

Information Architecture (IA) Change Process

How to make data changes clear



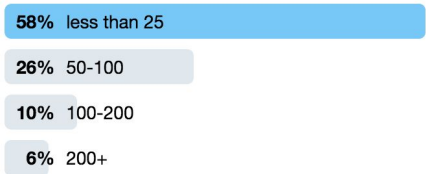
Why?



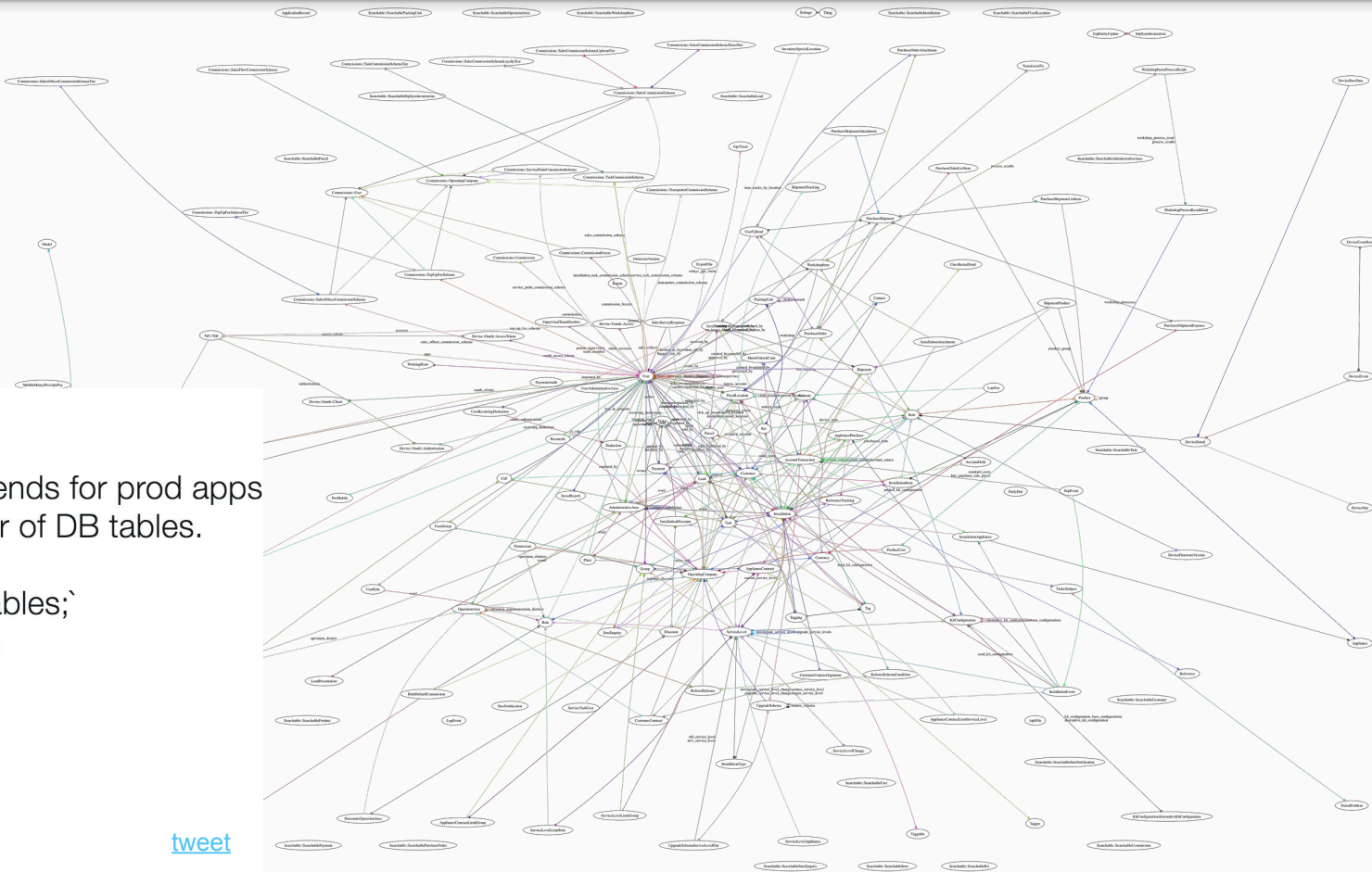
Dan Mayer
@danmayer

Quick survey [#developer](#) friends for prod apps you work on what is number of DB tables.

```
`select count(*)  
from information_schema.tables;`
```



31 votes • Final results



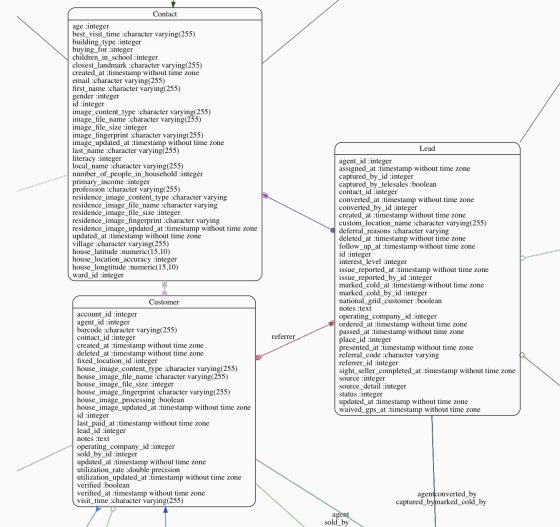
Why?

