

# Scaling the API Economy With Scale-Free Networks

Mike Amundsen  
Principal API Architect  
Layer 7 Technologies  
@mamund



@mamund  
#ScaleFreeAPIs

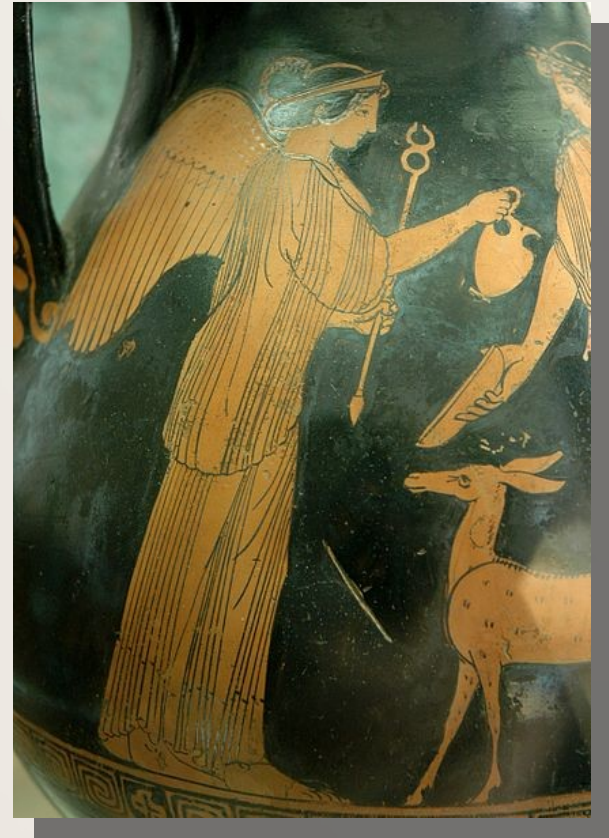
# I come as a messenger...



@mamund  
#ScaleFreeAPIs

# The Goddess Iris

- Iris is associated with
  - communication
  - new endeavors



# I bring hard news...



@mamund  
#ScaleFreeAPIs

# Iridium (Ir)

- Very hard & brittle
- Second-densest element
- Most corrosion-resistant
- Often found within impact craters
- Used in deep space satellites



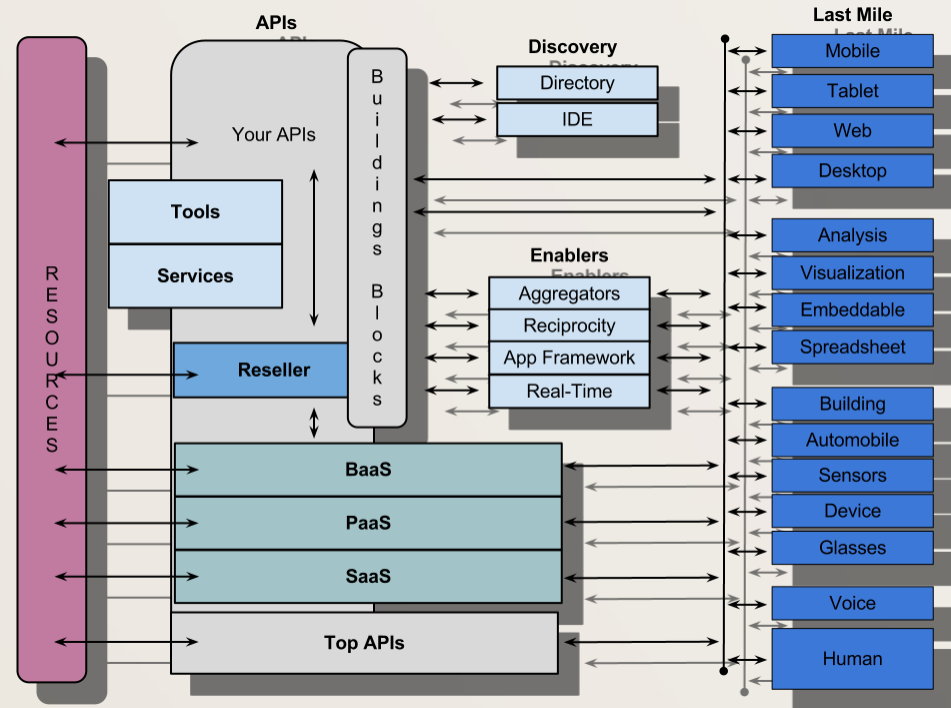
# My news is not good...



@mamund  
#ScaleFreeAPIs

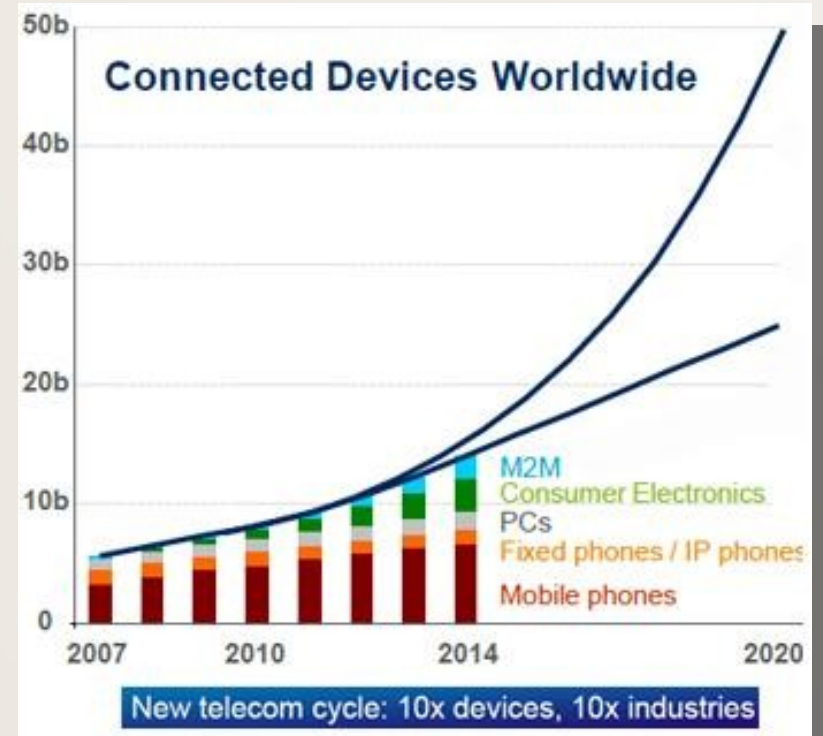
# My news is not good...

- The most common approach to Internet-based APIs runs counter to a fundamental property the Web



# My news is not good...

- At our current pace of growth, we may not be able to keep this up for much longer.



