto`?

When you run the protocol buffer compiler (#generating) on a `.proto`, the compiler generates the code in your chosen language you'll need to work with the message types you've described in the file, including getting and setting field values, serializing your messages to an output stream, and parsing your messages from an input stream.

- For **C++**, the compiler generates a `.h` and `.cc` file from each `.proto`, with a class for each message type described in your file.
- For **Java**, the compiler generates a `.java` file with a class for each message type, as well as special `Builder` classes for creating message class instances.
- **Python** is a little different – the Python compiler generates a module with a static descriptor of each message type in your `.proto`, which is then used with a *metaclass* to create the necessary Python data access class at runtime.
- For **Go**, the compiler generates a `.pb.go` file with a type for each message type in your file.

You can find out more about using the APIs for each language by following the tutorial for your chosen language. For even more API details, see the relevant API reference (https://developers.google.com/protocol-buffers/docs/reference/overview?hl=zh-cn).

Scalar Value Types

A scalar message field can have one of the following types – the table shows the type specified in the `.proto` file, and the corresponding type in the automatically generated class:

.proto Type	Notes	C++ Type	Java Type	Python Type[2]	Go Type
double		double	double	float	*float64

float		float	float	float	*float32
int32	Uses variable-length encoding. Inefficient for encoding negative numbers – if your field is likely to have negative values, use sint32 instead.	int32	int	int	*int32
int64	Uses variable-length encoding. Inefficient for encoding negative numbers – if your field is likely to have negative values, use sint64 instead.	int64	long	int/long[3]	*int64
uint32	Uses variable-length encoding.	uint32	int[1]	int/long[3]	*uint32
uint64	Uses variable-length encoding.	uint64	long[1]	int/long[3]	*uint64
sint32	Uses variable-length encoding. Signed int value. These more efficiently encode negative numbers than regular int32s.	int32	int	int	*int32
sint64	Uses variable-length encoding. Signed int value. These more efficiently encode negative numbers than regular int64s.	int64	long	int/long[3]	*int64
fixed32	Always four bytes. More efficient than uint32 if values are often greater than 2^{28}.	uint32	int[1]	int/long[3]	*uint32
fixed64	Always eight bytes. More efficient than uint64 if values are often greater than 2^{56}.	uint64	long[1]	int/long[3]	*uint64
sfixed32	Always four bytes.	int32	int	int	*int32
sfixed64	Always eight bytes.	int64	long	int/long[3]	*int64
bool		bool	boolean	bool	*bool
string	A string must always contain UTF-8 encoded or 7-bit ASCII text.	string	String	unicode (Python 2) or str (Python 3)	*string
bytes	May contain any arbitrary sequence of bytes.	string	ByteString	bytes	[]byte

You can find out more about how these types are encoded when you serialize your message in Protocol Buffer Encoding (https://developers.google.com/protocol-buffers/docs/encoding?hl=zh-cn).

[1] In Java, unsigned 32-bit and 64-bit integers are represented using their signed counterparts, with the top bit simply being stored in the sign bit.

[2] In all cases, setting values to a field will perform type checking to make sure it is valid.

[3] 64-bit or unsigned 32-bit integers are always represented as long when decoded, but can be an int if an int is given when setting the field. In all cases, the value must fit in the type represented when set. See [2].

Optional Fields And Default Values

As mentioned above, elements in a message description can be labeled `optional`. A well-formed message may or may not contain an optional element. When a message is parsed, if it does not contain an optional element, the corresponding field in the parsed object is set to the default value for that field. The default value can be specified as part of the message description. For example, let's say you want to provide a default value of 10 for a `SearchRequest`'s `result_per_page` value.

```
nal int32 result_per_page = 3 [default = 10];
```

If the default value is not specified for an optional element, a type-specific default value is used instead: for strings, the default value is the empty string. For bytes, the default value is the empty byte string. For bools, the default value is false. For numeric types, the default value is zero. For enums, the default value is the first value listed in the enum's type definition. This means care must be taken when adding a value to the beginning of an enum value list. See the Updating A Message Type (#updating) section for guidelines on how to safely change definitions.

Enumerations

When you're defining a message type, you might want one of its fields to only have one of a pre-defined list of values. For example, let's say you want to add a `corpus` field for each `SearchRequest`, where the corpus can be `UNIVERSAL`, `WEB`, `IMAGES`, `LOCAL`, `NEWS`, `PRODUCTS` or `VIDEO`. You can do this very simply by adding an `enum` to your message definition - a field with an `enum` type can only have one of a specified set of constants as its value (if you try to provide a different value, the parser will treat it like an unknown field). In the following example we've added an `enum` called `Corpus` with all the possible values, and a field of type `Corpus`:

```
ge SearchRequest {
uired string query = 1;
```

```
:ional int32 page_number = 2;
:ional int32 result_per_page = 3 [default = 10];
ım Corpus {
JNIVERSAL = 0;
/EB = 1;
IMAGES = 2;
.OCAL = 3;
IEWS = 4;
PRODUCTS = 5;
/IDEO = 6;

:ional Corpus corpus = 4 [default = UNIVERSAL];
```