- As DB complexity grows it's more difficult to make a correct schema.
 - Particularly taking into account multiple aspects of the business
- Many DBs have legacy tables & fields that can further confuse conversations around data
- As DBs grow teams need an easy way to focus in on only the models under discussion
- Larger changes to a DB require understanding & agreement across
 - Product Management
 - Data (BI & DI)
 - Software
 - & often Operations / Stakeholders
- Teams need communication tools to show data change & intents
- A change process that allows for incremental change during discussion

How?

- Integrated into development & PR flow
- Supports Async discussion & feedback
- Both visual & textual diffs
 - DOT <-> PNG
- Faster to express than full code change
 - DB changes don't actually need to be applied for visualization
 - Migrations don't need to be written or run
- BI & PMs can edit DOT files to express intents
- Fast iteration with images can be updated during live workgroup discussions

```
1 digraph models_diagram {
2   graph[overlap=false, splines=true, bgcolor="none"]
3   "Customer" [shape=Record, label="{Customer}account_id :integer\l
4   agent_id :integer\l
5   barcode :character varying(255)\l
6   contact_id :integer\l
7   created_at :timestamp without time zone\l
8   deleted_at :timestamp without time zone\l
9   email :character varying(255)\l
10  first_name :character varying\l
11  fixed_location_id :integer\l
12  gender :integer\l
13  house_image_content_type :character varying(255)\l
14  house_image_file_name :character varying(255)\l
15  house_image_file_size :integer\l
16  house_image_fingerprint :character varying(255)\l
17  house_image_processing :boolean\l
18  house_image_updated_at :timestamp without time zone\l
19  house_latitude :numeric(15,10)\l
```



When to follow new IA change process

For large data changes, not for small schema changes!

DB Change	Follow IA Change Process?
Delete a completely deprecated table	No
Removing completely deprecated fields	No
Adding small table for internal use not needed for reporting	No
Adding a few additional fields	Probably, use good judgement
Refactor: moving actively used columns to a different table	Yes
Adding complex tables that represent core business data	Yes
Refactor: renaming and replacing active used columns	Yes

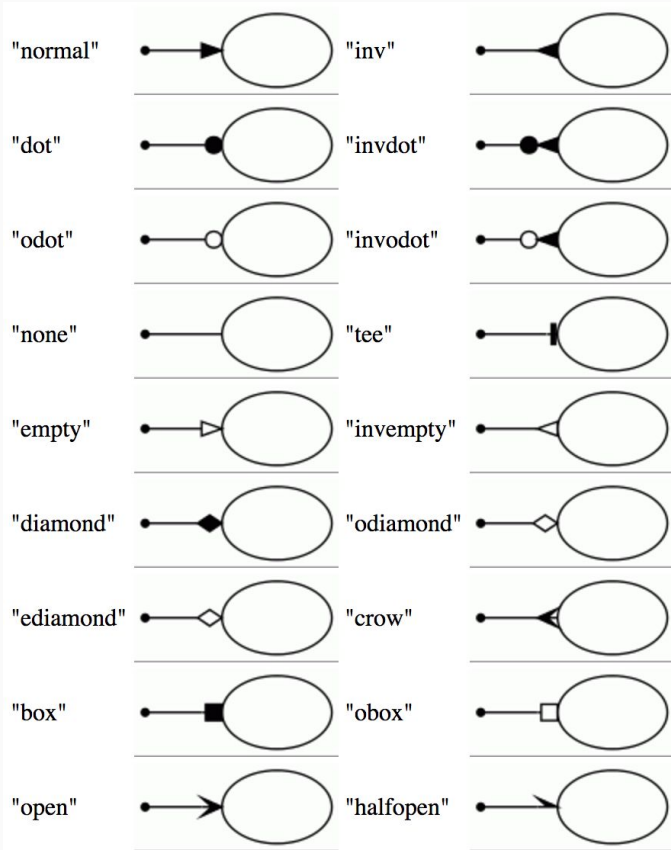
What is a dot file?

- [DOT \(graph description language\)](#)
 - [Node, Edge and Graph Attributes](#)
 - [Basic of UML with DOT](#)
 - Connections
 - [Dir](#)
 - [Arrows](#)
- Easy to generate from Rails Model files
- Easy to generate in multiple other frameworks
- Easy to build by hand
- Easy to “zoom” in on focus areas

```
1 digraph models_diagram {
2   graph[overlap=false, splines=true, bgcolor="none"]
3   "Customer" [shape=Mrecord, label="{Customer|account_id :integer\\l
4   agent_id :integer\\l
5   barcode :character varying(255)\\l
6   contact_id :integer\\l
7   created_at :timestamp without time zone\\l
8   deleted_at :timestamp without time zone\\l
9   email :character varying(255)\\l
10  first_name :character varying\\l
11  fixed_location_id :integer\\l
12  gender :integer\\l
13  house_image_content_type :character varying(255)\\l
14  house_image_file_name :character varying(255)\\l
15  house_image_file_size :integer\\l
16  house_image_fingerprint :character varying(255)\\l
17  house_image_processing :boolean\\l
18  house_image_updated_at :timestamp without time zone\\l
19  house_latitude :numeric(15,10)\\l
```

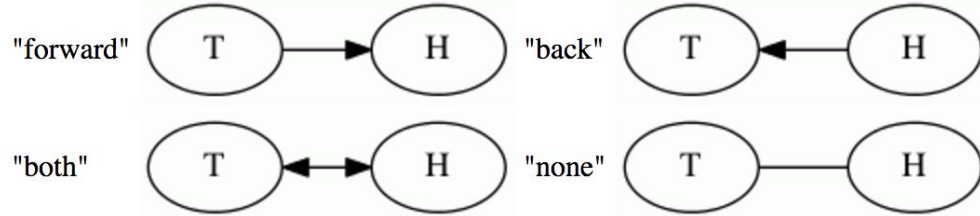
What is a dot file?

