

SDForum Event

Building Search Engine in the Social Media Era

AJ Chen

ajchen@web2express.org

twitter: @web2express

<http://web2express.org>

3/24/2010

Agenda

- 1. The changing search environment**
2. Traditional search engine architectures
3. New architectures for real time search and social search
4. Semantic search
5. Emerging real time web monitoring
6. Q/A

The Changing Search Environment

Users expect more from search box:

- find documents
- discover information
- answer specific questions

**Web of documents:
web pages, blogs,
Q/A, forums**

The Changing Search Environment

Users expect more from search box:

- find documents
- discover information
- answer specific questions
- check real-time status update
- check my social circle
- monitor social media

Web of documents:
web pages, blogs,
Q/A, forums

Status updates and
Social networks:
facebook, twitter,
google, ...

Real-time data
streams (API)

Web of data:
linked semantic data,
databases

The Changing Search Environment

Users expect more from search box:

- find documents
- discover information
- answer specific questions
- check real-time status update
- check my social circle
- monitor social media

- Full Text Search Technologies
- Semantic Search Technologies

Computing Clouds:
Google, Salesforce,
Facebook, ...

Web of documents:
web pages, blogs,
Q/A, forums

Status updates and
Social networks:
facebook, twitter,
google, ...

Real-time data
streams (API)

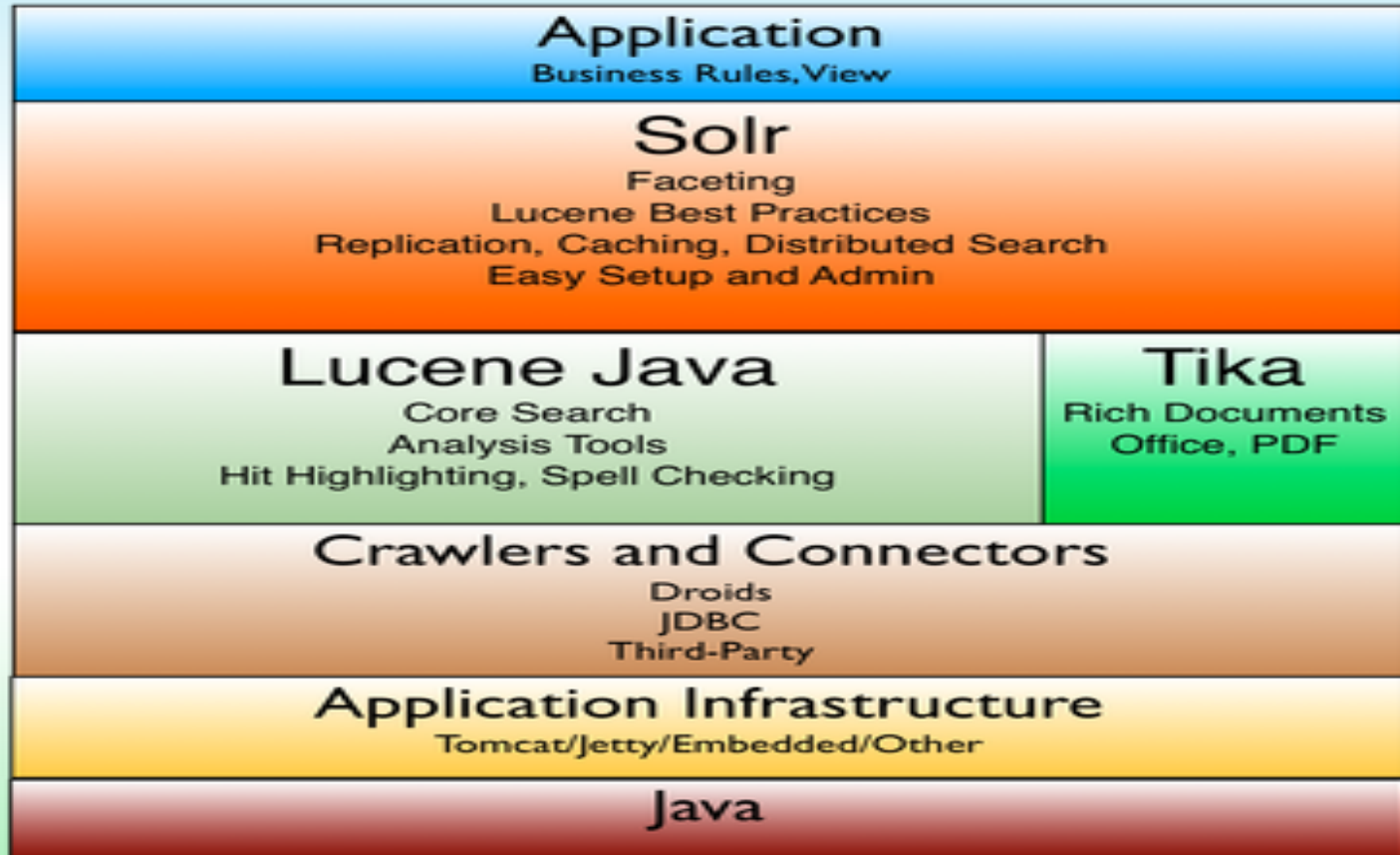
Web of data:
linked semantic data,
databases

Agenda

1. The changing environment
- 2. Traditional search engine technologies and architecture**
3. New architectures for supporting real time search and social search
4. Semantic search
5. Emerging real time web monitoring system
6. Q/A

