



Big Data EcoSystem @ LinkedIn

October 20, 2012



Sunil Shirguppi

Head of Data Services- International
LinkedIn Corporation

<http://www.linkedin.com/in/sunilshirguppi>

Outline

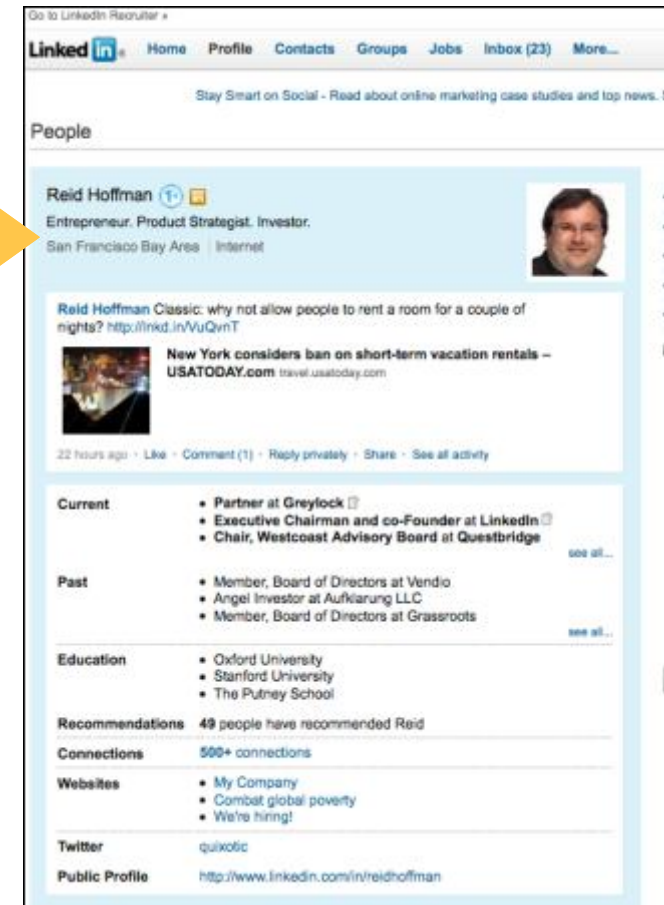
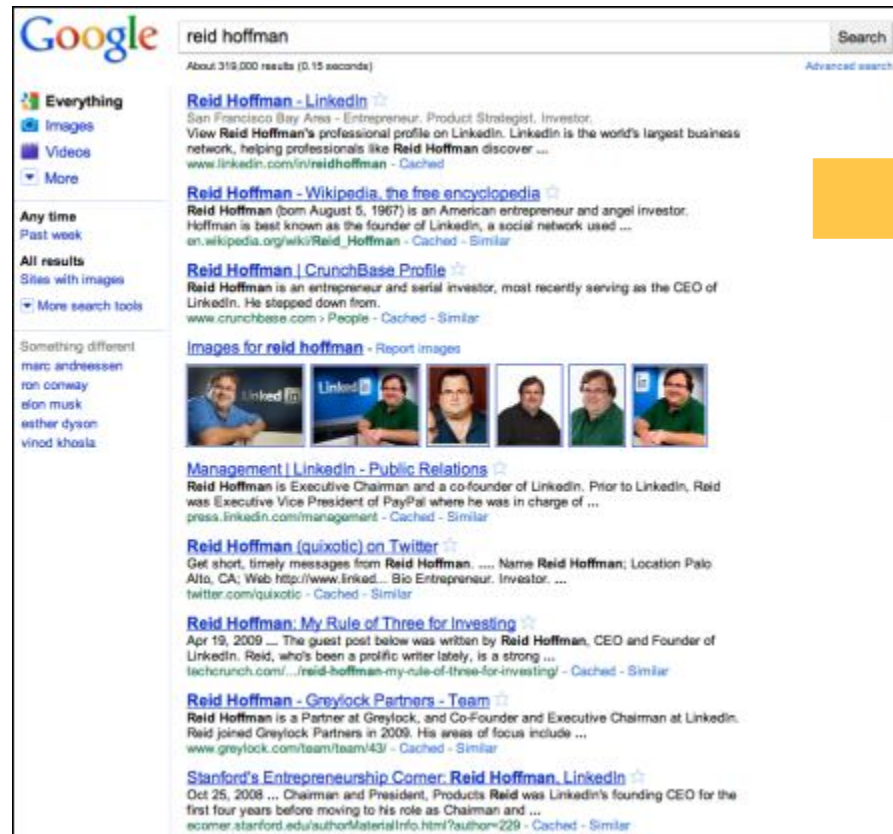
- LinkedIn Overview
- Data Science
- Big Data Eco-System
- Learnings

Our Mission

Connect the world's professionals
to make them more productive and successful

Googled yourself lately?