Arrows (we need agreed on legend 1:1, 1:Many)



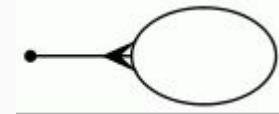
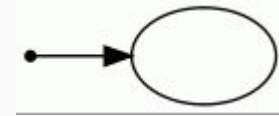
Dir

For an edge $T \rightarrow H$;

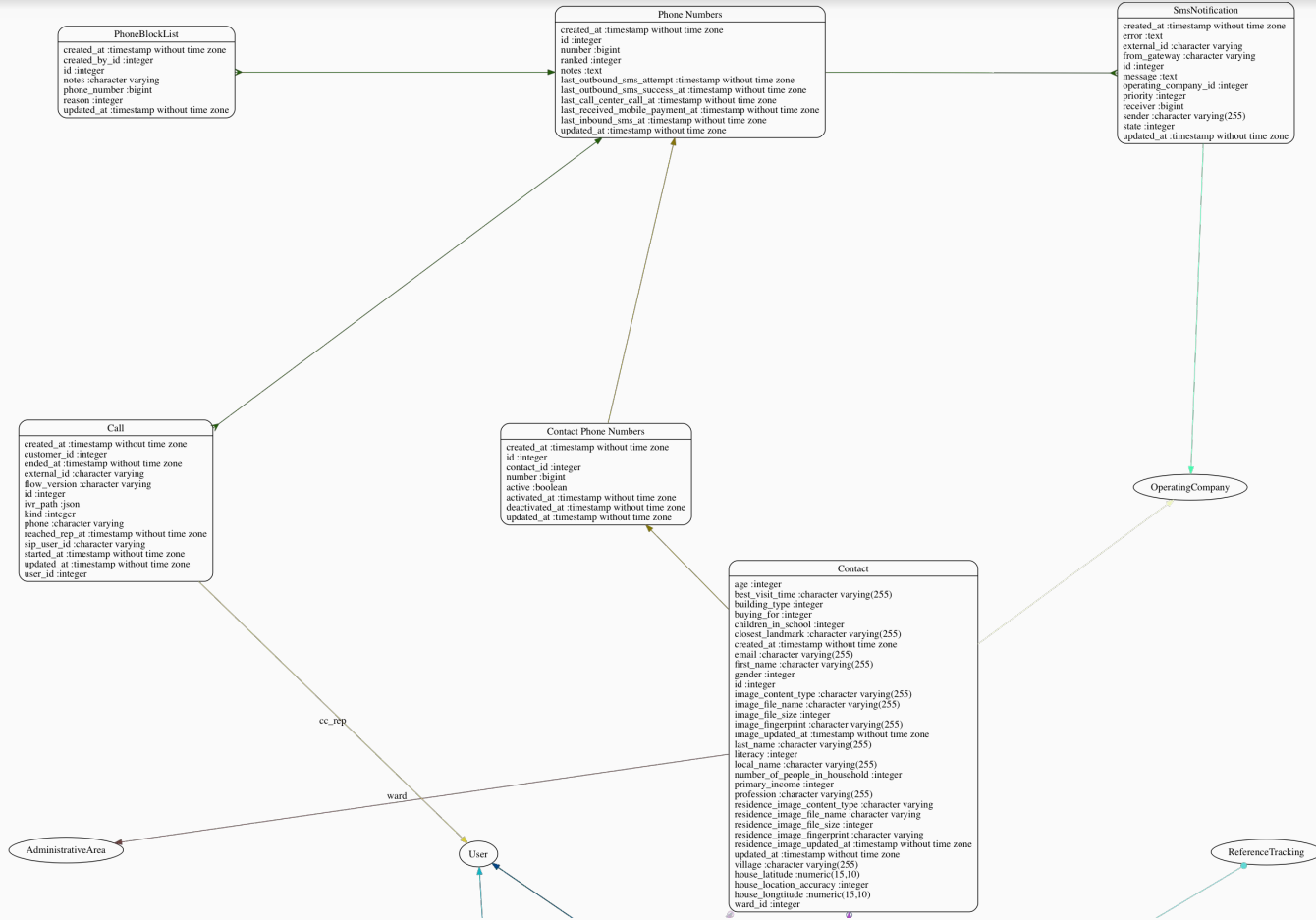


Proposed Legend

- 1:1 -> "normal"
- 1:Many -> "crow"