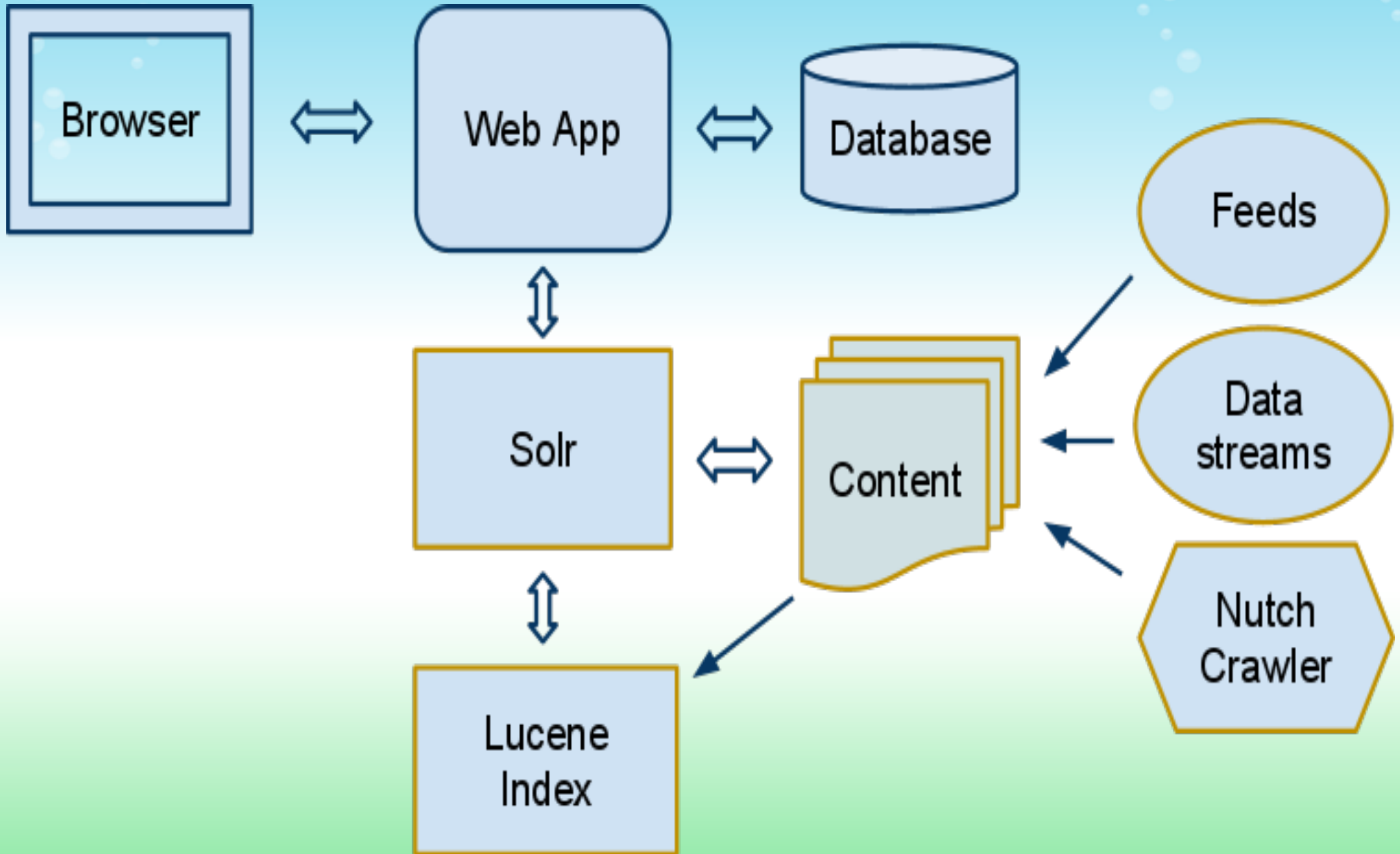
Open Source Lucene/Solr

The Lucene Stack

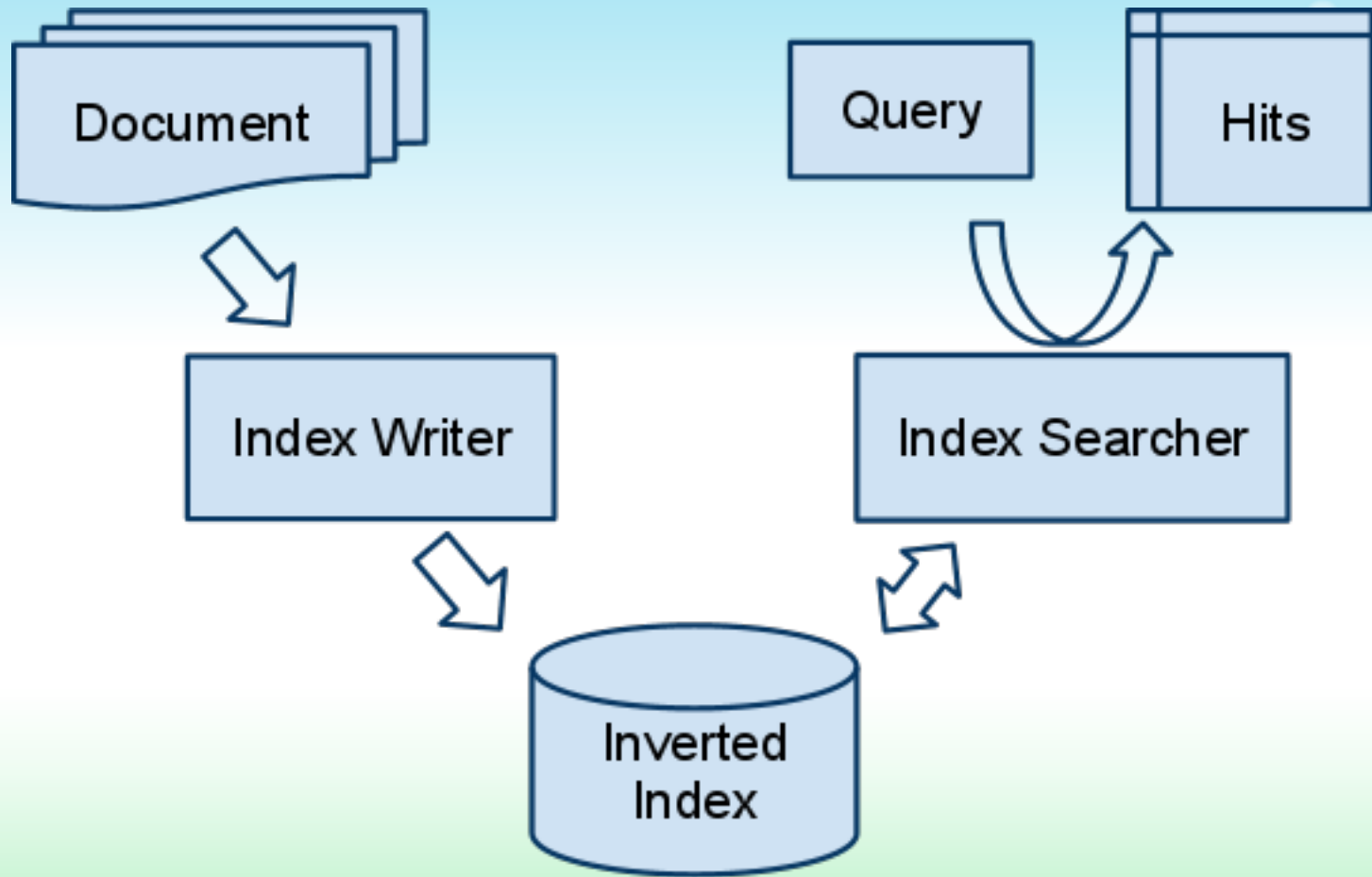


from Grant Ingersoll

Lucene/Solr Search for Web App

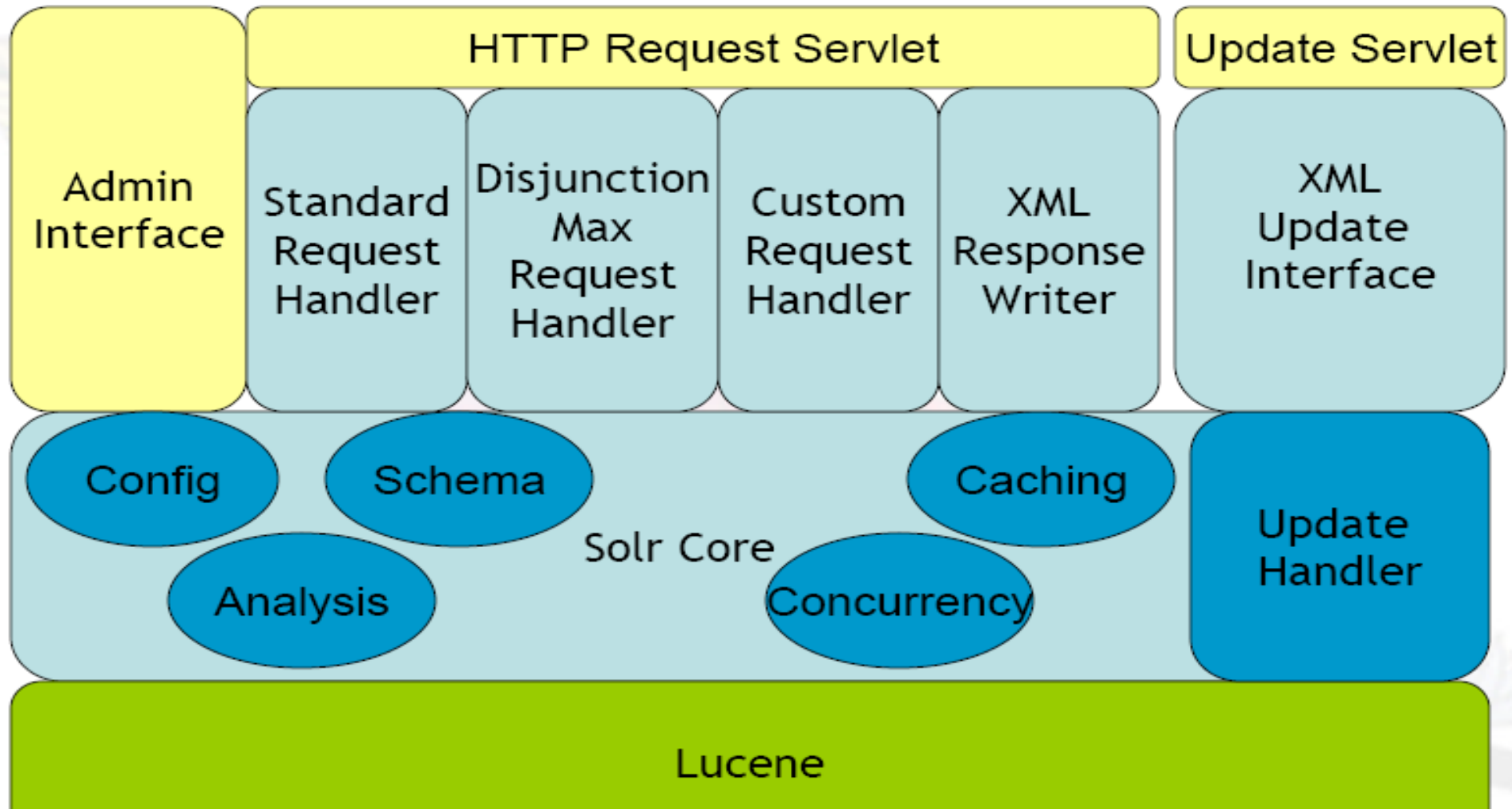


Lucence Indexing/Searching



Solr Server Architecture

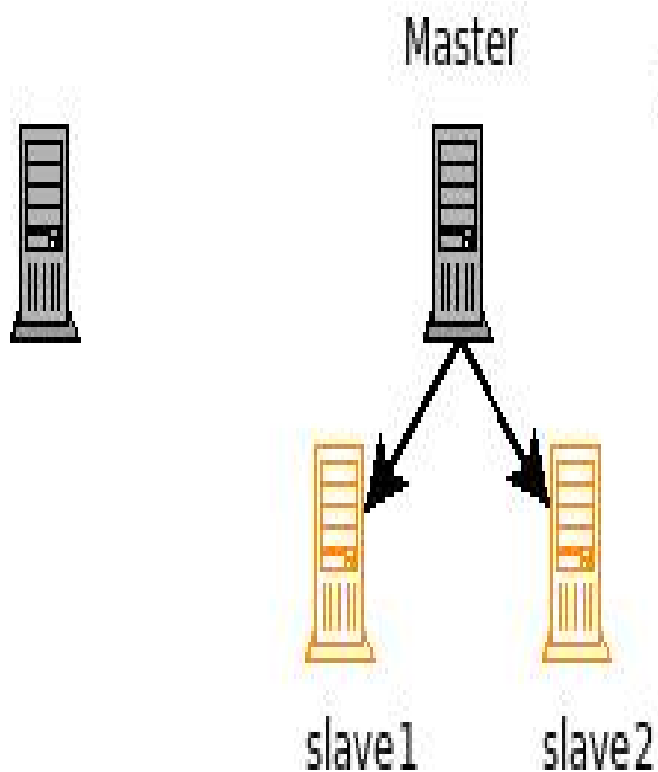
Architecture



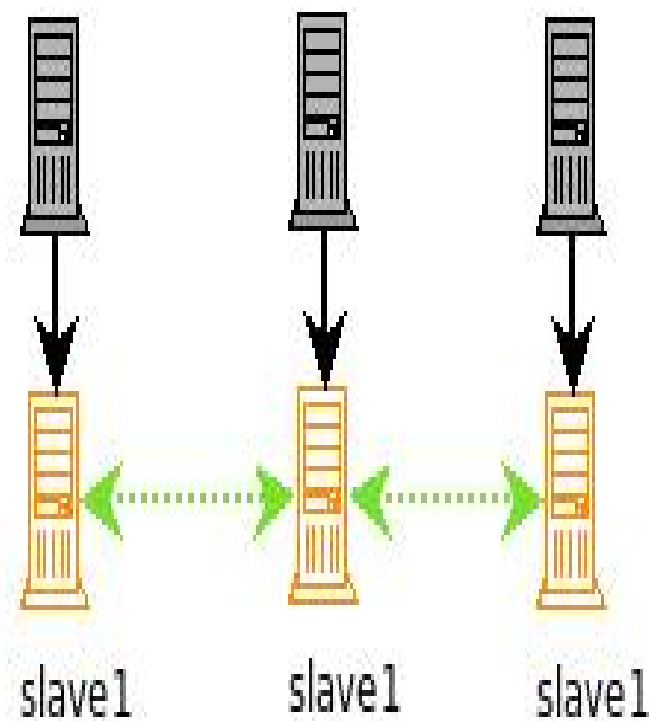
Scaling Lucene and Solr

Single Server Replication Distributed

Distributed + Replication



Shard 1 Master Shard 2 Master Shard 3 Master



Agenda

1. The changing environment
2. Traditional search engine technologies and architecture
- 3. New architectures for supporting real time search and social search**
4. Semantic search
5. Emerging real time web monitoring system
6. Q/A

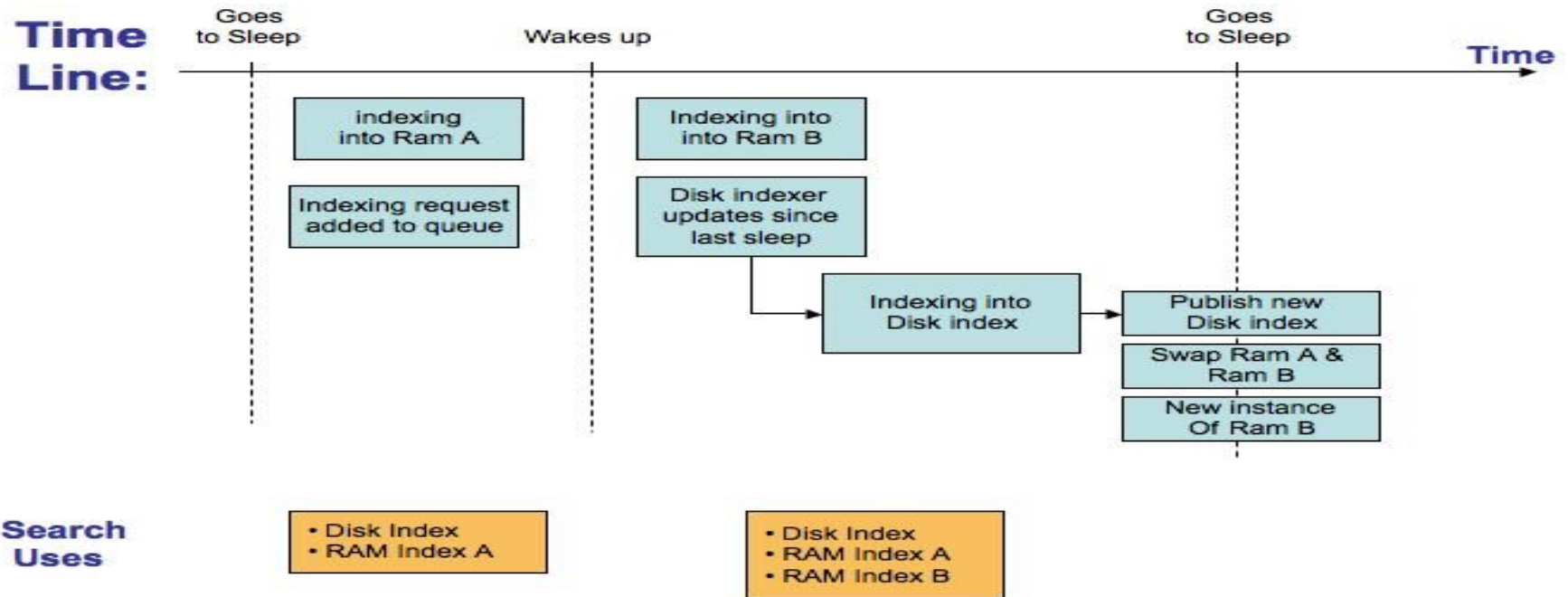
Near Real Time Search

Lucene 2.9:

IndexReader (old + new records) <-- IndexWriter

Zoie (LinkedIn):

Swap Indexes in memory



Google Real-time Search

The image shows a screenshot of the Google Real-time Search interface. At the top, the Google logo is on the left, followed by a search bar containing the text "social media monitoring" and a "Search" button on the right. Below the search bar, there are navigation options: "Web" (selected), "Latest", and "Hide options".

On the left side, there is a sidebar with the following sections:

- All results**
 - [Images](#)
 - [Videos](#)
 - [News](#)
 - [Blogs](#)
 - [Updates](#)