You can define aliases by assigning the same value to different enum constants. To do this you need to set the `allow_alias` option to `true`, otherwise protocol compiler will generate an error message when aliases are found.

```
 EnumAllowingAlias {
:ion allow_alias = true;
(NOWN = 0;
ıRTED = 1;
INING = 1;

 EnumNotAllowingAlias {
(NOWN = 0;
ıRTED = 1;
 RUNNING = 1;  // Uncommenting this line will cause a compile error inside Goo
```

Enumerator constants must be in the range of a 32-bit integer. Since `enum` values use varint encoding (https://developers.google.com/protocol-buffers/docs/encoding?hl=zh-cn) on the wire, negative values are inefficient and thus not recommended. You can define `enum`s within a message definition, as in the above example, or outside – these `enum`s can be reused in any message definition in your `.proto` file. You can also use an `enum` type declared in one message as the type of a field in a different message, using the syntax *`MessageType.EnumType`*.

When you run the protocol buffer compiler on a `.proto` that uses an `enum`, the generated code will have a corresponding `enum` for Java or C++, or a special `EnumDescriptor` class for Python that's used to create a set of symbolic constants with integer values in the runtime-generated class.

For more information about how to work with message `enum`s in your applications, see the generated code guide (https://developers.google.com/protocol-buffers/docs/reference/overview?hl=zh-cn) for your chosen language.

Reserved Values

If you update (#updating) an enum type by entirely removing an enum entry, or commenting it out, future users can reuse the numeric value when making their own updates to the type. This can cause severe issues if they later load old versions of the same `.proto`, including data corruption, privacy bugs, and so on. One way to make sure this doesn't happen is to specify that the numeric values (and/or names, which can also cause issues for JSON serialization) of your deleted entries are `reserved`. The protocol buffer compiler will complain if any future users try to use these identifiers. You can specify that your reserved numeric value range goes up to the maximum possible value using the `max` keyword.

```
 Foo {
served 2, 15, 9 to 11, 40 to max;
served "FOO", "BAR";
```

Note that you can't mix field names and numeric values in the same `reserved` statement.

Using Other Message Types

You can use other message types as field types. For example, let's say you wanted to include `Result` messages in each `SearchResponse` message – to do this, you can define a `Result` message type in the same `.proto` and then specify a field of type `Result` in `SearchResponse`:

```
ge SearchResponse {
eated Result result = 1;


ge Result {
uired string url = 1;
ional string title = 2;
eated string snippets = 3;
```

Importing Definitions

In the above example, the `Result` message type is defined in the same file as `SearchResponse` – what if the message type you want to use as a field type is already defined in another `.proto` file?

You can use definitions from other `.proto` files by *importing* them. To import another `.proto`'s definitions, you add an import statement to the top of your file:

```
rt "myproject/other_protos.proto";
```

By default you can only use definitions from directly imported `.proto` files. However, sometimes you may need to move a `.proto` file to a new location. Instead of moving the `.proto` file directly and updating all the call sites in a single change, now you can put a dummy `.proto` file in the old location to forward all the imports to the new location using the `import public` notion. `import public` dependencies can be transitively relied upon by anyone importing the proto containing the `import public` statement. For example:

```
ew.proto
ll definitions are moved here
```

```
ld.proto
his is the proto that all clients are importing.
rt public "new.proto";
rt "other.proto";
```

```
lient.proto
rt "old.proto";
ou use definitions from old.proto and new.proto, but not other.proto
```

The protocol compiler searches for imported files in a set of directories specified on the protocol compiler command line using the `-I`/`--proto_path` flag. If no flag was given, it looks in the directory in which the compiler was invoked. In general you should set the `--proto_path` flag to the root of your project and use fully qualified names for all imports.

Using proto3 Message Types

It's possible to import proto3 (https://developers.google.com/protocol-buffers/docs/proto3?hl=zh-cn) message types and use them in your proto2 messages, and vice versa. However, proto2 enums cannot be used in proto3 syntax.

Nested Types

You can define and use message types inside other message types, as in the following example – here the `Result` message is defined inside the `SearchResponse` message:

```
ge SearchResponse {
ssage Result {
equired string url = 1;
ptional string title = 2;
epeated string snippets = 3;

eated Result result = 1;
```

If you want to reuse this message type outside its parent message type, you refer to it as *`Parent.Type`*:

```
ge SomeOtherMessage {
ional SearchResponse.Result result = 1;
```

You can nest messages as deeply as you like:

```
ge Outer {                  // Level 0
ssage MiddleAA {  // Level 1
essage Inner {   // Level 2
 required int64 ival = 1;
 optional bool  booly = 2;


ssage MiddleBB {  // Level 1
essage Inner {   // Level 2
```

```
required int32 ival = 1;
optional bool  booly = 2;
```

Groups

Note that this feature is deprecated and should not be used when creating new message types – use nested message types instead.