What is a dot file? Text Input -> Image Diagram Output



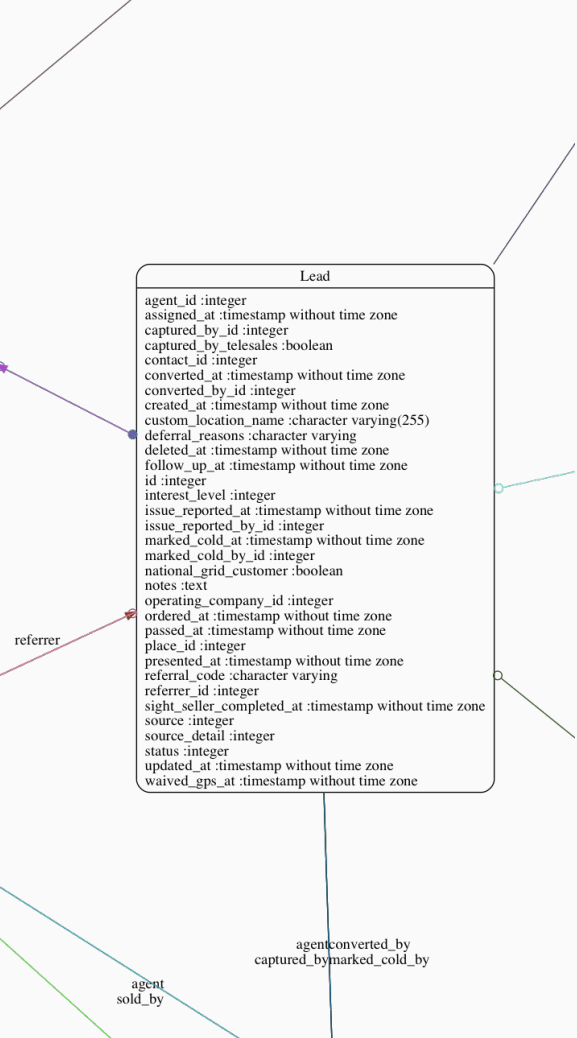
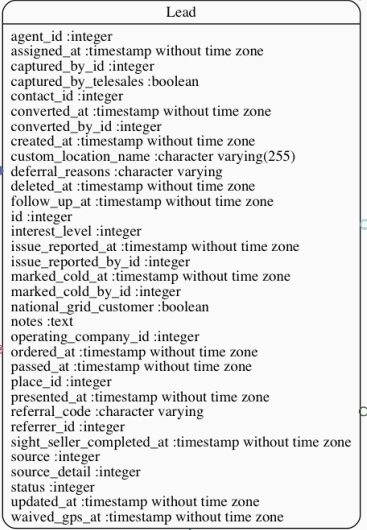
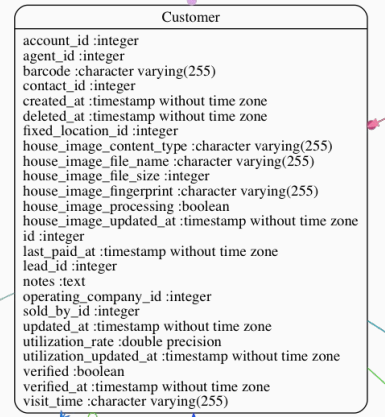
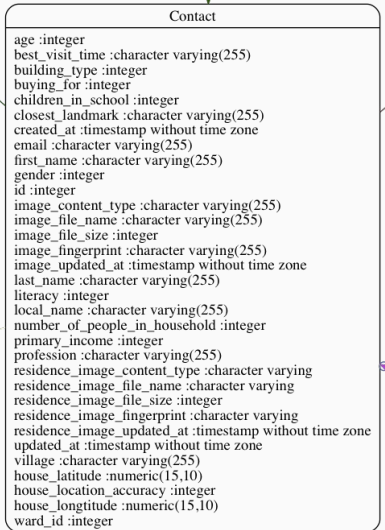
Example: Contact Cleanup

