

Taming the API Scraper from Hell

analyzing virtual markets for fun and profit

Maybe you came across one of these

GAMES 1/29/2014 @ 4:55PM | 85,525 views

Massive 'EVE Online' Battle Could Cost \$300,000 In Real Money [Update]

[+ Comment Now](#) [+ Follow Comments](#)

EVE Online is fascinating.

It's one of those games I'm too timid to actually venture into, but that I watch from the sidelines from time to time simply because it's such a marvelous economic experiment.

The space MMO has a fully functioning economy, with regular trade of in-game currency and real money, which leads to a uniquely organic universe unlike any other MMO on the market.

Of course, with real money comes real losses, and when a huge space battle broke out in *EVE Online* recently, sparked over an unpaid bill, the ensuing damage was enormous.

Estimates of the total losses incurred by players during the height of the battle range from \$300,000 to \$500,000, most of that coming from the wreckage of massive "Titan" space-ships, which cost between \$3,000 and \$3,500 a piece in real money. (Exchange rates vary in *EVE*.)

Biggest EVE Online scam ever recorded nets over a trillion ISK

127

by Brendan Drain  on Aug 12th 2011 3:45PM

Sci-Fi, EVE Online, Culture, Economy, Events (In-Game), Lore, Professions, PvP, News Items



Of all the stories that come out of *EVE Online*'s colossal sandbox, it's the tales of criminality and dirty dealings that grab our attention the most. Thefts and scams would be cause for account closure in most MMOs, but they form a legitimate and dark part of the *EVE* universe. The famous [Guiding Hand Social Club heist](#) of 2005, Ricdic's EBANK scandal and last year's massive 800 billion ISK Titans4U scam are just a few of the most [high-profile thefts and scams](#) to hit the *EVE* community over the years.



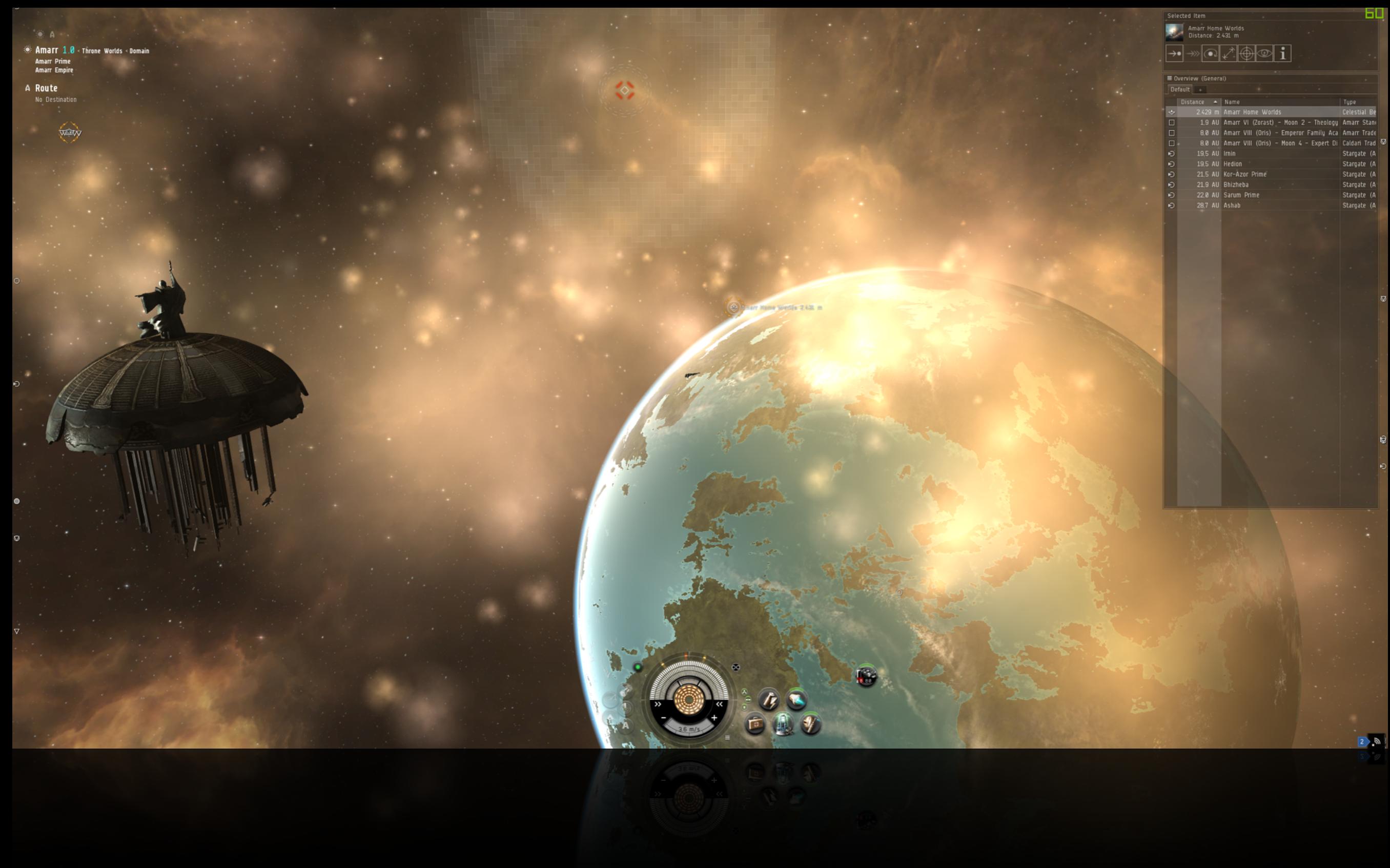


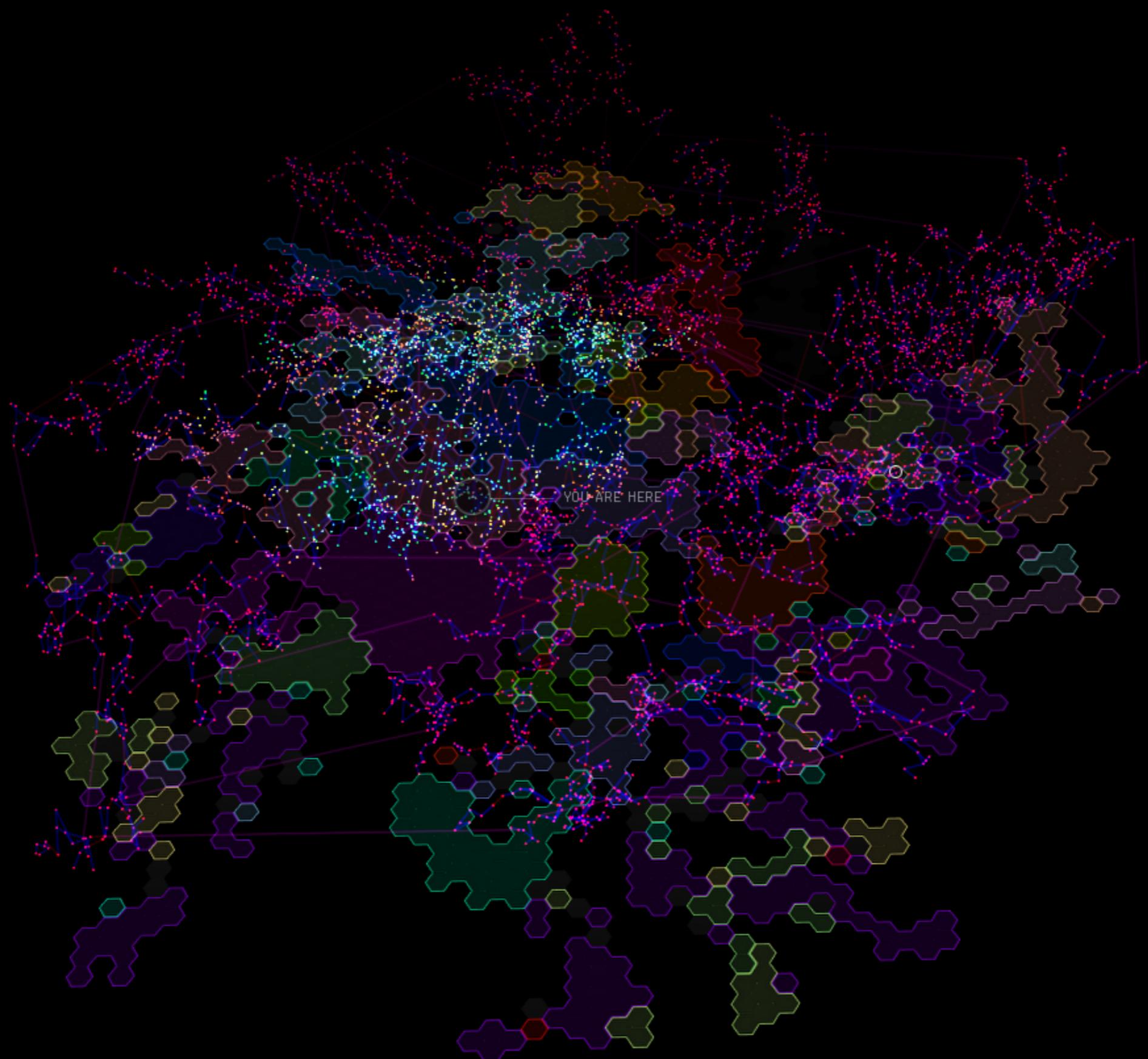
WTF IS EVE?

