

# **File-Format Migrations**

on the Type-Level!

Victor Miraldo

January 5, 2024







SaaS company in Utrecht, with  $\sim\!300$  employees.





SaaS company in Utrecht, with  $\sim$ 300 employees.

We develop a tool that empowers advertisers and online businesses to manage, scale, and optimize their marketing.





SaaS company in Utrecht, with  $\sim$ 300 employees.

We develop a tool that empowers advertisers and online businesses to manage, scale, and optimize their marketing.

This tool is composed by a number of services in the backend, most of which are written in Haskell.



#### **Haskell at Channable**

- We have 11 (out of 15) services written in Haskell,





#### Haskell at Channable

- We have 11 (out of 15) services written in Haskell,
- using about 1000 source files (not counting tests),





#### Haskell at Channable

- We have 11 (out of 15) services written in Haskell,
- using about 1000 source files (not counting tests),
- which wc -1 to around 200k lines!





## **The Problem**





#### **Context**

Per-user dynamic configuration, stored in a disk not too far away:

```
reasonable_programming_languages="functional" type_system_and_eval=strong_lazy
```





#### **Context**

Per-user dynamic configuration, stored in a disk not too far away:

reasonable\_programming\_languages="functional"
type\_system\_and\_eval=strong\_lazy

Two services, A and B, rely on these configs





#### A Wild Change Appears

Now we want to split type\_system\_and\_eval into two settings:

```
reasonable_programming_languages="functional"
type_system=strong
eval=lazy
```





#### A Wild Change Appears

Now we want to split type\_system\_and\_eval into two settings:

reasonable\_programming\_languages="functional"
type\_system=strong
eval=lazy

How do we update A and B to use this new format?











1. Pause A and B; change all user settings, deploy new versions of A and B.





. .

1. Pause A and B; change all user settings, deploy new versions of A and B. Not great! Downtime, difficult to coordinate, prone to errors.





- 1. Pause A and B; change all user settings, deploy new versions of A and B. Not great! Downtime, difficult to coordinate, prone to errors.
- 2. Make versions of A and B that support both formats; read format **0**, write format **1**.





- 1. Pause A and B; change all user settings, deploy new versions of A and B. Not great! Downtime, difficult to coordinate, prone to errors.
- 2. Make versions of A and B that support both formats; read format 0, write format 1. Better... not ideal... Say A is in Haskell, B is in Python: need four parsers for your config file...





- 1. Pause A and B; change all user settings, deploy new versions of A and B. Not great! Downtime, difficult to coordinate, prone to errors.
- 2. Make versions of A and B that support both formats; read format 0, write format 1. Better... not ideal... Say A is in Haskell, B is in Python: need four parsers for your config file...
- 3. Serve the config file in a given format.





- 1. Pause A and B; change all user settings, deploy new versions of A and B. Not great! Downtime, difficult to coordinate, prone to errors.
- 2. Make versions of A and B that support both formats; read format 0, write format 1. Better... not ideal... Say A is in Haskell, B is in Python: need four parsers for your config file...
- 3. Serve the config file in a given format. Can update A independently: use config in format 1 while B can still use format 0.



20

- 1. Pause A and B; change all user settings, deploy new versions of A and B. Not great! Downtime, difficult to coordinate, prone to errors.
- 2. Make versions of A and B that support both formats; read format 0, write format 1. Better... not ideal... Say A is in Haskell, B is in Python: need four parsers for your config file...
- 3. Serve the config file *in a given format*. Can update A independently: use config in format 1 while B can still use format 0.

Going for the last option!