Refactoring data, attributes, & associations of Contact Models



Textual Diff

```
60 doc/customers.dot View
@@ -6,54 +6,33 @@ barcode :character varying(255)\l
6 6 contact_id :integer\l
7 7 created_at :timestamp without time zone\l
8 8 deleted_at :timestamp without time zone\l
9 -email :character varying(255)\l
10 -first_name :character varying\l
11 0 fixed_location_id :integer\l
12 -gender :integer\l
13 10 house_image_content_type :character varying(255)\l
14 11 house_image_file_name :character varying(255)\l
15 12 house_image_file_size :integer\l
16 13 house_image_fingerprint :character varying(255)\l
17 14 house_image_processing :boolean\l
18 15 house_image_updated_at :timestamp without time zone\l
19 -house_latitude :numeric(15,10)\l
20 -house_location_accuracy :integer\l
21 -house_longitude :numeric(15,10)\l
22 16 id :integer\l
23 -image_content_type :character varying(255)\l
24 -image_file_name :character varying(255)\l
25 -image_file_size :integer\l
26 -image_fingerprint :character varying(255)\l
27 -image_processing :boolean\l
28 -image_updated_at :timestamp without time zone\l
29 -landmark :character varying(255)\l
30 -last_name :character varying\l
31 17 last_paid_at :timestamp without time zone\l
32 18 lead_id :integer\l
33 -local_name :character varying(255)\l
34 19 notes :text\l
35 20 operating_company_id :integer\l
36 -primary_mobile :bigint\l
37 -profession :character varying(255)\l
38 -secondary_mobile :bigint\l
39 21 sold_by_id :integer\l
40 -spouse_mobile :bigint\l
41 22 updated_at :timestamp without time zone\l
42 23 utilization_rate :double precision\l
43 24 utilization_updated_at :timestamp without time zone\l
44 25 verified :boolean\l
45 26 verified_at :timestamp without time zone\l
46 -village :character varying(255)\l
47 -visit_time :character varying(255)\l
48 -ward_id :integer\l
49 -~*]
27 ~*visit_time :character varying(255)\l~*]
50 28 "Contact" [shape=Record, label="(Contact)age :integer\l
51 29 best_visit_time :character varying(255)\l
52 30 building_type :integer\l
53 31 buying_for :integer\l
54 32 children_in_school :integer\l
55 33 closest_landmark :character varying(255)\l
56 34 created_at :timestamp without time zone\l
57 35 email :character varying(255)\l
58 36 first_name :character varying(255)\l
59 37 gender :integer\l
60 38 id :integer\l
@@ -67,49 +46,36 @@ literacy :integer\l
67 46 local_name :character varying(255)\l
68 47 number_of_people_in_household :integer\l
69 48 primary_income :integer\l
100 -house_location_accuracy :integer\l
101 -house_longitude :numeric(15,10)\l
102 73 id :integer\l
103 74 interest_level :integer\l
104 75 issue_reported_at :timestamp without time zone\l
105 76 issue_reported_by_id :integer\l
106 -last_name :character varying(255)\l
107 -latitude :numeric(15,10)\l
108 -local_name :character varying(255)\l
109 -longitude :numeric(15,10)\l
110 77 marked_cold_at :timestamp without time zone\l
111 78 marked_cold_by_id :integer\l
112 -mobile :character varying(255)\l
113 79 national_grid_customer :boolean\l
114 80 notes :text\l
115 81 operating_company_id :integer\l
@@ -119,16 +85,12 @@ place_id :integer\l
119 85 presented_at :timestamp without time zone\l
120 86 referral_code :character varying\l
121 87 referrer_id :integer\l
122 -service_level_id :integer\l
123 88 sight_seller_completed_at :timestamp without time zone\l
124 89 source :integer\l
125 90 source_detail :integer\l
126 91 status :integer\l
127 92 updated_at :timestamp without time zone\l
128 -village :character varying(255)\l
129 -waived_gps_at :timestamp without time zone\l
130 -ward_id :integer\l
131 -~*]
93 ~*waived_gps_at :timestamp without time zone\l~*]
132 94 "SmsNotification" [shape=Record, label="(SmsNotification)created_at :timestamp without time zone\l
133 95 error :text\l
134 96 external_id :character varying\l
@@ -165,9 +127,7 @@ started_at :timestamp without time zone\l
165 127 updated_at :timestamp without time zone\l
166 128 user_id :integer\l
167 129 ~*]
168 - "Customer" -> "OperatingCompany" [arrowtail=none, arrowhead=normal, dir=both color="#E9FF63"]
169 130 "Customer" -> "FixedLocation" [arrowtail=none, arrowhead=normal, dir=both color="#6998B0"]
170 - "Customer" -> "AdministrativeArea" [label="ward", arrowtail=none, arrowhead=normal, dir=both color=
171 131 "Customer" -> "Account" [arrowtail=none, arrowhead=normal, dir=both color="#37DF17"]
172 132 "Customer" -> "User" [label="agent", arrowtail=none, arrowhead=normal, dir=both color="#82A8CA"]
173 133 "Customer" -> "User" [label="sold_by", arrowtail=none, arrowhead=normal, dir=both color="#3164C3"]
@@ -177,15 +137,13 @@ user_id :integer\l
177 137 "Customer" -> "InstallationItem" [arrowtail=crow, arrowhead=crow, dir=both color="#2179E1"]
178 138 "Contact" -> "Lead" [arrowtail=odot, arrowhead=dot, dir=both color="#526D98"]
179 139 "Contact" -> "Customer" [arrowtail=odot, arrowhead=dot, dir=both color="#DA9E33"]
180 - "Lead" -> "OperatingCompany" [arrowtail=none, arrowhead=normal, dir=both color="#EABAC0"]
181 + "Contact" -> "OperatingCompany" [arrowtail=none, arrowhead=normal, dir=both color="#E9FF63"]
182 + "Contact" -> "AdministrativeArea" [label="ward", arrowtail=none, arrowhead=normal, dir=both color=
181 142 "Lead" -> "ReferenceTracking" [arrowtail=odot, arrowhead=dot, dir=both color="#51D7D6"]
182 143 "Lead" -> "Customer" [arrowtail=dot, arrowhead=dot, dir=both color="#984680"]
183 - "Lead" -> "Installation" [arrowtail=odot, arrowhead=dot, dir=both color="#9D7499"]
184 144 "Lead" -> "Place" [arrowtail=none, arrowhead=normal, dir=both color="#4C356C"]
185 145 "Lead" -> "Contact" [arrowtail=none, arrowhead=normal, dir=both color="#A4CC77"]
186 146 "Lead" -> "User" [label="agent", arrowtail=none, arrowhead=normal, dir=both color="#925AE9"]
187 - "Lead" -> "Group" [arrowtail=none, arrowhead=normal, dir=both color="#865878"]
188 - "Lead" -> "AdministrativeArea" [label="ward", arrowtail=none, arrowhead=normal, dir=both color="#69
189 147 "Lead" -> "User" [label="captured_by", arrowtail=none, arrowhead=normal, dir=both color="#9398AE"]
190 148 "Lead" -> "User" [label="converted_by", arrowtail=none, arrowhead=normal, dir=both color="#06CB2F"]
191 149 "Lead" -> "User" [label="marked_cold_by", arrowtail=none, arrowhead=normal, dir=both color="#074481]
```

ZOOMED IN!