 - [Books](#)
 - [Discussions](#)
- Any time**
- Latest**
 - [Past 24 hours](#)
 - [Past week](#)
 - [Past year](#)
 - [Specific date range](#)
- [Reset options](#)

The main content area displays the following results:

New results will appear below as they become available. [Pause](#)

[Social Media Explorer — Social Media Consulting, Public S](#)
Social media monitoring firm Sysomos is launching a free location-ba today it's calling FourWhere. The tool mines publicly available ...
www.socialmediaexplorer.com/ - 4 minutes ago

[social media monitoring - Socially Minded](#)
However, in today's world of the Internet and the growth of **social media**, does get it wrong, dissatisfied customers can tell everyone about ...
www.sociallyminded.co.uk/?cat=103 - 4 minutes ago

[Social Media Examiner, Your Guide to the Social Media Jur](#)
Monitoring the innovations in the **social media** space also gives you a market is changing and will evolve over the next few years. ...
www.socialmediaexaminer.com/ - 4 minutes ago

media monitoring: : Guest post: Email Marketing + Social Intelligence
[monitoring the media monitoring](#) - barnish.org

 [barnishhj](#) - [twitter.com](#) - 31 minutes ago

Google Social Search

Google [Advanced Search](#)

[Web](#) > [Social](#) [Hide options](#) Results 1 - 3 of 3 for

› **All results**
[Images](#)
[Videos](#)
[News](#)
[Blogs](#)
[Updates](#)
[Books](#)
[Discussions](#)