- User settings are seen as a resource

22





- User settings are seen as a resource
- The format of a resource changes over time





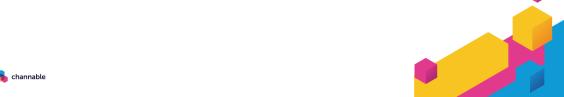


- User settings are seen as a resource
- The format of a resource changes over time
- The format depends on the service using the resource



2/

- User settings are seen as a resource
- The format of a resource changes over time
- The format depends on the service using the resource
- Multiple services use resources concurrently



25







#### Resource

Information seen under different views (formats)





#### Resource

Information seen under different views (formats)

Loosely:

$$Resource = \mathbb{N} \to \{0,1\}^* + \bot$$

Takes the desired format; returns bytes of respective view, if any.





Families of Formats

20





Families of Formats

```
class Resource r where
  type Formats r :: '[ Type ]
```





Families of Formats

```
class All1 ValidFormat (Formats r) => Resource r where
  type Formats r :: '[ Type ]
```





Families of Formats

```
class All1 ValidFormat (Formats r) => Resource r where
  type Formats r :: '[ Type ]
```

```
type family All1 c xs where
All1 c '[] = TypeError "Need at least one element"
All1 c '[x] = ()
All1 c (x ': xs) = (c x, All1 c xs)
```



3

#### **Formats**

3.





```
class (...) => ValidFormat (fmt :: Type) where
```

type FormatBody fmt :: Type type FormatErrors fmt :: Type





#### **Formats**

The view over some information:

```
class (...) => ValidFormat (fmt :: Type) where
  type FormatErrors fmt :: Type
  type FormatErrors fmt :: Type
```

```
parser :: ByteString -> Either (FormatErrors fmt) (FormatBody fmt)
```





### **Formats: Example**

3





```
data Settings0 = Settings0 {
    reasonable_languages :: String,
    type_system_and_eval :: TSEOpts
} deriving (...)

data F0
class ValidFormat F0 where
  type FormatBody F0 = Settings0
  type FormatErrors F0 = ()
```





```
data Settings0 = Settings0 {
    reasonable_languages :: String,
    type_system_and_eval :: TSEOpts
} deriving (...)

data F0
class ValidFormat F0 where
    type FormatBody F0 = Settings0
    type FormatErrors F0 = ()

parser = parseFromJSON
```





```
data Settings0 = Settings0 {
    reasonable_languages :: String,
    type_system_and_eval :: TSEOpts
} deriving (...)

data F0
class ValidFormat F0 where
    type FormatBody F0 = Settings0
    type FormatErrors F0 = ()

parser = parseFromJSON
```

```
data Settings1 = Settings1 {
    reasonable_languages :: String,
    type_system :: TSOpts,
    eval :: EOpts
} deriving (...)

data F1
class ValidFormat F1 where
    type FormatBody F1 = Settings1
    type FormatErrors F1 = ()
```



```
data Settings0 = Settings0 {
    reasonable_languages :: String,
    type_system_and_eval :: TSEOpts
  } deriving (...)
data F0
class ValidFormat F0 where
  type FormatBody F0 = Settings0
  type FormatErrors F0 = ()
  parser = parseFromJSON
data MvSettings
class Resource MySettings where
  type Formats MySettings = '[ F0. F1 ]
```

```
data Settings1 = Settings1 {
    reasonable_languages :: String,
    type_system :: TSOpts,
    eval :: EOpts
} deriving (...)

data F1
class ValidFormat F1 where
    type FormatBody F1 = Settings1
    type FormatErrors F1 = ()
```







Someone comes along and adds F2: data F2





Someone comes along and adds F2:

```
data F2 -- new format, added without 'ValidFormat' instance !
```





```
Someone comes along and adds F2:
data F2 -- new format, added without 'ValidFormat' instance !!
instance Resource MySettings where
  type Formats MyFormats = '[ F0, F1, F2 ]
```





22 | instance Resource 'MySettings where

```
Someone comes along and adds F2:
data F2 -- new format, added without 'ValidFormat' instance !!
instance Resource MySettings where
  type Formats MyFormats = '[ F0, F1, F2 ]
your/path/to/Module.hs:42:22: error:
   • No instance for (ValidFormat F2)
        arising from the superclasses of an instance declaration
   • In the instance declaration for Resource 'MySettings
```



Receiving settings from A or B: which format?