Sandbox space MMORPG by CCP Games

500,000 active subscriptions

One single server cluster (except for China)





Ship Fitting

Fiesertyp Lennelluc's Catalyst

Calibration 100,0 / 400,0

Capacitor Depletes in 00:08.04
780 GJ / 146.16 s
Δ -0,4 GJ/s (-2,7%)

Offense 100,9 dps
100,9 dps 0 dps 0 dps

Defense 3.091 hp
4 hp/s 900 hp 500 s 960 hp 1.125 hp

0 % 20 % 40 % 50 %
50 % 35 % 35 % 10 %
0 % 0 % 0 % 0 %

Targeting 33,00 km
11,00 points 600 mm
68 m 7x

Navigation 331,3 m/s
1.550,00 t 2,0314x
4,50 AU/s

CPU 70,5/222,5
Power Grid 31,5/100,0

Browse Save Strip

THE ECONOMY OF EVE



ECONOMY

Open economy

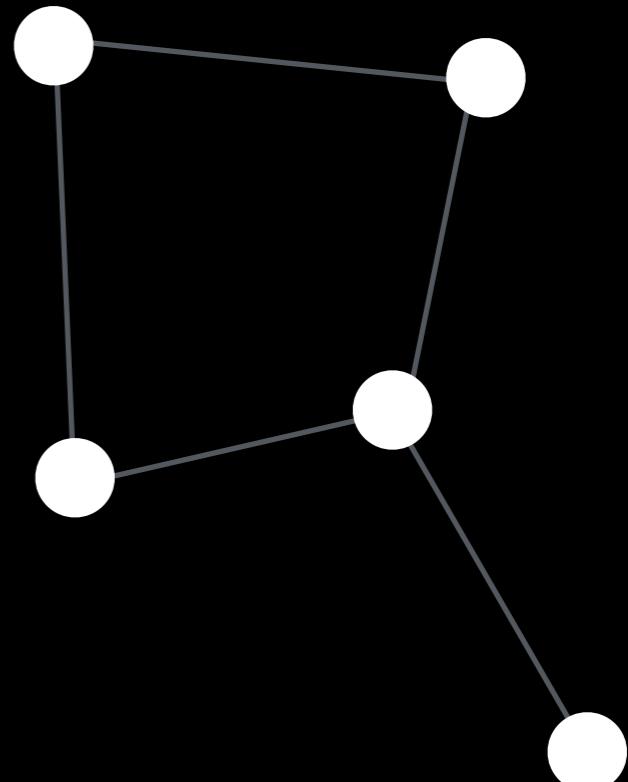
Ask/bid order system

Regional markets

Systems

Constellations

Regions



Regional Market

Domain Regional Market

Details Groups My Orders

Pilot's Services / Pilot's License Extensions (PLEX) / **30 Day Pilot's License Extension (PLEX)** ⓘ

Concord Partners

Browse Quickbar

Search term Search ⚙

- ▶ Ammunition & Charges
- ▶ Apparel
- ▶ Blueprints
- ▶ Drones
- ▶ Implants & Boosters
- ▶ Infantry Gear
- ▶ Manufacture & Research
- ▶ Pilot's Services
- ▶ Planetary Infrastructure
- ▶ Reactions
- ▶ Ship Equipment
- ▶ Ship Modifications
- ▶ Ships
- ▶ Skills
- ▶ Special Edition Assets
- ▶ Structures
- ▶ Trade Goods

Market Data Price History

Sellers

Jumps	Quantity	Price	Location	Expires in
Station	2	867.995.992,99 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 23h 59m 49s
Station	2	867.995.993,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 23h 51m 45s
Station	2	867.999.999,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 22h 45m 6s
Station	1	868.000.000,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 14h 58m 45s
Station	3	868.994.993,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 23h 53m 56s
Station	3	868.994.993,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 23h 57m 45s
Station	2	868.995.000,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 23h 20m 15s
Station	2	869.000.000,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 13h 46m 46s
Station	1	869.000.000,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 14h 58m 51s
Station	1	869.895.992,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	14h 7m 19s
Station	1	869.896.000,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 11h 4m 52s
Station	2	869.896.993,97 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 10h 3m 28s
Station	2	869.897.995,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 8h 33m 49s
Station	1	869.900.000,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	3h 49m 15s
Station	5	869.999.999,99 ISK	Amarr VIII (Oris) - Emperor Family Academy	89d 12h 2m 25s
Station	5	870.000.000,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	26d 4h 42m 49s
Station	1	870.000.000,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	27d 2h 14m 51s
Station	6	870.000.000,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	28d 9h 48m 18s

Buyers

Jumps	Quantity	Price	Location	Range	Min volume	Expires in
Station	1	845.813.359,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	Station	1	89d 23h 57m 28s
4	4	845.720.000,00 ISK	Gosalav VII - Moon 1 - Quafe Company Factory	Station	1	29d 23h 28m 8s
Station	1	845.702.000,42 ISK	Amarr VIII (Oris) - Emperor Family Academy	Station	1	89d 23h 56m 24s
2	9	845.701.126,00 ISK	Penirgman IX - Moon 11 - Ishukone Corporation Factory	20 Jumps	1	89d 23h 51m 31s
Station	5	845.701.125,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	Station	1	89d 23h 35m 55s
Station	1	845.701.123,43 ISK	Amarr VIII (Oris) - Emperor Family Academy	Station	1	89d 23h 30m 32s
Station	4	845.701.123,02 ISK	Amarr VIII (Oris) - Emperor Family Academy	20 Jumps	1	89d 22h 48m 53s
Station	6	845.701.123,02 ISK	Amarr VIII (Oris) - Emperor Family Academy	20 Jumps	1	89d 22h 48m 59s
Station	6	845.701.123,02 ISK	Amarr VIII (Oris) - Emperor Family Academy	20 Jumps	1	89d 22h 49m 2s
Station	6	845.701.123,02 ISK	Amarr VIII (Oris) - Emperor Family Academy	20 Jumps	1	89d 22h 49m 5s
2	6	845.701.123,02 ISK	Penirgman IX - Moon 11 - Ishukone Corporation Factory	20 Jumps	1	89d 22h 48m 57s
Station	44	845.701.123,01 ISK	Amarr VIII (Oris) - Emperor Family Academy	Station	1	89d 22h 48m 20s
Station	8	845.700.007,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	Station	1	29d 18h 25m 5s
Station	5	845.700.000,02 ISK	Amarr VIII (Oris) - Emperor Family Academy	Station	1	89d 18h 25m 4s
Station	4	845.602.017,00 ISK	Amarr VIII (Oris) - Emperor Family Academy	Station	1	13d 14h 33m 13s
Station	1	845.602.015,01 ISK	Amarr VIII (Oris) - Emperor Family Academy	Station	1	89d 13h 1m 2s
Station	4	845.601.999,99 ISK	Amarr VIII (Oris) - Emperor Family Academy	Station	1	89d 12h 31s