› **Any time**
[Latest](#)
[Past 24 hours](#)
[Past week](#)
[Past year](#)
[Specific date range](#)

[All results](#)
› **Social**
[Nearby](#)

› **All people**
[Mike Volpe](#)
[Web2express New...](#)
[David Meerman S...](#)

› **All results**
[Visited pages](#)
[Not yet visited](#)


› **Standard view**


Social Media Monitoring Sponsored Links
www.social360.co.uk What's being said about your brand? **Monitor** blogs & forums

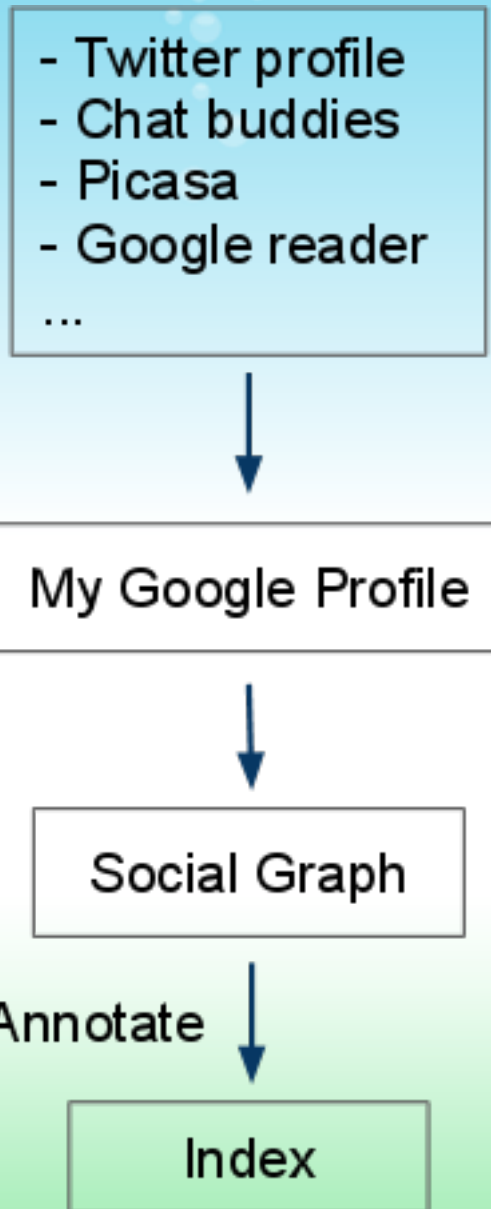
[Track Online Buzz](#)
www.CustomScoop.com What are blogs & Twitter saying? Take a Free Trial to find out!

Social Media Monitoring
www.vocus.com **Monitor** Your Traditional, Online & **Social Media** Efforts with Vocus!

Results from people in your social circle for **social media monitoring** - BETA - [My social circle](#) - [My social content](#)

[How to Monitor Your Social Media Presence in 10 Minutes a Day](#)
 [Mike Volpe](#) - connected via Gmail
How to **monitor** your **social media** presence on Twitter, Facebook, LinkedIn, Flickr and more in 10 minutes a day.
blog.hubspot.com/blog/tabid/6307/bid/4663/How-to-Monitor-Your-Social-Media-Presence-in-10-Minutes-a-Day.aspx
[How to Monitor Your Brand on the Web for Free](#)
But unsurprisingly, I'm going to respectfully disagree with Paul up ...
blog.hubspot.com/blog/tabid/6307/bid/5498/How-to-Monitor-Your-Brand-on-the-Web-for-Free.aspx
[More results from Mike Volpe »](#)

[Web2express Monitoring - Social Media and Real Time Web](#)
 [Web2express News](#) - Monitoring **social media** like twitter, facebook, myspace, blogs, and news in real time; Analyzing and digesting fresh web content to help people and ...
web2express.org/
web2express.org/
Who will benefit from **social media monitoring**? Marketers can use ...
web2express.org/openlab/
[More results from Web2express News »](#)