Receiving settings from A or B: which format? Whichever was sent! Receive a (Int, ByteString),





Receiving settings from A or B: which format? Whichever was sent! Receive a (Int, ByteString),





Receiving settings from A or B: which format? Whichever was sent! Receive a (Int, ByteString),





Receiving settings from A or B: which format? Whichever was sent! Receive a (Int, ByteString),

- Reify fmtNum to the type-level
- Lookup the correct ValidFormat instance











Cleared the problem: serve different views of the "same" information





\_\_

Cleared the problem: serve different views of the "same" information

To do so: Defined formats and resources





Cleared the problem: serve different views of the "same" information

To do so: Defined formats and resources

Accept either format F0 or F1





Cleared the problem: serve different views of the "same" information

To do so: Defined formats and resources

Accept either format F0 or F1

Next up: Ability to migrate between F0 and F1





-







reasonable\_programming\_languages="functional"
type\_system\_and\_eval=strong\_and\_lazy

#### Should become:

reasonable\_programming\_languages="functional"
type\_system=strong
eval=lazy





Example

```
splitSettings :: TSEOpts -> (TSOpts, EOpts)

upgrade :: FormatBody F0 -> FormatBody F1

upgrade Settings0 {..}

= let (type_system, eval) = splitSettigs type_system_and_eval
    in Settings1 {..}
```

















Representing Migrations

- Composable: Migration b c -> Migration a b -> Migration a c.



- Composable: Migration b c -> Migration a b -> Migration a c.
- Can represent absence of migrations:
   Migration (const Nothing) (const Nothing).



Representing Migrations

Composable: Migration b c -> Migration a b -> Migration a c.

- Can represent *absence* of migrations:
  - Migration (const Nothing) (const Nothing).
- Programmer has to be explicit.



Keeping Track of Migrations





Keeping Track of Migrations

```
instance Resource 'MySettings where
  type Formats 'MySettings = [F0, F1, F2, F3]
```





Keeping Track of Migrations

```
instance Resource 'MySettings where
  type Formats 'MySettings = [F0, F1, F2, F3]
```

### Should be three migrations:

```
m01 :: Migration F0 F1
m12 :: Migration F1 F2
m23 :: Migration F2 F3
```





Keeping Track of Migrations

```
instance Resource 'MySettings where
  type Formats 'MySettings = [F0, F1, F2, F3]
```

Should be three migrations:

m01 :: Migration F0 F1
m12 :: Migration F1 F2
m23 :: Migration F2 F3

In general, for a list of formats  $[f_1, \dots, f_n]$ , we need:

(Migration  $f_1 f_2, \dots, Migration f_{n-1} f_n$ )





Keeping Track of Migrations





Keeping Track of Migrations

```
data Pairwise (f :: k -> k -> Type) (ns :: [k]) :: Type where
PNil :: Pairwise f '[k]
(:::) :: f a b -> Pairwise f (b ': ks) -> Pairwise f (a ': b ': ks)
```

/(





Keeping Track of Migrations

```
data Pairwise (f :: k -> k -> Type) (ns :: [k]) :: Type where
  PNil :: Pairwise f '[k]
  (:::) :: f a b -> Pairwise f (b ': ks) -> Pairwise f (a ': b ': ks)

class Resource r where
  type Formats r :: '[ Type ]
```

migrations :: Pairwise Migration (Formats r)





Computing a Migration Path



# **Migrations**

Computing a Migration Path

Say A wants settings in format F0, but storage has F2.