Export to File Place Buy Order

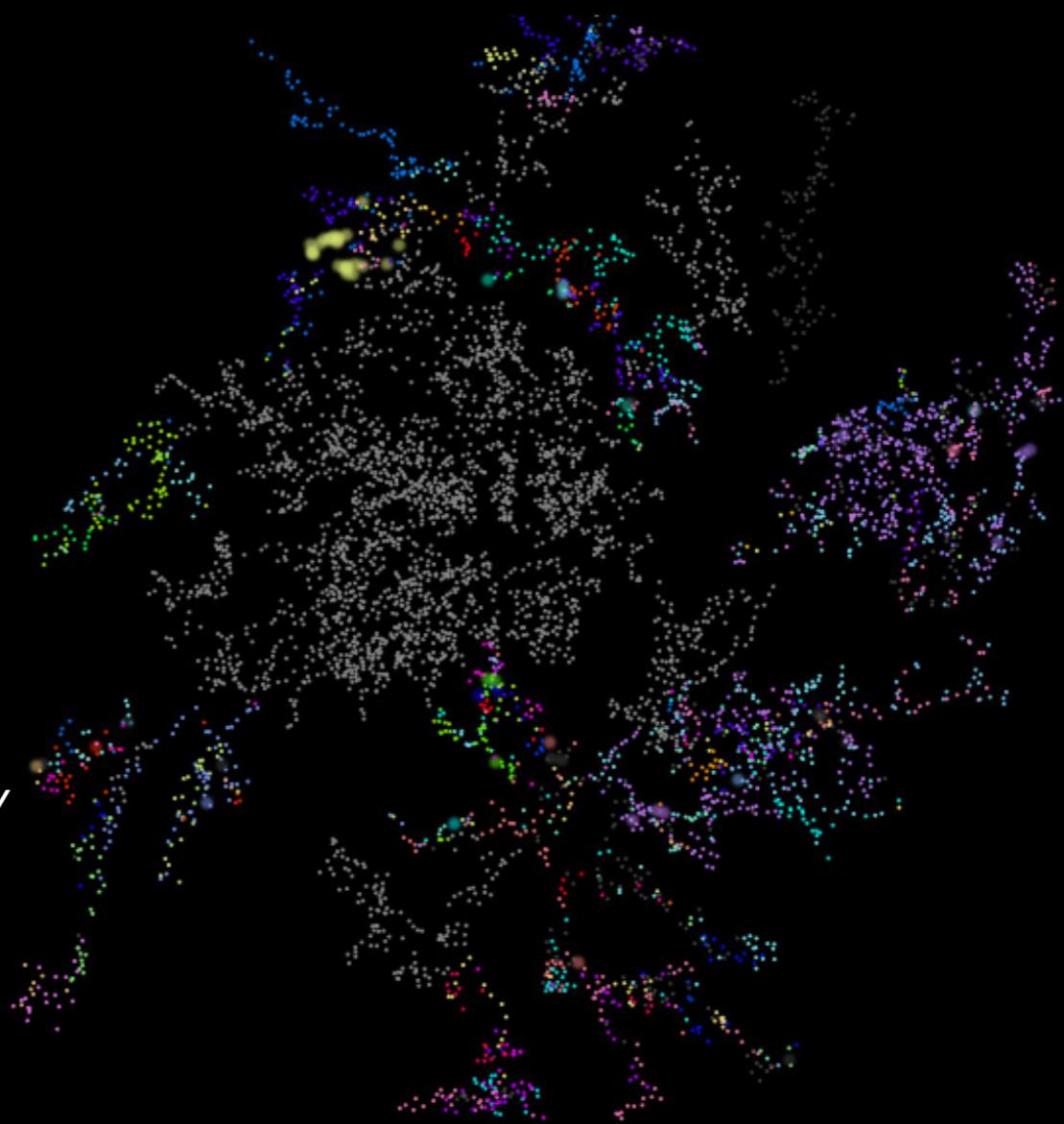
ECONOMY

Around 5,000 systems

97 regions

Problem

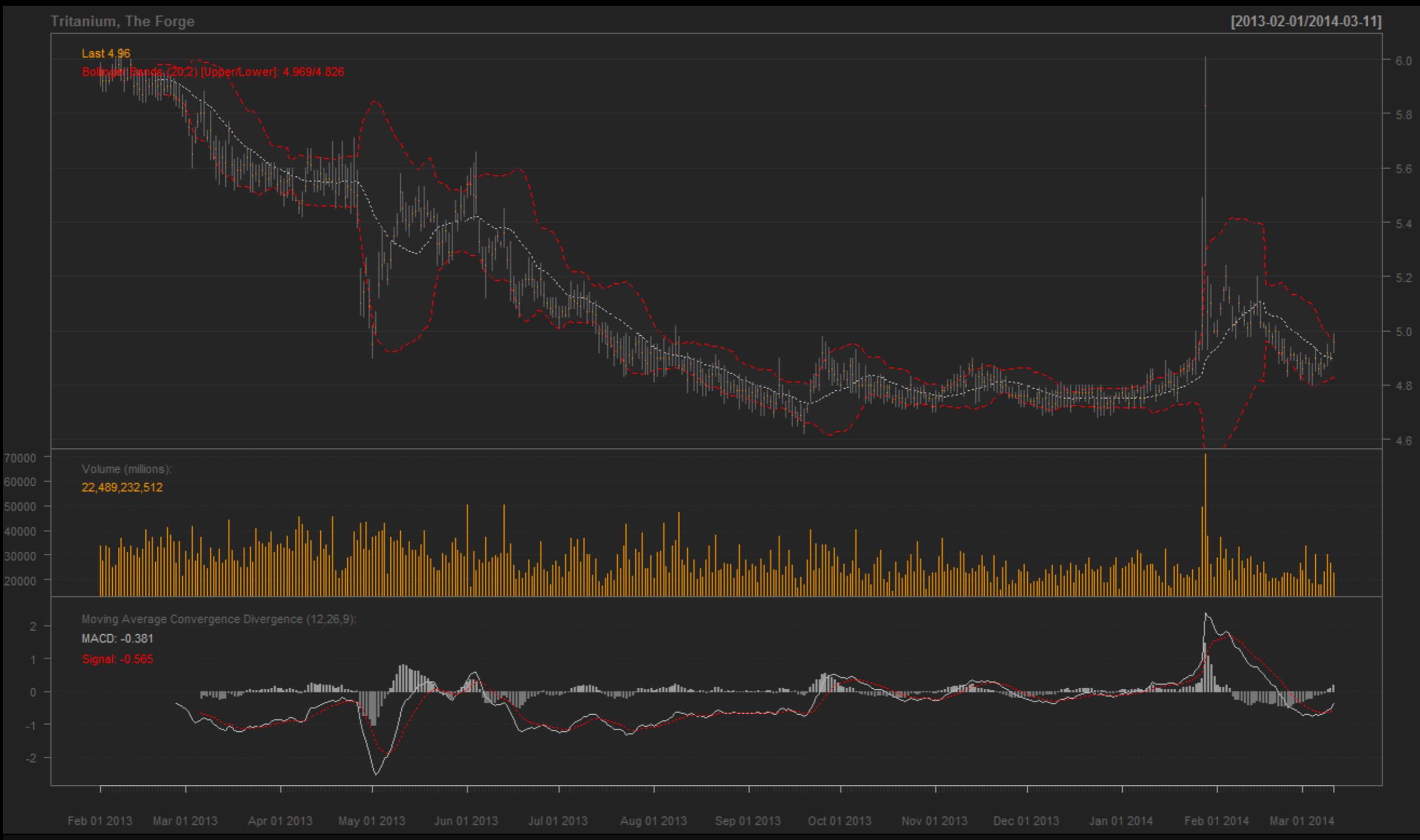
I can only see the market of my current location



SOLUTION

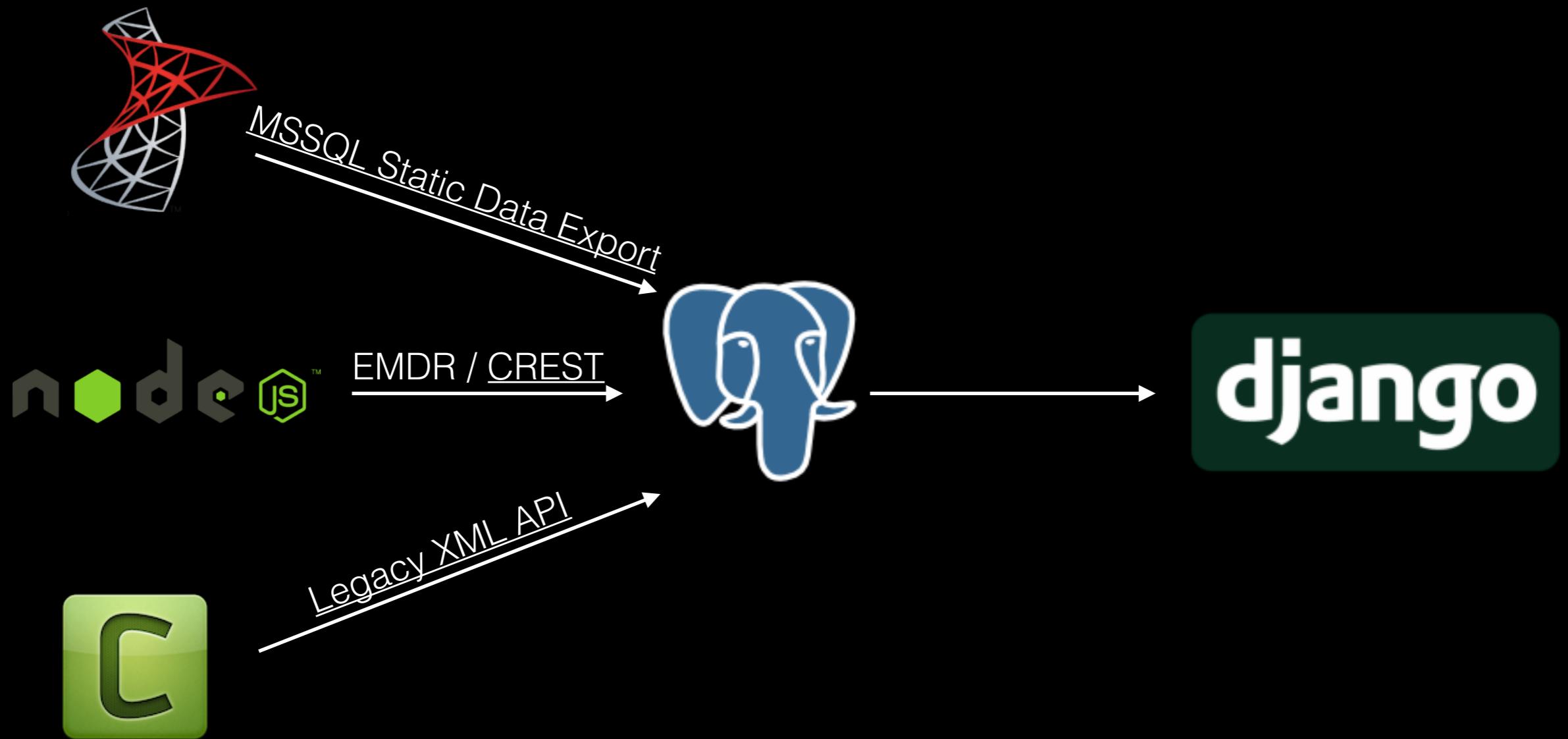
Crowdsourcing market data!

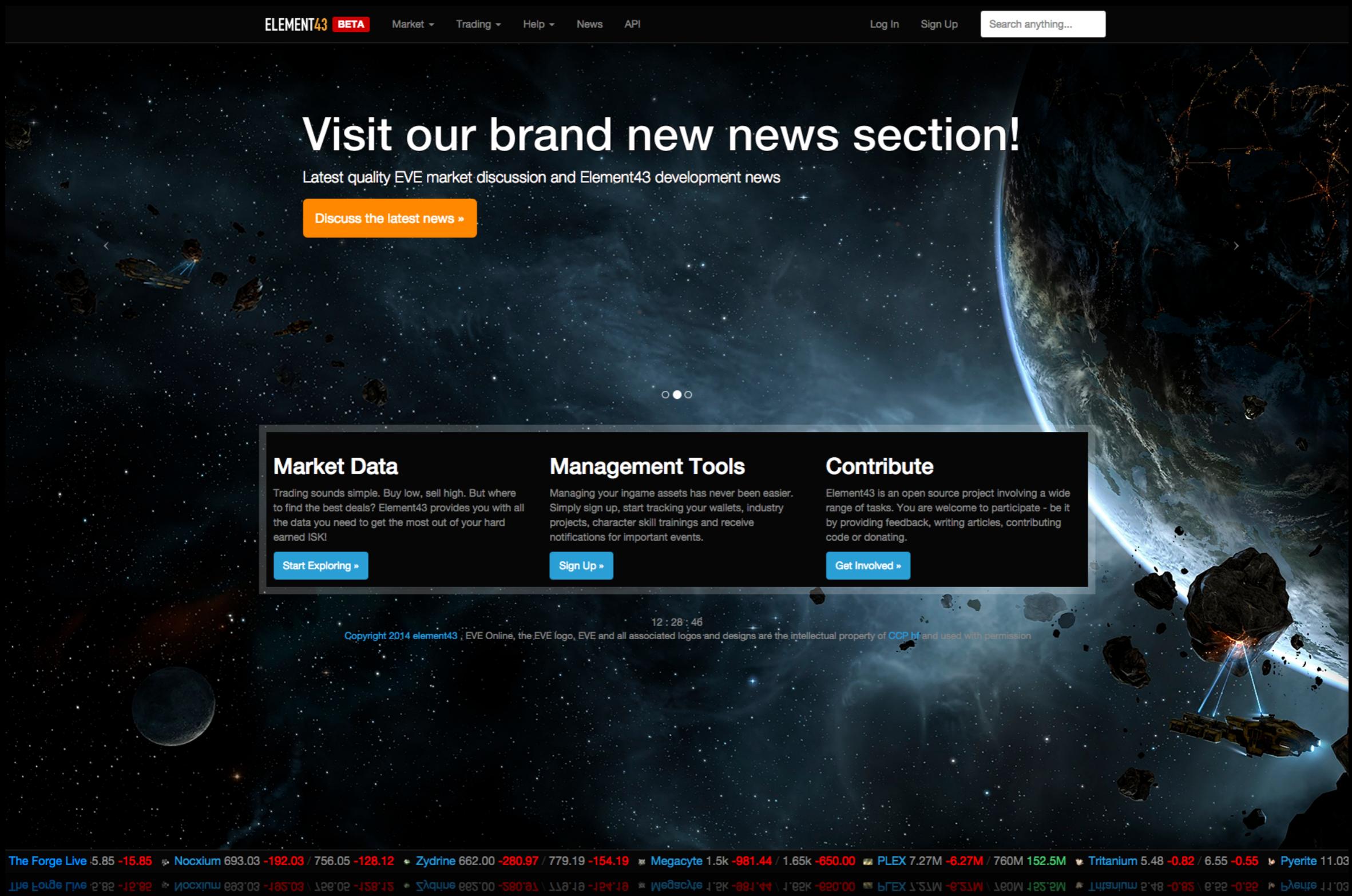




EVEN BETTER

Open-source web applications!





