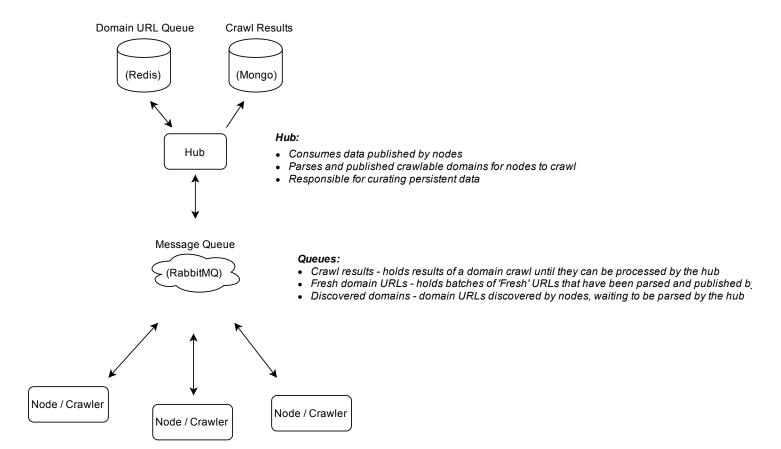
6/8/2015 Preview



Crawler node:

- Crawls domains in parallel (thread count specified at runtime)
- Adheres a domain's robots.txt rules (crawl rate, disallows, sitemaps)
- Compresses results and sends them up to message queue service to be stored by the hub

https://www.draw.io/