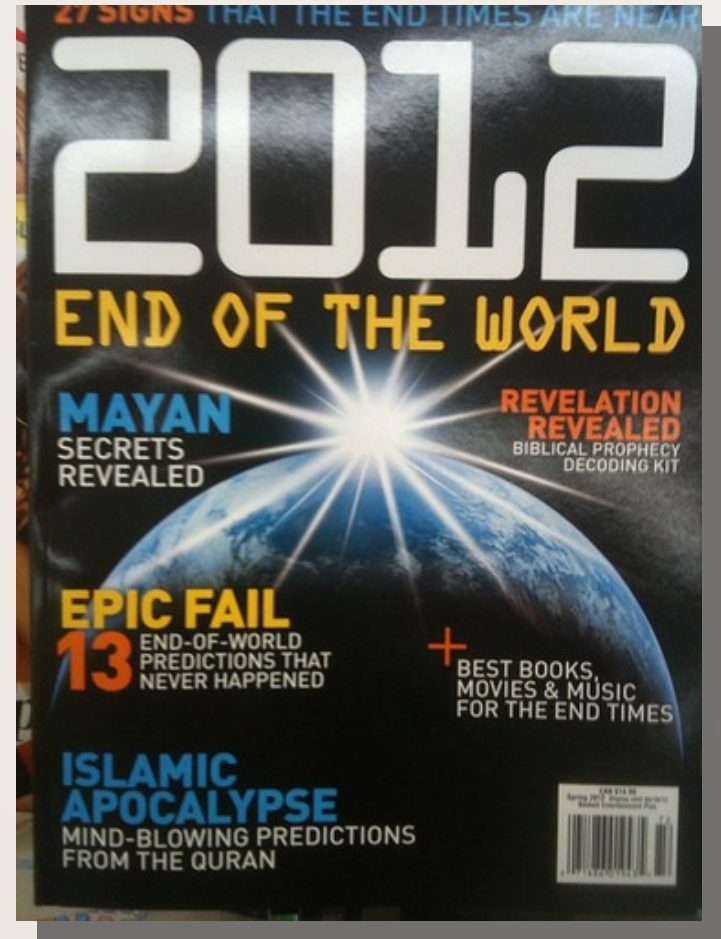
@mamund  
#ScaleFreeAPIs



# My news is not good...

- We may be headed for:

***“Scale-pocalypse!”***



@mamund  
#ScaleFreeAPIs

*But wait,  
I'm getting ahead of myself...*



@mamund  
#ScaleFreeAPIs

# *Scale-pocalypse 1998*



@mamund  
#ScaleFreeAPIs

# 1973 – Lansing, Michigan

- I start my first year at Michigan State University
- The same year, same city, Larry Page is born.



@mamund  
#ScaleFreeAPIs

# 1994 – Stanford University

- Jerry Yang & David Filo

- *"Jerry's guide to the world wide web"*

- **YAHOO!**

- Curated, hierarchical index



@mamund  
#ScaleFreeAPIs

# 1998 – Stanford University

- Sergey Brin & Larry Page
- **The Anatomy of a Large-Scale Hypertextual Web Search Engine (1998)**
- *“...due to rapid advance in technology and web proliferation, creating a web search engine today is very different from three years ago.”*

## The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page

Computer Science Department,  
Stanford University, Stanford, CA 94305, USA  
sergey@cs.stanford.edu and page@cs.stanford.edu

### Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at <http://google.stanford.edu/>. To engineer a search engine is a challenging task. Search engines index tens to hundreds of millions of web pages involving a comparable number of distinct terms. They answer tens of millions of queries every day. Despite the importance of large-scale search engines on the web, very little academic research has been done on them. Furthermore, due to rapid advance in technology and web proliferation, creating a web search engine today is very different from three years ago. This paper provides an in-depth description of our large-scale web search engine -- the first such detailed public description we know of to date. Apart from the problems of scaling traditional search techniques to data of this magnitude, there are new technical challenges involved with using the additional information present in hypertext to produce better search results. This paper addresses this question of how to build a practical large-scale system which can exploit the additional information present in hypertext. Also we look at the problem of how to effectively deal with uncontrolled hypertext collections: where anyone can publish anything they want.

### Keywords

World Wide Web, Search Engines, Information Retrieval, PageRank, Google

### 1. Introduction

*(Note: There are two versions of this paper -- a longer full version and a shorter printed version. The full version is available on the web and the conference CD-ROM.)*

The web creates new challenges for information retrieval. The amount of information on the web is growing rapidly, as well as the number of new users inexperienced in the art of web research. People are likely to surf the web using its link graph, often starting with high quality human maintained indices such as Yahoo! or with search engines. Human maintained lists cover popular topics effectively but are subjective, expensive to build and maintain, slow to improve, and cannot cover all esoteric topics. Automated search engines that rely on keyword matching usually return too many low quality matches. To make matters worse, some advertisers attempt to gain people's attention by taking measures meant to mislead automated search engines. We have built a large-scale search engine which addresses many of the problems of existing systems. It makes especially heavy use of the additional structure present in hypertext to provide much higher quality search results. We chose our system name, Google, because it is a common spelling of googol, or  $10^{100}$  and fits well with our goal of building very large-scale search



@mamund  
#ScaleFreeAPIs

# What was different?

- What did Brin & Page perceive that others had missed?
- How did Google avoid the “Scale-pocalypse”?



***“Free” as in “Scale-Free”***



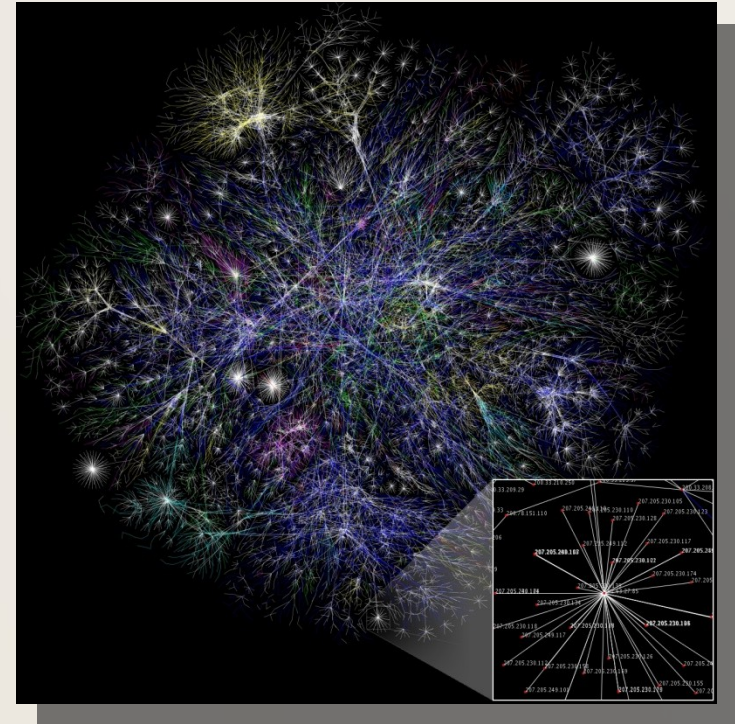
@mamund  
#ScaleFreeAPIs



# Scale-Free Network

- “A network whose degree distribution follows a power law.”

