

ELASTIC @ MAPZEN

geocoding in the open



Google HQ uses goats to
mow their lawn



“Real Fact” #326

On average, you'll
spend a year of
your life looking
for misplaced
objects.

Find more “Real Facts” at snapple.com



DIANA SHKOLNIKOV



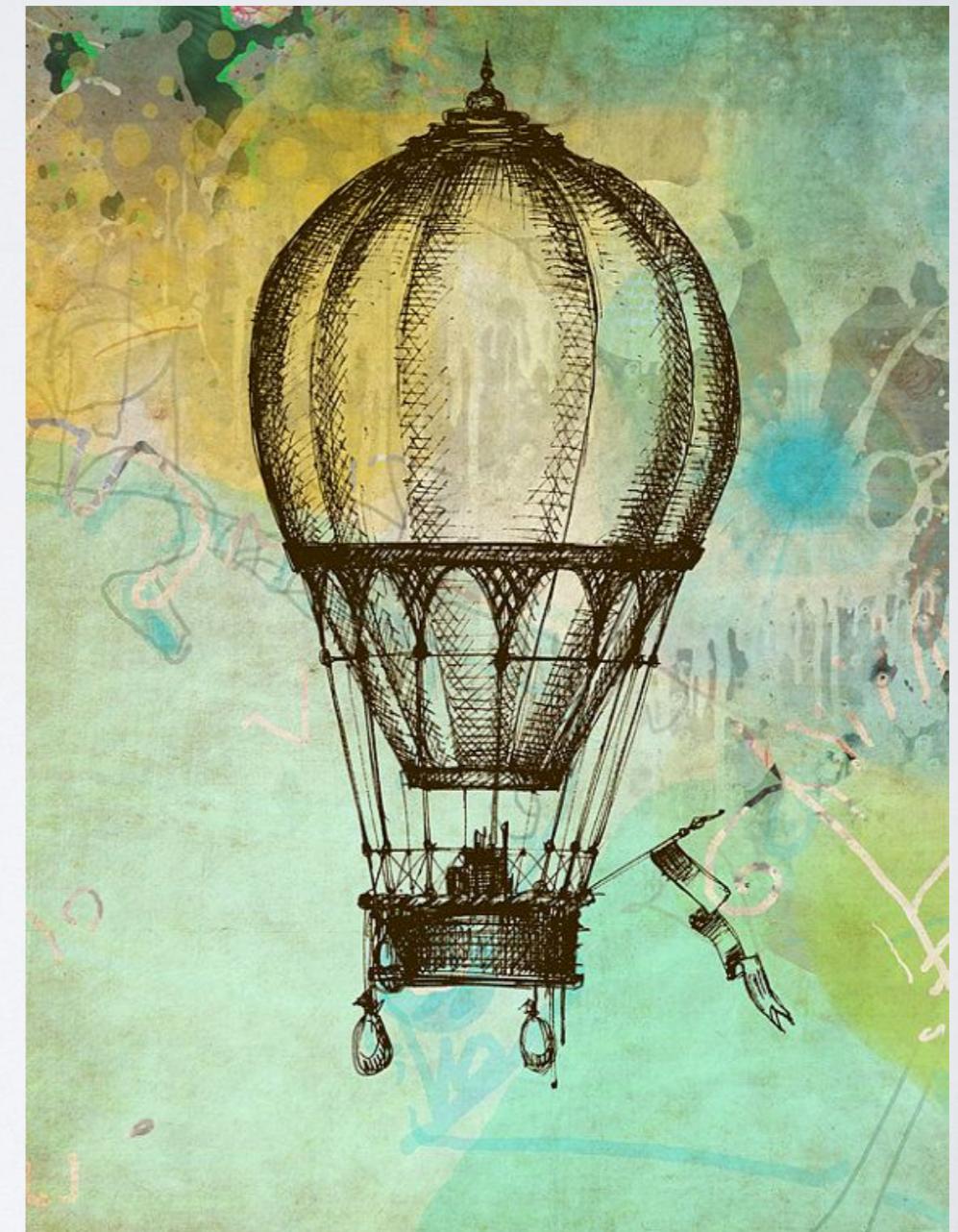
diana@mapzen.com



@dianashk

AGENDA

1. WHAT IS GEOCODING
2. WHAT IS PELIAS
3. PELIAS <3 ELASTIC
 - NGRAMS
 - SHINGLES
 - GEOHASH
4. PELIAS ARCHITECTURE
5. HANDS-ON W/ SENSE





GEOCODING

expectations & obstacles



github.com/pelias



@dianashk

“

Geocoding (sometimes called **forward geocoding**) is the process of enriching a description of a location, most typically a postal address or place name, with geographic coordinates from spatial reference data such as building polygons, land parcels, street addresses, postal codes (e.g. ZIP codes, CEDEX) and so on.

— wikipedia



github.com/pelias



@dianashk

WELCOME HOME

Rent unique places to stay from local hosts in 190+ countries.

How It Works

Philadelphia, PA, United States Check In Check Out 1 Guest Search

Philadelphia PA, United States

Philadelphia County PA, United States

Philadelphia Museum of Art Benjamin Franklin Parkway, Philadelphia, PA, United States

Philmont Avenue Philadelphia, PA, United States

Sign Up Log In Help ▾ List Your Space



github.com/pelias



@dianashk

ENDS TONIGHT! • 20% OFF \$50+ 25% OFF \$100+ 30% OFF \$150+ • USE CODE **BIGRED** BY 11:59PM ET 04/15 (see details)



24/7 CUSTOMER SERVICE: [LIVE CHAT](#) [HELP](#)



FREE SHIPPING ON ORDERS \$75+
OR FLAT \$6.95 U.S. STANDARD SHIPPING ([click here for details](#))

SHOP BY CATEGORY ▾

SHOP BY INTEREST ▾

Search...

HAI, DIANASHK
YOUR ACCOUNT ▾

1 ITEM

WHAT'S NEW

BESTSELLERS

EXCLUSIVES

GIFTS

T-SHIRTS & APPAREL

ELECTRONICS & GADGETS

GIFT CERTIFICATES

ON SALE

GEEKPOINTS

WARP SPEED CHECKOUT!

Billing Address:

Keep typing your address to display more results

100 Union Sq, Randolph, MA

100 Union Sq, Forest City, NC

Country:

100 Union Sq, Milford, NH

First name:

100 Union St S, Kennett Sq, PA

Last name:

100 Union Square Dr, New Hope, PA

Company:

Capture+ Select Country ▾

Address:

100 union sq

Apt/Suite #:

City:

Shipping Address

Same as billing or

Choose a saved address ▾

Country:

United States ▾

* (required)

[Country not listed? Click Here.](#)

First name:

Diana

*

Last name:

Shkolnikov

*

Company:

Address:

*

Apt/Suite #:

*



github.com/pelias



@dianashk

Google Search Calendar +Diana    

Party

4/15/2015 10:30pm to 11:30pm 4/15/2015 Time zone

All day Repeat...

Event details [Find a time](#)

Where Zivte
Zivtech Walnut Street, Philadelphia, PA, United States
Zivtech, LLC Strawberry Street, Philadelphia, PA, United States
Zivtex Ltd Harpenden, United Kingdom

Add guests
Enter email addresses [Add](#)

Guests can
 modify event
 invite others
 see guest list

Event color 

Notifications
Email 1 days 
SMS 1 days 
Pop-up 1 days 
[Add a notification](#)