Market / Pilot's License Extensions (PLEX) / Pilot's License Extensions (PLEX) / 30 Day Pilot's License Extension (PLEX)

Overview

Regions

Ask / Sell

Bid / Buy

30 Day Pilot's License Extension (PLEX)

Volume: 0.01m³

 PLEX - Pilot License Extension PLEX can be used to add game time to your account or to unlock special character and account services. Account time: One PLEX is equal to 30 days of game time which can be added to your account by using the PLEX directly from your in-game inventory. Multiple character training: Activate 30 days of passive skill gain on additional characters on your account with PLEX. Character transfer: Transferring characters between two accounts can be paid for with PLEX. Character resculpt: Customize and recreate your characters facial appearance. Convert to AURUM: PLEX can be exchanged for AURUM, the currency used in the New Eden Store where unique fashion accessories and cosmetic ship customizations are sold. Buying PLEX You can buy PLEX on the EVE market using ISK or securely on https://secure.eveonline.com/PLEX/ using conventional payment methods.

Zoom 1m 3m 6m YTD 1y All

Average Prices in Main Trading Hubs



Price

800M
700M
600M
500M

Dec '13 Jan '14 Feb '14 Mar '14 Apr '14 May '14 Jun '14 Jul '14 Aug '14 Sep '14 Oct '14 Nov '14

2011 2012 2013 2014

Highcharts.com

Ask / Sell

Location	Price	Quantity	Last Updated
• 0.0 The Spire - K-BBYU III - AND Holdings - 2Skulls Paradise	840,000,000.00	1	1 day, 13 hours ago
• 0.5 Khanid - Palas II - Royal Khanid Navy Assembly Plant	848,969,428.00	1	1 week, 5 days ago
• 0.0 Delve - JP4-AA V - JAVAH Inc	849,999,999.97	1	3 hours ago
• 0.0 Outer Passage - SN-DZ6 V - More Like SMD	850,000,000.00	2	2 weeks ago
• 0.5 Khanid - Palas II - Royal Khanid Navy Assembly Plant	850,000,900.22	2	1 week, 5 days ago

More

Bid / Buy

Location	Price	Quantity	Last Updated
• 1.0 Domain - Amarr VIII (Oris) - Emperor Family Academy	845,701,000.00	2	4 minutes ago
• 1.0 Domain - Amarr VIII (Oris) - Emperor Family Academy	845,700,018.00	6	4 minutes ago
• 1.0 Domain - Amarr VIII (Oris) - Emperor Family Academy	845,700,007.00	8	4 minutes ago
• 1.0 Domain - Amarr VIII (Oris) - Emperor Family Academy	845,700,000.02	5	4 minutes ago
• 1.0 Domain - Amarr VIII (Oris) - Emperor Family Academy	845,700,000.01	6	4 minutes ago

More

12 : 30 : 46

Copyright 2014 element43 , EVE Online, the EVE logo, EVE and all associated logos and designs are the intellectual property of CCP hf and used with permission

Dashboard

[Dashboard](#)[Wallet](#)[Settings](#)[Feedback and Support](#)[Donate](#)

Black Legion, and Wormholes Kill 13 BOT Ratting carriers.

1 week, 6 days ago

Monday night late EU time zone saw 13 Brothers of Tangra ratting carriers die at the hands of Black Legion, and several large wormhole entities including: Irresponsible Use of Capital, ... [read on »](#)



EVE Online forum update on October 21, 2014

2 weeks ago

The EVE Online forum will be updated on Tuesday, October 21st, with bug fixes, an improved user interface, and better functionality. The deployment of this update is scheduled for 14:00 ... [read on »](#)



Join the mass test on Tuesday, October 14th

3 weeks ago

Join the mass test on our test server Singularity on October 14 at 17:00 UTC for testing the jump changes for the Phoebe release! We will take a closer look at ... [read on »](#)



EVE Vegas 2014 is sold out!

3 weeks ago

We are proud to announce that we are sold out for this year's EVE Vegas convention! Thank you everyone that has attended and supported the event in the past years, ... [read on »](#)



EVE Vegas Alliance Prize Draw Winner

Notifications Panel

Notifications not yet available

now

[View All Alerts](#)

Characters

 Character Name
10,702,842 ISK

○ Update: 2 weeks ago

 Character Name
156,202,069 ISK

○ Update: 2 weeks ago

[Manage Characters](#)

Overview

	Profit	Brokers Fee	Taxes
Month	None	None	None
Week	None	None	None
Day	None	None	None

Latest Sales

	1	Propulsion Jamming	148,000.0	148,000.0
	100	Iridium Charge S	10.39	1,039.0
	873	Cap Booster 400	4,189.77	3,657,669.21
	20	Cap Booster 400	4,189.79	83,795.8

STATUS

1,200 Users

Tracking several million active orders

54,000,000 archived orders

114,000,000 OHLC datapoints

130GB database

... on a 20€/month server

region = collection of systems

type = item traded on market

market = unique region/type combination

order = individual order in market

bid order = buy order

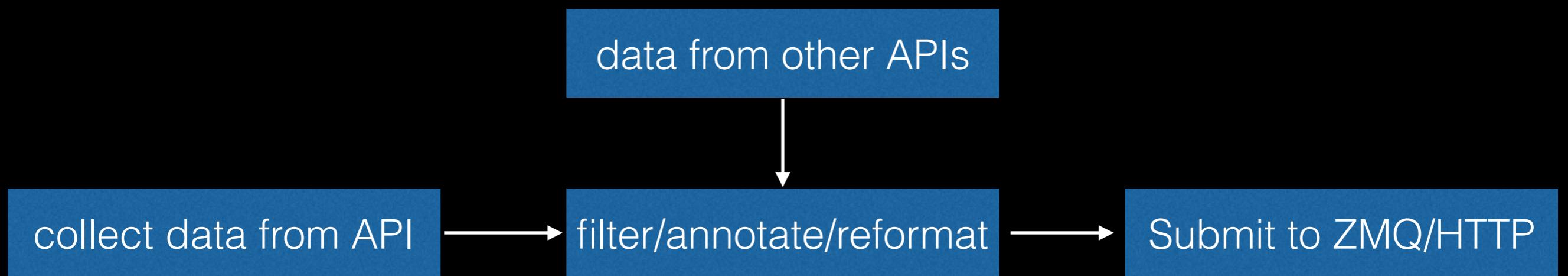
ask order = sell order

„Generate a ,closest to realtime as possible‘ feed of the current state changes of the global market.“

situation

- oAuth'ed JSON Hypermedia API
 - token expiration after a couple of minutes
- Guaranteed daily downtime of around 15 minutes
- „*The actual rate limit is set to something ridiculous, like 100,000 req/s.*“
- Very limited budget for scraping server

```
{  
  - motd: {  
    - dust: {  
      href: "http://newsfeed.eveonline.com/articles/71"  
    },  
    - eve: {  
      href: "http://client.eveonline.com/motd/"  
    },  
    - server: {  
      href: "http://client.eveonline.com/motd/"  
    }  
  },  
  - crestEndpoint: {  
    href: "https://crest-tq.eveonline.com/"  
  },  
  - corporationRoles: {  
    href: "https://crest-tq.eveonline.com/corporations/roles/"  
  },  
  - constellations: {  
    href: "https://crest-tq.eveonline.com/constellations/"  
  },  
  - itemGroups: {  
    href: "https://crest-tq.eveonline.com/inventory/groups/"  
  },  
  - . . .  
}
```



```
{  
    totalCount_str: "48",  
    - items: [  
        - {  
            volume_str: "224206141",  
            buy: true,  
            issued: "2016-05-23T02:22:59",  
            price: 5.8,  
            volumeEntered: 374000000,  
            minVolume: 1,  
            volume: 224206141,  
            range: "2",  
            href: "https://crest-tq.eveonline.com/market/10000002/orders/4537189778/",  
            duration_str: "90",  
            - location: {  
                id_str: "60003061",  
                href: "https://crest-tq.eveonline.com/universe/locations/60003061/",  
                id: 60003061,  
                name: "Mitsolen I - Moon 1 - Expert Housing Warehouse"  
            },  
            duration: 90,  
            minVolume_str: "1",  
            volumeEntered_str: "374000000",  
            - type: {  
                id_str: "34",  
                href: "https://crest-tq.eveonline.com/types/34/",  
                id: 34,  
                name: "Tritanium"  
            },  
            id: 4537189778,  
            id_str: "4537189778"  
        },  
    ]  
}
```

```
{
  "resultType" : "orders",
  "version" : "0.1",
  "uploadKeys" : [
    { "name" : "emk", "key" : "abc" },
    { "name" : "ec", "key" : "def" }
  ],
  "generator" : { "name" : "Yapeal", "version" : "11.335.1737" },
  "currentTime" : "2011-10-22T15:46:00+00:00",
  "columns" : [ "price", "volRemaining", "range", "orderID", "volEntered", "minVolume", "bid", "issueDate", "duration" ],
  "rowsets" : [
    {
      "generatedAt" : "2011-10-22T15:43:00+00:00",
      "regionID" : 10000065,
      "typeID" : 11134,
      "rows" : [
        [ 8999, 1, 32767, 2363806077, 1, 1, false, "2011-12-03T08:10:59+00:00", 90, 60008692, 30005038 ],
        [ 11499.99, 10, 32767, 2363915657, 10, 1, false, "2011-12-03T10:53:26+00:00", 90, 60006970, null ],
        [ 11500, 48, 32767, 2363413004, 50, 1, false, "2011-12-02T22:44:01+00:00", 90, 60006967, 30005039 ]
      ]
    },
    {
      "generatedAt" : "2011-10-22T15:43:00+00:00",
      "regionID" : 10000065,
      "typeID" : 11134,
      "rows" : [
        [ 8999, 1, 32767, 2363806077, 1, 1, false, "2011-12-03T08:10:59+00:00", 90, 60008692, 30005038 ],
        [ 11499.99, 10, 32767, 2363915657, 10, 1, false, "2011-12-03T10:53:26+00:00", 90, 60006970, null ],
        [ 11500, 48, 32767, 2363413004, 50, 1, false, "2011-12-02T22:44:01+00:00", 90, 60006967, 30005039 ]
      ]
    }
  ]
}
```

some simple math...