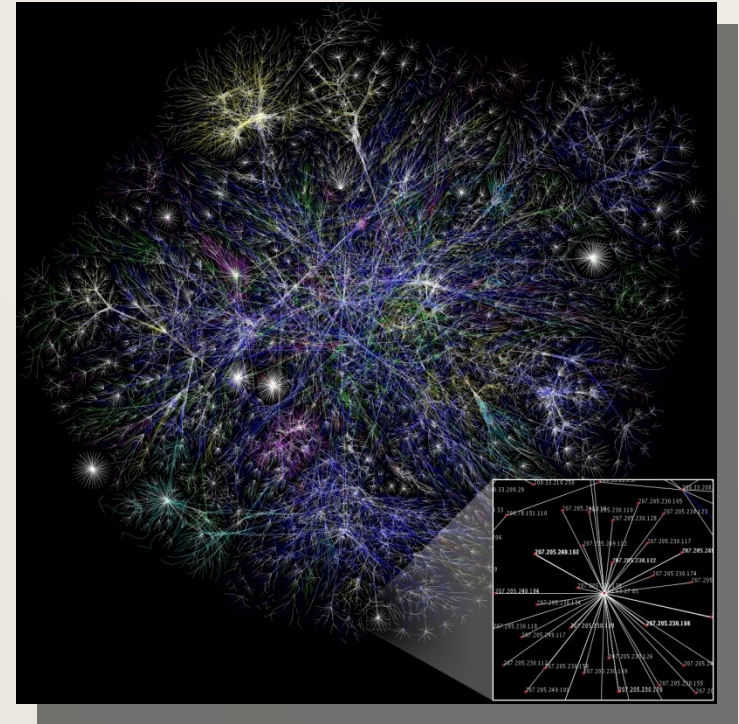
$$P(k) \sim k^{-\gamma}$$



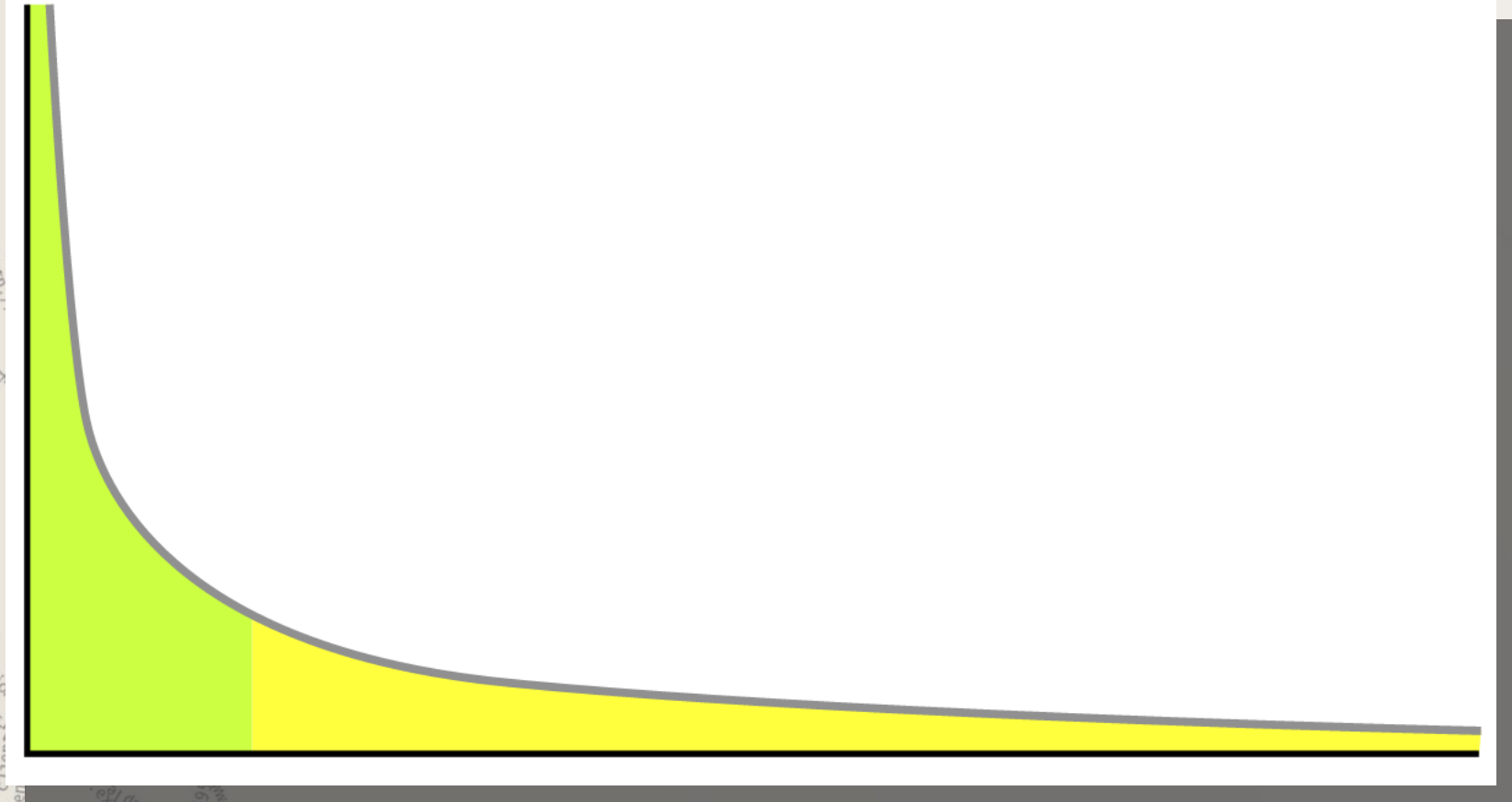
# Scale-Free Network

- “A network whose degree distribution follows a **power law**.”