Show me as Available Busy

Visibility Calendar default Public Private

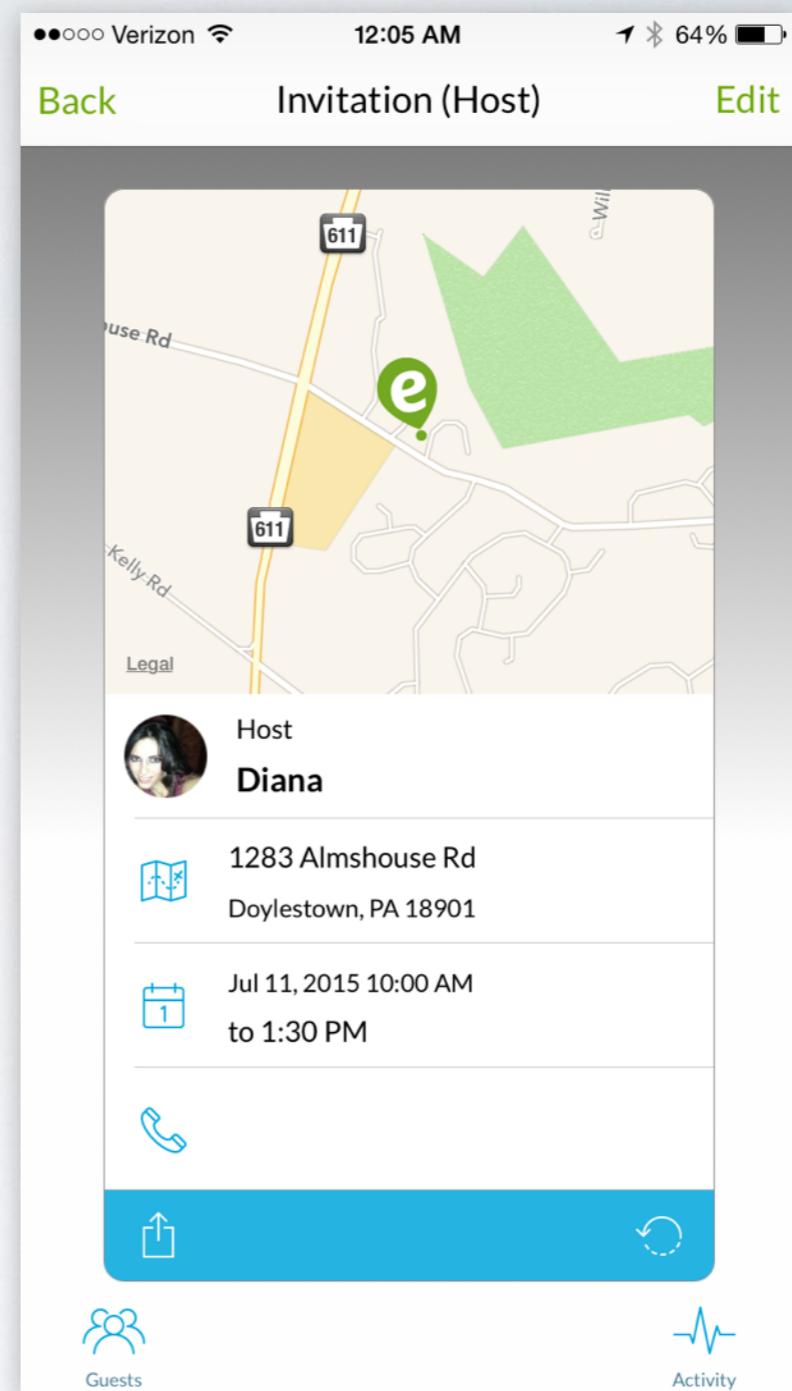
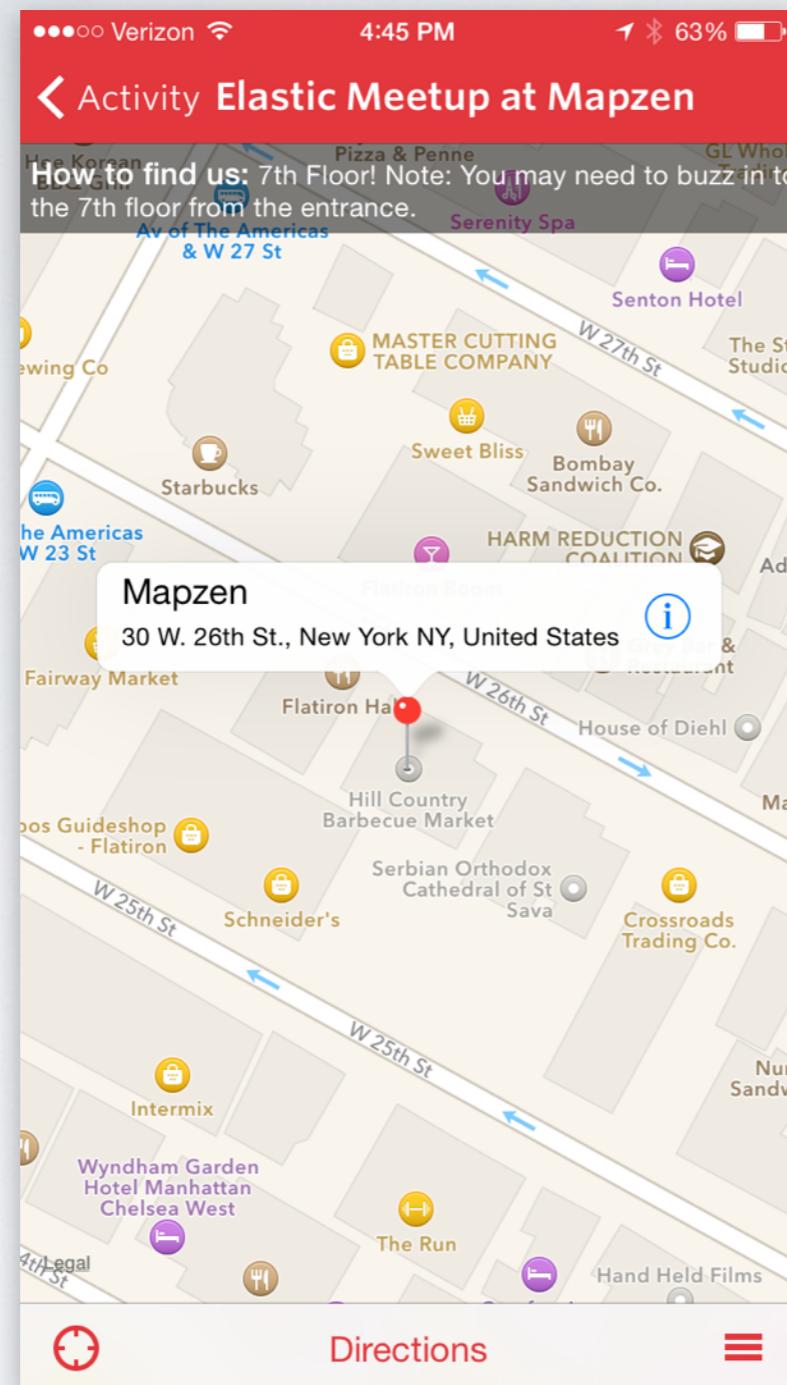
By default this event will follow the [sharing settings](#) of this calendar: event details will be visible to anyone who can see details of other events in this calendar. [Learn more](#)



github.com/pelias



@dianashk



github.com/pelias



@dianashk

KEY GEOCODER FUNCTIONALITY



github.com/pelias



@dianashk

30 W 26th St, NY

SEARCH



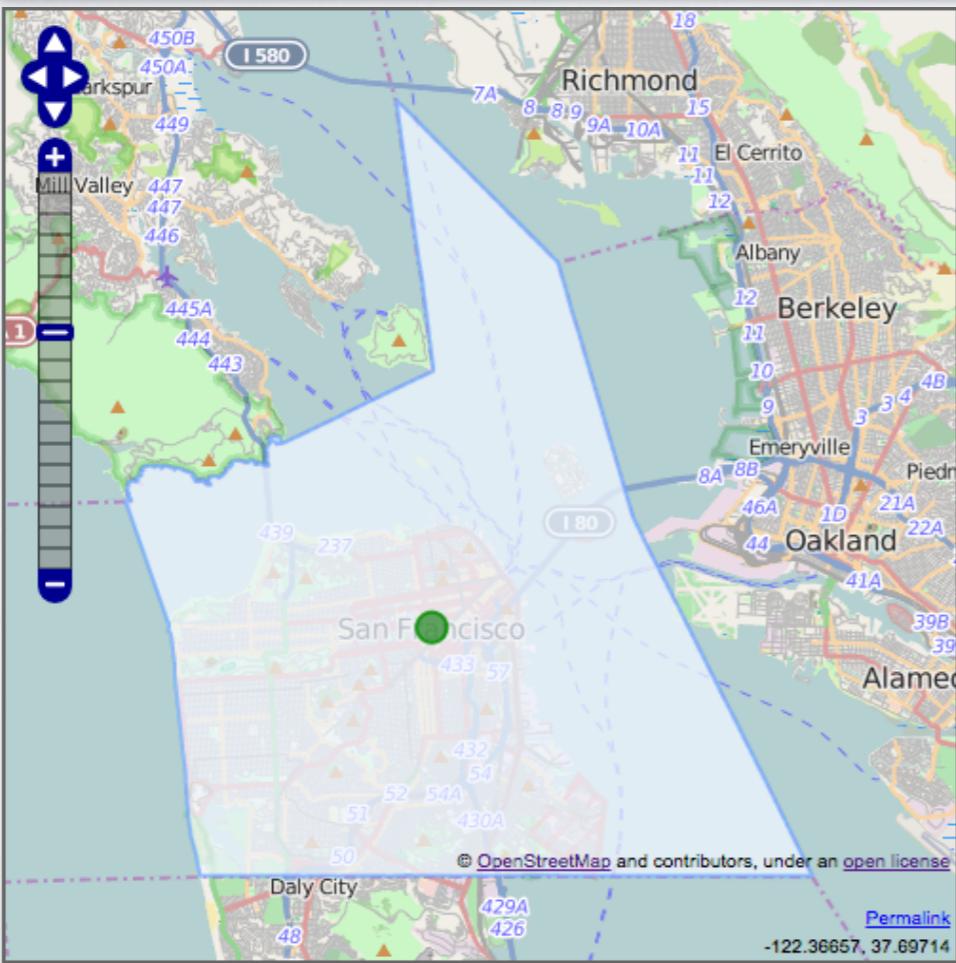
github.com/pelias



@dianashk

San Francisco

SEARCH



github.com/pelias



@dianashk

new yo

SEARCH

New York, NY

Pennsylvania Station New York, NY

New York University, New York, NY

New York Penn Station, New York, NY



github.com/pelias



@dianashk

italian restaurant in Philadelphia, PA

SEARCH

Bistro Romano

Vetri

Spasso Italian Grill

Dante & Luigi's



github.com/pelias



@dianashk

итальянский ресторан в Москве

SEARCH



github.com/pelias



@dianashk

 your current location

SEARCH

30 West 26th Street, Manhattan, NY

Hill Country Barbecue Market, Manhattan, NY

Samsung Accelerator, Manhattan, NY

Hanjan, Manhattan, NY



github.com/pelias



@dianashk

bagel shop around the corner

SEARCH



github.com/pelias



@dianashk

- 
- batch processing
 - pagination
 - result formatting



github.com/pelias



@dianashk

“

You are what you eat!