- 67 Regions
- 13,030 Types
- 5 minutes cache expiration and 2 requests per market
- $67 * 13,030 * 2 / 5 / 60 = 5,820 \text{ req/s}$

... okay, maybe that's a little too much

- only about 400k markets actually have orders
- $380,000 * 2 / 5 / 60 = 2,533 \text{ req/s}$

basic considerations

- How to only submit changed markets?
 - Generate hash of response and only submit changed markets.
- How to handle inactive markets which might become active?
 - Scrape those at way lower rate to save requests.

„We ran a million processes (sequentially) in about 5 1/2 seconds. And, as this graph shows, the time per process was pretty much linear once we overcame the startup time.

This kind of performance is stunning, and it changes the way we design code. We can now create hundreds of little helper processes. And each process can contain its own state—in a way, processes in Elixir are like objects in an object-oriented system (but they have a better sense of humor).“

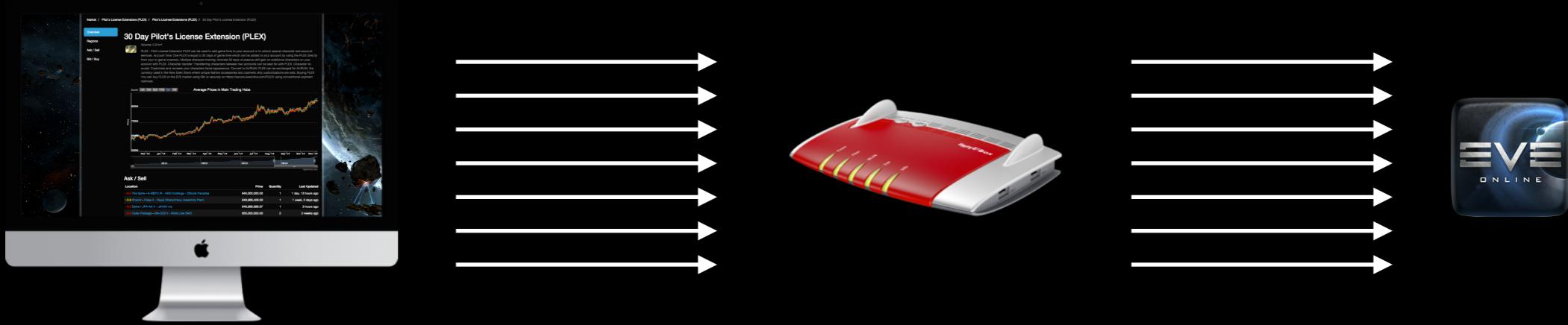
– *Dave Thomas in „Programming Elixir“*

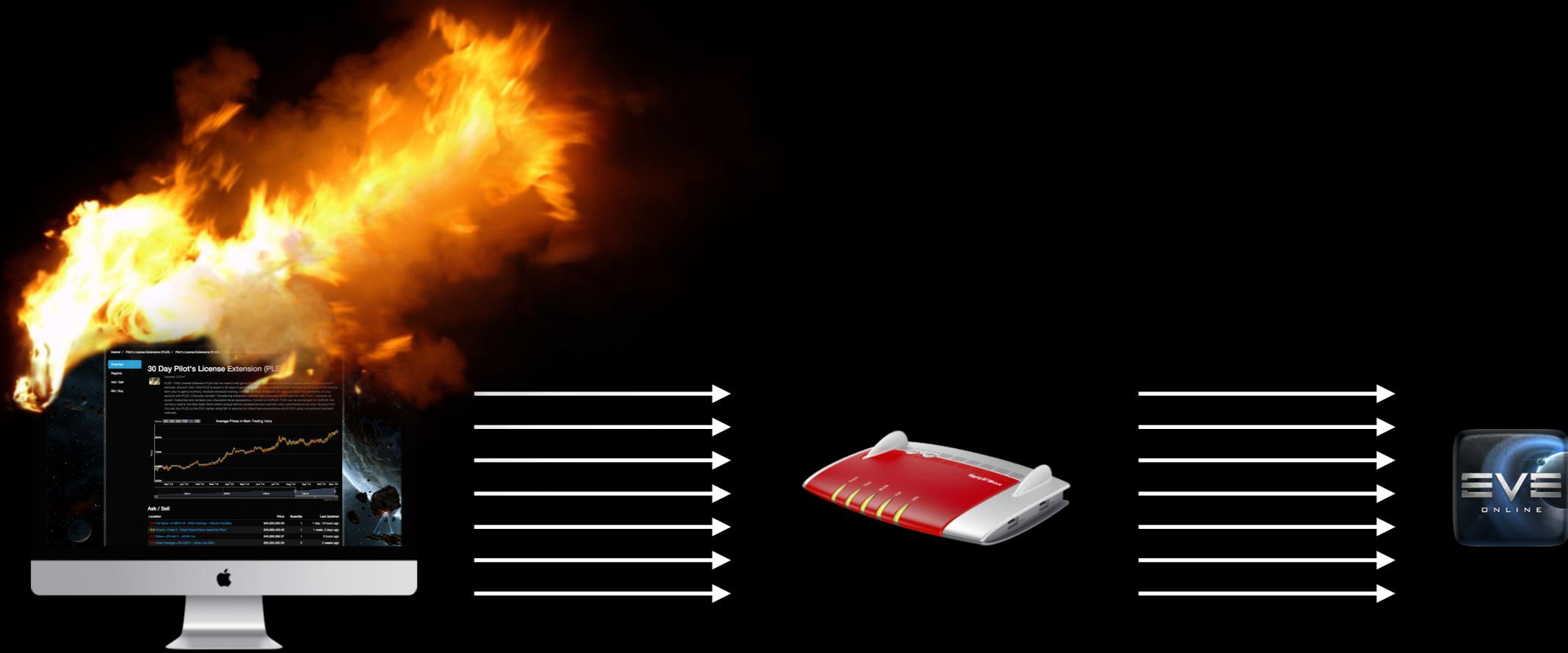
the naïve implementation

- one GenServer per market
- spread out startup time
- re-schedules itself
 - **Lesson #1:** Do not use the :timer module for these scenarios
 - Better: Process.send_after/3

the naïve implementation

- Hash stored in state parameter
 - we'll be doing lots of hashing
 - erlang-murmurhash is a reasonably fast NIF
 - **Lesson #2:** NIFs can bring down your whole VM!
- If market is empty, retry in 30 minutes, else in 5 minutes