$$P(k) \sim k^{-\gamma}$$



# Power Law

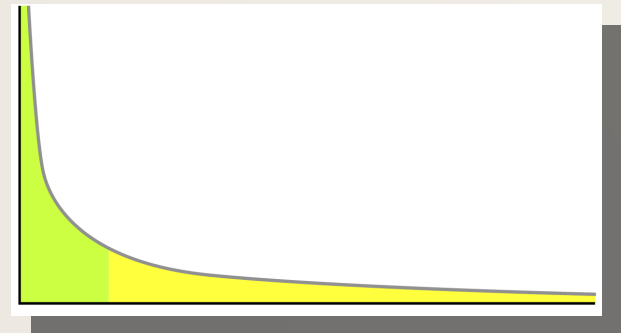


# “Long Tail”



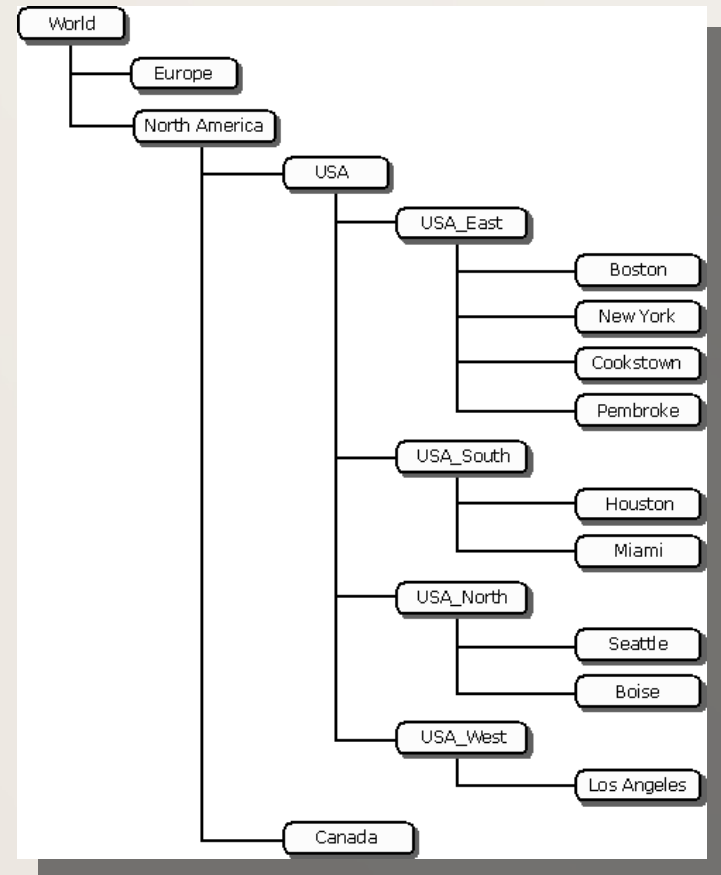
# “Long Tail”

- Not an equal distribution
- A few nodes have many of links
- Many nodes have a few links
- “Preferential Attachment”
  - Barabási–Albert (1999)



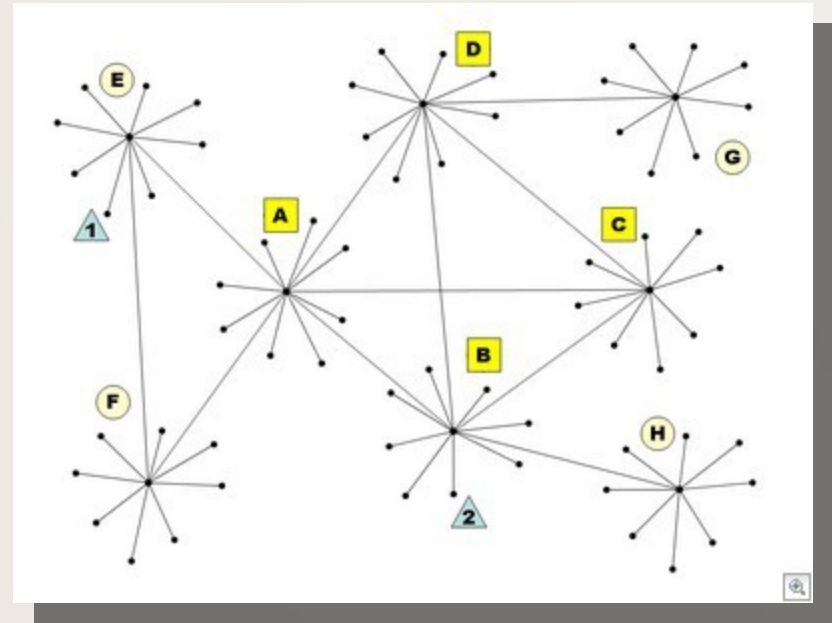
# Power Law Web Search

- Brin & Page recognized that
- curation and aggregation
- would not scale on the WWW



# Power Law Web Search

- Instead, they decided to
- allow users
- *and the links they create*
- to inform the search engine.



# Power Law Success

- Using this feature of the WWW allowed Google to
- gain the leading role in search
- and build a multi-billion dollar company.