Google Aardvark Social Q/A Engine

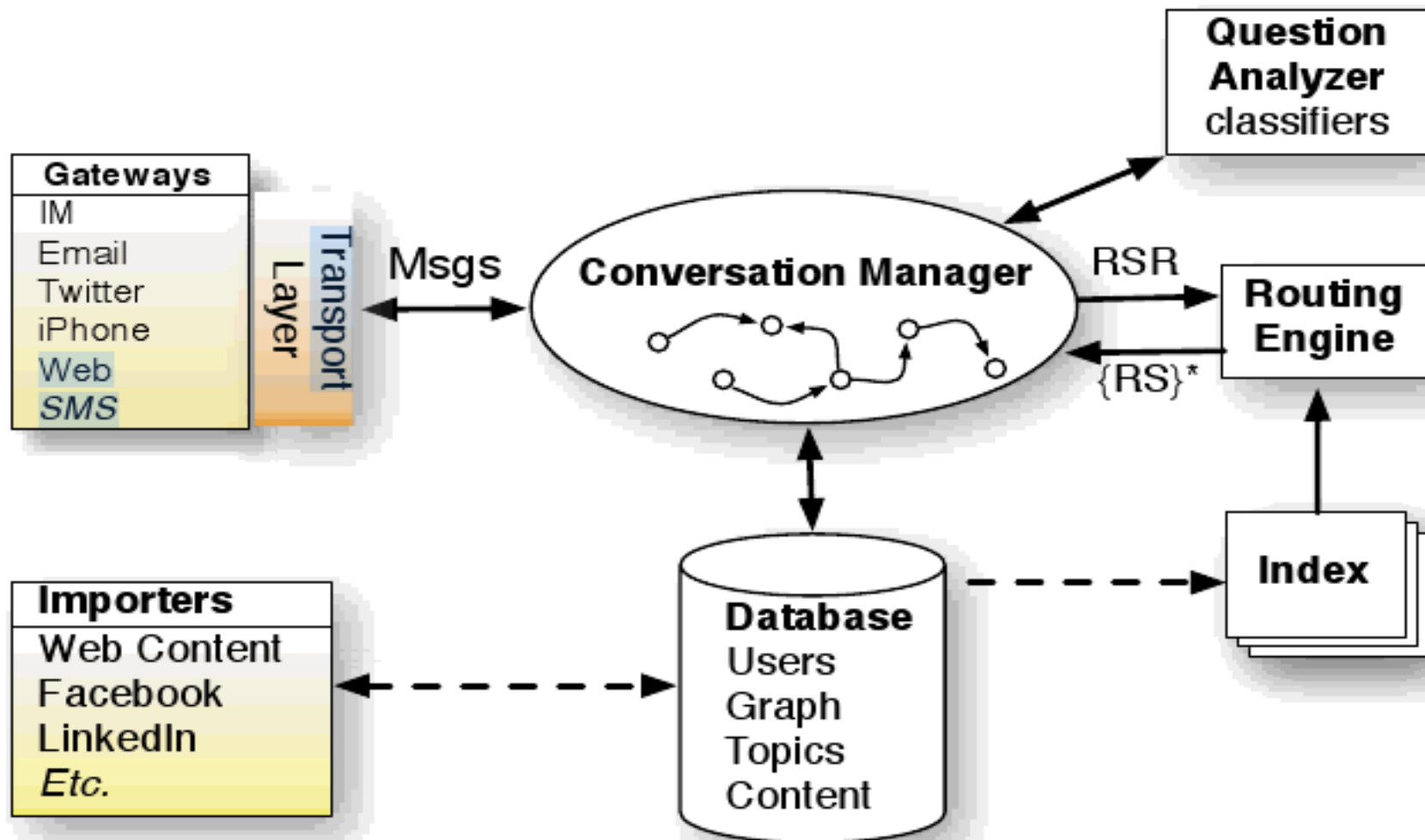


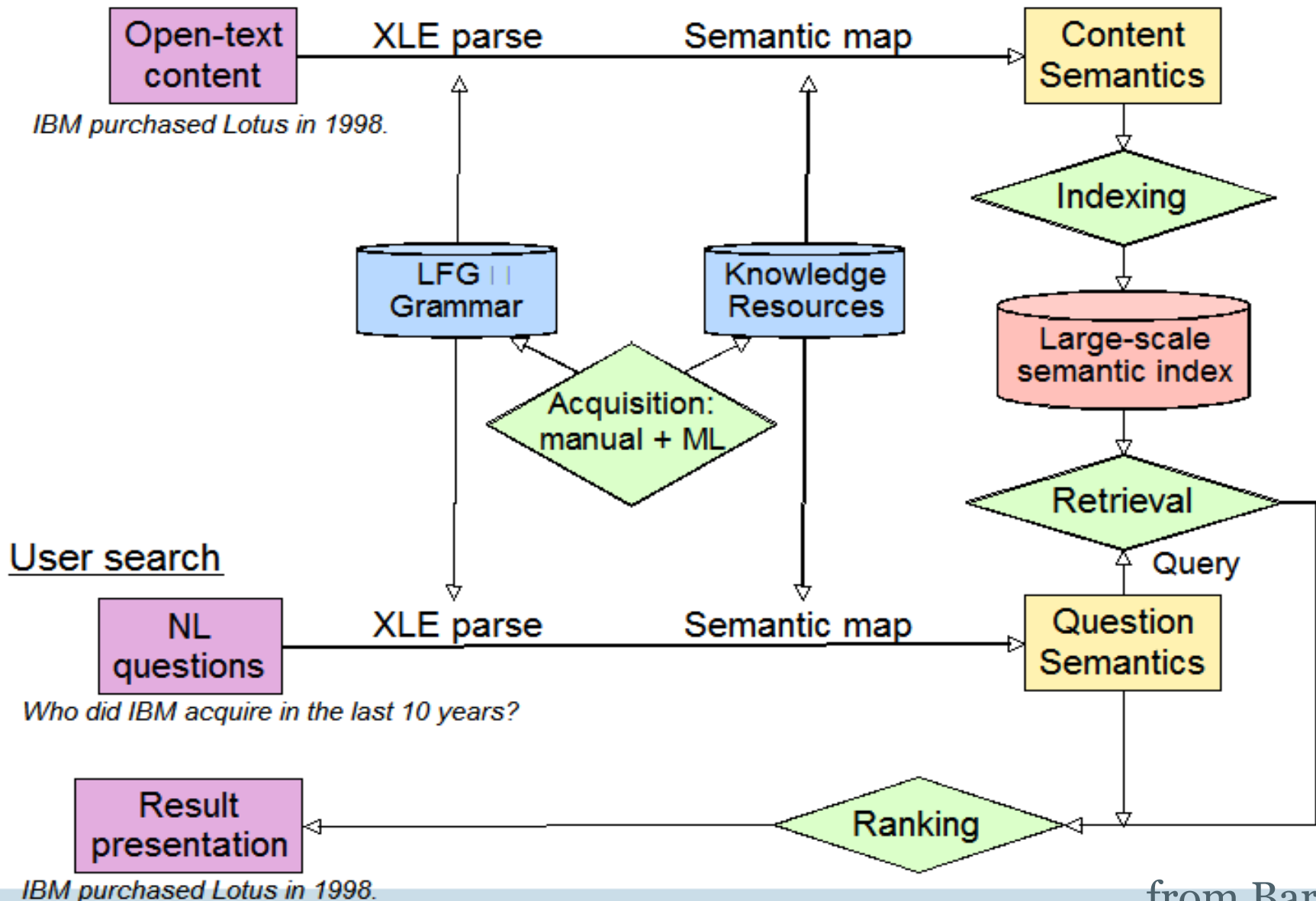
Figure 1: Schematic of the architecture of Aardvark

Agenda

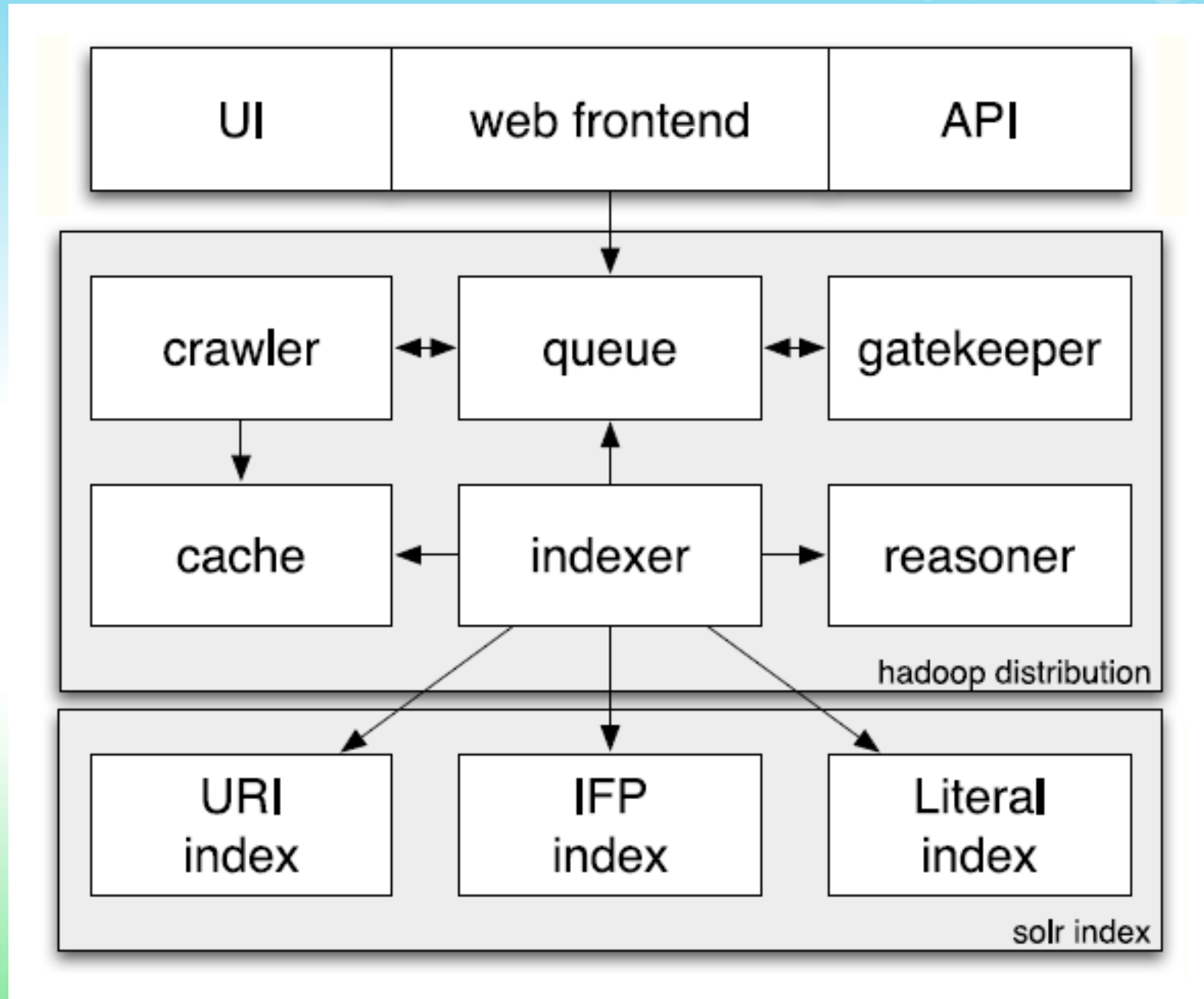
1. The changing environment
2. Traditional search engine technologies and architecture
3. New architectures for supporting real time search and social search
- 4. Semantic search**
5. Emerging real time web monitoring system
6. Q/A

Bing/Powerset NLP Search Engine

Content Acquisition



Sindice: Retrieving semantic data



index API:
sindice.com
search engine:
sig.ma

Google Semantic Search

Google

Web Show options...

[SDForum - Home](#) ☆

If you are an SDForum Member and interested email webmgr@sdforum.org for the discount www.sdforum.org/ - [Cached](#) - [Similar](#)

[Calendar](#) [Membership Types](#)
[Login](#) [Membership Benefits](#)
[Overview](#) [Contact Us](#)
[Join](#) [Technology](#)

[More results from sdforum.org »](#)

[VentureWiki : Negotiating a Venture C](#)

Apr 20, 2006 ... As a result, it is common for them to block certain actions or **events**. ... wiki.sdforum.org/tiki-index.php?page=Negoti

[SDForum - Quarterly Venture Breakfa](#)

SDForum is the leading Silicon Valley non-pr **events** are organized each month, reaching 1 www.sdforum.com/index.cfm?fuseaction=Pa

[SDForum Hosts Korean Delegation f](#)

The evening **event** will be held at Orrick, 1000 information, visit <http://www.sdforum.org/Inte> www.redorbit.com/news/technology/.../sdforu

Google [Advanced Se](#)

[Web](#) > [Wonder wheel](#) [Hide options](#)

Results 1 - 10 of about

> All results

[Images](#)
[Videos](#)
[News](#)
[Blogs](#)
[Updates](#)
[Books](#)
[Discussions](#)

> Any time

[Latest](#)
[Past 24 hours](#)
[Past week](#)
[Past year](#)
[Specific date range](#)