# **Migrations**

Computing a Migration Path

Say A wants settings in format F0, but storage has F2. We can try!

```
p = d10 <=< d21 :: FormatBody F2 -> Maybe (FormatBody F0)
```



Computing a Migration Path

foldSection





Computing a Migration Path

```
foldSection :: (Monoid a) => Int -> Int -> [a] -> a
```



🐴 channable

```
foldSection :: (Monoid a) => Int -> Int -> [a] -> a
foldSection offset len =
  foldr mappend mempty . take len . drop offset
```



Computing a Migration Path

```
foldSection :: (Monoid a) => Int -> Int -> [a] -> a
foldSection offset len =
  foldr mappend mempty . take len . drop offset
```

foldSection 2 3 [10, 11, 12, 13, 14, 15] :: Sum Int





```
foldSection :: (Monoid a) => Int -> Int -> [a] -> a
foldSection offset len =
  foldr mappend mempty . take len . drop offset

foldSection 2 3 [10, 11, 12, 13, 14, 15] :: Sum Int
  == foldr mappend mempty (take 3 (drop 2 [10, 11, 12, 13, 14, 15]))
```





Computing a Migration Path

```
foldSection :: (Monoid a) => Int -> Int -> [a] -> a
foldSection offset len =
  foldr mappend mempty . take len . drop offset

foldSection 2 3 [10, 11, 12, 13, 14, 15] :: Sum Int
  == foldr mappend mempty (take 3 (drop 2 [10, 11, 12, 13, 14, 15]))
  == foldr mappend mempty (take 3 [12, 13, 14, 15])
```



```
foldSection :: (Monoid a) => Int -> Int -> [a] -> a
foldSection offset len =
  foldr mappend mempty . take len . drop offset

foldSection 2 3 [10, 11, 12, 13, 14, 15] :: Sum Int
  == foldr mappend mempty (take 3 (drop 2 [10, 11, 12, 13, 14, 15]))
  == foldr mappend mempty (take 3 [12, 13, 14, 15])
  == foldr mappend mempty [12, 13, 14]
```





```
foldSection :: (Monoid a) => Int -> Int -> [a] -> a
foldSection offset len =
  foldr mappend mempty . take len . drop offset
foldSection 2 3 [10, 11, 12, 13, 14, 15] :: Sum Int
  == foldr mappend mempty (take 3 (drop 2 [10, 11, 12, 13, 14, 15]))
  == foldr mappend mempty (take 3 [12, 13, 14, 15])
  == foldr mappend mempty [12, 13, 14]
 == 12 + (13 + (14 + 0))
p = let path :: Migration F0 F3
        path = pairwiseFoldrSection1 comp 0 2 (migrations @'MySettings)
    in downstream path
```



Computing a Migration Path: The Ugly Bit

pairwiseFoldSection







Computing a Migration Path: The Ugly Bit

pairwiseFoldSection1

8.





Computing a Migration Path: The Ugly Bit

pairwiseFoldSection1

:: (forall 
$$x y z$$
.  $f x y \rightarrow f y z \rightarrow f x z$ )





Computing a Migration Path: The Ugly Bit

#### pairwiseFoldSection1

```
:: (forall x y z. f x y -> f y z -> f x z)
```

- -> SNat offset
- -> SNat (S len)



-

Computing a Migration Path: The Ugly Bit

### pairwiseFoldSection1

```
:: (forall x y z. f x y -> f y z -> f x z)
```

- -> **SNat** offset
- -> SNat (S len) -- Trick: non-empty section !