— Mom



github.com/pelias



@dianashk



github.com/pelias



@dianashk

PELIAS



pelias@mapzen.com



mapzen.com



github.com/pelias



+



github.com/pelias



@dianashk

PELIAS <3



N-GRAMS
SHINGLES
GEOHASH



github.com/pelias



@dianashk

FLATIRON

N=1

F

L

A

T

I

R

O

N

N=2

FL

LA

AT

TI

IR

RO

ON

N=3

FLA

LAT

ATI

TIR

IRO

RON

N=4

FLAT

LATI

ATIR

TIRO

IRON

N-GRAMS



github.com/pelias



@dianashk

FLATIRON

F
FL
FLA
FLAT
FLATI
FLATIR
FLATIRO
FLATIRON

EDGE N-GRAMS



github.com/pelias



@dianashk

I HAVE AN ELASTIC HEART

N=2

I HAVE
HAVE AN
AN ELASTIC
ELASTIC HEART

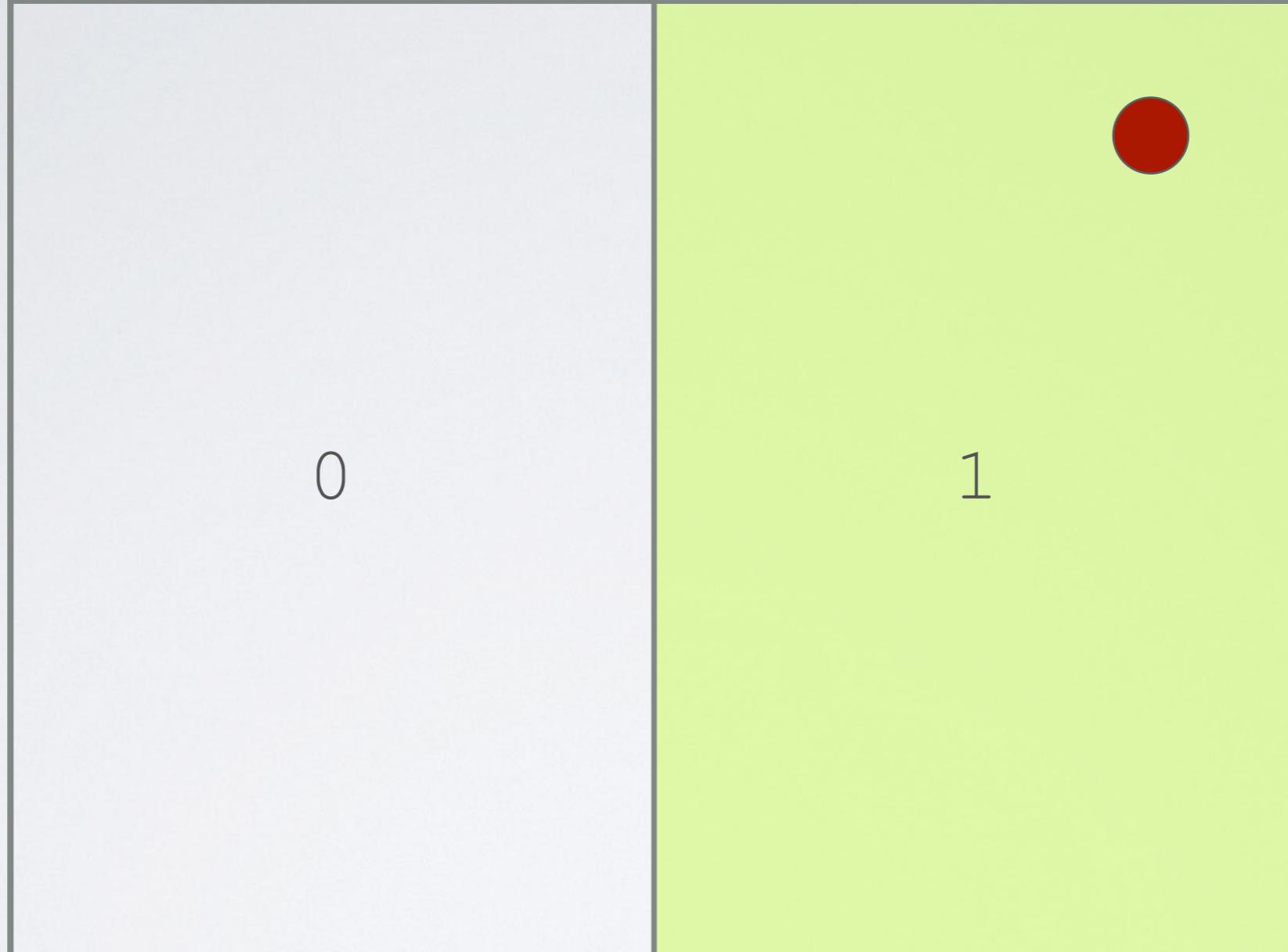
SHINGLES



github.com/pelias



@dianashk



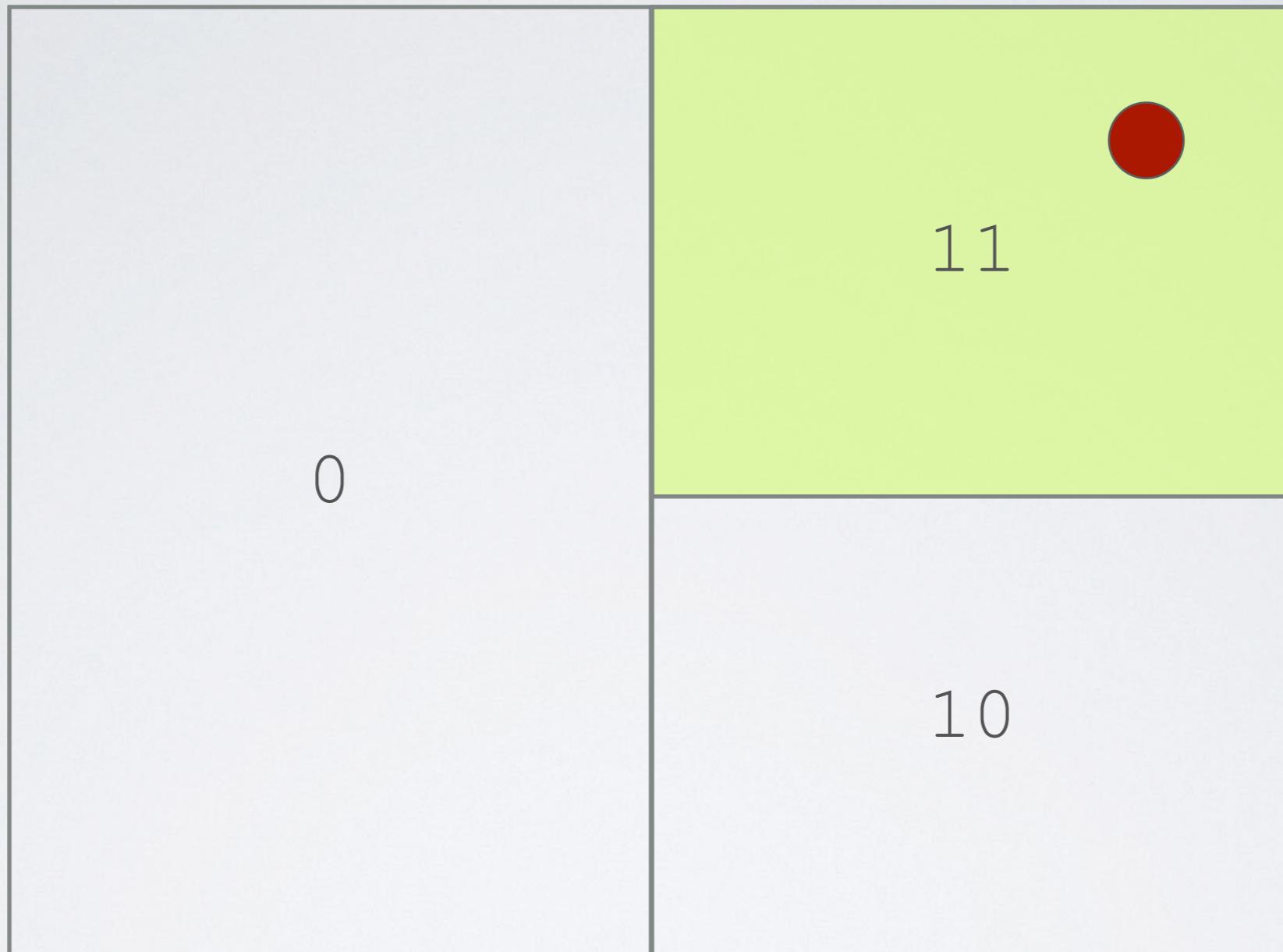
GEOHASH



github.com/pelias



@dianashk



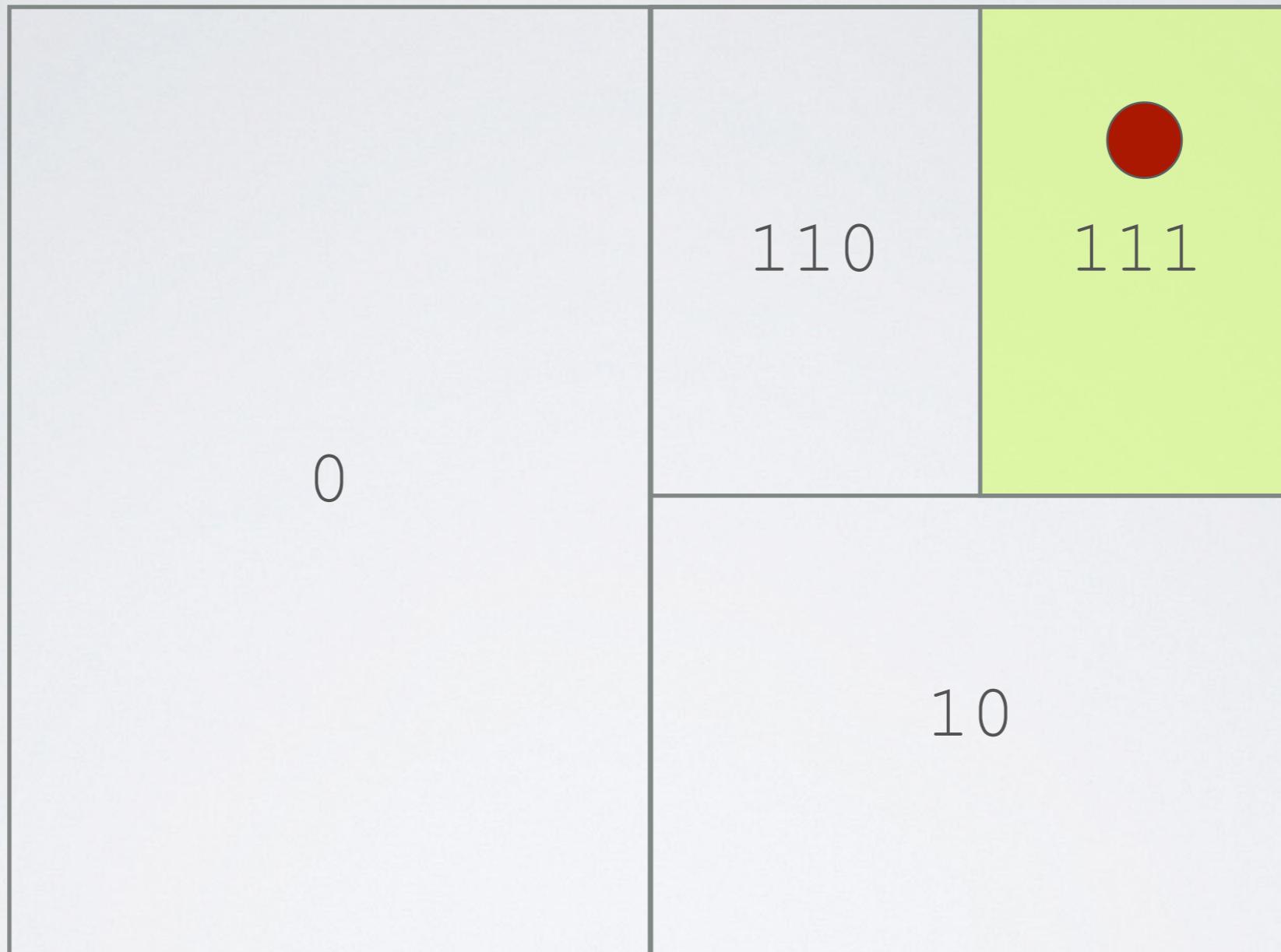
GEOHASH



github.com/pelias



@dianashk



GEOHASH



github.com/pelias



@dianashk

011011101100101110111110000110001101011
01101 11011 00101 11011 11110 00011 00011 01011