Don't feel bad, we all do it.



We are the professional profile of record

Executives from all

FORTUNE

MAY 4, 2008 JULY \$4.99

500

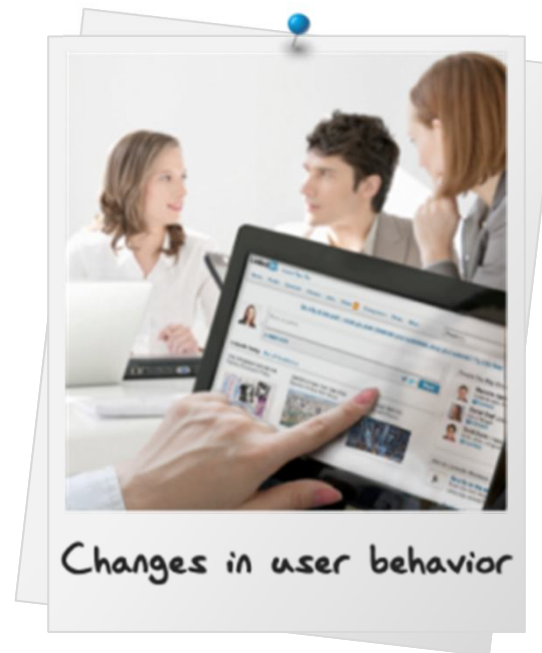
Companies are
LinkedIn members

The LinkedIn Opportunity

Connect talent with opportunity at **massive scale**

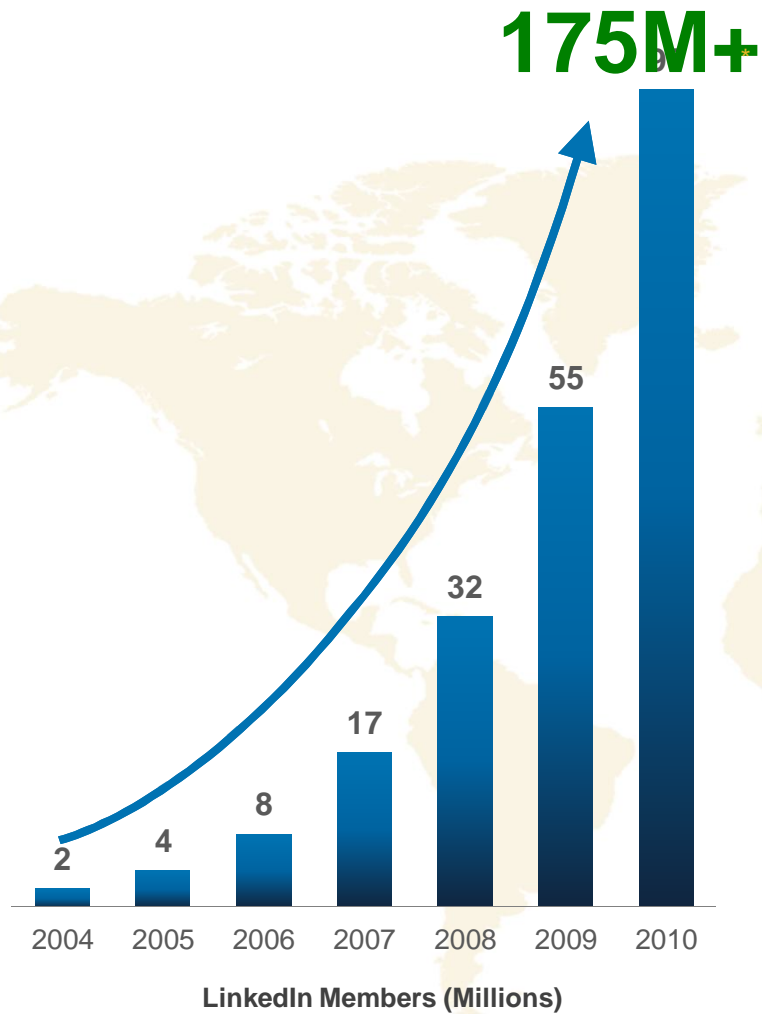


+



Fundamentally transforming the way the world works

The World's Largest Professional Network



~2/sec

New Members joining



>2M

Company Pages



82%

Fortune 100 Companies
use LinkedIn to hire



~4.2B

Professional
searches in 2011

*as of Nov 4, 2011
**as of June 30, 2011

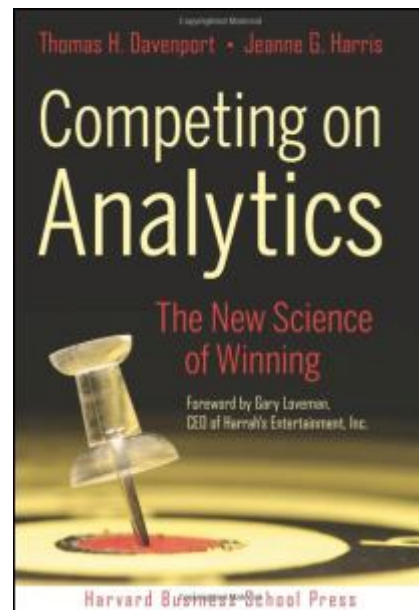
Multiple revenue channels

- Premium Subscriptions
- Self Serve Ads
- Hiring Solutions
- Marketing Solutions

Let's talk Data...