Groups are another way to nest information in your message definitions. For example, another way to specify a `SearchResponse` containing a number of `Results` is as follows:

```
ıge SearchResponse {
ıeated group Result = 1 {
'equired string url = 2;
ıptional string title = 3;
'epeated string snippets = 4;
```

A group simply combines a nested message type and a field into a single declaration. In your code, you can treat this message just as if it had a `Result` type field called `result` (the latter name is converted to lower-case so that it does not conflict with the former). Therefore, this example is exactly equivalent to the `SearchResponse` above, except that the message has a different wire format (https://developers.google.com/protocol-buffers/docs/encoding?hl=zh-cn).

Updating A Message Type

If an existing message type no longer meets all your needs – for example, you'd like the message format to have an extra field – but you'd still like to use code created with the old format, don't worry! It's very simple to update message types without breaking any of your existing code. Just remember the following rules:

- Don't change the field numbers for any existing fields.

- Any new fields that you add should be `optional` or `repeated`. This means that any messages serialized by code using your "old" message format can be parsed by your new generated code, as they won't be missing any `required` elements. You should set up sensible default values (#optional) for these elements so that new code can properly interact with messages generated by old code. Similarly, messages created by your new code can be parsed by your old code: old binaries simply ignore the new field when parsing. However, the unknown fields are not discarded, and if the message is later serialized, the unknown fields are serialized along with it – so if the message is passed on to new code, the new fields are still available.
- Non-required fields can be removed, as long as the field number is not used again in your updated message type. You may want to rename the field instead, perhaps adding the prefix "OBSOLETE_", or make the field number reserved (#reserved), so that future users of your `.proto` can't accidentally reuse the number.
- A non-required field can be converted to an extension (#extensions) and vice versa, as long as the type and number stay the same.
- `int32`, `uint32`, `int64`, `uint64`, and `bool` are all compatible – this means you can change a field from one of these types to another without breaking forwards- or backwards-compatibility. If a number is parsed from the wire which doesn't fit in the corresponding type, you will get the same effect as if you had cast the number to that type in C++ (e.g. if a 64-bit number is read as an int32, it will be truncated to 32 bits).
- `sint32` and `sint64` are compatible with each other but are *not* compatible with the other integer types.
- `string` and `bytes` are compatible as long as the bytes are valid UTF-8.
- Embedded messages are compatible with `bytes` if the bytes contain an encoded version of the message.
- `fixed32` is compatible with `sfixed32`, and `fixed64` with `sfixed64`.
- `optional` is compatible with `repeated`. Given serialized data of a repeated field as input, clients that expect this field to be `optional` will take the last input value if it's a primitive type field or merge all input elements if it's a message type field.
- Changing a default value is generally OK, as long as you remember that default values are never sent over the wire. Thus, if a program receives a message in which a particular field isn't set, the program will see the default value as it was defined in that program's version of the protocol. It will NOT see the default value that was defined in the sender's code.

- `enum` is compatible with `int32`, `uint32`, `int64`, and `uint64` in terms of wire format (note that values will be truncated if they don't fit), but be aware that client code may treat them differently when the message is deserialized. Notably, unrecognized `enum` values are discarded when the message is deserialized, which makes the field's `has..` accessor return false and its getter return the first value listed in the `enum` definition, or the default value if one is specified. In the case of repeated enum fields, any unrecognized values are stripped out of the list. However, an integer field will always preserve its value. Because of this, you need to be very careful when upgrading an integer to an `enum` in terms of receiving out of bounds enum values on the wire.
- In the current Java and C++ implementations, when unrecognized `enum` values are stripped out, they are stored along with other unknown fields. Note that this can result in strange behavior if this data is serialized and then reparsed by a client that recognizes these values. In the case of optional fields, even if a new value was written after the original message was deserialized, the old value will be still read by clients that recognize it. In the case of repeated fields, the old values will appear after any recognized and newly-added values, which means that order will not be preserved.
- Changing a single `optional` value into a member of a **new** `oneof` is safe and binary compatible. Moving multiple `optional` fields into a new `oneof` may be safe if you are sure that no code sets more than one at a time. Moving any fields into an existing `oneof` is not safe.

Extensions

Extensions let you declare that a range of field numbers in a message are available for third-party extensions. An extension is a placeholder for a field whose type is not defined by the original `.proto` file. This allows other `.proto` files to add to your message definition by defining the types of some or all of the fields with those field numbers. Let's look at an example:

```
ige Foo {
 ...
:ensions 100 to 199;
```

This says that the range of field numbers [100, 199] in `Foo` is reserved for extensions. Other users can now add new fields to `Foo` in their own `.proto` files that import your `.proto`, using field numbers within your specified range – for example:

```
ıd Foo {
:ional int32 bar = 126;
```

This adds a field named `bar` with the field number 126 to the original definition of `Foo`.