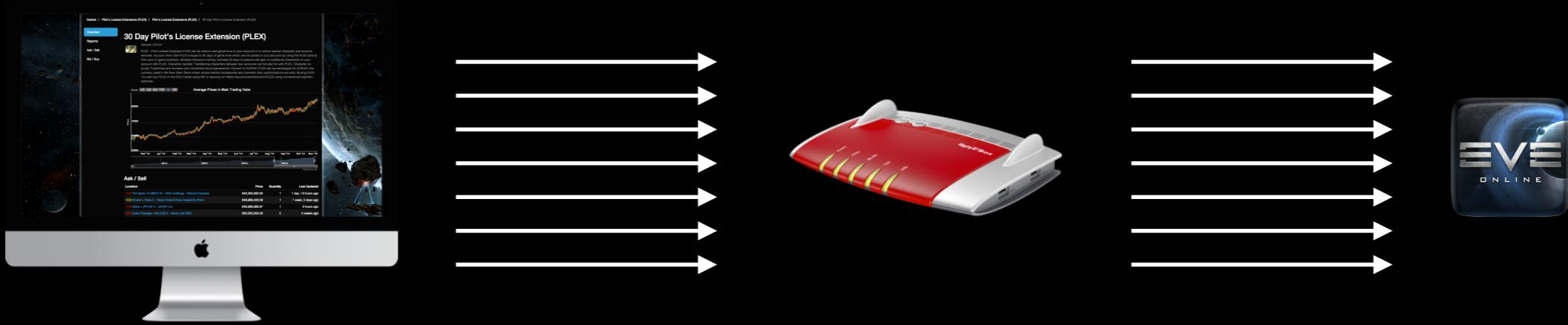


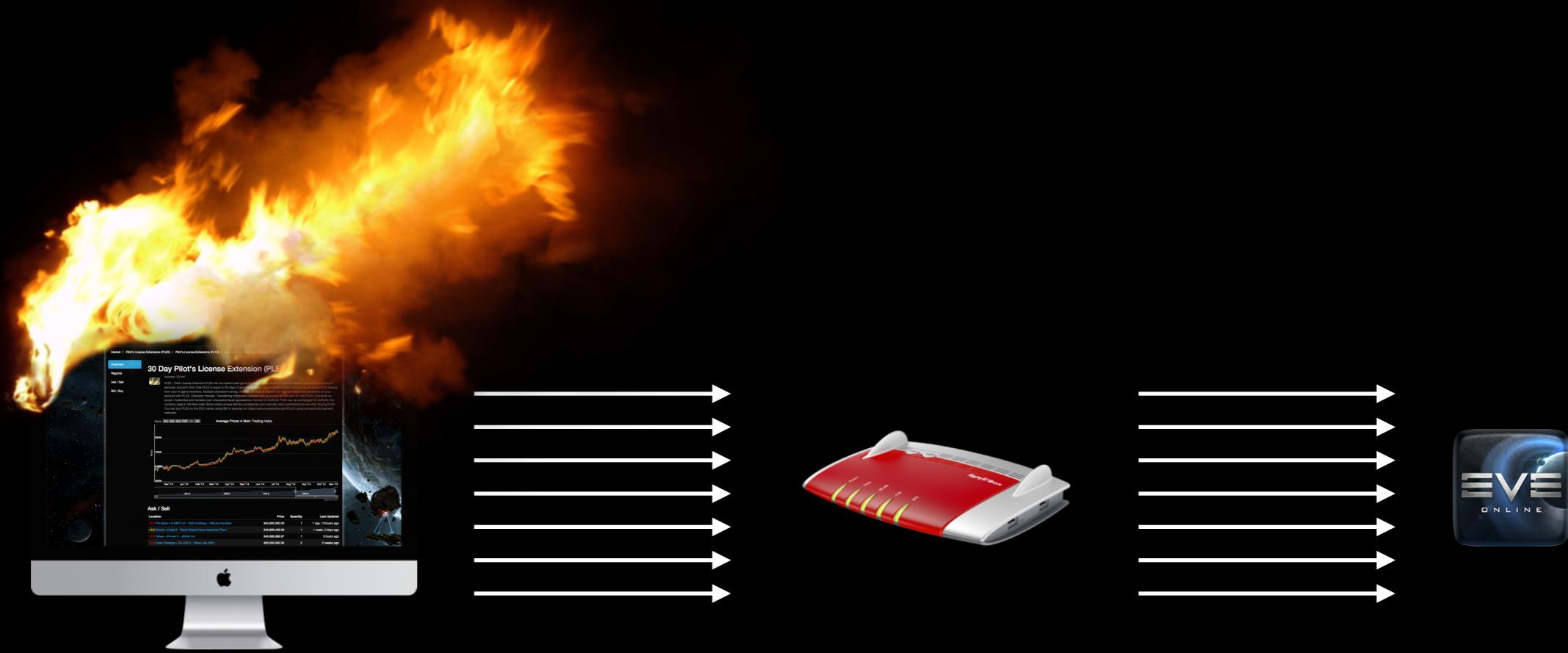


oh well...

What happened?

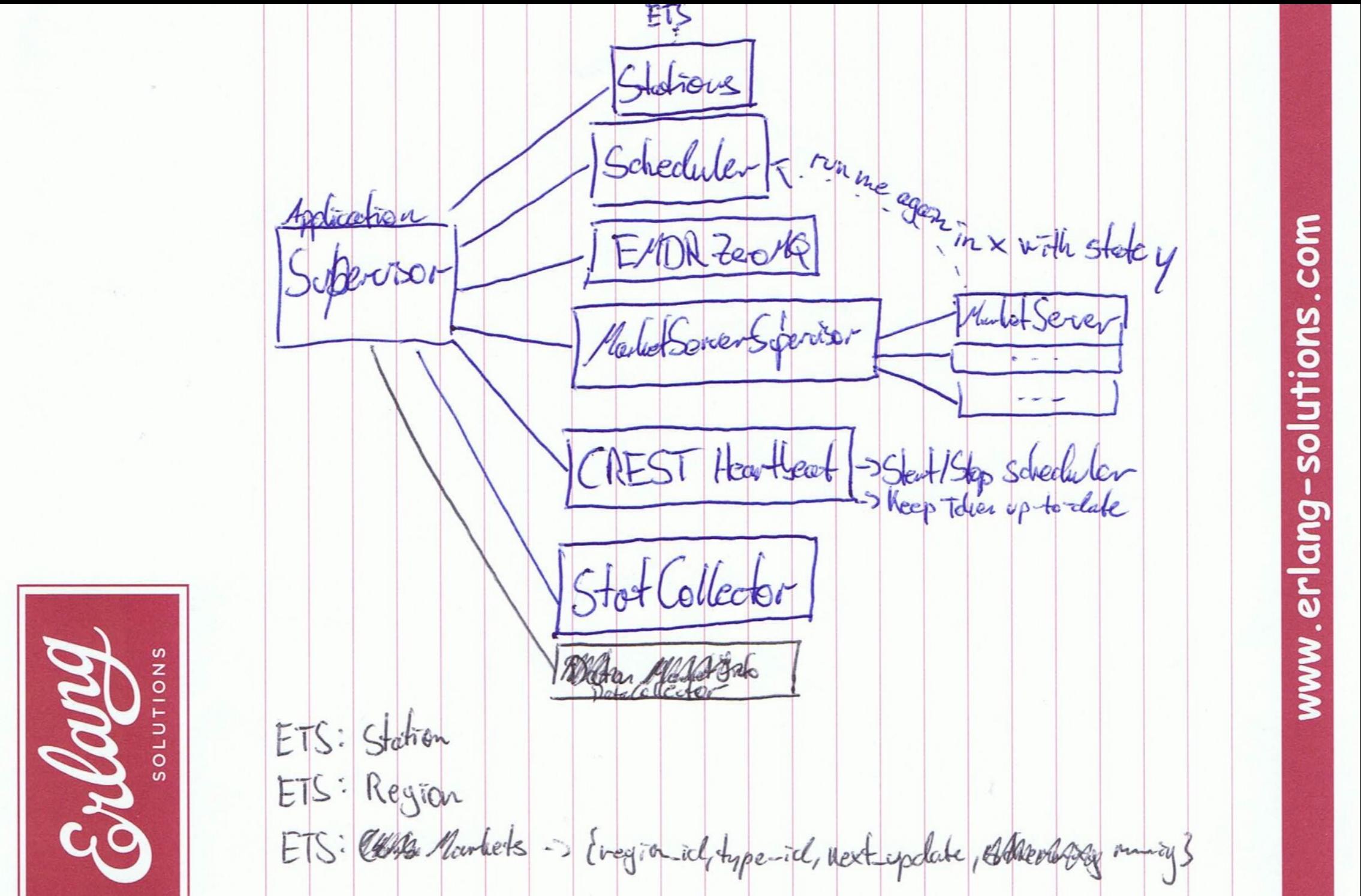
- **Lesson #3:** Turns out 830k processes use quite a lot of RAM
- **Lesson #4:** You can let processes „sleep“ for saving memory, once they receive a message, they wake up
 - erlang:hibernate/3





maybe take a step back

- What happens if a process dies mid-execution?
 - trap exits?, central state-keeping instance?
- How to deal with oAuth token expiration?
 - Maybe running into authentication errors at 6k req/s is not that nice
- Using observer @830k processes is no fun
- Time for serious software engineering

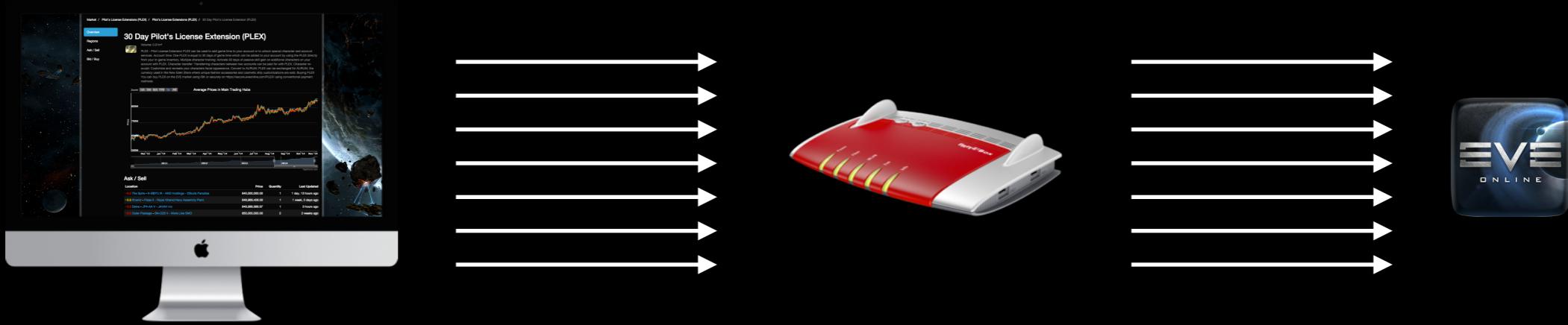


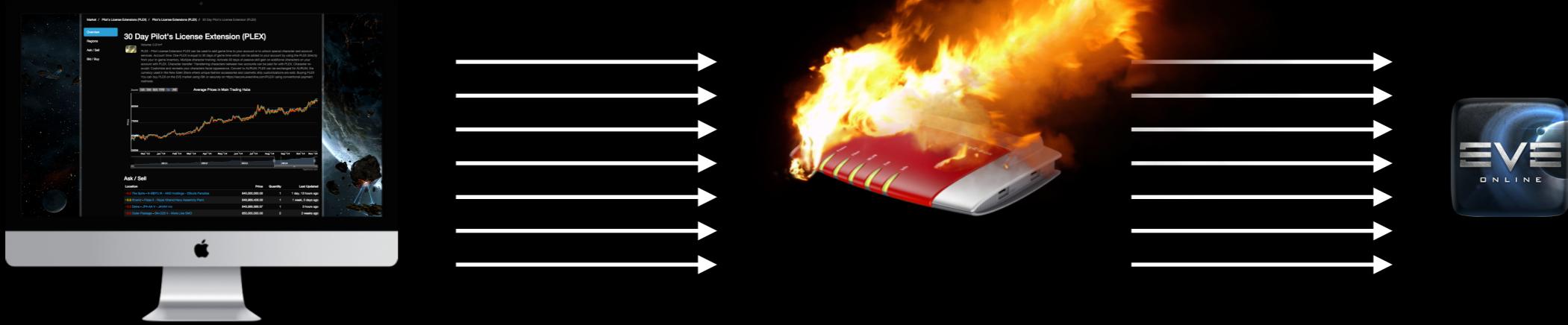
```
# Get all expired, non-started region/type pairs
query = [
    {
        {:_, :"$1", :_ , :"$2"},

        [{:andalso,
            {:<, :"$1", now},
            {==, :"$2", false}}
        ],
        [":$_"]
    }
]
```

writing ETS queries

- **Lesson #5:** You do not need to write simple match specs by hand
- [ets:fun2ms\(\)](#)
 - Turns functions into match_specs - magic!

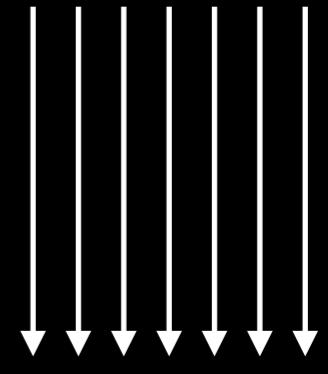
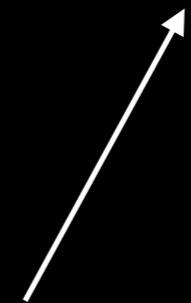




nooo - why again?