> All results

[Social](#)
[Nearby](#)

> All results

[Visited pages](#)
[Not yet visited](#)

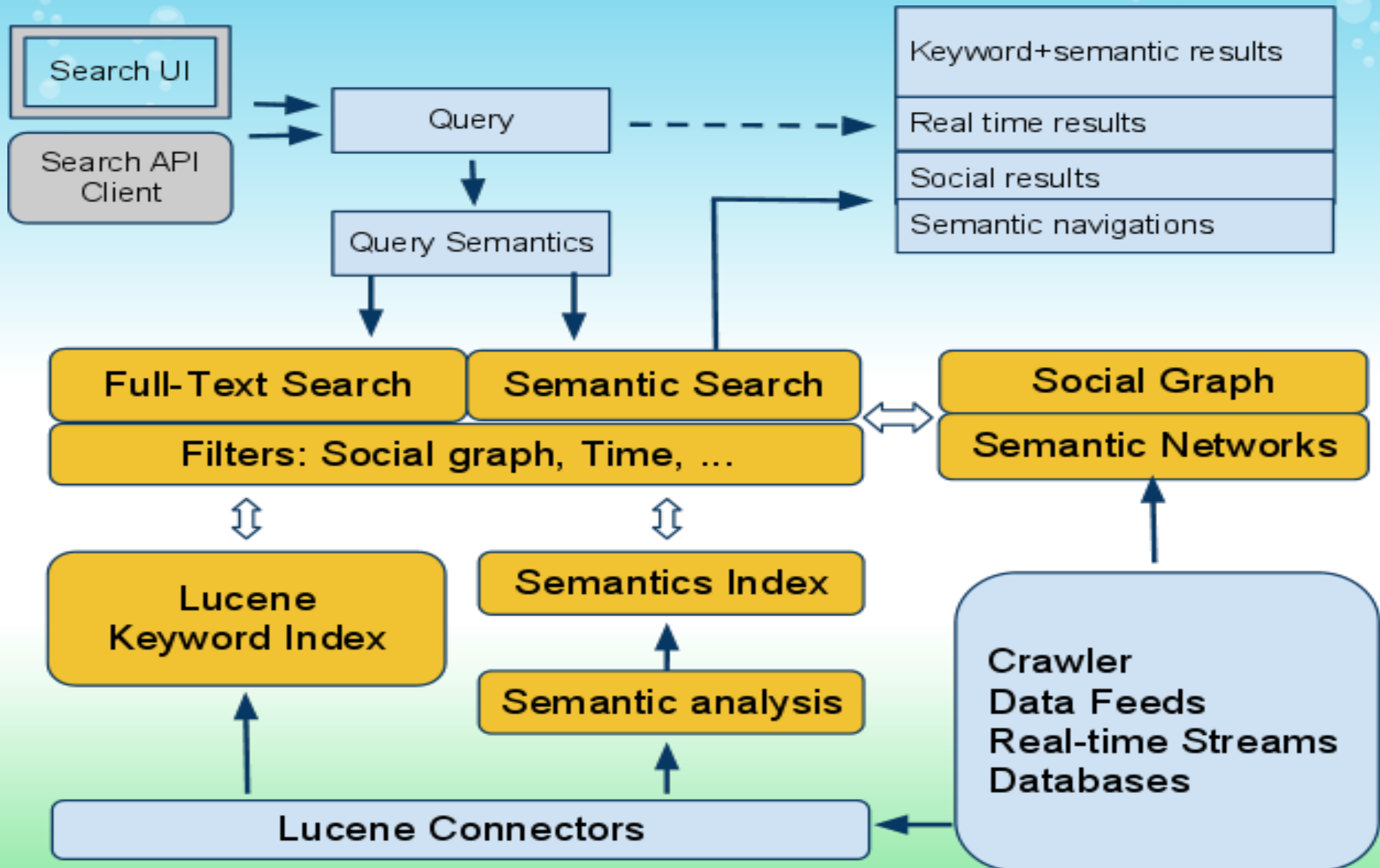
[Standard view](#)
[Related searches](#)

> Wonder wheel

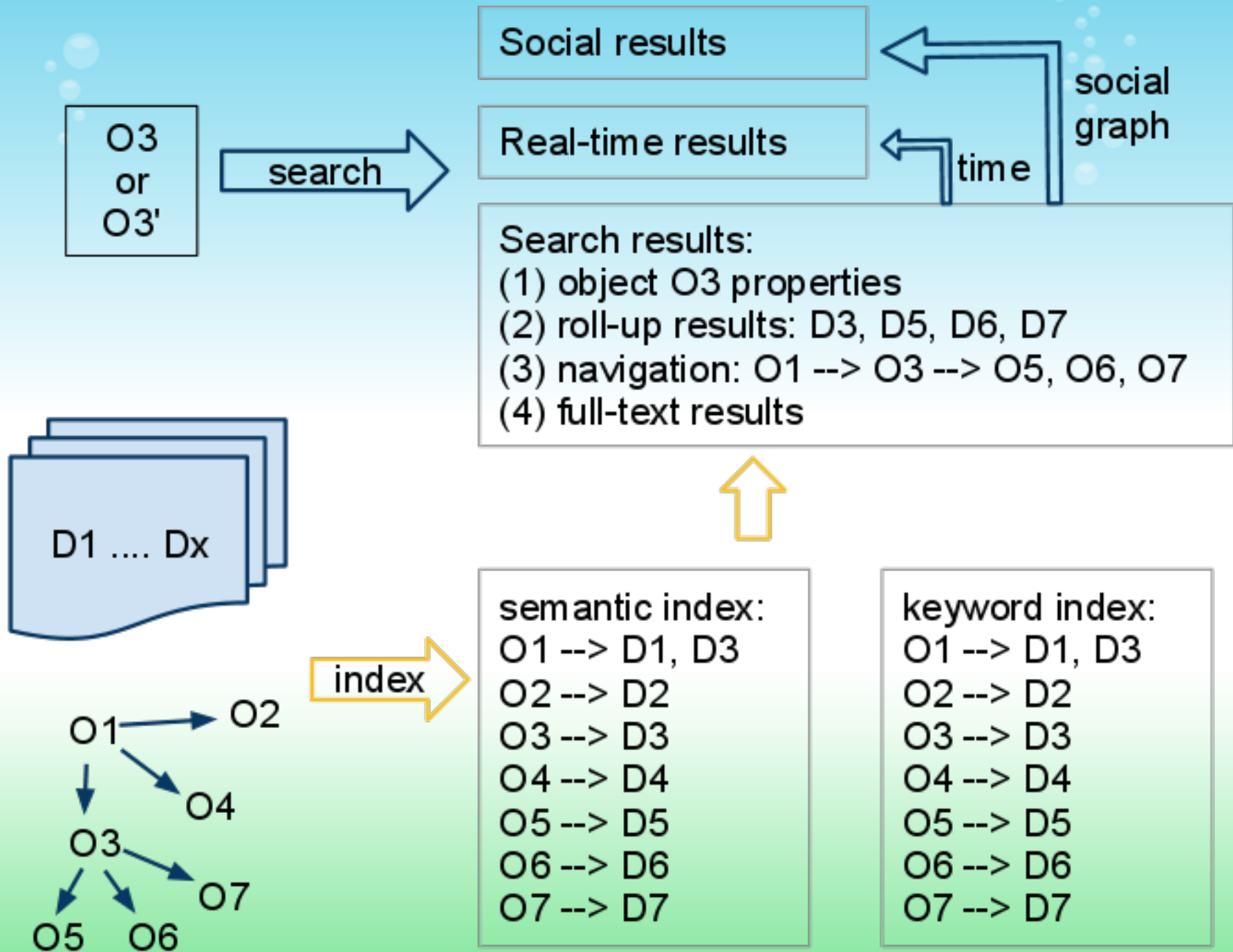
[Timeline](#)



Hybrid Search Engine



Hybrid (Universal) Search



Agenda

1. The changing environment
2. Traditional search engine technologies and architecture
3. New architectures for supporting real time search and social search
4. Semantic search
- 5. Emerging real time web monitoring system**
6. Q/A