When your user's `Foo` messages are encoded, the wire format is exactly the same as if the user defined the new field inside `Foo`. However, the way you access extension fields in your application code is slightly different to accessing regular fields – your generated data access code has special accessors for working with extensions. So, for example, here's how you set the value of `bar` in C++:

```
'oo;
;etExtension(bar, 15);
```

Similarly, the `Foo` class defines templated accessors `HasExtension()`, `ClearExtension()`, `GetExtension()`, `MutableExtension()`, and `AddExtension()`. All have semantics matching the corresponding generated accessors for a normal field. For more information about working with extensions, see the generated code reference for your chosen language.

Note that extensions can be of any field type, including message types, but cannot be oneofs or maps.

Nested Extensions

You can declare extensions in the scope of another type:

```
ıge Baz {
:end Foo {
ɔptional int32 bar = 126;
```

In this case, the C++ code to access this extension is:

```
'oo;
;etExtension(Baz::bar, 15);
```

In other words, the only effect is that `bar` is defined within the scope of `Baz`.

a common source of confusion: Declaring an **`extend`** block nested inside a message type *does not* imply nship between the outer type and the extended type. In particular, the above example *does not* mean that rt of subclass of **`Foo`**. All it means is that the symbol **`bar`** is declared inside the scope of **`Baz`**; it's simply a er.

A common pattern is to define extensions inside the scope of the extension's field type – for example, here's an extension to `Foo` of type `Baz`, where the extension is defined as part of `Baz`:

```
ıge Baz {
:end Foo {
ɪptional Baz foo_ext = 127;
```

However, there is no requirement that an extension with a message type be defined inside that type. You can also do this:

```
ıge Baz {


ıis can even be in a different file.
ıd Foo {
:ional Baz foo_baz_ext = 127;
```

In fact, this syntax may be preferred to avoid confusion. As mentioned above, the nested syntax is often mistaken for subclassing by users who are not already familiar with extensions.

Choosing Extension Numbers

It's very important to make sure that two users don't add extensions to the same message type using the same field number – data corruption can result if an extension is accidentally interpreted as the wrong type. You may want to consider defining an extension numbering convention for your project to prevent this happening.

If your numbering convention might involve extensions having very large field numbers, you can specify that your extension range goes up to the maximum possible field number using the `max` keyword:

```
ige Foo {
:ensions 1000 to max;
```

`max` is 2^{29} - 1, or 536,870,911.

As when choosing field numbers in general, your numbering convention also needs to avoid field numbers 19000 though 19999 (`FieldDescriptor::kFirstReservedNumber` through `FieldDescriptor::kLastReservedNumber`), as they are reserved for the Protocol Buffers implementation. You can define an extension range that includes this range, but the protocol compiler will not allow you to define actual extensions with these numbers.

Oneof

If you have a message with many optional fields and where at most one field will be set at the same time, you can enforce this behavior and save memory by using the oneof feature.

Oneof fields are like optional fields except all the fields in a oneof share memory, and at most one field can be set at the same time. Setting any member of the oneof automatically clears all the other members. You can check which value in a oneof is set (if any) using a special `case()` or `WhichOneof()` method, depending on your chosen language.

Using Oneof

To define a oneof in your `.proto` you use the `oneof` keyword followed by your oneof name, in this case `test_oneof`:

```
ige SampleMessage {
eof test_oneof {
 string name = 4;
 SubMessage sub_message = 9;
```

You then add your oneof fields to the oneof definition. You can add fields of any type, but cannot use the `required`, `optional`, or `repeated` keywords. If you need to add a repeated field to a oneof, you can use a message containing the repeated field.

In your generated code, oneof fields have the same getters and setters as regular `optional` methods. You also get a special method for checking which value (if any) in the oneof is set. You can find out more about the oneof API for your chosen language in the relevant [API reference](https://developers.google.com/protocol-buffers/docs/reference/overview?hl=zh-cn) (https://developers.google.com/protocol-buffers/docs/reference/overview?hl=zh-cn).