Computing a Migration Path: The Ugly Bit

```
pairwiseFoldSection1
```

```
:: (forall x y z. f x y \rightarrow f y z \rightarrow f x z)
```

- -> **SNat** offset
- -> SNat (S len) -- Trick: non-empty section !!
- -> Pairwise f ks





Computing a Migration Path: The Ugly Bit

```
pairwiseFoldSection1
:: (forall x y z. f x y -> f y z -> f x z)
-> SNat offset
-> SNat (S len) -- Trick: non-empty section !!
-> Pairwise f ks
-> f (Lookup ks offset) (Lookup ks (offset :+: S len))
```





Computing a Migration Path: The Ugly Bit

```
pairwiseFoldSection1
```

```
:: (forall x y z. f x y \rightarrow f y z \rightarrow f x z)
```

- -> SNat offset
- -> SNat (S len) -- Trick: non-empty section !!
- -> Pairwise f ks
- -> f (Lookup ks offset) (Lookup ks (offset :+: S len))

#### Not too different from:

```
foldSection :: (Monoid a) => Int -> Int -> [a] -> a
```





#### Pros:

- Cannot forget a migraton.





#### Pros:

- Cannot forget a migraton. Even if it's (const Nothing).

9.





#### Pros:

- Cannot forget a migraton. Even if it's (const Nothing).
- Centralized handling of formats:





#### Pros:

- Cannot forget a migraton. Even if it's (const Nothing).
- Centralized handling of formats: A and B can request different formats; helpful for no-downtime deploys.





#### Pros:

- Cannot forget a migraton. Even if it's (const Nothing).
- Centralized handling of formats: A and B can request different formats; helpful for no-downtime deploys.
- Handling of formats is independent of the Resource r





#### Pros:

- Cannot forget a migraton. Even if it's (const Nothing).
- Centralized handling of formats: A and B can request different formats; helpful for no-downtime deploys.
- Handling of formats is independent of the Resource r
   Cons:
  - Non-trivial amount of type-level programming.



#### Pros:

- Cannot forget a migraton. Even if it's (const Nothing).
- Centralized handling of formats: A and B can request different formats; helpful for no-downtime deploys.
- Handling of formats is independent of the Resource r
   Cons:
  - Non-trivial amount of type-level programming.
  - Only support linear migration paths





#### Pros:

- Cannot forget a migraton. Even if it's (const Nothing).
- Centralized handling of formats: A and B can request different formats; helpful for no-downtime deploys.
- Handling of formats is independent of the Resource r
   Cons:
  - Non-trivial amount of type-level programming.
  - Only support linear migration paths (exercise to the reader?)





# **File-Format Migrations**

on the Type-Level!

Victor Miraldo

January 5, 2024



# Things I didn't show you!





Remember definition of Resource:

```
class All1 ValidFormat (Formats r) => Resource r where
 type Formats r :: '[ Type ]
type family All1 c xs where
 All1 c '[] = TypeError "Need at least one element"
 All1 c \Gamma
 All1 c (x ': xs) = (c x, All1 c xs)
```





# LIES!





For withFormat, need witnesses:

```
class All c xs => All1 c xs where
  witness :: SNat m -> (c (Lookup xs m) => r) -> r
```





```
For withFormat, need witnesses:
```

```
class All c xs => All1 c xs where
  witness :: SNat m -> (c (Lookup xs m) => r) -> r
```

Then, withFormat





```
For withFormat, need witnesses:
```

```
class All c xs => All1 c xs where
  witness :: SNat m -> (c (Lookup xs m) => r) -> r
```

#### Then,

withFormat

:: Resource r

=> Int





```
For withFormat, need witnesses:
class All c xs => All1 c xs where
  witness :: SNat m \rightarrow (c (Lookup xs m) \Rightarrow r) \rightarrow r
Then.
withFormat
  :: Resource r
 => Int
  -> (forall n. (ValidFormat (Lookup (Formats r) n)) => SNat n -> res)
  -> res
withFormat format act
```



```
For withFormat, need witnesses:
class All c xs => All1 c xs where
  witness :: SNat m \rightarrow (c (Lookup xs m) \Rightarrow r) \rightarrow r
Then.
withFormat
  :: Resource r
 => Int
  -> (forall n. (ValidFormat (Lookup (Formats r) n)) => SNat n -> res)
  -> res
withFormat format act =
  withSomeSNat (fromIntegral format) (\sn ->
    witness @ValidFormat @(Formats r) sn (act sn))
```