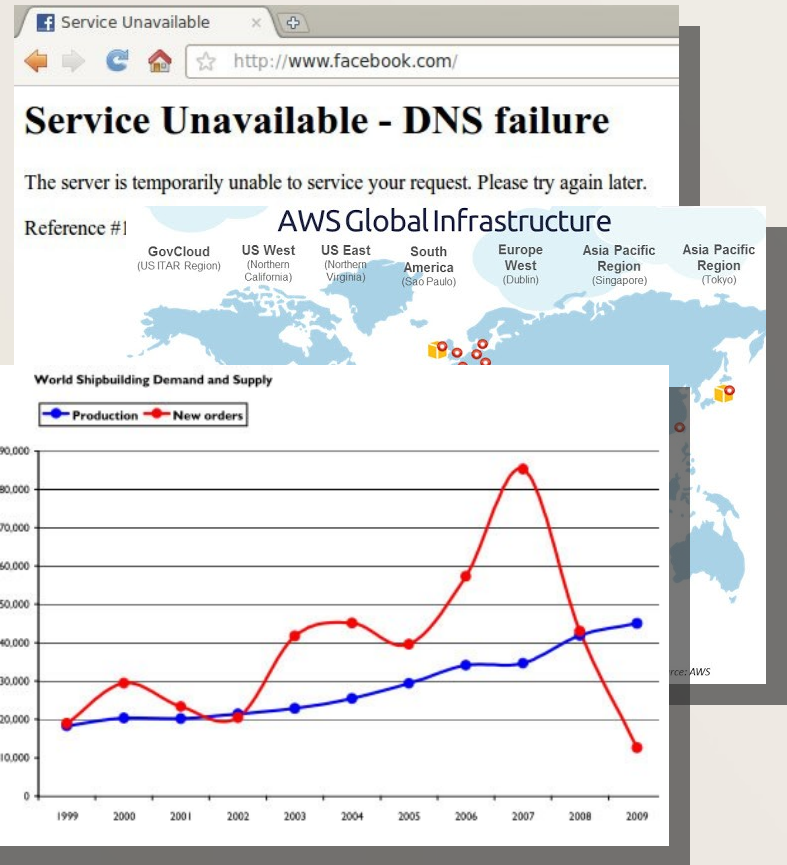
*Have we forgotten so soon?*



@mamund  
#ScaleFreeAPIs

# Hub-Vulnerability

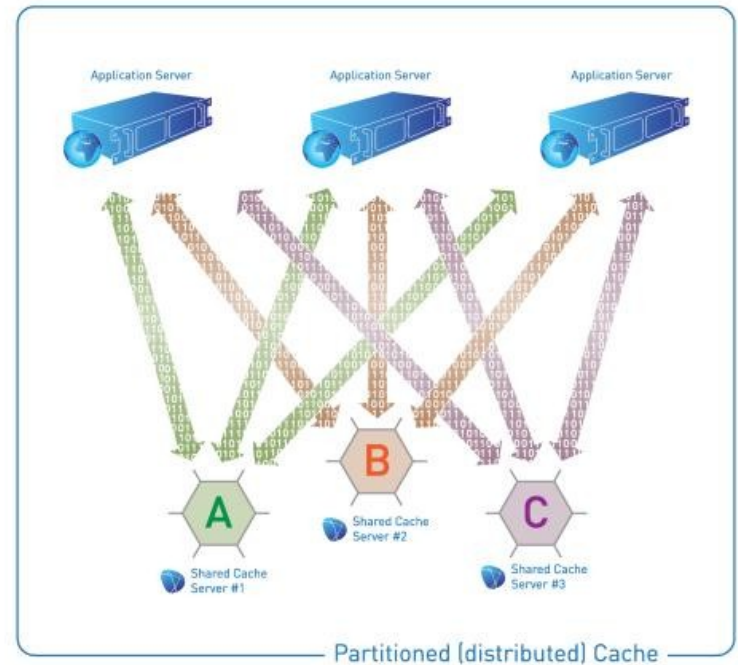
- Unexpected disasters can cause complex systems to fail.
- DNS failures of the 2000s
- Recent Amazon AWS failures
- 2008 Monetary system failures
- **“Too big to fail”**



@mamund  
#ScaleFreeAPIs

# Node-Resiliency

- Highly distributed systems reduce risk of system-wide failures.



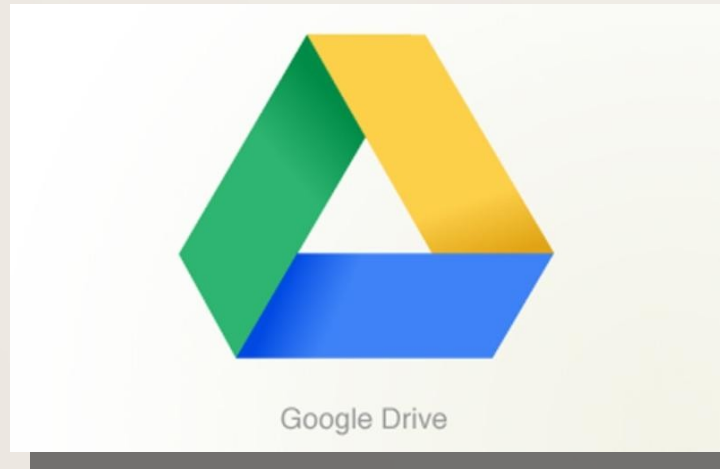
# *Does power-law only apply to search?*



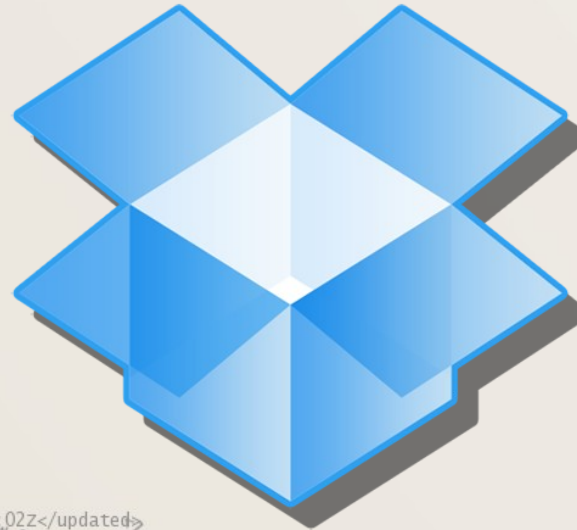
@mamund  
#ScaleFreeAPIs

# Storage

- Google Drive



- Dropbox



# Collaboration

- SVN



- git



@mamund  
#ScaleFreeAPIs

# Identity



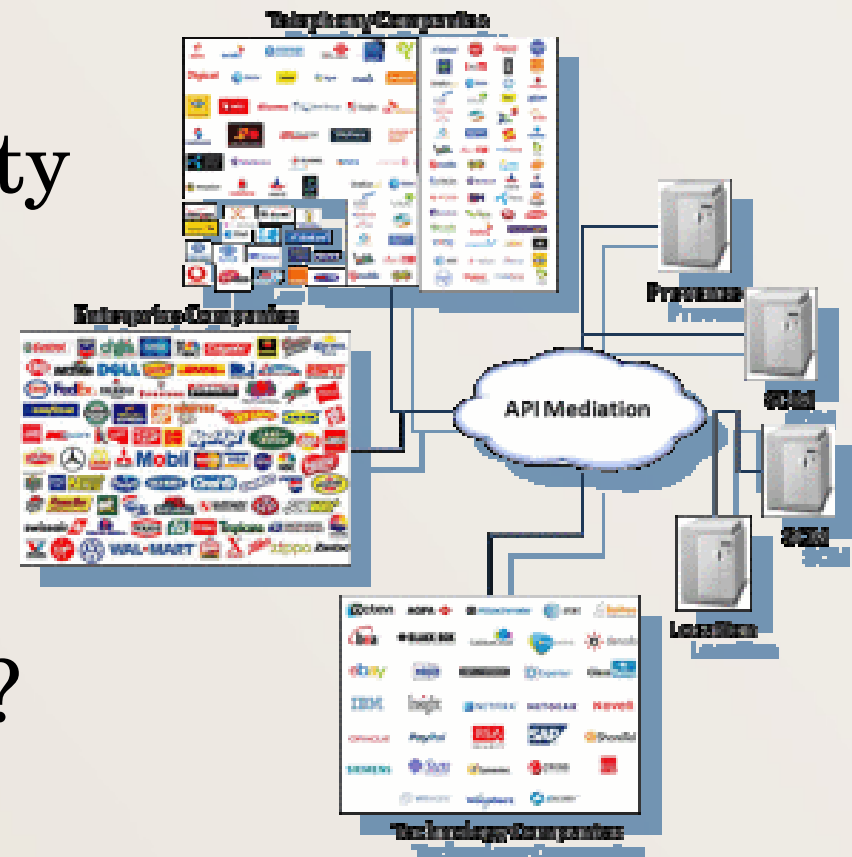
- Captive User Stores

- Multiple OAuth Stores



# APIs?

- Do we really want to add more vulnerability to the WWW?
- Can we afford to rely on curation and aggregation for APIs?