GEOHASH



github.com/pelias



@dianashk



OPERATIONS

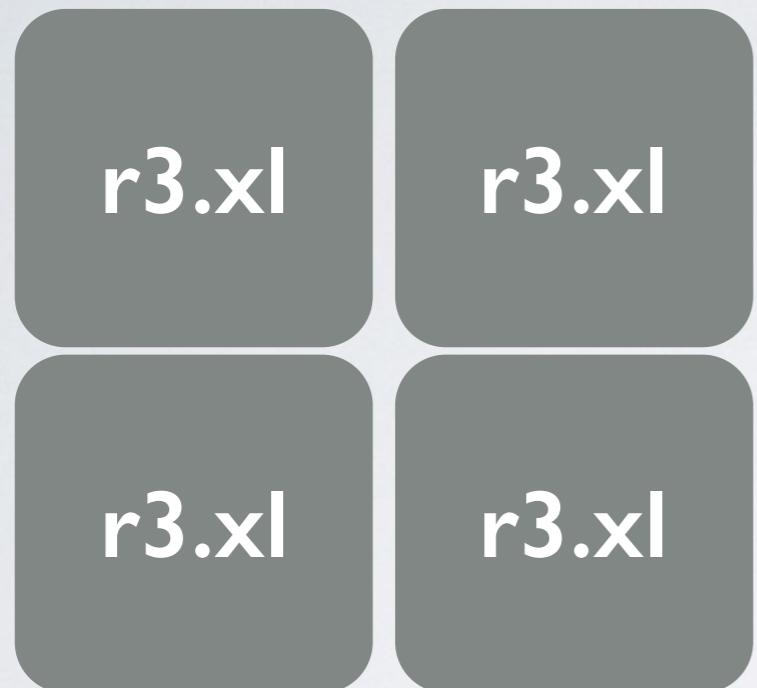


github.com/pelias

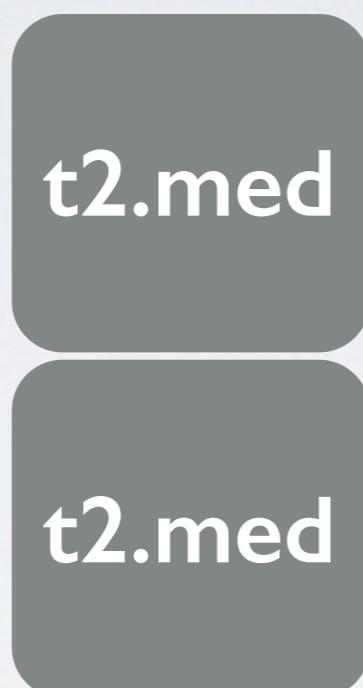


@dianashk

data



routing



master



Data size ~65GB

Data size + 1 replica ~127GB

Field data ~38GB

Filter cache ~8GB

1 shard