Social Media Monitoring

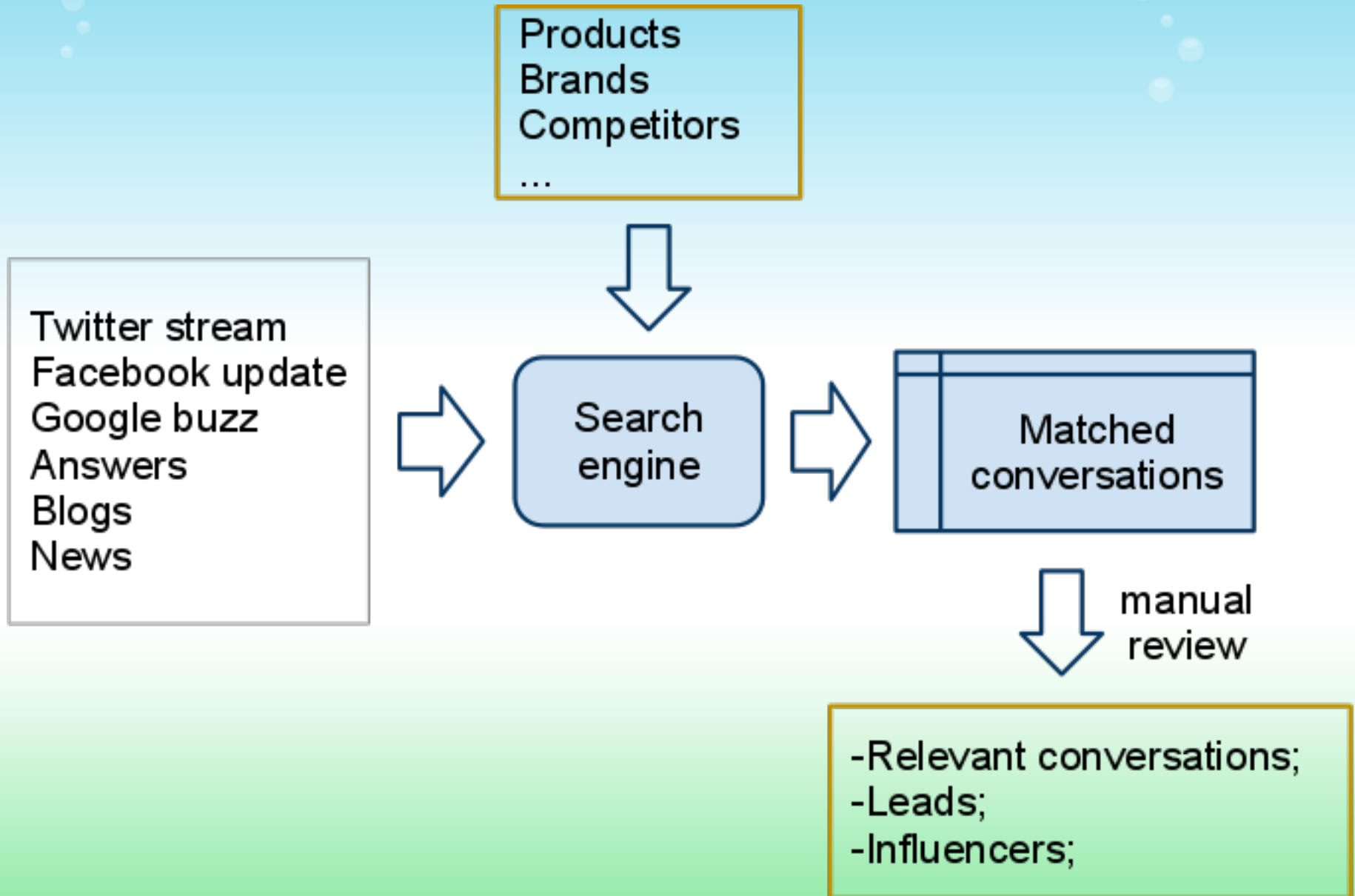
Benefits for businesses:

- understand what people are talking about the company, brands, products, and competitions, etc.
- identify leads for marketing and sales;
- engage with customer and community conversations;
- support customers;
- cultivate product advocates;
- use customer feedback to improve products and brands;

Differences from search engine:

- search engine for filtering
- real-time text analysis and semantic analysis
- auto-discovery of trends/patterns
- collective intelligence
- rule engine for routing right information to right person

Basic Monitoring



Scaling up Social CRM

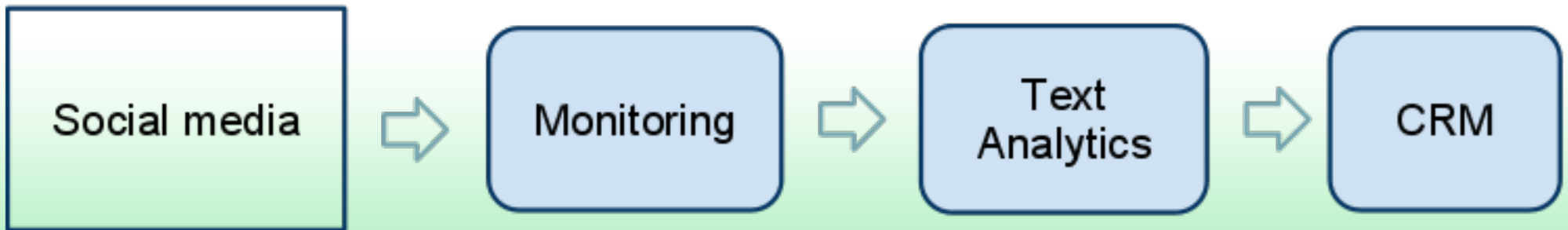
Problem:

Today's social media program does not scale

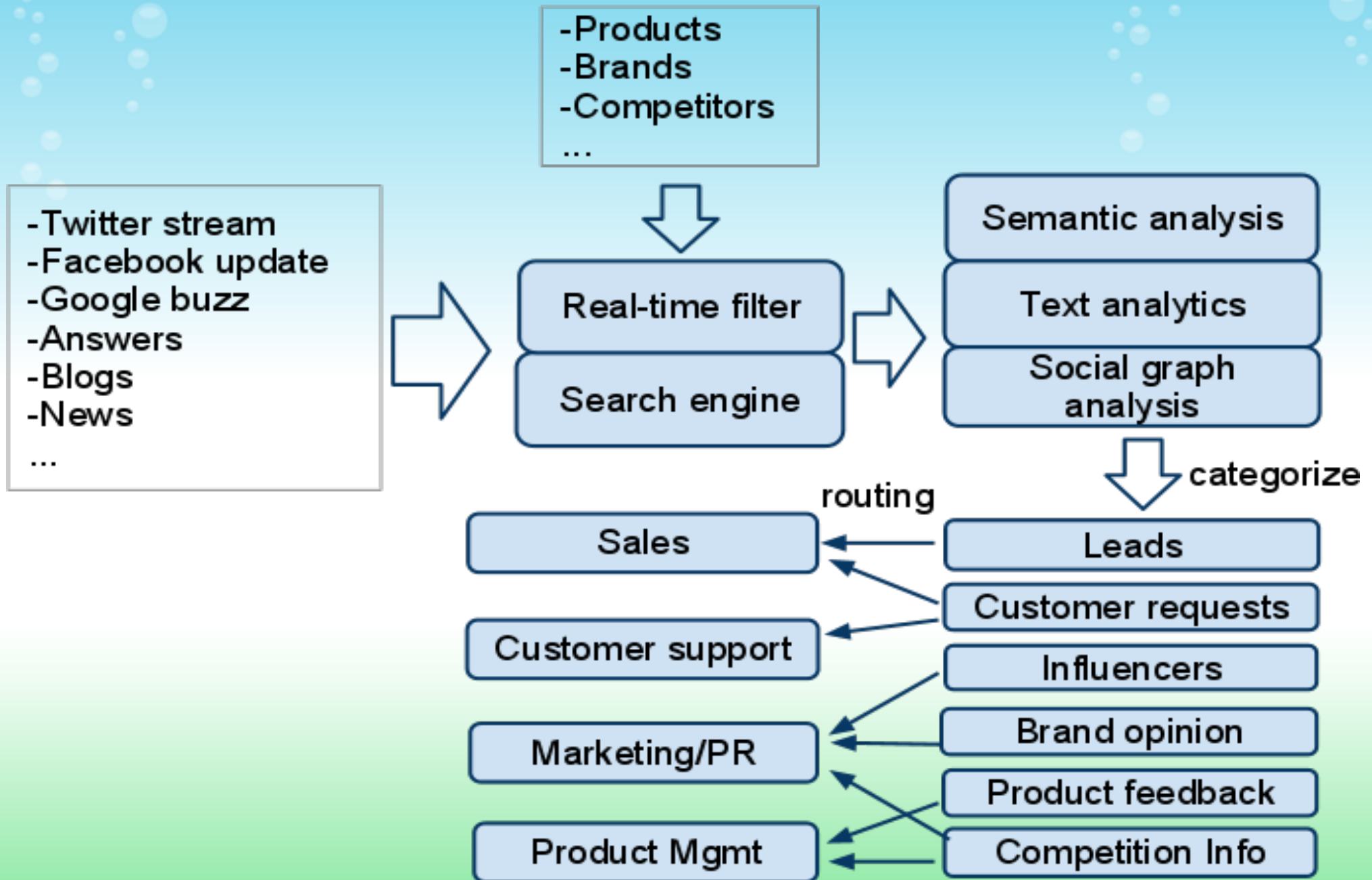
Solution:

Build a pipeline of social context flow and then apply advanced text analytics to automate routing of social contexts.