Refactoring out Phone Numbers

```
21 doc/customers.dot View
@@ -46,15 +46,12 @@ literacy :integer\
46 46 local_name :character varying(255)\
47 47 number_of_people_in_household :integer\
48 48 primary_income :integer\
49 -primary_mobile :bigint\
50 49 profession :character varying(255)\
51 50 residence_image_content_type :character varying\
52 51 residence_image_file_name :character varying\
53 52 residence_image_file_size :integer\
54 53 residence_image_fingerprint :character varying\
55 54 residence_image_updated_at :timestamp without time zone\
56 -secondary_mobile :bigint\
57 -third_mobile :bigint\
58 55 updated_at :timestamp without time zone\
59 56 village :character varying(255)\
60 57 house_latitude :numeric(15,10)\
@@ -94,6 +91,18 @@ source_detail :integer\
94 91 status :integer\
95 92 updated_at :timestamp without time zone\
96 93 waived_gps_at :timestamp without time zone\
94 + "PhoneNumber" [shape=Mrecord, label="(Phone Numbers)created_at :timestamp without time zone\
95 +id :integer\
96 +contact_id :integer\
97 +number :bigint\
98 +ranked :integer\
99 +notes :text\
100 +last_outbound_sms_attempt :timestamp without time zone\
101 +last_outbound_sms_success_at :timestamp without time zone\
102 +last_call_center_call_at :timestamp without time zone\
103 +last_received_mobile_payment_at :timestamp without time zone\
104 +last_inbound_sms_at :timestamp without time zone\
105 +updated_at :timestamp without time zone\]"
97 106 "SmsNotification" [shape=Mrecord, label="(SmsNotification)created_at :timestamp without time zone\
98 107 error :text\
99 108 external_id :character varying\
@@ -142,6 +151,7 @@ user_id :integer\
142 151 "Contact" -> "Customer" [arrowtail=odot, arrowhead=dot, dir=both color="#DA9E3"]
143 152 "Contact" -> "OperatingCompany" [arrowtail=none, arrowhead=normal, dir=both color="#EFF6C3"]
144 153 "Contact" -> "AdministrativeArea" [label="ward", arrowtail=none, arrowhead=normal, dir=both color=
154 + "Contact" -> "PhoneNumber" [arrowtail=crow, arrowhead=none, dir=both color="#24580C"]
145 155 "Lead" -> "ReferenceTracking" [arrowtail=odot, arrowhead=dot, dir=both color="#61D706"]
146 156 "Lead" -> "Customer" [arrowtail=odot, arrowhead=dot, dir=both color="#984680"]
147 157 "Lead" -> "Place" [arrowtail=none, arrowhead=normal, dir=both color="#4C356"]
@@ -153,9 +163,10 @@ user_id :integer\
153 163 "Lead" -> "Customer" [label="referrer", arrowtail=none, arrowhead=normal, dir=both color="#EF7CA1"]
154 164 "Lead" -> "LeadPresentation" [arrowtail=odot, arrowhead=crow, dir=both color="#24580C"]
155 165 "SmsNotification" -> "OperatingCompany" [arrowtail=none, arrowhead=normal, dir=both color="#49FCAE"]
156 - "SmsNotification" -> "Contact" [arrowtail=crow, arrowhead=none, dir=both color="#24580C"]
157 - "PhoneBlockList" -> "Contact" [arrowtail=crow, arrowhead=normal, dir=both color="#24580C"]
166 + "SmsNotification" -> "PhoneNumber" [arrowtail=crow, arrowhead=none, dir=both color="#24580C"]
167 + "PhoneBlockList" -> "PhoneNumber" [arrowtail=crow, arrowhead=normal, dir=both color="#24580C"]
158 168 "Call" -> "Customer" [arrowtail=none, arrowhead=normal, dir=both color="#1B4EEB"]
169 + "Call" -> "PhoneNumber" [arrowtail=crow, arrowhead=normal, dir=both color="#24580C"]
159 170 "Call" -> "User" [arrowtail=none, arrowhead=normal, dir=both color="#3438E8"]
160 171 "Call" -> "User" [label="cc_rep", arrowtail=none, arrowhead=normal, dir=both color="#DC3B34"]
161 172 "Call" -> "IvrStep" [arrowtail=odot, arrowhead=crow, dir=both color="#0C6371"]
```


The Previous Iteration is Wrong

Issues:

- DB currently has multiple users with the same primary number
 - shouldn't be allowed, but exists
- DB intends to have many accounts with secondary numbers that are other's primary numbers.
 - this can & should happen for friends / family accounts
- Previous design would have duplicate phone number records per unique number.
- Therefore we couldn't appropriately link *SMSs* and *Calls* to *PhoneNumbers* without creating multiple records

Improvements

- *Contact* -> *ContactPhoneNumbers* -> *Phonenumber* <- [*SMSs*, *PhoneBlockList*, *Calls*]
- Remove *Contact_Id* and now *PhoneNumbers* is correct can map directly to tables with a number *SMSs*, *Calls*, & *PhoneBlockList*.
- We can now see messages that went to individual users after a phone number changes as we can query between *activate_at* / *deactivated_at*.
- We can also show when one message now goes to many contacts who all are sharing a number (hopefully as secondary numbers)