Oneof Features

- Setting a oneof field will automatically clear all other members of the oneof. So if you set several oneof fields, only the *last* field you set will still have a value.

  ```
  SampleMessage message;
  message.set_name("name");
  CHECK(message.has_name());
  message.mutable_sub_message();   // Will clear name field.
  CHECK(!message.has_name());
  ```

- If the parser encounters multiple members of the same oneof on the wire, only the last member seen is used in the parsed message.
- Extensions are not supported for oneof.
- A oneof cannot be `repeated`.
- Reflection APIs work for oneof fields.
- If you set a oneof field to the default value (such as setting an int32 oneof field to 0), the "case" of that oneof field will be set, and the value will be serialized on the wire.
- If you're using C++, make sure your code doesn't cause memory crashes. The following sample code will crash because `sub_message` was already deleted by calling the `set_name()` method.

  ```
  SampleMessage message;
  SubMessage* sub_message = message.mutable_sub_message();
  ```

```
message.set_name("name");      // Will delete sub_message
sub_message->set_...            // Crashes here
```

- Again in C++, if you `Swap()` two messages with oneofs, each message will end up with the other's oneof case: in the example below, `msg1` will have a `sub_message` and `msg2` will have a `name`.

```
SampleMessage msg1;
msg1.set_name("name");
SampleMessage msg2;
msg2.mutable_sub_message();
msg1.swap(&msg2);
CHECK(msg1.has_sub_message());
CHECK(msg2.has_name());
```

Backwards-compatibility issues

Be careful when adding or removing oneof fields. If checking the value of a oneof returns `None`/`NOT_SET`, it could mean that the oneof has not been set or it has been set to a field in a different version of the oneof. There is no way to tell the difference, since there's no way to know if an unknown field on the wire is a member of the oneof.

Tag Reuse Issues

- **Move optional fields into or out of a oneof**: You may lose some of your information (some fields will be cleared) after the message is serialized and parsed. However, you can safely move a single field into a **new** oneof and may be able to move multiple fields if it is known that only one is ever set.
- **Delete a oneof field and add it back**: This may clear your currently set oneof field after the message is serialized and parsed.
- **Split or merge oneof**: This has similar issues to moving regular `optional` fields.

Maps

If you want to create an associative map as part of your data definition, protocol buffers provides a handy shortcut syntax:

```
:ey_type, value_type> map_field = N;
```

...where the `key_type` can be any integral or string type (so, any scalar (#scalar) type except for floating point types and `bytes`). Note that enum is not a valid `key_type`. The `value_type` can be any type except another map.

So, for example, if you wanted to create a map of projects where each `Project` message is associated with a string key, you could define it like this:

```
;tring, Project> projects = 3;
```

The generated map API is currently available for all proto2 supported languages. You can find out more about the map API for your chosen language in the relevant API reference (https://developers.google.com/protocol-buffers/docs/reference/overview?hl=zh-cn).

Maps Features

- Extensions are not supported for maps.
- Maps cannot be `repeated`, `optional`, or `required`.
- Wire format ordering and map iteration ordering of map values is undefined, so you cannot rely on your map items being in a particular order.
- When generating text format for a `.proto`, maps are sorted by key. Numeric keys are sorted numerically.
- When parsing from the wire or when merging, if there are duplicate map keys the last key seen is used. When parsing a map from text format, parsing may fail if there are duplicate keys.

Backwards compatibility

The map syntax is equivalent to the following on the wire, so protocol buffers implementations that do not support maps can still handle your data:

```
ige MapFieldEntry {
:ional key_type key = 1;
:ional value_type value = 2;
```

```
ıted MapFieldEntry map_field = N;
```

Any protocol buffers implementation that supports maps must both produce and accept data that can be accepted by the above definition.

Packages

You can add an optional `package` specifier to a `.proto` file to prevent name clashes between protocol message types.

```
ıge foo.bar;
ıge Open { ... }
```

You can then use the package specifier when defining fields of your message type:

```
ıge Foo {

ıuired foo.bar.Open open = 1;
```

The way a package specifier affects the generated code depends on your chosen language:

- In **C++** the generated classes are wrapped inside a C++ namespace. For example, `Open` would be in the namespace `foo::bar`.
- In **Java**, the package is used as the Java package, unless you explicitly provide a `option java_package` in your `.proto` file.
- In **Python**, the `package` directive is ignored, since Python modules are organized according to their location in the file system.
- In **Go**, the `package` directive is ignored, and the generated `.pb.go` file is in the package named after the corresponding `go_proto_library` rule.

Note that even when the `package` directive does not directly affect the generated code, for example in Python, it is still strongly recommended to specify the package for the `.proto` file, as otherwise it may lead to naming conflicts in descriptors and make the proto not portable for other languages.

Packages and Name Resolution

Type name resolution in the protocol buffer language works like C++: first the innermost scope is searched, then the next-innermost, and so on, with each package considered to be "inner" to its parent package. A leading '.' (for example, `.foo.bar.Baz`) means to start from the outermost scope instead.

The protocol buffer compiler resolves all type names by parsing the imported `.proto` files. The code generator for each language knows how to refer to each type in that language, even if it has different scoping rules.