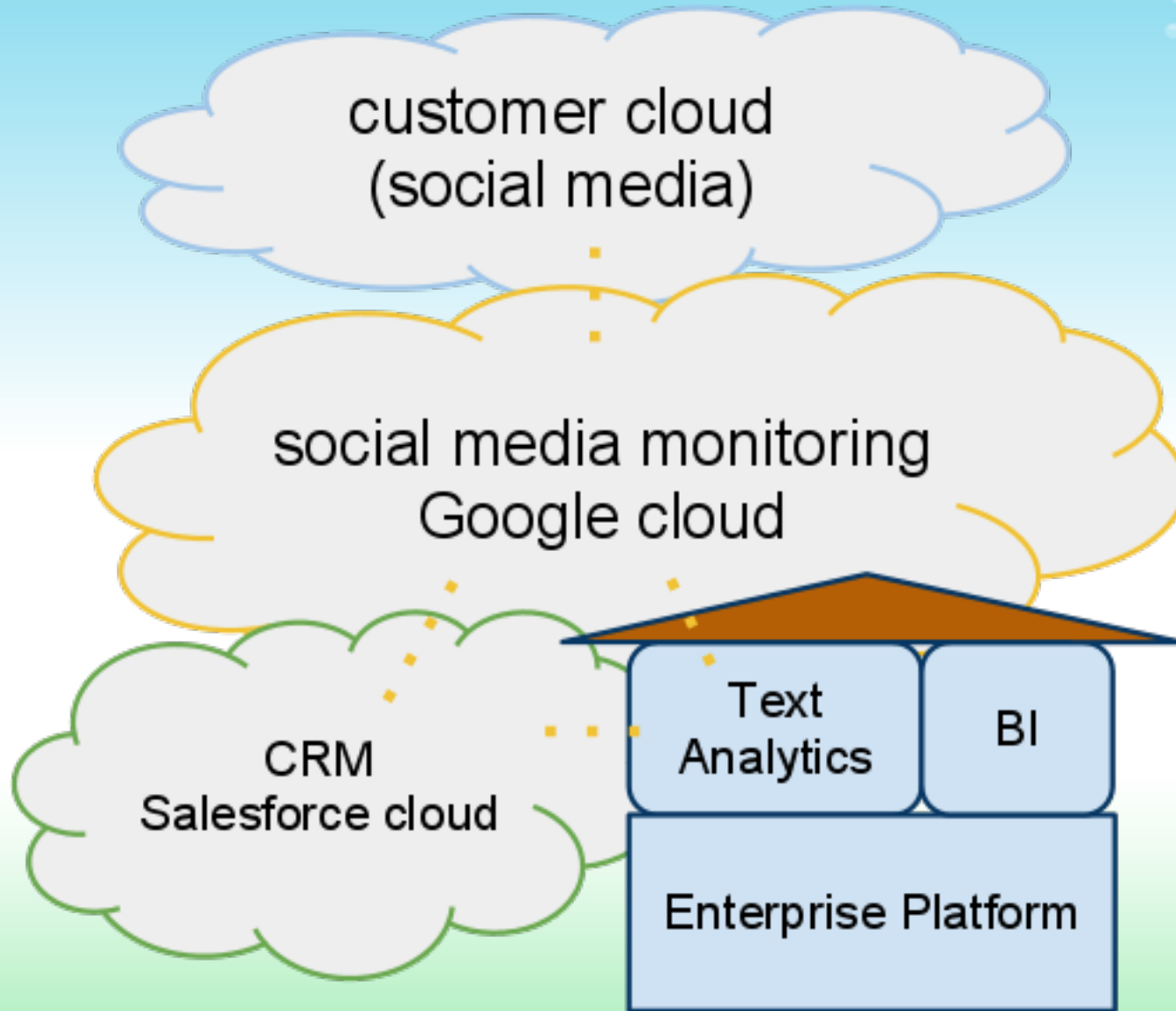
Component Architecture:



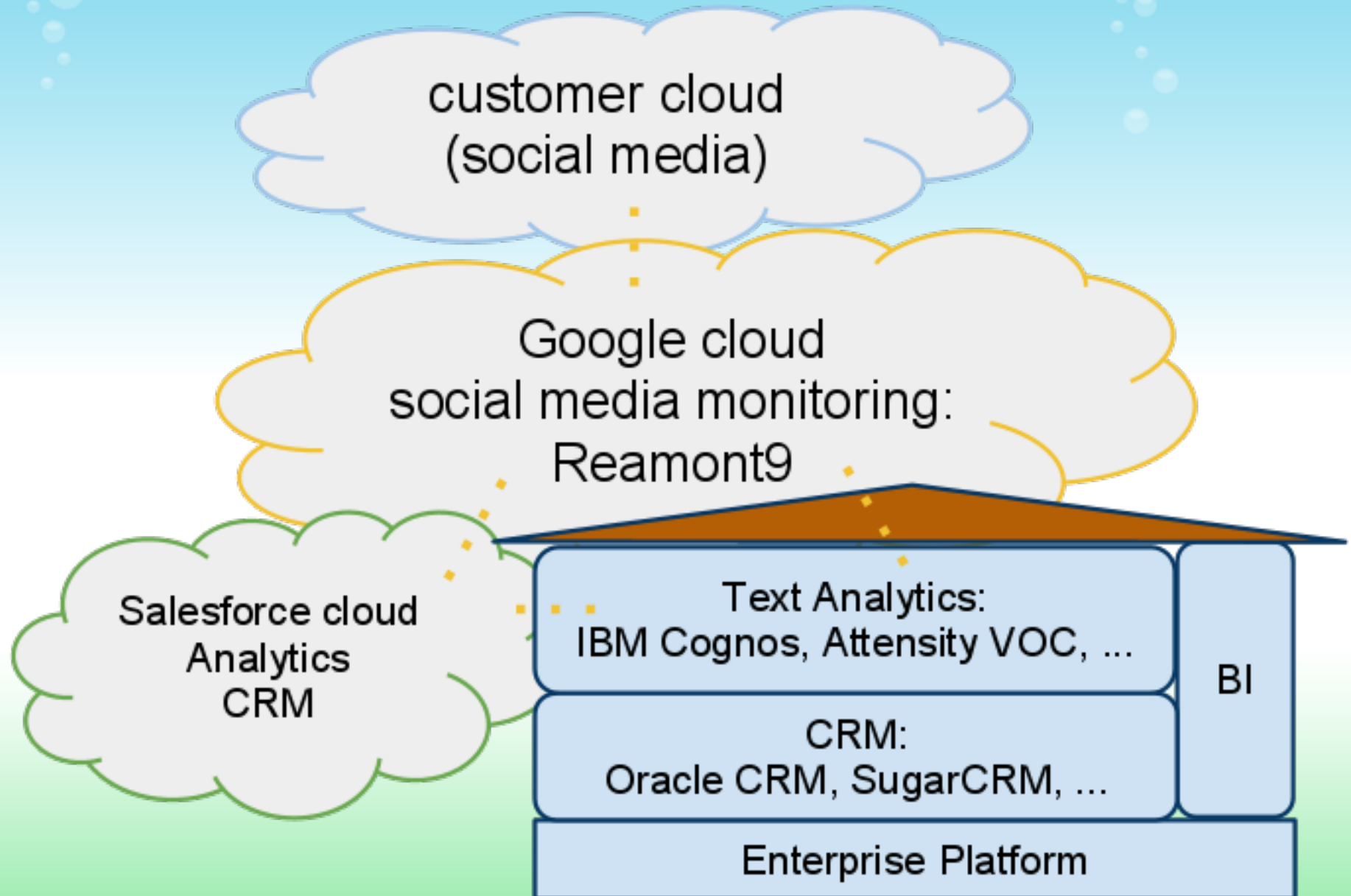
Intelligent Monitoring Pipeline



Social CRM Integration in the Clouds



Turning CRM into Social CRM



Agenda

1. The changing environment
2. Traditional search engine technologies and architecture
3. New architectures for supporting real time search and social search
4. Semantic search
5. Emerging real time web monitoring system
- 6. Q/A**