Business is recognizing the importance of analytics



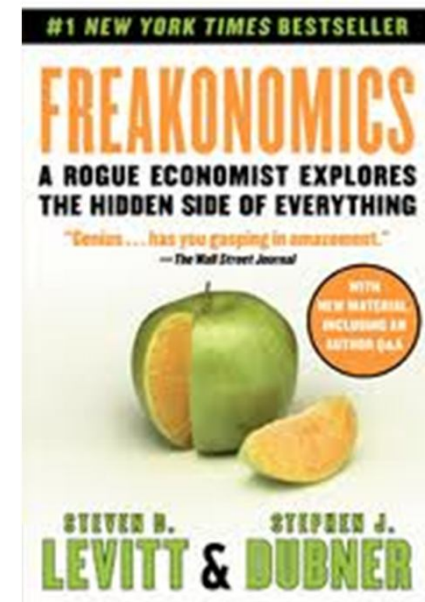
The Deciding Factor



The Power of ANALYTICS to Make Every Decision a Winner

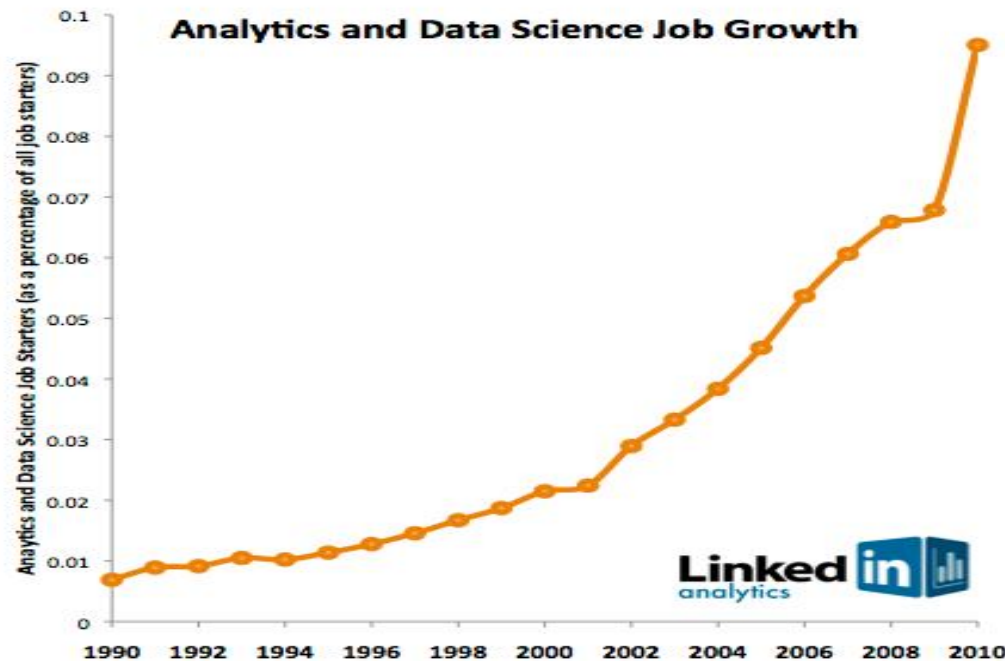
Larry E. Rosenberger
John Nash

WITH ANN GRAHAM

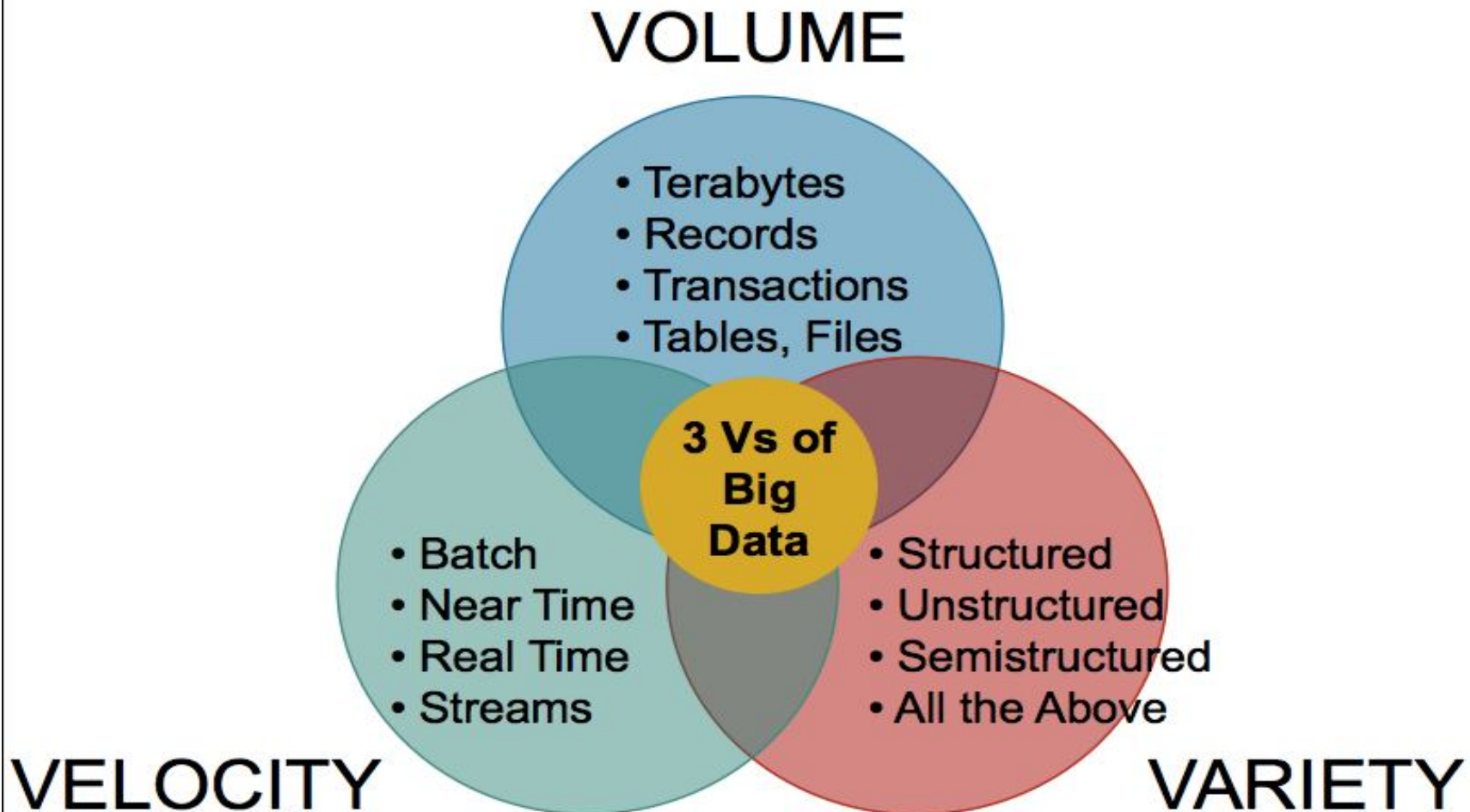


What makes a Data Scientist?

Data Scientist = Curiosity + Intuition + Data gathering + Standardization + Statistics + Modeling + Visualization + Communication



Big Data at LinkedIn



** Chart from Philip Russom- Research Director: TDWI*

What do we do with Data?

- Data Standardization
- Build innovative data products to help professionals
- Draw insights
- Drive the business

Before we can do that...