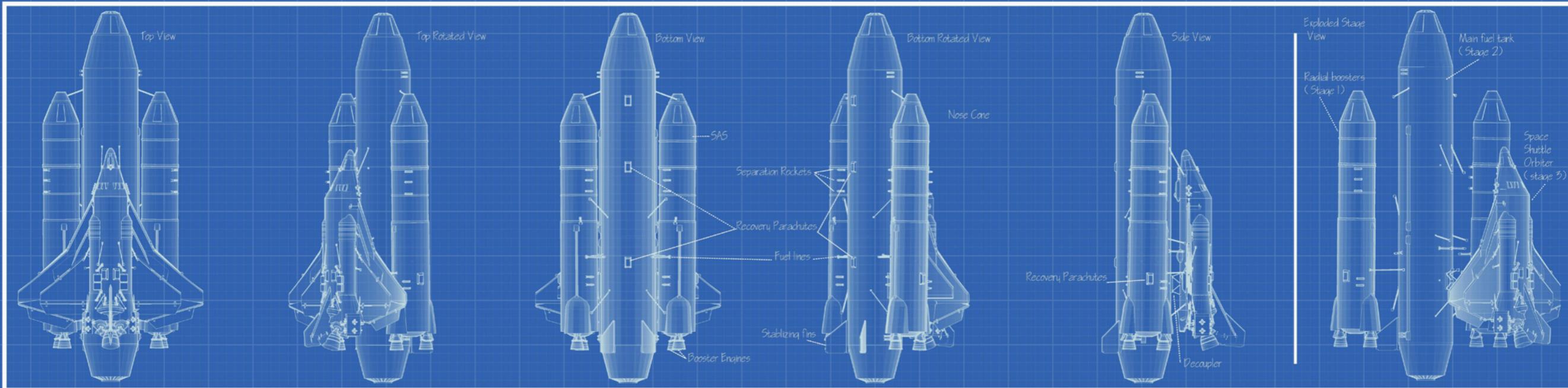
github.com/pelias



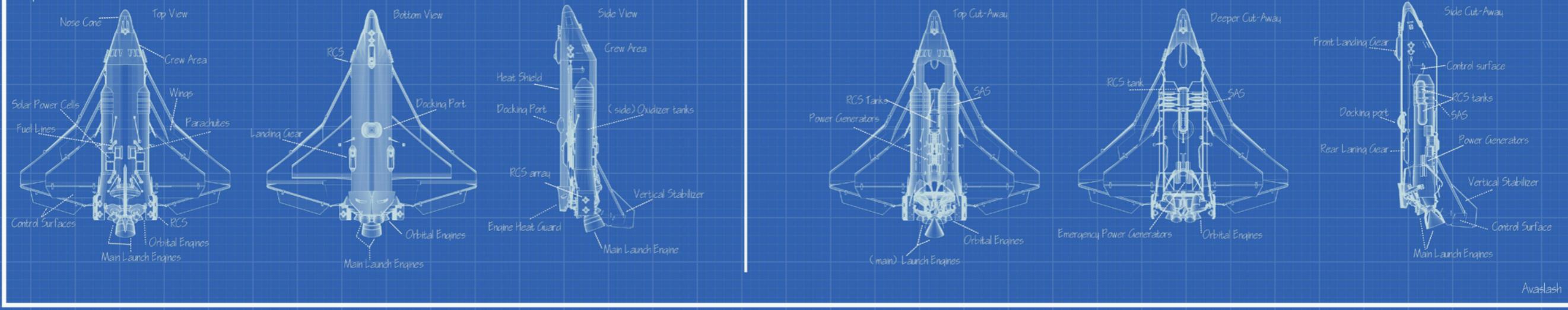
@dianashk

Space Shuttle Blueprints

Space Shuttle Launch Stages



Space Shuttle Orbiter



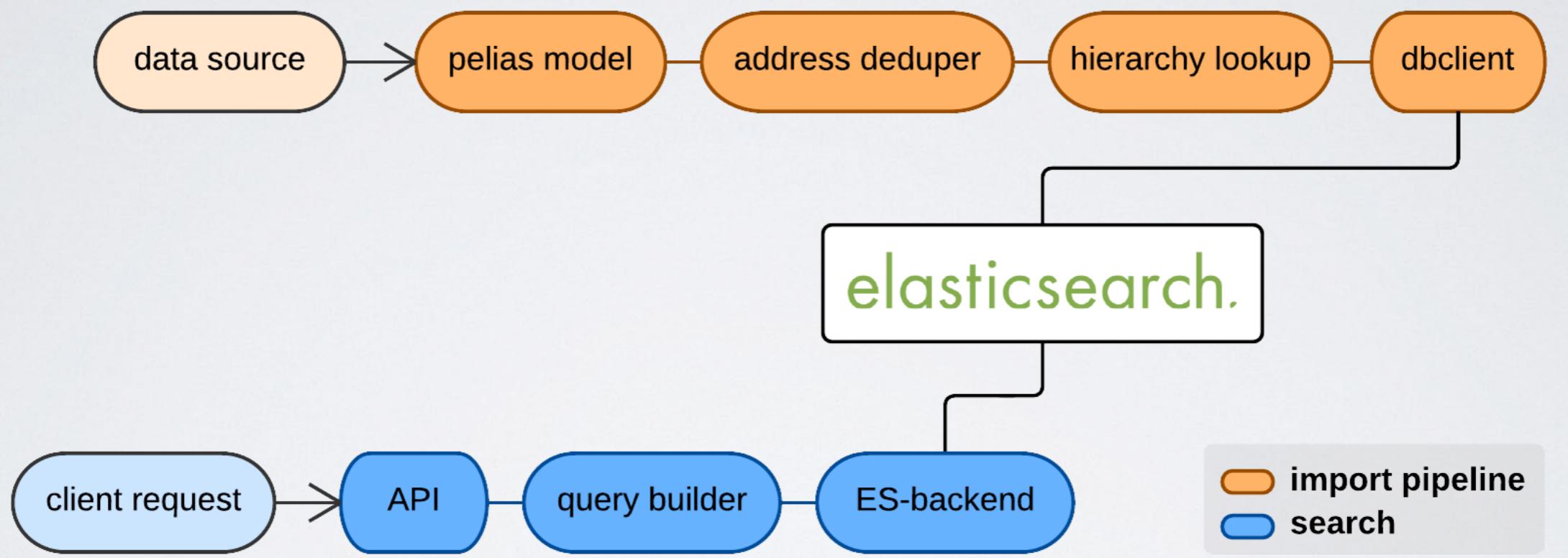
ARCHITECTURE



github.com/pelias



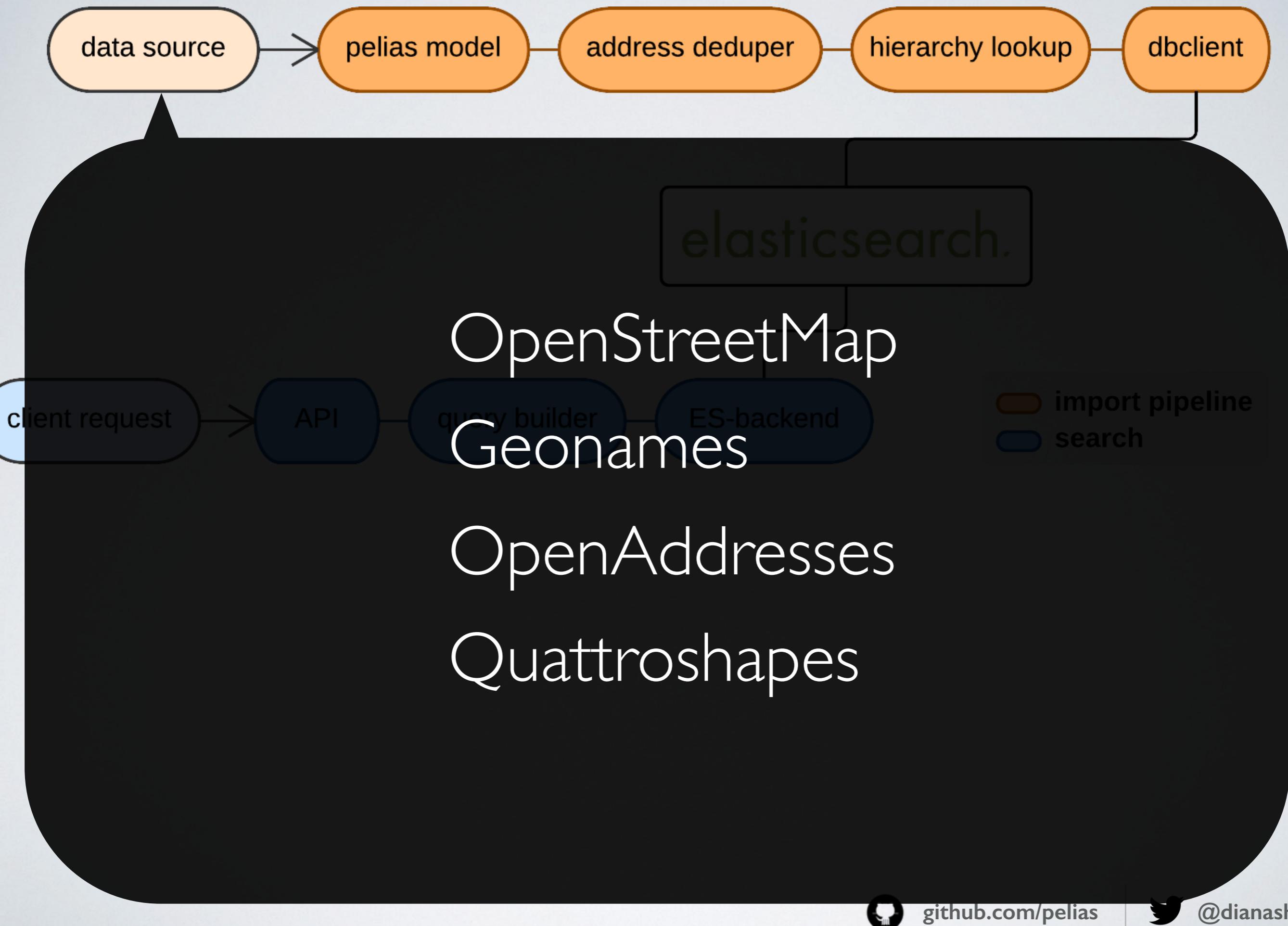