- **Lesson #6:** Consumer-grade hardware is not built for handling that many connections at that rate
 - What I did not know was that I was using HTTPoison incorrectly (no pooling? -> many connections)
 - But that could not stop me - any ideas?

some VPN



SUCCESS!

- Lesson #7: You can use VPNs to ,bypass' some network limitations
- Stuff crashes horribly once the oAuth token expires and there are a lot of HTTP errors but apart from that - nice
- Maybe compression would be nice
 - Leads to Lesson #8 and #9

```
# gunzip response
response = case response.headers["Content-Encoding"]
  "gzip" -> %{response | body: :zlib.gunzip(response.body)}
  nil      -> response
end
```

inflate != unzip != gunzip

months pass

developments

- Old data source broke so the scraper was needed
- API no longer authenticated
 - ... but limited to 20 connections @ 150 r/s
- Still getting a lot of HTTP errors

Switched to HTTPotion and the errors vanished

```
:ibrowse.set_dest(@crest_base_host |> String.to_char_list,  
                  443,  
                  max_sessions: @crest_pool_size,  
                  max_pipeline_size: @crest_pipeline_size)
```

Lesson #10: ibrowse's connection pooling/pipelining
is easy to set up

...how to get around rate limits

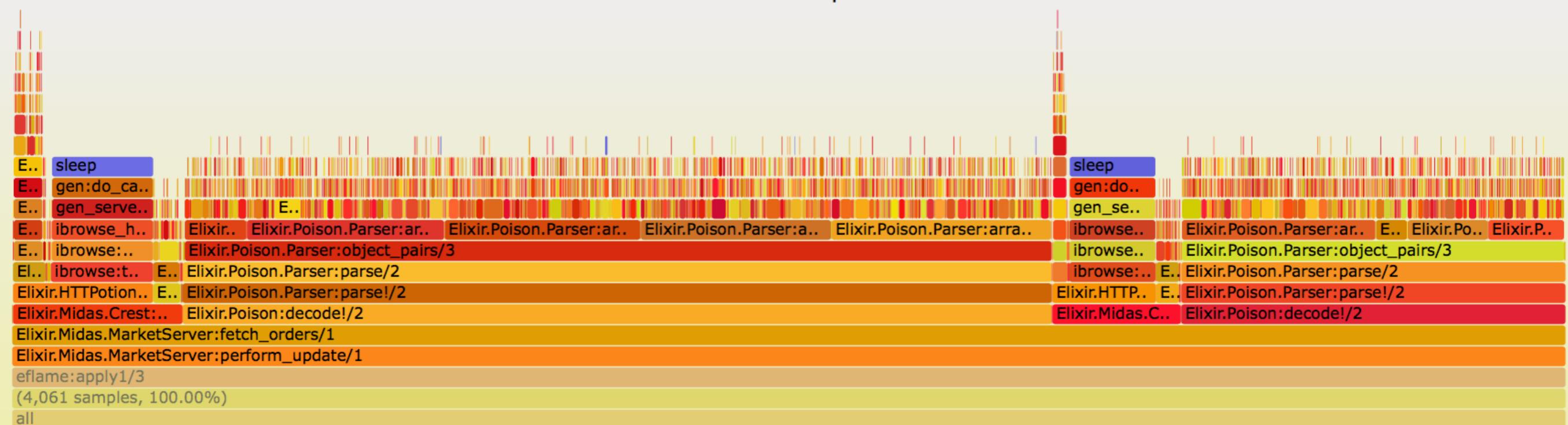
- Lesson #11 - bypassing rate limits for public APIs
 - Get VPS, add some IPv4 IPs
 - Perform some iptables magic
 - Have a different source IP for each *n-th* connection
- Lesson #12: ...or just talk to the real nice devs of the API you're consuming - communication is key!
- Lesson #13: tcptrack is a very nice tool for checking your connection pool

CPU optimization

- Most of CPU time spent on checking ETS
 - :set, :ordered_set
 - :read_concurrency
- Started profiling using flame
- **Lesson #15:** Actually thinking about your data structures helps a lot. Data is your problem - not code!

Reset Zoom

Flame Graph

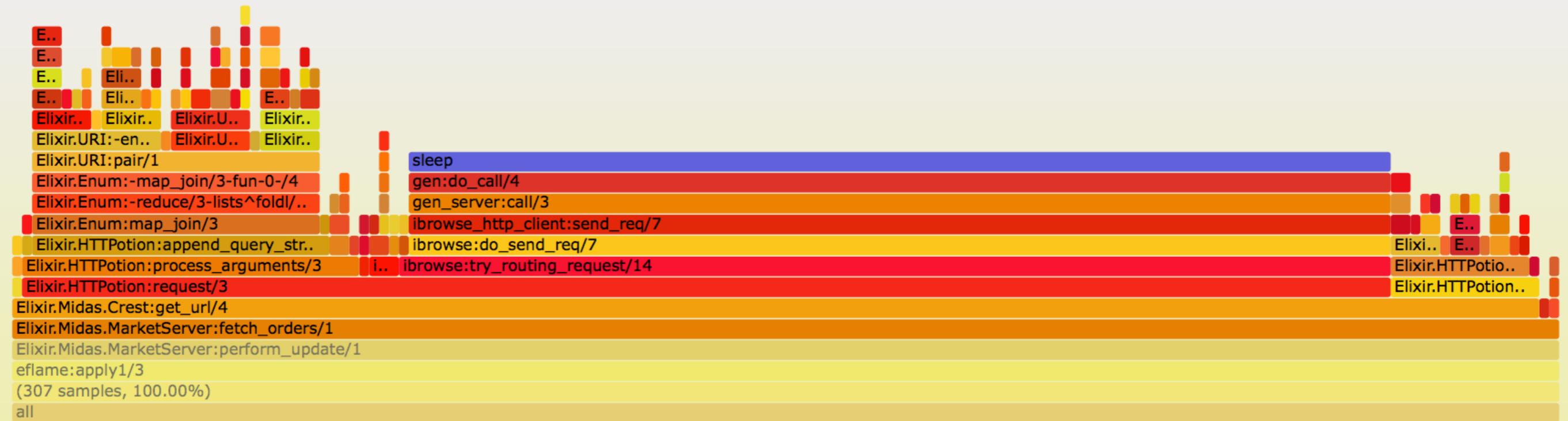


... when JSON parsing is your bottleneck

- „*Poison is a new JSON library for Elixir focusing on wicked-fast speed without sacrificing simplicity, completeness, or correctness.*“
- Maybe I was doing it wrong
- Switched JSON parsing from Poison to Jiffy
 - Again: NIF
 - **Lesson #16:** When in doubt - NIF and hope it does not crash, also Jiffy can output nice maps!
 - `response.body |> :jiffy.decode([:return_maps])`

Reset Zoom

Flame Graph

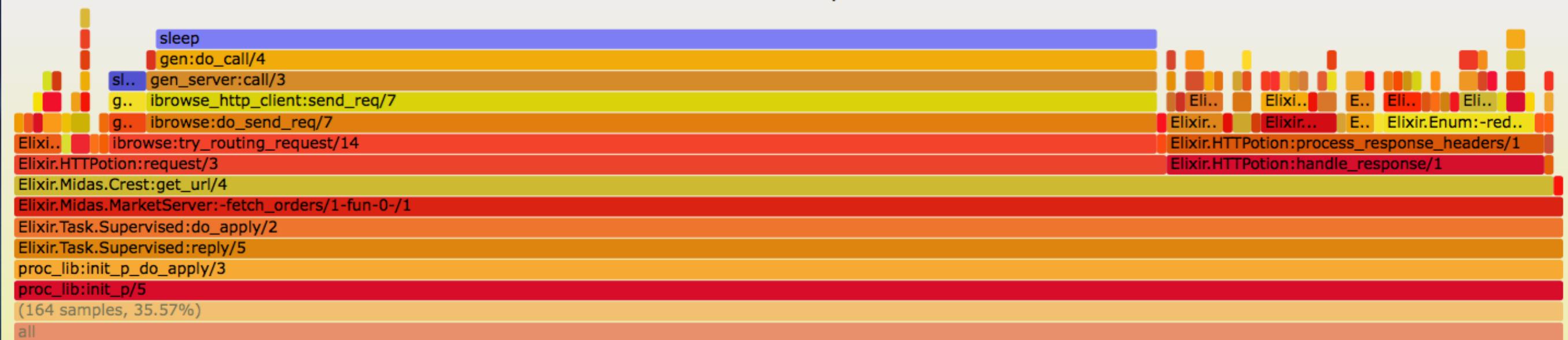


Lesson #17: Appending to the query string can be surprisingly slow

```
# Not via params, because it is way faster
orders_url = region[:orders_url] <> "?type=#{type[:url]}"
```