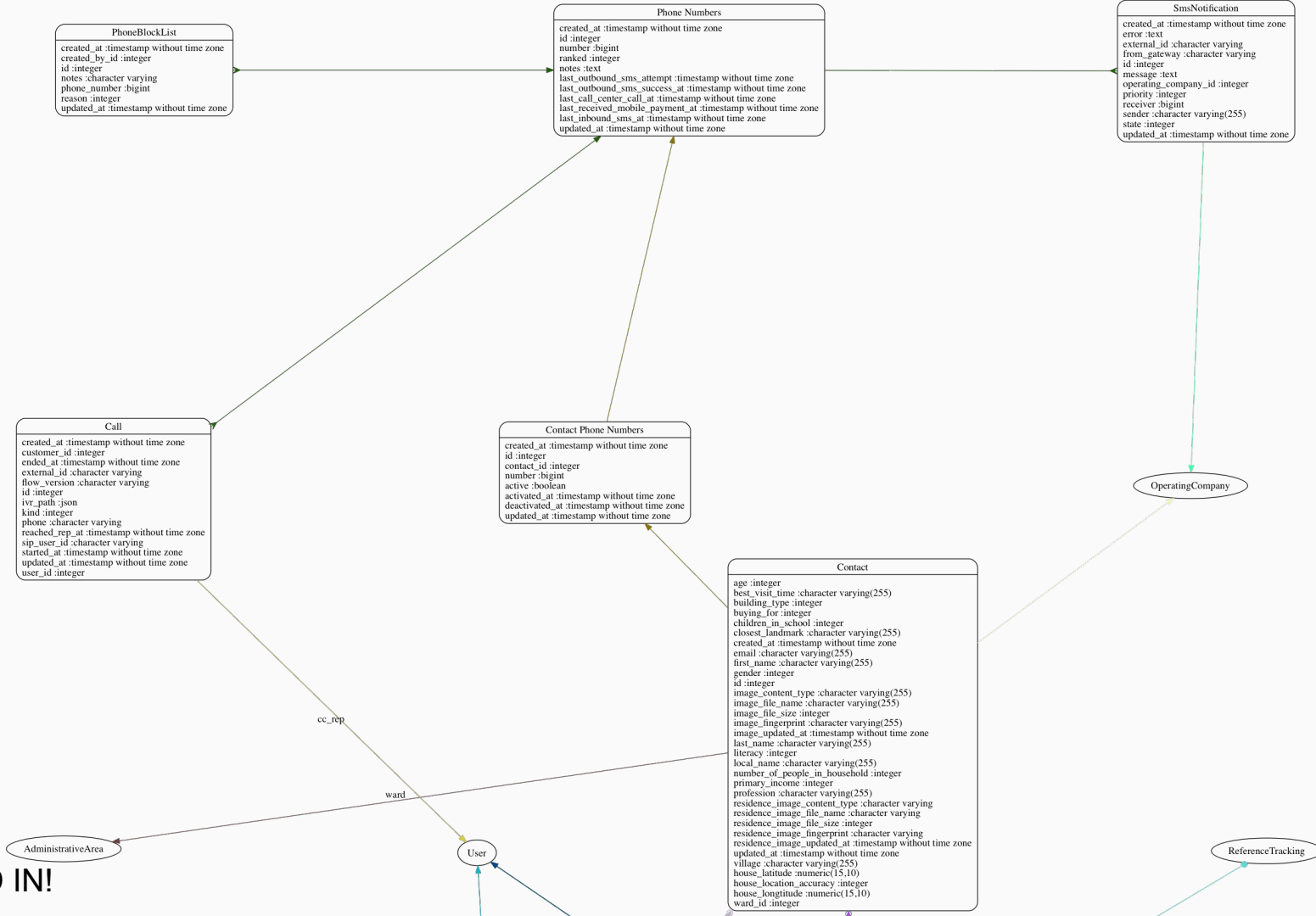
```

@ -93,7 +93,6 @@ updated_at :timestamp without time zone\l
93 93 waived_gps_at :timestamp without time zone\l")
94 94   "PhoneNumber" [shape=Mrecord, label="{Phone Numbers|created_at :timestamp without time zone\l
95 95   id :integer\l
96 96   -contact_id: integer\l
97 96   number :bigint\l
98 97   ranked :integer\l
99 98   notes :text\l

@ -139,6 +138,14 @@ started_at :timestamp without time zone\l
139 138 updated_at :timestamp without time zone\l
140 139 user_id :integer\l
141 140   }"]
141 +["ContactPhoneNumber" [shape=Mrecord, label="{Contact Phone Numbers|created_at :timestamp without time zone\l
142 +id :integer\l
143 +contact_id :integer\l
144 +number :bigint\l
145 +active :boolean\l
146 +activated_at :timestamp without time zone\l
147 +deactivated_at :timestamp without time zone\l
148 +updated_at :timestamp without time zone\l)"]
142 149   "Customer" -> "FixedLocation" [arrowtail=none, arrowhead=normal, dir=both color="#6980B0"]
143 150   "Customer" -> "Account" [arrowtail=none, arrowhead=normal, dir=both color="#37DF17"]
144 151   "Customer" -> "User" [label="agent", arrowtail=none, arrowhead=normal, dir=both color="#82A8CA"]

@ -151,7 +158,8 @@ user_id :integer\l
151 158   "Contact" -> "Customer" [arrowtail=odot, arrowhead=dot, dir=both color="#DAA9E3"]
152 159   "Contact" -> "OperatingCompany" [arrowtail=none, arrowhead=normal, dir=both color="#EFF6C3"]
153 160   "Contact" -> "AdministrativeArea" [label="ward", arrowtail=none, arrowhead=normal, dir=both color="#6A4341"]
154 -   "Contact" -> "PhoneNumber" [arrowtail=crow, arrowhead=none, dir=both color="#24580C"]
161 +   "Contact" -> "ContactPhoneNumber" [arrowtail=none, arrowhead=normal, dir=both color="#857308"]
162 +   "ContactPhoneNumber" -> "PhoneNumber" [arrowtail=none, arrowhead=normal, dir=both color="#857308"]
155 163   "Lead" -> "ReferenceTracking" [arrowtail=odot, arrowhead=dot, dir=both color="#61D7D6"]
156 164   "Lead" -> "Customer" [arrowtail=odot, arrowhead=dot, dir=both color="#8B4680"]
157 165   "Lead" -> "Place" [arrowtail=none, arrowhead=normal, dir=both color="#4C356C"]

```

ZOOMED IN!