@dianashk



github.com/pelias



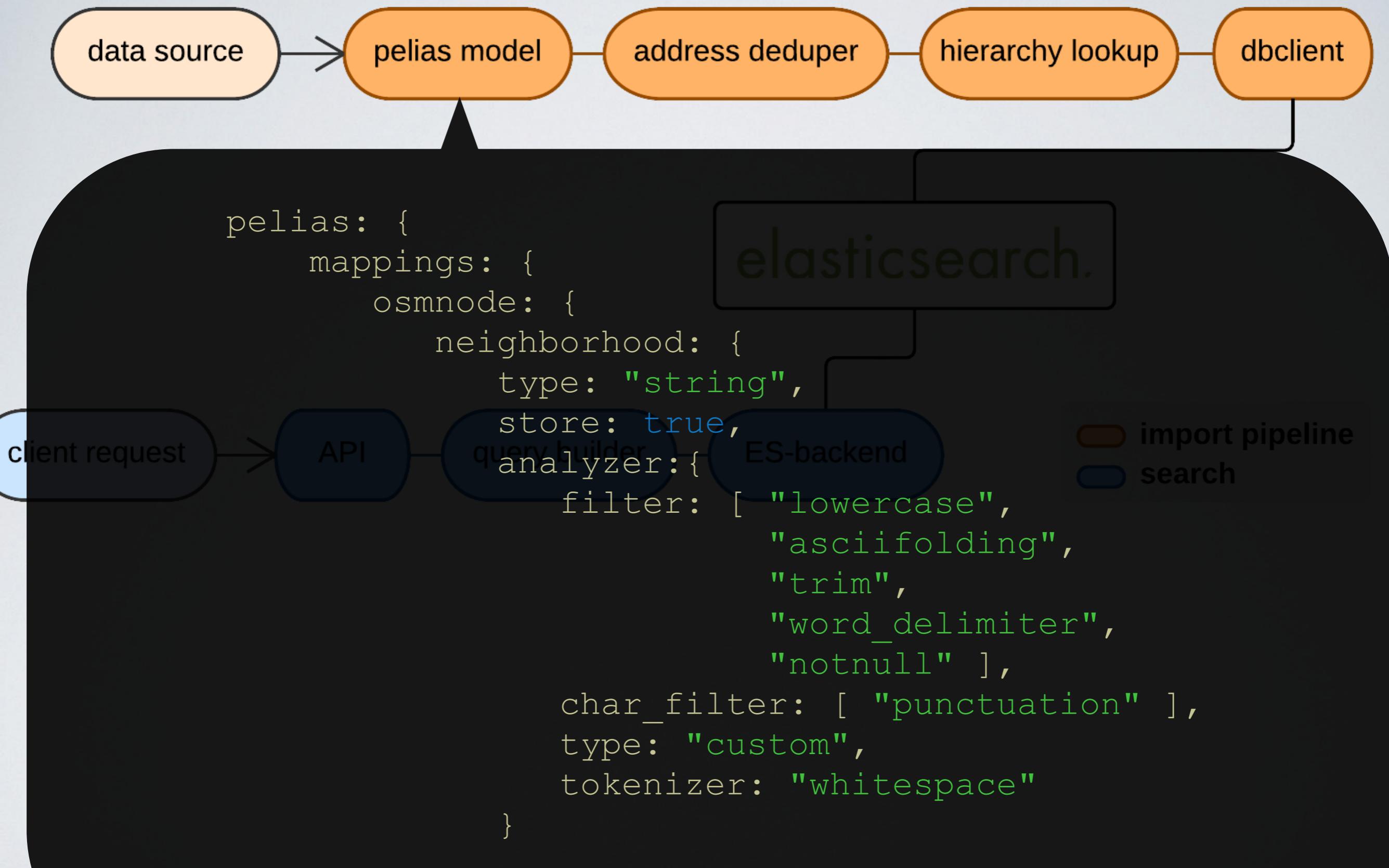
@dianashk



github.com/pelias



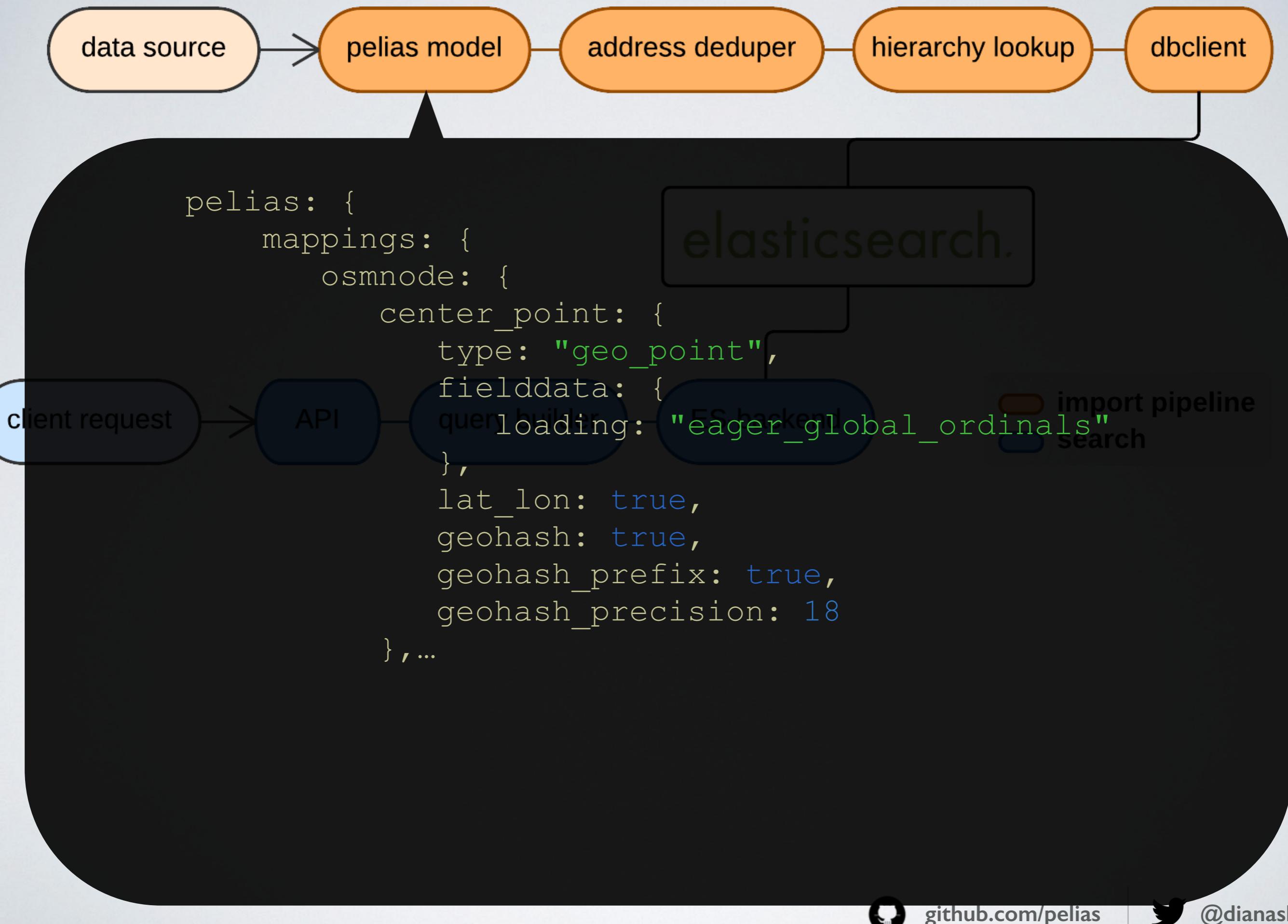
@dianashk



github.com/pelias



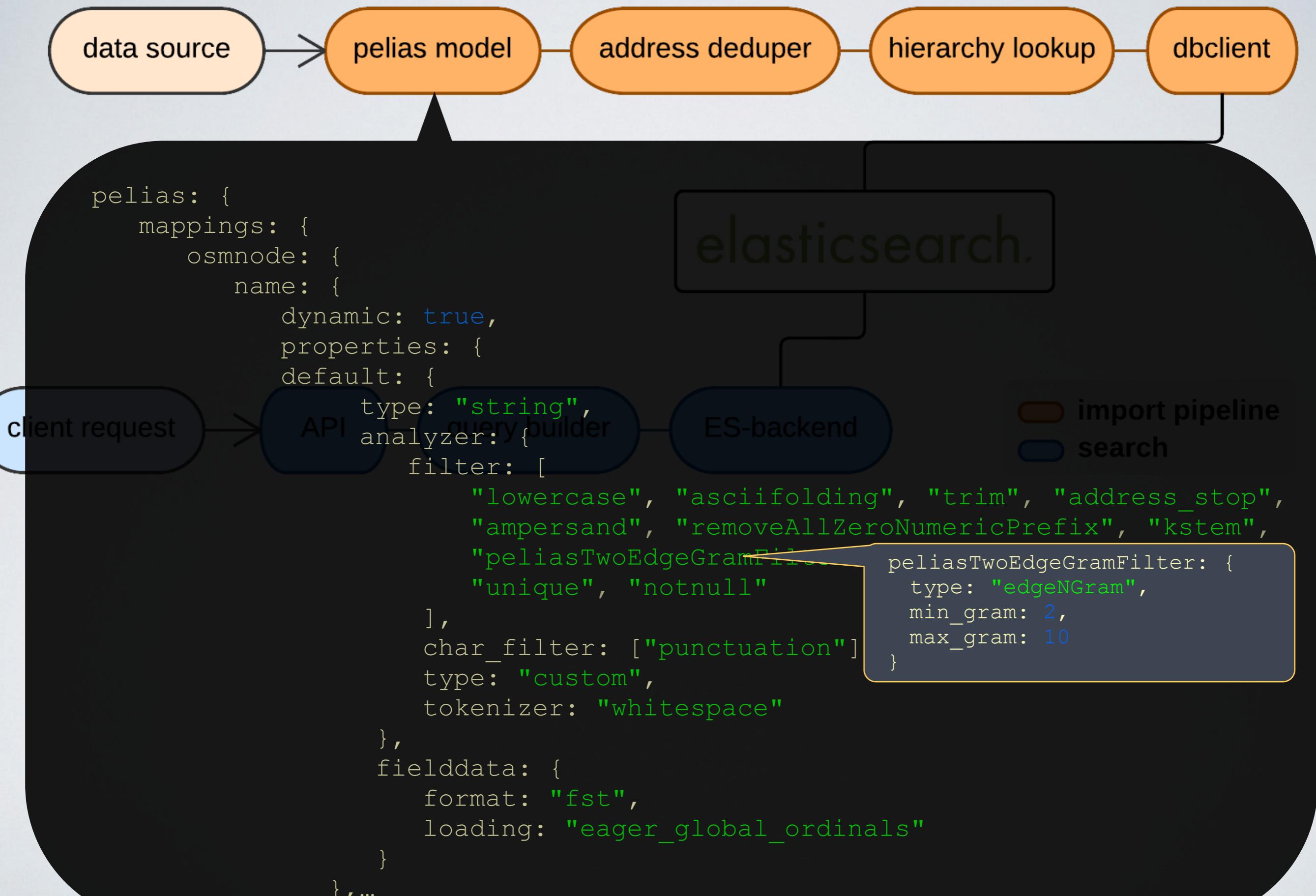
@dianashk



github.com/pelias



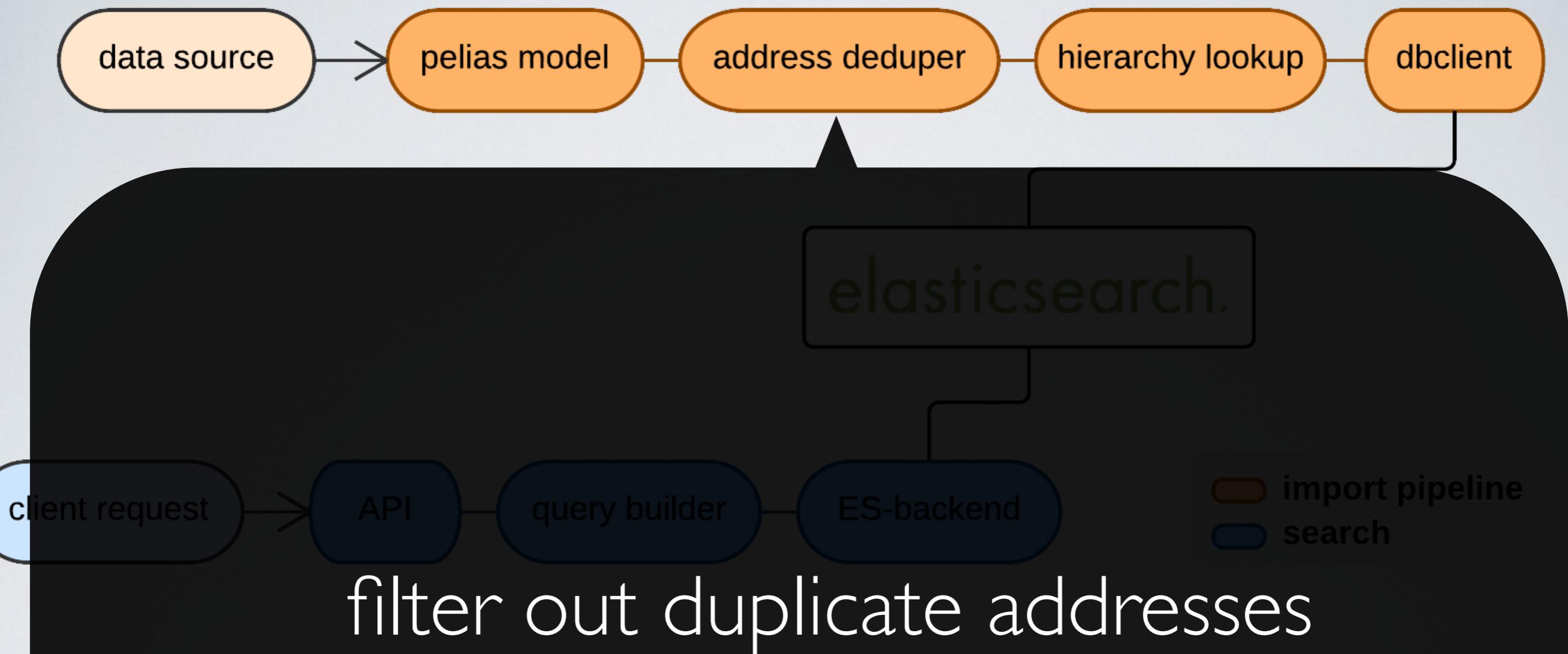