### Our SNat Datatype

Peano naturals are easier to work with.

```
data N = Z | S N
```

data SNat (n :: N) where

ZZ :: SNat Z

SS :: SNat n -> SNat (S n)



We wanted GHC. TypeNats. SNat, but writing an eliminator is not simple:

```
natElim :: f \emptyset -> (forall k . SNat k -> f (k + 1)) -> SNat n -> f n
```





77.0

We wanted GHC. TypeNats. SNat, but writing an eliminator is not simple:

```
natElim :: f 0 -> (forall k . SNat k -> f (k + 1)) -> SNat n -> f n natElim base ind = aux base ind . unsafeCoerce
```





We wanted GHC. TypeNats. SNat, but writing an eliminator is not simple:

```
natElim :: f 0 -> (forall k . SNat k -> f (k + 1)) -> SNat n -> f n
natElim base ind = aux base ind . unsafeCoerce -- UnsafeSNat is not exposed :/
where
   aux b i (UUnsafeSNat 0) = unsafeCoerce b
   aux b i (UUnsafeSNat n) = unsafeCoerce $ i $ unsafeCoerce (USNat n)
newtype USNat n = UUnsafeSNat Natural -- made to mirror SNat
```





We wanted GHC. TypeNats. SNat, but writing an eliminator is not simple:

```
natElim :: f 0 -> (forall k . SNat k -> f (k + 1)) -> SNat n -> f n
natElim base ind = aux base ind . unsafeCoerce -- UnsafeSNat is not exposed :/
where
   aux b i (UUnsafeSNat 0) = unsafeCoerce b
   aux b i (UUnsafeSNat n) = unsafeCoerce $ i $ unsafeCoerce (USNat n)
newtype USNat n = UUnsafeSNat Natural -- made to mirror SNat
```

We couldn't escape the unsafeCoerce though...



Had to use unsafeCoerce when reifying a run-time **Int8** to tye type-level.

```
withSomeMySNat :: forall rep (r :: TYPE rep). Int8 -> (forall n. SNat n -> r) -> r
withSomeMySNat n k = k @(Any @N) (unsafeMySNat n)
where
unsafeMySNat :: Int8 -> SNat (Any @N)
unsafeMySNat 0 = unsafeCoerce ZZ
unsafeMySNat sn = unsafeCoerce (SS (unsafeMySNat (sn - 1)))
```





Had to use unsafeCoerce when reifying a run-time Int8 to tye type-level.

```
withSomeMySNat :: forall rep (r :: TYPE rep). Int8 -> (forall n. SNat n -> r) -> r
withSomeMySNat n k = k @(Any @N) (unsafeMySNat n)
where
  unsafeMySNat :: Int8 -> SNat (Any @N)
  unsafeMySNat 0 = unsafeCoerce ZZ
  unsafeMySNat sn = unsafeCoerce (SS (unsafeMySNat (sn - 1)))
```

Like GHC. TypeNats.withSomeSNat, needs NOINLINE; We can't use GHC. TypeNats yet, still in base-4.17.



Had to use unsafeCoerce when reifying a run-time **Int8** to tye type-level.

```
withSomeMySNat :: forall rep (r :: TYPE rep). Int8 -> (forall n. SNat n -> r) -> r
withSomeMySNat n k = k @(Any @N) (unsafeMySNat n)
where
unsafeMySNat :: Int8 -> SNat (Any @N)
unsafeMySNat 0 = unsafeCoerce ZZ
unsafeMySNat sn = unsafeCoerce (SS (unsafeMySNat (sn - 1)))
```

Like GHC. TypeNats.withSomeSNat, needs NOINLINE; We can't use GHC. TypeNats yet, still in base-4.17.

No big deal in practice: format integers quite small (< 10).





# **File-Format Migrations**

on the Type-Level!

Victor Miraldo

January 5, 2024