Defining Services

If you want to use your message types with an RPC (Remote Procedure Call) system, you can define an RPC service interface in a `.proto` file and the protocol buffer compiler will generate service interface code and stubs in your chosen language. So, for example, if you want to define an RPC service with a method that takes your `SearchRequest` and returns a `SearchResponse`, you can define it in your `.proto` file as follows:

```
.ce SearchService {
: Search (SearchRequest) returns (SearchResponse);
```

By default, the protocol compiler will then generate an abstract interface called `SearchService` and a corresponding "stub" implementation. The stub forwards all calls to an `RpcChannel`, which in turn is an abstract interface that you must define yourself in terms of your own RPC system. For example, you might implement an `RpcChannel` which serializes the message and sends it to a server via HTTP. In other words, the generated stub provides a type-safe interface for making protocol-buffer-based RPC calls, without locking you into any particular RPC implementation. So, in C++, you might end up with code like this:

```
| google::protobuf;

ıbuf::RpcChannel* channel;
```

```
buf::RpcController* controller;
hService* service;
hRequest request;
hResponse response;

DoSearch() {
You provide classes MyRpcChannel and MyRpcController, which implement
the abstract interfaces protobuf::RpcChannel and protobuf::RpcController.
nnel = new MyRpcChannel("somehost.example.com:1234");
troller = new MyRpcController;

The protocol compiler generates the SearchService class based on the
definition given above.
vice = new SearchService::Stub(channel);

Set up the request.
uest.set_query("protocol buffers");

Execute the RPC.
vice->Search(controller, request, response, protobuf::NewCallback(&Done));


Done() {
ete service;
ete channel;
ete controller;
```

All service classes also implement the `Service` interface, which provides a way to call specific methods without knowing the method name or its input and output types at compile time. On the server side, this can be used to implement an RPC server with which you could register services.

```
 google::protobuf;

 ExampleSearchService : public SearchService {
ic:
d Search(protobuf::RpcController* controller,
         const SearchRequest* request,
         SearchResponse* response,
         protobuf::Closure* done) {
f (request->query() == "google") {
 response->add_result()->set_url("http://www.google.com");
 else if (request->query() == "protocol buffers") {
 response->add_result()->set_url("http://protobuf.googlecode.com");
```

```
lone->Run();



ıain() {
 You provide class MyRpcServer.  It does not have to implement any
 particular interface; this is just an example.
RpcServer server;

otobuf::Service* service = new ExampleSearchService;
rver.ExportOnPort(1234, service);
rver.Run();

lete service;
turn 0;
```

If you don't want to plug in your own existing RPC system, you can now use gRPC (https://github.com/grpc/grpc-common): a language- and platform-neutral open source RPC system developed at Google. gRPC works particularly well with protocol buffers and lets you generate the relevant RPC code directly from your `.proto` files using a special protocol buffer compiler plugin. However, as there are potential compatibility issues between clients and servers generated with proto2 and proto3, we recommend that you use proto3 for defining gRPC services. You can find out more about proto3 syntax in the Proto3 Language Guide (https://developers.google.com/protocol-buffers/docs/proto3?hl=zh-cn). If you do want to use proto2 with gRPC, you need to use version 3.0.0 or higher of the protocol buffers compiler and libraries.

In addition to gRPC, there are also a number of ongoing third-party projects to develop RPC implementations for Protocol Buffers. For a list of links to projects we know about, see the third-party add-ons wiki page (https://github.com/protocolbuffers/protobuf/blob/master/docs/third_party.md).

Options

Individual declarations in a `.proto` file can be annotated with a number of *options*. Options do not change the overall meaning of a declaration, but may affect the way it is handled in a particular context. The complete list of available options is defined in `google/protobuf/descriptor.proto`.

Some options are file-level options, meaning they should be written at the top-level scope, not inside any message, enum, or service definition. Some options are message-level options, meaning they should be written inside message definitions. Some options are field-level options, meaning they should be written inside field definitions. Options can also be written on enum types, enum values, service types, and service methods; however, no useful options currently exist for any of these.

Here are a few of the most commonly used options:

- `java_package` (file option): The package you want to use for your generated Java classes. If no explicit `java_package` option is given in the `.proto` file, then by default the proto package (specified using the "package" keyword in the `.proto` file) will be used. However, proto packages generally do not make good Java packages since proto packages are not expected to start with reverse domain names. If not generating Java code, this option has no effect.

```
option java_package = "com.example.foo";
```

- `java_outer_classname` (file option): The class name for the outermost Java class (and hence the file name) you want to generate. If no explicit `java_outer_classname` is specified in the `.proto` file, the class name will be constructed by converting the `.proto` file name to camel-case (so `foo_bar.proto` becomes `FooBar.java`). If not generating Java code, this option has no effect.

```
option java_outer_classname = "Ponycopter";
```

- `optimize_for` (file option): Can be set to `SPEED`, `CODE_SIZE`, or `LITE_RUNTIME`. This affects the C++ and Java code generators (and possibly third-party generators) in the following ways:
  - `SPEED` (default): The protocol buffer compiler will generate code for serializing, parsing, and performing other common operations on your message types. This code is highly optimized.
  - `CODE_SIZE`: The protocol buffer compiler will generate minimal classes and will rely on shared, reflection-based code to implement serialialization, parsing, and various other operations. The generated code will thus be much smaller than with `SPEED`, but operations will be slower. Classes will still implement exactly the same public API as they do in `SPEED` mode. This mode is most useful in apps that contain a very large number `.proto` files and do not need all of them to be blindingly fast.