@dianashk



github.com/pelias



@dianashk



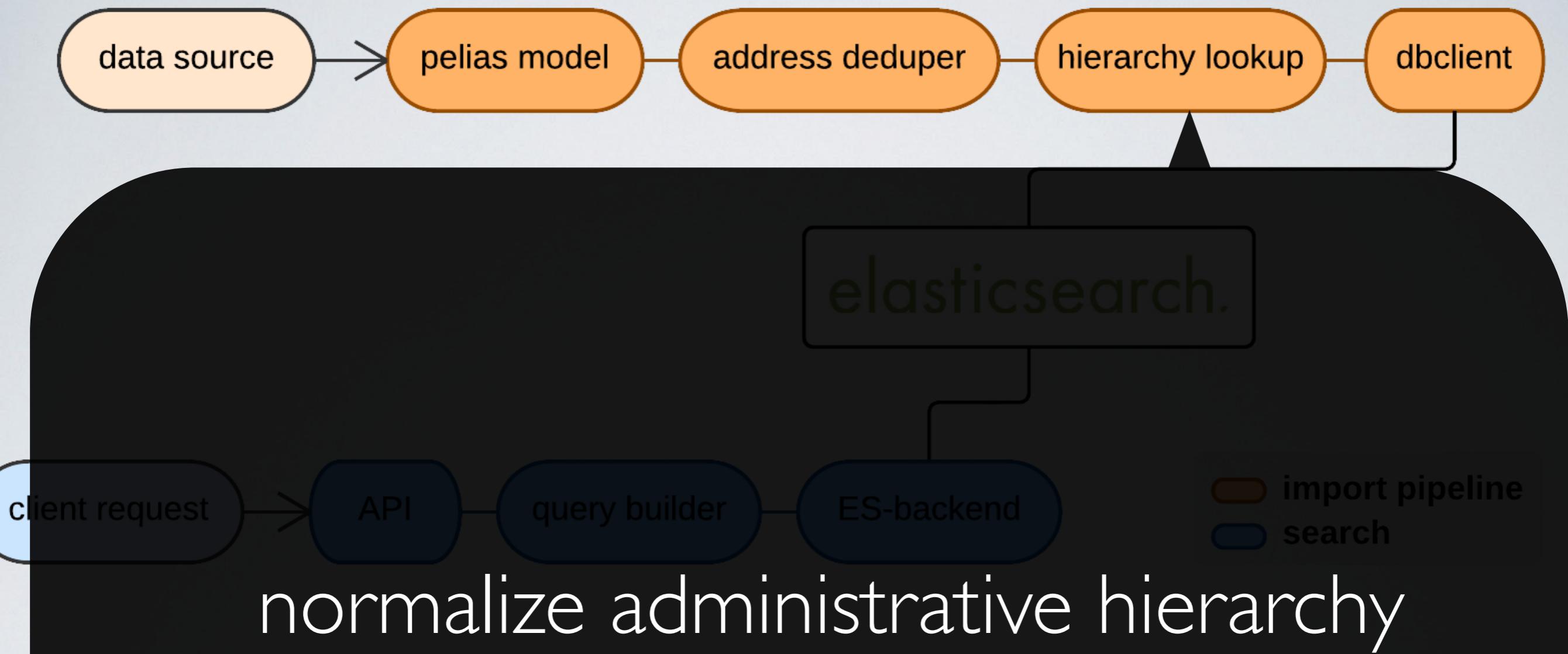
filter out duplicate addresses



github.com/pelias



@dianashk



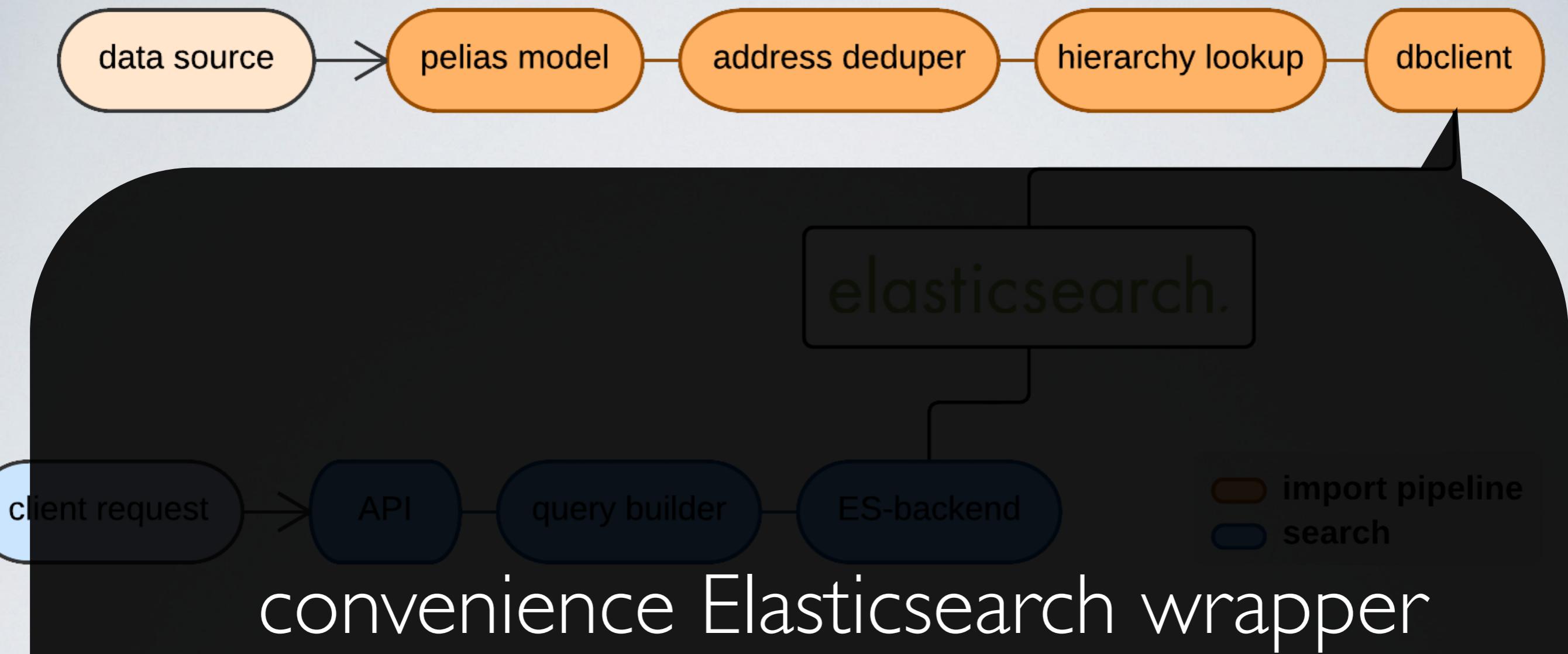
normalize administrative hierarchy



github.com/pelias



@dianashk

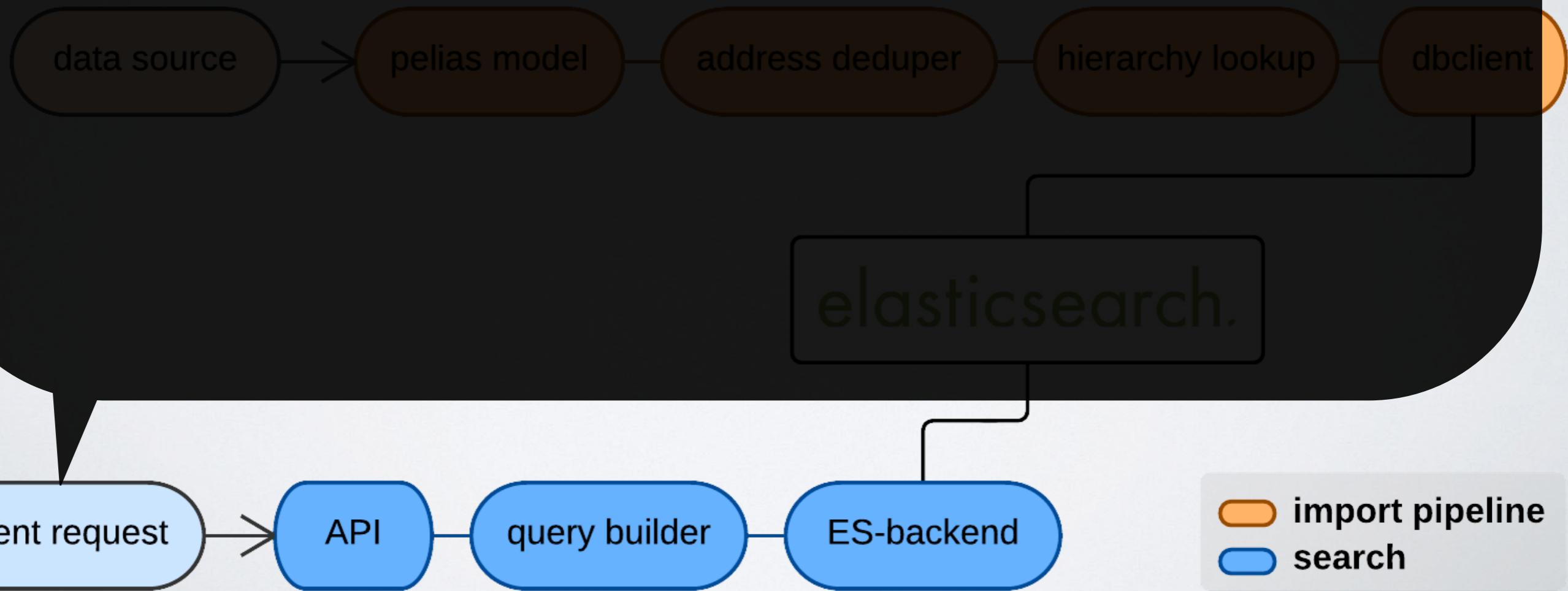