*“We can rebuild [it].*

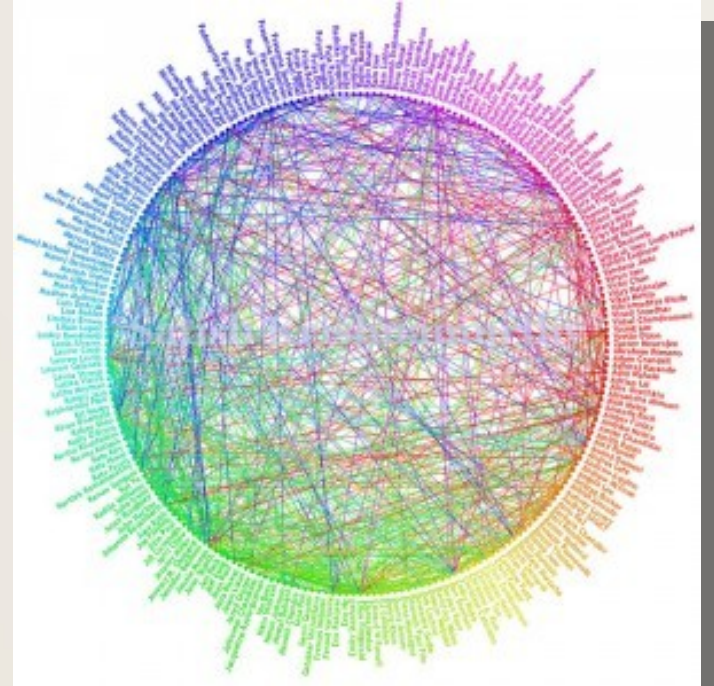
*We have the technology.”*



@mamund  
#ScaleFreeAPIs

# What if we changed our model?

- Could we create node-based services?
- What would we gain?
- What would we lose?



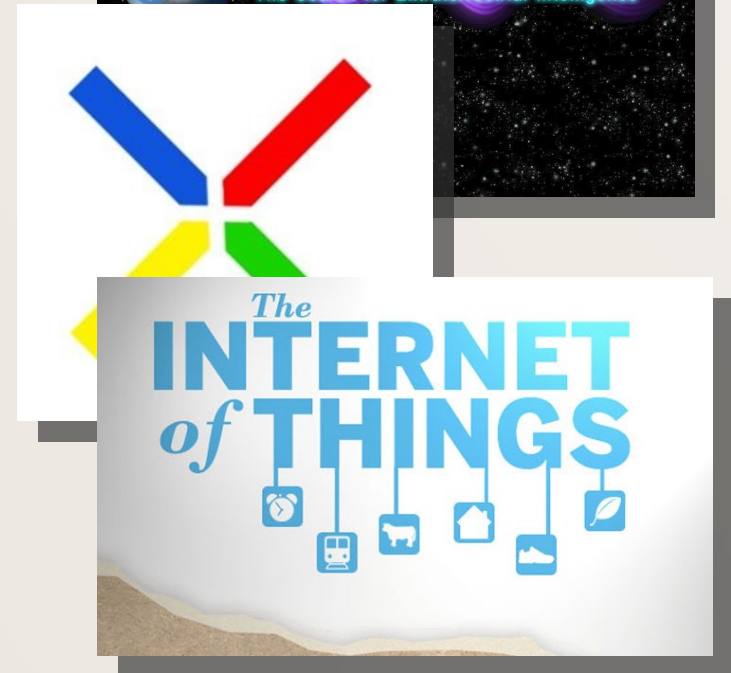
# Business models based on nodes

- Powerful client aggregators
- User-centric
  - Discovery
  - Linking
  - Sharing
- Allow users to identify new, popular services



# Advantages of node-based models

- Increased processing power (SETI)
- Access to more metadata (EXIF)
- User-driven selection models (reverse-flow)
- Reverse scale/vulnerability challenge
- Increase “network” intelligence
- Mix of “smart” and “dumb” clients (IoT)



@mamund  
#ScaleFreeAPIs

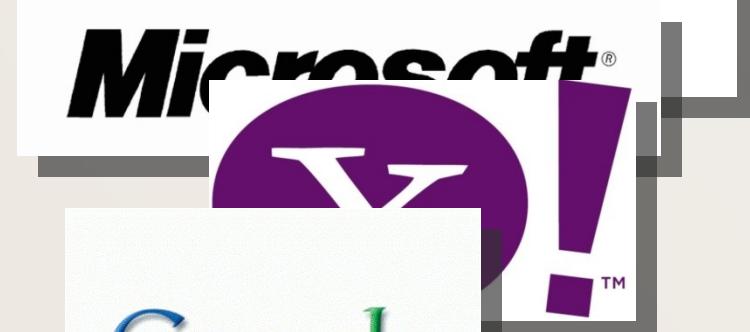
*Yes, but can you make money this way?*



@mamund  
#ScaleFreeAPIs

# Models/Paradigms

- IBM gladly let Microsoft own the OS (*IBM knew hardware was the real profit source*)
- Yahoo knew the value of curated indexes (*Google had another idea*)
- Few thought there was a business in Open Source movement.



@mamund  
#ScaleFreeAPIs

*Be a Node and not a Hub!*



@mamund  
#ScaleFreeAPIs

# Node-based?

- Twitter
- Facebook
- Google Docs
- Wordpress





# Node-based?

- I post messages to my own machine
- The client sends it to one or more selected services of my choice
- Publishers compete for my content
- Publishers distribute to other nodes



# Avoiding the Scale-pocalypse

- **API providers:**
- Consider hub vulnerability problem
- Explore node-based advantages
- Empower users and reap rewards







# So, my message is a positive one...

- We have an opportunity for new endeavors



@mamund  
#ScaleFreeAPIs

# So, my message is a positive one...

- We can build strong, vulnerability-resistant systems



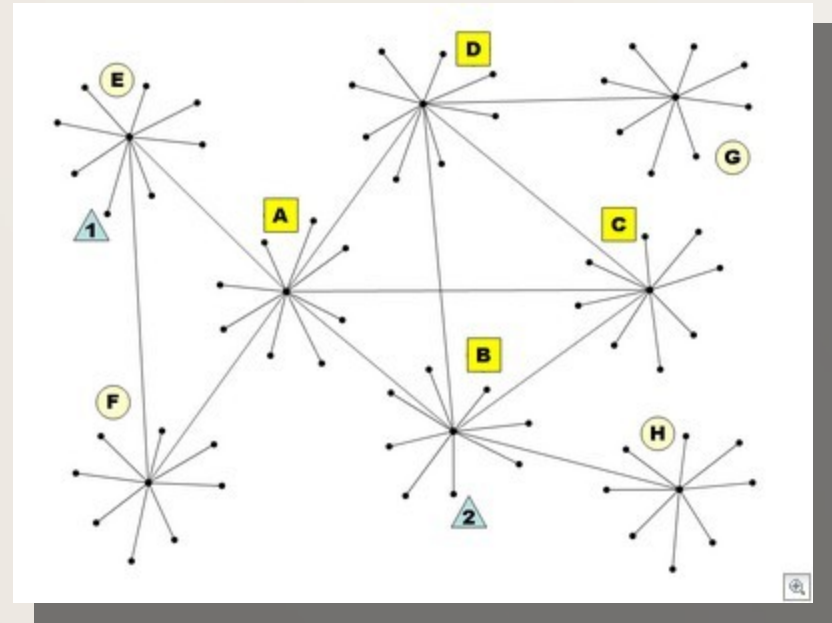
@mamund  
#ScaleFreeAPIs

# The news is not so bad...



# The news is not so bad...

- There is a way to build systems that take advantage of a key principal of the WWW

