

# HOW ARBISOFT HELPED AEE REDUCE CRAWLTIME FROM 168 HOURS TO LESS THAN 4 HOURS



Data Analytics



Tailor-made data extraction



Semantic analysis



Large scale crawls

## OVERVIEW

Advanced Energy Economy is a group of businesses on a mission to nurture a prosperous economy powered by secure, sustainable, clean and affordable energy. They do this using policy advocacy, analysis and education--and to do that they need access to data. A lot of data. Their product, PowerSuite, collects vast amounts of public data from Public Utility Commission websites for each U.S. state and then provides it to PowerSuite subscribers in an easy-to-understand aggregate they can use to obtain updated information, track long-term trends, and most importantly: make better decisions.

## THE PROBLEM

To be competitive, PowerSuite required daily data crawls of 50 large PUC websites. This entailed massive overhead costs in time, performance, and API service expenses. There had to be a more efficient way of doing this--and they came to Arbisoft to find it.

## THE SOLUTION

We implemented a custom-designed distributed crawling mechanism that used automated intelligent data extractors capable of semantic analysis. We optimized their system and added the ability to clean up data in realtime, further cutting down time between data collection and data usability. We also provided a REST API that fed data to their user-facing Ruby on Rails app. The high-performance API was initially implemented using Google App Engine and later transitioned to the Django REST Framework app.

## Summary

- Industry: Energy sector
- Client relationship: Since 2013
- Challenge: Daily data crawls of 50 large PUC websites
- Solution: customised crawlers, high performance API for Ruby on Rails app
- Impact: 97.6% decrease in crawl duration, \$600 monthly savings, \$7,200 in savings per year

Level 21 Tower 2 Darling  
Park, 201 Sussex Street,  
Sydney NSW  
**Australia**

2035 Central Cir, Suite  
#201 McKinney, TX,  
75069,  
**USA**

Cisma RKT, 3rd Floor,  
Block C Jalan Raja  
Abdullah, 50300, KL  
**Malaysia**

25 Westwood  
Colony, Lahore,  
54000  
**Pakistan**

# HOW ARBISOFT HELPED AEE REDUCE CRAWLTIME FROM 168 HOURS TO LESS THAN 4 HOURS

## Immediate Results

Our work made data fetches hyperefficient, reducing crawler runtime from over a week (168+ hours) to less than 4 hours. This represented a whopping decrease of 97.6% in minimum runtime duration, significantly enhancing the crawlers' ability to capture and present new information to subscribers as quickly as possible.

Despite the massive increase in performance, the expense of data fetches also reduced to half of that incurred by the conventional method. Shifting the API from Google App Engine to the Django REST Framework also yielded an additional \$600 in savings per month for an approximate \$7200 in savings per year.

Reach Out to Us Today to Learn What Arbisoft Can Do For You

[CONTACT US](#)

Level 21 Tower 2 Darling  
Park, 201 Sussex Street,  
Sydney NSW  
**Australia**

2035 Central Cir, Suite  
#201 McKinney, TX,  
75069,  
**USA**

Cisma RKT, 3rd Floor,  
Block C Jalan Raja  
Abdullah, 50300, KL  
**Malaysia**

25 Westwood  
Colony, Lahore,  
54000  
**Pakistan**