github.com/pelias



@dianashk

`pelias.mapzen.com/search?input=flatiron`

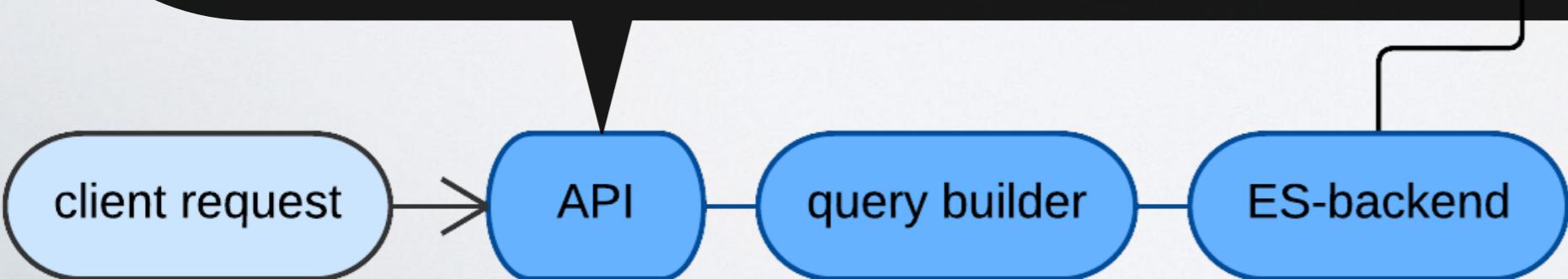


slim express.js server

/search

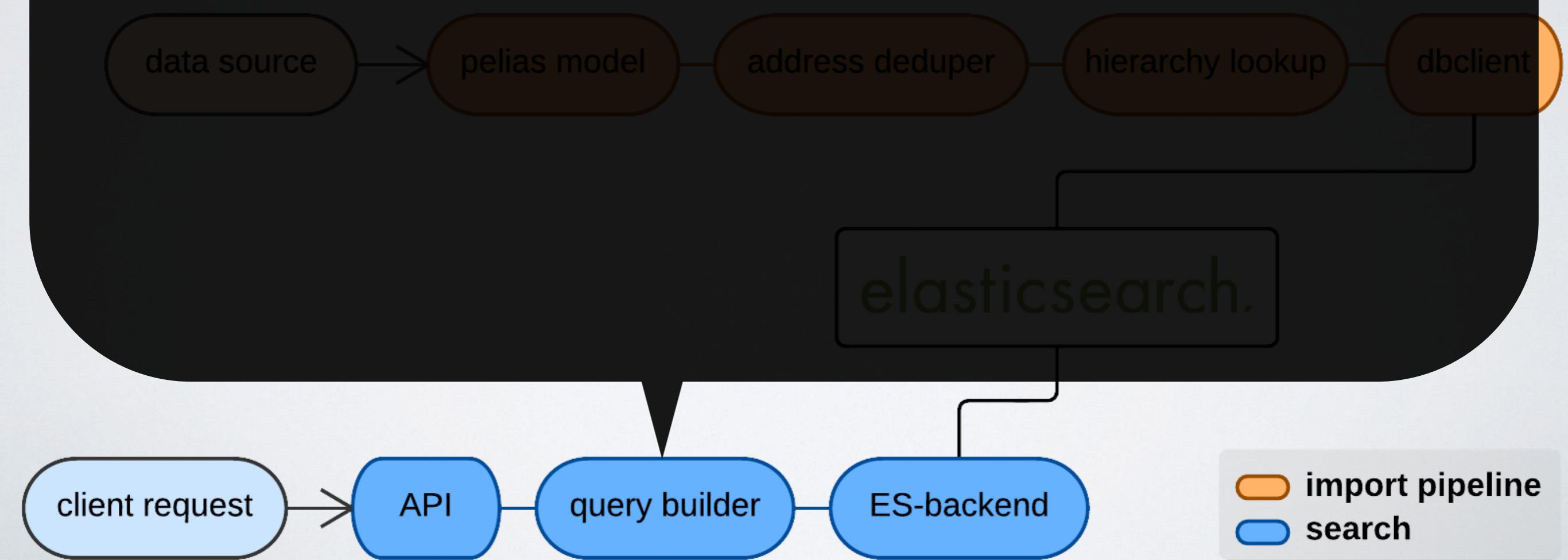


/suggest



import pipeline
search

query builder logic abstracted



convenience Elasticsearch wrapper