Workflow

1. Generate or update initial model on dev branch
 - * generate dot file: ``FOCUS=true VERBOSE=true OUTPUT=doc/customers.dot rake diagram:models:customers``
 - * generate image: ``neato -Tpng doc/customers.dot > doc/customers.png``
 - * check in the un-edited DOT & image files
 - * commit to dev branch
 - * this will serve as your before to diff against later
2. Create a feature branch to work on your changes
 - * new branch: ``git checkout -b feature/db_phone_change``
 - * edit to reflect your changes: ``open doc/customer.dot``
 - * convert updated DOT file to updated image: ``neato -Tpng doc/customer.dot > doc/customers.png``
 - * view PNG & Refactor until happy
 - * create a PR and upload the before and after images, diff will be automatically created for the DOT file changes.

Workflow: Initial Model

- If there is a suitable DOT file use it, if not modify `dev.rake` to create one

```
desc 'Generates an class diagram for all customer models.'
task :customers do
  output = ENV['OUTPUT'] || 'tmp/customers.png'
  customer_ignores = if ENV['FOCUS']
    %w(Tagging AppliancePurchase Tag Ticket InstallationItems Task).join(',')
  else
    ''
  end
  ENV['EXCLUDE_MODELS'] = ENV['EXCLUDE_MODELS'].to_s + customer_ignores
  files = %w(
    app/models/customer.rb
    app/models/contact.rb
    app/models/lead.rb
    app/models/sms_notification.rb
    app/models/phone_block_list.rb
    app/models/call.rb
  )
  generate_for_files(files, output)
end
```

DB Refactoring

The image shows a code editor interface with a search function. The search results are as follows:

File	Line	Code Snippet
customers.dot	157	"Contact" -> "Lead" [arrowtail=odot, arrowhead=dot, dir=both color="#526D9B"]
customers.dot	158	"Contact" -> "Customer" [arrowtail=odot, arrowhead=dot, dir=both color="#DAA9E3"]
customers.dot	159	"Contact" -> "OperatingCompany" [arrowtail=none, arrowhead=normal, dir=both color="#EFF6C3"]
customers.dot	160	"Contact" -> "AdministrativeArea" [label="ward", arrowtail=none, arrowhead=normal, dir=both color="#6A4341"]
customers.dot	161	"Contact" -> "ContactPhoneNumber" [arrowtail=none, arrowhead=normal, dir=both color="#85730B"]
customers.dot	162	"ContactPhoneNumber" -> "PhoneNumber" [arrowtail=none, arrowhead=normal, dir=both color="#85730B"]
customers.dot	163	"Lead" -> "ReferenceTracking" [arrowtail=odot, arrowhead=dot, dir=both color="#61D7D6"]
customers.dot	164	"Lead" -> "Customer" [arrowtail=odot, arrowhead=dot, dir=both color="#B84680"]
customers.dot	165	"Lead" -> "Place" [arrowtail=none, arrowhead=normal, dir=both color="#4C356C"]
customers.dot	166	"Lead" -> "Contact" [arrowtail=none, arrowhead=normal, dir=both color="#AA4CC7"]
customers.dot	167	"Lead" -> "User" [label="agent", arrowtail=none, arrowhead=normal, dir=both color="#025AE9"]
customers.dot	168	"Lead" -> "User" [label="captured_by", arrowtail=none, arrowhead=normal, dir=both color="#9398AE"]
customers.dot	169	"Lead" -> "User" [label="converted_by", arrowtail=none, arrowhead=normal, dir=both color="#06CB2F"]
customers.dot	170	"Lead" -> "User" [label="marked_cold_by", arrowtail=none, arrowhead=normal, dir=both color="#074489"]
customers.dot	171	"Lead" -> "Customer" [label="referrer", arrowtail=none, arrowhead=normal, dir=both color="#EF7CA1"]
customers.dot	172	"Lead" -> "LeadPresentation" [arrowtail=odot, arrowhead=crow, dir=both color="#245B0C"]
customers.dot	173	"SmsNotification" -> "OperatingCompany" [arrowtail=none, arrowhead=normal, dir=both color="#49FCAE"]
customers.dot	174	"SmsNotification" -> "PhoneNumber" [arrowtail=crow, arrowhead=none, dir=both color="#245B0C"]
customers.dot	175	"PhoneBlockList" -> "PhoneNumber" [arrowtail=crow, arrowhead=normal, dir=both color="#245B0C"]
customers.dot	176	"Call" -> "Customer" [arrowtail=none, arrowhead=normal, dir=both color="#1B4EE8"]
customers.dot	177	"Call" -> "PhoneNumber" [arrowtail=crow, arrowhead=normal, dir=both color="#245B0C"]
customers.dot	178	"Call" -> "User" [arrowtail=none, arrowhead=normal, dir=both color="#343BE0"]
customers.dot	179	"Call" -> "User" [label="cc_rep", arrowtail=none, arrowhead=normal, dir=both color="#DCCB34"]
customers.dot	180	"Call" -> "IvrStep" [arrowtail=odot, arrowhead=crow, dir=both color="#DC6371"]
customers.dot	181	}
customers.dot	182	

Search details: 8 results found for 'phonenum'. Finding with Options: Case Insensitive. * A a [icons].

Buttons: Find, Find All, Replace, Replace All.