    - `LITE_RUNTIME`: The protocol buffer compiler will generate classes that depend only on the "lite" runtime library (`libprotobuf-lite` instead of `libprotobuf`). The lite runtime is much smaller than the full library (around an order of magnitude smaller) but omits certain features like descriptors and reflection. This is particularly useful for apps running on constrained platforms like mobile phones. The compiler will still generate fast implementations of all methods as it does in `SPEED` mode. Generated classes will only implement the `MessageLite` interface in each language, which provides only a subset of the methods of the full `Message` interface.

```
option optimize_for = CODE_SIZE;
```

- `cc_generic_services`, `java_generic_services`, `py_generic_services` (file options): Whether or not the protocol buffer compiler should generate abstract service code based on services definitions (#services) in C++, Java, and Python, respectively. For legacy reasons, these default to `true`. However, as of version 2.3.0 (January 2010), it is considered preferrable for RPC implementations to provide code generator plugins (https://developers.google.com/protocol-buffers/docs/reference/cpp/google.protobuf.compiler.plugin.pb?hl=zh-cn) to generate code more specific to each system, rather than rely on the "abstract" services.

```
// This file relies on plugins to generate service code.
option cc_generic_services = false;
option java_generic_services = false;
option py_generic_services = false;
```

- `cc_enable_arenas` (file option): Enables arena allocation (https://developers.google.com/protocol-buffers/docs/reference/arenas?hl=zh-cn) for C++ generated code.
- `message_set_wire_format` (message option): If set to `true`, the message uses a different binary format intended to be compatible with an old format used inside Google called `MessageSet`. Users outside Google will probably never need to use this option. The message must be declared exactly as follows:

```
message Foo {
  option message_set_wire_format = true;
```

```
  extensions 4 to max;
}
```

- **`packed`** (field option): If set to `true` on a repeated field of a basic numeric type, a more compact encoding (https://developers.google.com/protocol-buffers/docs/encoding.html?hl=zh-cn#packed) is used. There is no downside to using this option. However, note that prior to version 2.3.0, parsers that received packed data when not expected would ignore it. Therefore, it was not possible to change an existing field to packed format without breaking wire compatibility. In 2.3.0 and later, this change is safe, as parsers for packable fields will always accept both formats, but be careful if you have to deal with old programs using old protobuf versions.

  ```
  repeated int32 samples = 4 [packed=true];
  ```

- **`deprecated`** (field option): If set to `true`, indicates that the field is deprecated and should not be used by new code. In most languages this has no actual effect. In Java, this becomes a `@Deprecated` annotation. In the future, other language-specific code generators may generate deprecation annotations on the field's accessors, which will in turn cause a warning to be emitted when compiling code which attempts to use the field. If the field is not used by anyone and you want to prevent new users from using it, consider replacing the field declaration with a reserved (#reserved) statement.

  ```
  optional int32 old_field = 6 [deprecated=true];
  ```

Custom Options

Protocol Buffers even allow you to define and use your own options. Note that this is an **advanced feature** which most people don't need. Since options are defined by the messages defined in `google/protobuf/descriptor.proto` (like `FileOptions` or `FieldOptions`), defining your own options is simply a matter of extending (#extensions) those messages. For example:

```
rt "google/protobuf/descriptor.proto";

nd google.protobuf.MessageOptions {
tional string my_option = 51234;
```

```
ıge MyMessage {
:ion (my_option) = "Hello world!";
```

Here we have defined a new message-level option by extending `MessageOptions`. When we then use the option, the option name must be enclosed in parentheses to indicate that it is an extension. We can now read the value of `my_option` in C++ like so:

```
ıg value = MyMessage::descriptor()->options().GetExtension(my_option);
```

Here, `MyMessage::descriptor()->options()` returns the `MessageOptions` protocol message for `MyMessage`. Reading custom options from it is just like reading any other extension (#extensions).