Reset Zoom

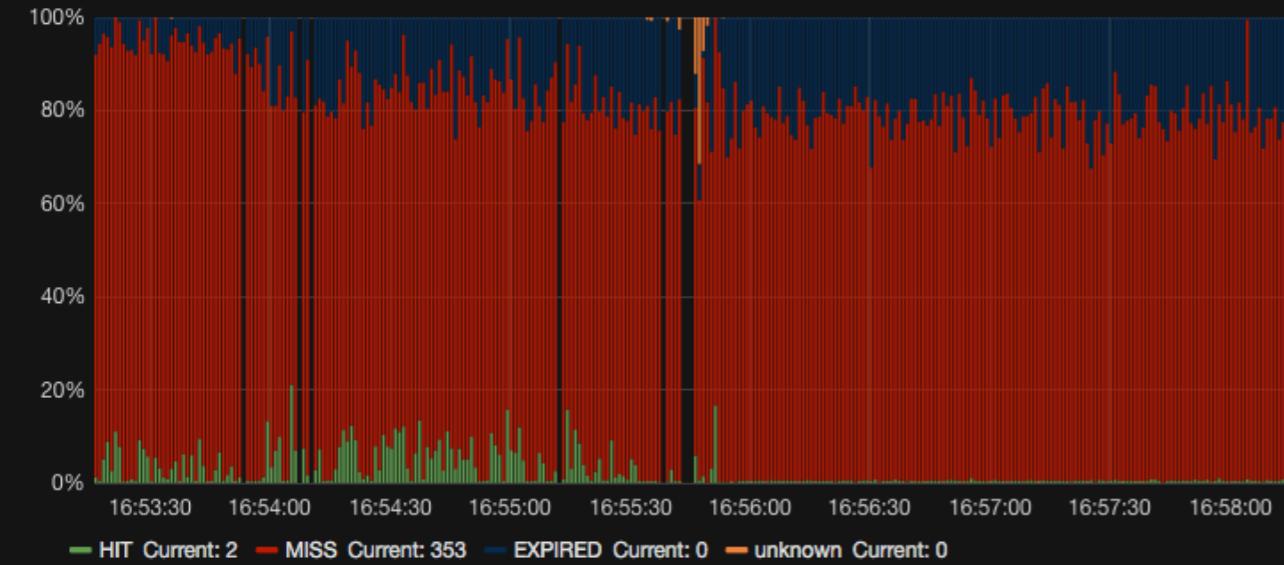
Flame Graph



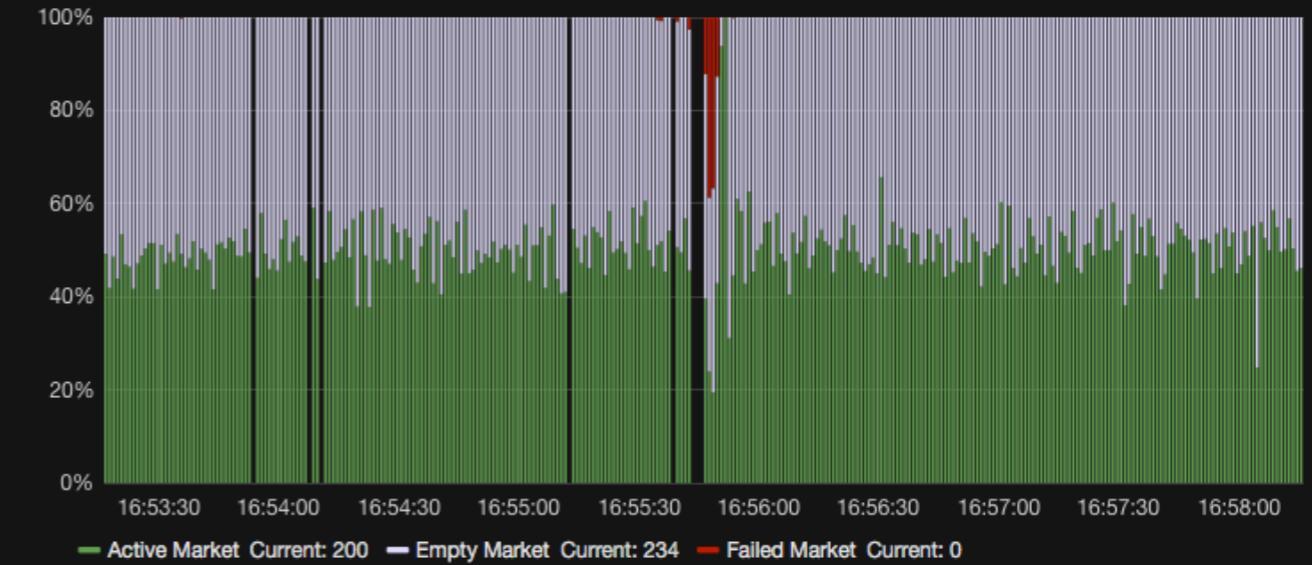
- Uncrashability - the final frontier
 - Added static data collection and updates
 - ...including custom CSV „parsing“
 - Adding influxDB metrics collection was simple
 - Added back-off for API errors
 - **Lesson #18:** Moved ETS tables to own process
 - Reworked supervision tree

CREST EMDR BRIDGE METRICS

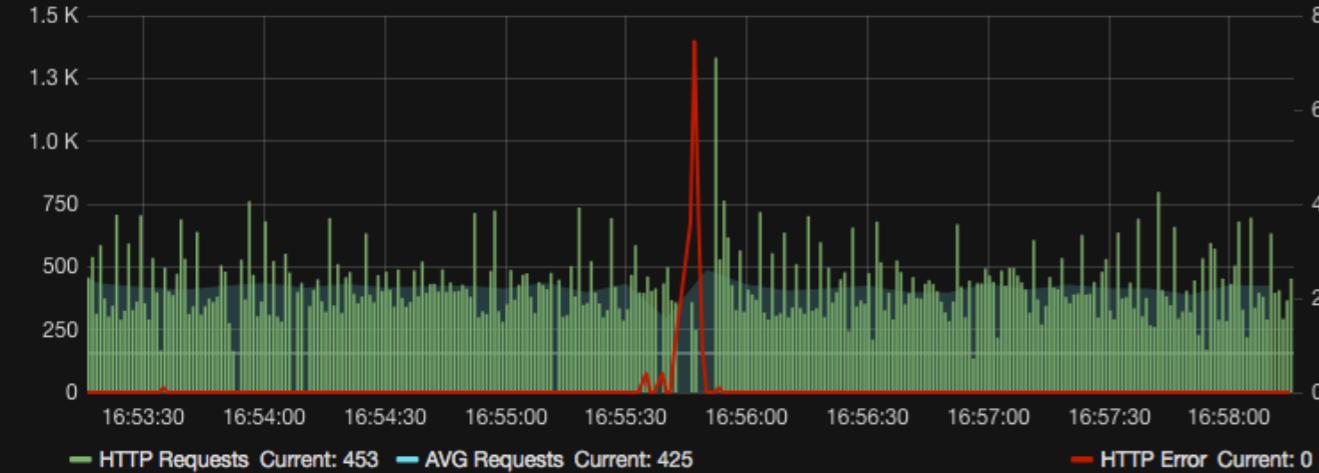
Caching Status



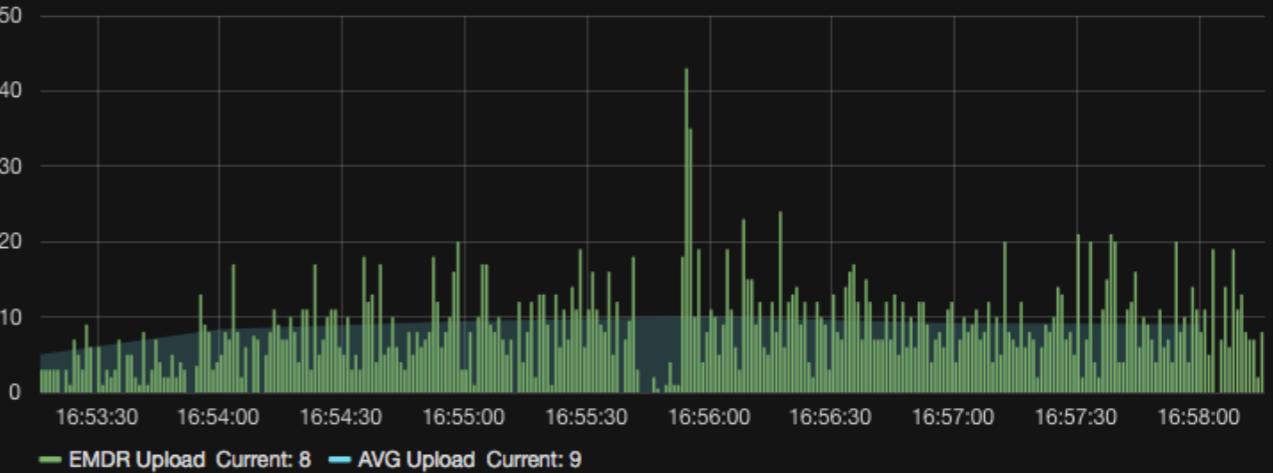
Market Activity



HTTP Requests



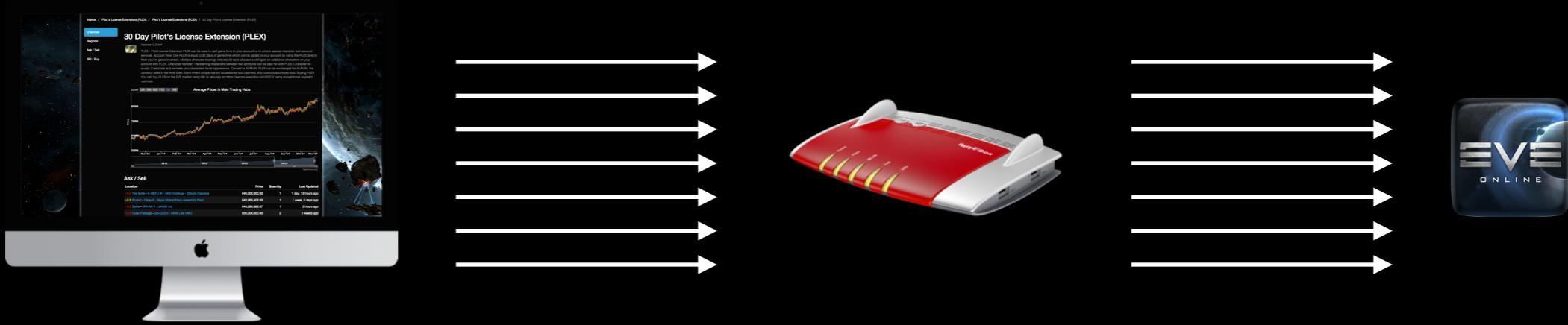
Upload Stats

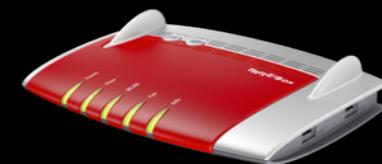
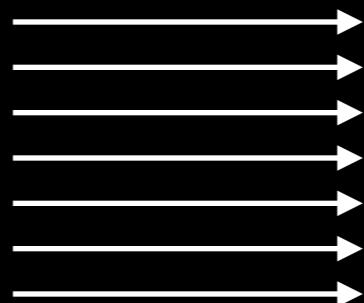


retrospective

- A lot of fun for the first project
- BEAM VM delivered stable multiple thousand requests per second and superior concurrency and availability
- New API makes our scraping process obsolete
 - Although useless now, I learned a ton
 - Next time I'll write tests first, I promise
 - Maybe try some type specs, despite mixed experiences

...one more thing





Thank you!