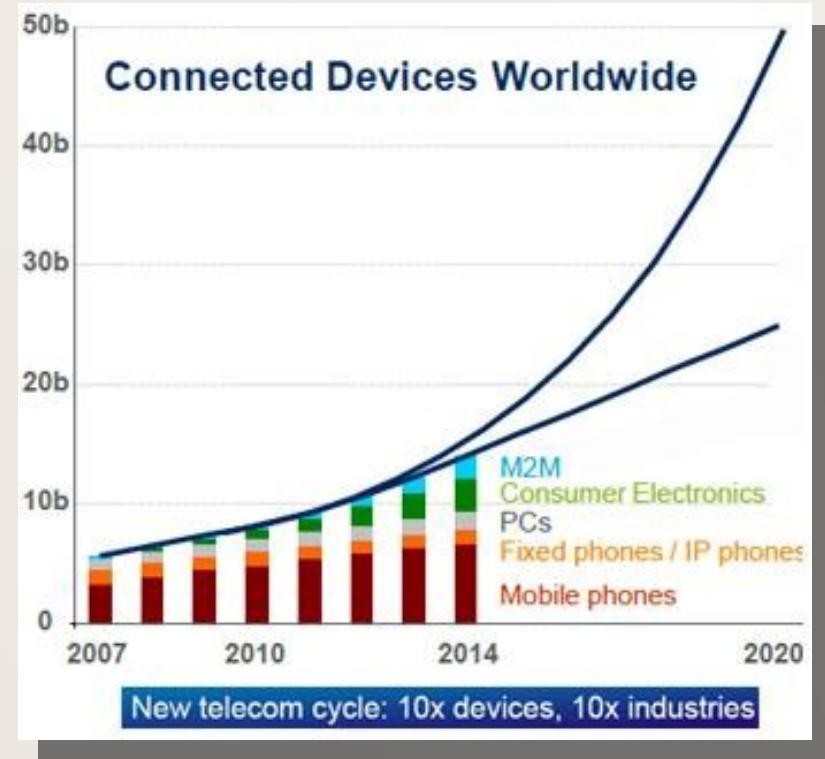


@mamund  
#ScaleFreeAPIs



# So, the news is no so bad...

- Using power-law and node-based systems, we can continue to meet growing demand.