Similarly, in Java we would write:

```
ıg value = MyProtoFile.MyMessage.getDescriptor().getOptions()
:tExtension(MyProtoFile.myOption);
```

In Python it would be:

```
: = my_proto_file_pb2.MyMessage.DESCRIPTOR.GetOptions()
:tensions[my_proto_file_pb2.my_option]
```

Custom options can be defined for every kind of construct in the Protocol Buffers language. Here is an example that uses every kind of option:

```
·t "google/protobuf/descriptor.proto";

ıd google.protobuf.FileOptions {
:ional string my_file_option = 50000;

ıd google.protobuf.MessageOptions {
:ional int32 my_message_option = 50001;

ıd google.protobuf.FieldOptions {
:ional float my_field_option = 50002;
```

```
nd google.protobuf.EnumOptions {
tional bool my_enum_option = 50003;

nd google.protobuf.EnumValueOptions {
tional uint32 my_enum_value_option = 50004;

nd google.protobuf.ServiceOptions {
tional MyEnum my_service_option = 50005;

nd google.protobuf.MethodOptions {
tional MyMessage my_method_option = 50006;


on (my_file_option) = "Hello world!";

age MyMessage {
tion (my_message_option) = 1234;

tional int32 foo = 1 [(my_field_option) = 4.5];
tional string bar = 2;


 MyEnum {
tion (my_enum_option) = true;

 = 1 [(my_enum_value_option) = 321];
 = 2;


age RequestType {}
age ResponseType {}

ce MyService {
tion (my_service_option) = FOO;

 MyMethod(RequestType) returns(ResponseType) {
/ Note:  my_method_option has type MyMessage.  We can set each field
/   within it using a separate "option" line.
ption (my_method_option).foo = 567;
ption (my_method_option).bar = "Some string";
```

Note that if you want to use a custom option in a package other than the one in which it was defined, you must prefix the option name with the package name, just as you would for type names. For example:

```
o.proto
't "google/protobuf/descriptor.proto";
ıge foo;
ıd google.protobuf.MessageOptions {
:ional string my_option = 51234;
```

```
ır.proto
't "foo.proto";
ıge bar;
ıge MyMessage {
:ion (foo.my_option) = "Hello world!";
```

One last thing: Since custom options are extensions, they must be assigned field numbers like any other field or extension. In the examples above, we have used field numbers in the range 50000-99999. This range is reserved for internal use within individual organizations, so you can use numbers in this range freely for in-house applications. If you intend to use custom options in public applications, however, then it is important that you make sure that your field numbers are globally unique. To obtain globally unique field numbers, please send a request to add an entry to protobuf global extension registry (https://github.com/protocolbuffers/protobuf/blob/master/docs/options.md). Usually you only need one extension number. You can declare multiple options with only one extension number by putting them in a sub-message:

```
ıge FooOptions {
:ional int32 opt1 = 1;
:ional string opt2 = 2;


ıd google.protobuf.FieldOptions {
:ional FooOptions foo_options = 1234;


ıage:
ıge Bar {
:ional int32 a = 1 [(foo_options).opt1 = 123, (foo_options).opt2 = "baz"];
 alternative aggregate syntax (uses TextFormat):
:ional int32 b = 2 [(foo_options) = { opt1: 123 opt2: "baz" }];
```

Also, note that each option type (file-level, message-level, field-level, etc.) has its own number space, so e.g. you could declare extensions of FieldOptions and MessageOptions with the same number.

Generating Your Classes

To generate the Java, Python, or C++ code you need to work with the message types defined in a `.proto` file, you need to run the protocol buffer compiler `protoc` on the `.proto`. If you haven't installed the compiler, download the package (https://developers.google.com/protocol-buffers/docs/downloads.html?hl=zh-cn) and follow the instructions in the README.

The Protocol Compiler is invoked as follows:

```
c --proto_path=IMPORT_PATH --cpp_out=DST_DIR --java_out=DST_DIR --python_out=
```

- `IMPORT_PATH` specifies a directory in which to look for `.proto` files when resolving `import` directives. If omitted, the current directory is used. Multiple import directories can be specified by passing the `--proto_path` option multiple times; they will be searched in order. `-I=IMPORT_PATH` can be used as a short form of `--proto_path`.
- You can provide one or more *output directives*:
  - `--cpp_out` generates C++ code in `DST_DIR`. See the C++ generated code reference (https://developers.google.com/protocol-buffers/docs/reference/cpp-generated?hl=zh-cn) for more.
  - `--java_out` generates Java code in `DST_DIR`. See the Java generated code reference (https://developers.google.com/protocol-buffers/docs/reference/java-generated?hl=zh-cn) for more.
  - `--python_out` generates Python code in `DST_DIR`. See the Python generated code reference (https://developers.google.com/protocol-buffers/docs/reference/python-generated?hl=zh-cn) for more.

  As an extra convenience, if the `DST_DIR` ends in `.zip` or `.jar`, the compiler will write the output to a single ZIP-format archive file with the given name. `.jar` outputs will also be given a manifest file as required by the Java JAR specification. Note that if the output archive already exists, it will be overwritten; the compiler is not smart enough to add files to an existing archive.

- You must provide one or more `.proto` files as input. Multiple `.proto` files can be specified at once. Although the files are named relative to the current directory, each file must reside in one of the `IMPORT_PATH`s so that the compiler can determine its canonical name.