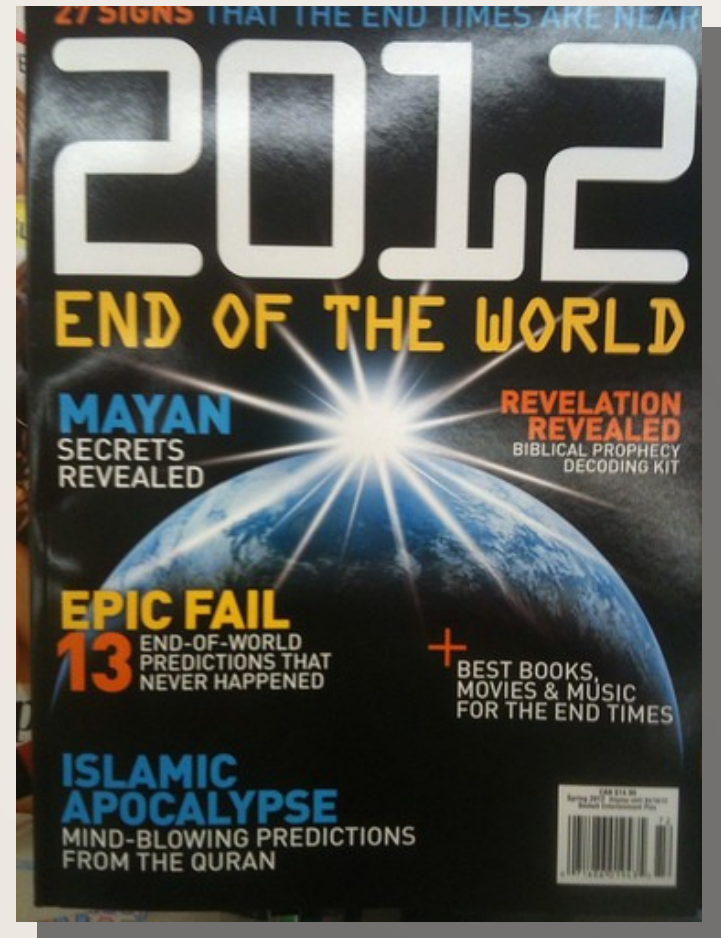


@mamund  
#ScaleFreeAPIs

# The news is not so bad...

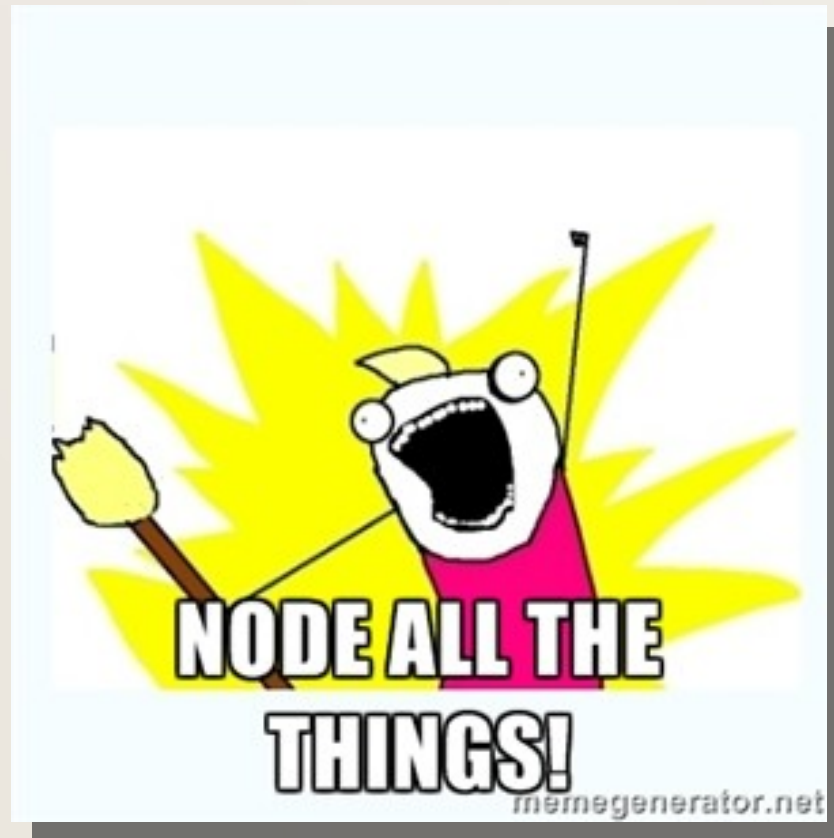
- And we can avoid

*“Scale-pocalypse!”*



@mamund  
#ScaleFreeAPIs

# Be a Node!



@mamund  
#ScaleFreeAPIs

# Scaling the API Economy With Scale-Free Networks

Mike Amundsen  
Principal API Architect  
Layer 7 Technologies

@mamund

@mamund  
#ScaleFreeAPIs



# References

- [http://en.wikipedia.org/wiki/File:Winged\\_goddess\\_Cdm\\_Paris\\_392.jpg](http://en.wikipedia.org/wiki/File:Winged_goddess_Cdm_Paris_392.jpg)
- <http://en.wikipedia.org/wiki/File:Iridium-2.jpg>
- <http://dynamicdecorator.files.wordpress.com/2013/01/fesvh.png>
- <http://www.flickr.com/photos/luvataciousskull/3374223404/in/photolist-696AXi-69aM4A-69aM51-69aM7h-69aM8o-69aM95-69aM9S-69aMaj-69aMbC-69aMdW-69aMeC-69aMgf-69aMgE-69aMij-6c5PSD-6kLGyn-6uuq6s-6CazEo-6CDvfQ-6PtMEG-6ZBpBx-7f9XeB-7keQgx-7kzMec-7kzMmB-7kDDmN-7kDDxb-7kDDDh-7rtS3a-ePziXc-9eQeak-9eTm8W-9eQdCr-9eQcYk-cwkcSy-eS88UX-8pb5fQ-9D4KUU-884Y12-eRZG7k-ejqih3-cnWzM3-k9G6Hdb/53qEWH-cu9GaC-84WtXY-88WN2g-bqxc3k-9LjgRw-8GtxN8-8Aw55>
- <http://3.bp.blogspot.com/VyTCyizrHs/S-Lp0UmtOVI/AAAAAAAAAHik/HSQ1KfRLRa0/s1600/ericsson1.jpg>
- <http://www.flickr.com/photos/34206904@N02/6841249195/in/photolist-bqxc3k-9LjgRw-8GtxN8-8Aw55k-9G6Hdb>
- <http://kinlane-productions.s3.amazonaws.com/api-evangelist-site/blog/apis-resource-to-last-mile.png>
- <http://blog.adw.org/wp-content/uploads/question-mark.jpg>
- [http://upload.wikimedia.org/wikipedia/commons/a/ac/Jerry\\_Yang\\_and\\_David\\_Filo.jpg](http://upload.wikimedia.org/wikipedia/commons/a/ac/Jerry_Yang_and_David_Filo.jpg)
- [http://en.wikipedia.org/wiki/File:Larry\\_Page\\_in\\_the\\_European\\_Parliament,\\_17.06.2009.jpg](http://en.wikipedia.org/wiki/File:Larry_Page_in_the_European_Parliament,_17.06.2009.jpg)
- [http://upload.wikimedia.org/wikipedia/commons/d/d2/Internet\\_map\\_1024.jpg](http://upload.wikimedia.org/wikipedia/commons/d/d2/Internet_map_1024.jpg)
- [http://upload.wikimedia.org/wikipedia/commons/8/8a/Long\\_tail.svg](http://upload.wikimedia.org/wikipedia/commons/8/8a/Long_tail.svg)
- <http://i.msdn.microsoft.com/dynimg/IC76141.gif>
- [http://dangerouslyirrelevant.org/images/various-small/Linked\\_thumb1.jpg](http://dangerouslyirrelevant.org/images/various-small/Linked_thumb1.jpg)
- <http://tctechmunch2011.files.wordpress.com/2009/08/google.jpg>
- [http://farm5.staticflickr.com/4150/5018184807\\_60be8b669e\\_z.jpg](http://farm5.staticflickr.com/4150/5018184807_60be8b669e_z.jpg)



@mamund  
#ScaleFreeAPIs

# References

- [http://www.cesa.eu/images/timeline/pictures/09\\_2008.jpg](http://www.cesa.eu/images/timeline/pictures/09_2008.jpg)
- <http://www.codeproject.com/KB/web-cache/AdvanceCaching/partioned.jpg>
- <http://blog.mybackupbox.com/wp-content/uplo>
- [https://si0.twimg.com/profile\\_images/480262550/box.png\\_ads/2012/06/google\\_drive.jpg](https://si0.twimg.com/profile_images/480262550/box.png_ads/2012/06/google_drive.jpg)
- [http://svn.apache.org/repos/asf/subversion/trunk/notes/logo/256-colour/subversion\\_logo-384x332.png](http://svn.apache.org/repos/asf/subversion/trunk/notes/logo/256-colour/subversion_logo-384x332.png)
- <https://secure.gravatar.com/avatar/878a0ea898da1701df8573ed64a5cc9d?s=420&d=https://a248.e.akamai.net/assets.github.com%2Fimages%2Fgravatars%2Fgravatar-user-420.png>
- <http://3.bp.blogspot.com/-ZeYZpE3tESU/UOz8-49WYvI/AAAAAAAAAP8/vt0-vDzXxwk/s72-c/database-developer-Eastbourne-Brighton-Sussex.png>
- <http://hueniverse.com/wp-content/uploads/2009/09/OAuth-Shine-300x298.png>
- [http://mindcommerce.files.wordpress.com/2013/05/telephonyapi\\_management.gif?w=286&h=300](http://mindcommerce.files.wordpress.com/2013/05/telephonyapi_management.gif?w=286&h=300)
- <http://blogs.bmj.com/bmj-journals-development-blog/files/2010/12/semantic-web.jpg>
- [http://blog.caplinked.com/wp-content/uploads/2011/09/iStock\\_000009230566Large.jpg](http://blog.caplinked.com/wp-content/uploads/2011/09/iStock_000009230566Large.jpg)
- [http://studsovet.ipsa.kpi.ua/uploads/posts/2013-05/1368982043\\_ibm-intel-logo.jpg](http://studsovet.ipsa.kpi.ua/uploads/posts/2013-05/1368982043_ibm-intel-logo.jpg)
- [http://www.shinyshiny.tv/microsoft-logo\\_111129012732.jpg](http://www.shinyshiny.tv/microsoft-logo_111129012732.jpg)
- <http://rack.2.mshcdn.com/media/ZgkyMDEyLzEyLzA0L2I1L3lhaG9vdm9pY2VzLmJrSC5qcGcKcAl0aHVtYgk5NTB4NTM0IwplCWpwZw/61883c7e/479/yahoo-voices-hacked-450-000-passwords-posted-online-7169a7e88d.jpg>
- <http://nyocommercialobserver.files.wordpress.com/2012/06/google-logo.gif?w=400&h=283>
- <http://upload.wikimedia.org/wikipedia/commons/thumb/4/42/Opensource.svg/220px-Opensource.svg.png>
- <http://www.salesnet.com/images/xml-web-services-api.jpg>



@mamund  
#ScaleFreeAPIs