There are a few challenges that we have to overcome

- Scale
- Standardization
- Infrastructure

Few Data-Driven Products

Jobs You May Be Interested In



Companies

Recommendations, similar companies search, peer companies, and company browse maps



Groups

Recommendations, similar groups search



Talent Match



Related search



Behind the Scenes



Similar jobs



CAP



Profile browse maps



Jobs browse maps



Ad matching engine

$pCTR = f(\text{member, creative, advertiser, context, inventory, CCTR})$

Referral Engine



Pandora Search for People



Events You May Be Interested In



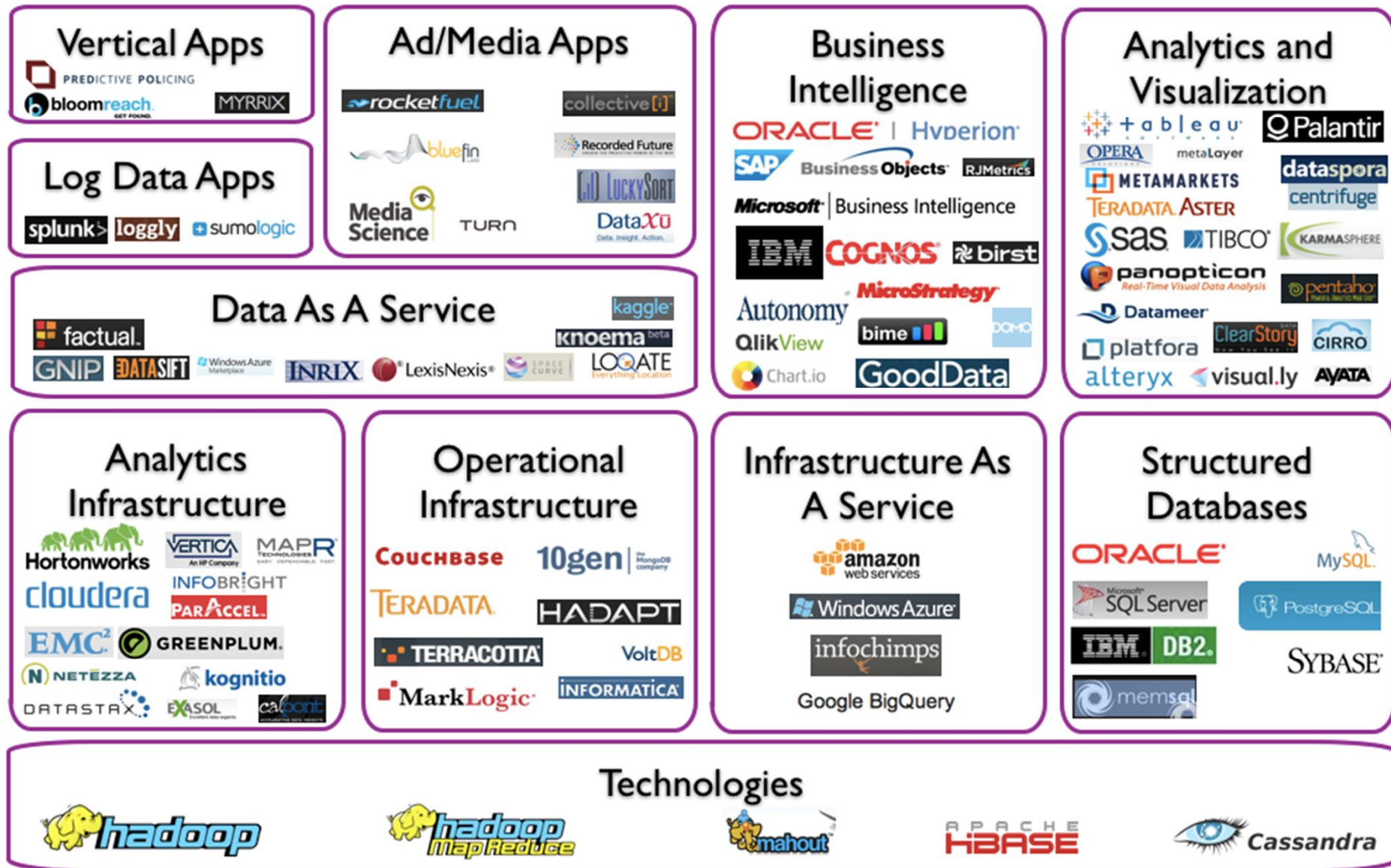
Groups browse maps



How do we do it?



Big Data Landscape



LinkedIn Sample Data Stack

ORACLE
TERADATA ASTER

Project Voldemort
A distributed database

 **hadoop**

Kafka

 **Azkaban**
Simple hadoop workflow.



 python™



Crowdsourcing

MicroStrategy
Best In Business Intelligence

SQL

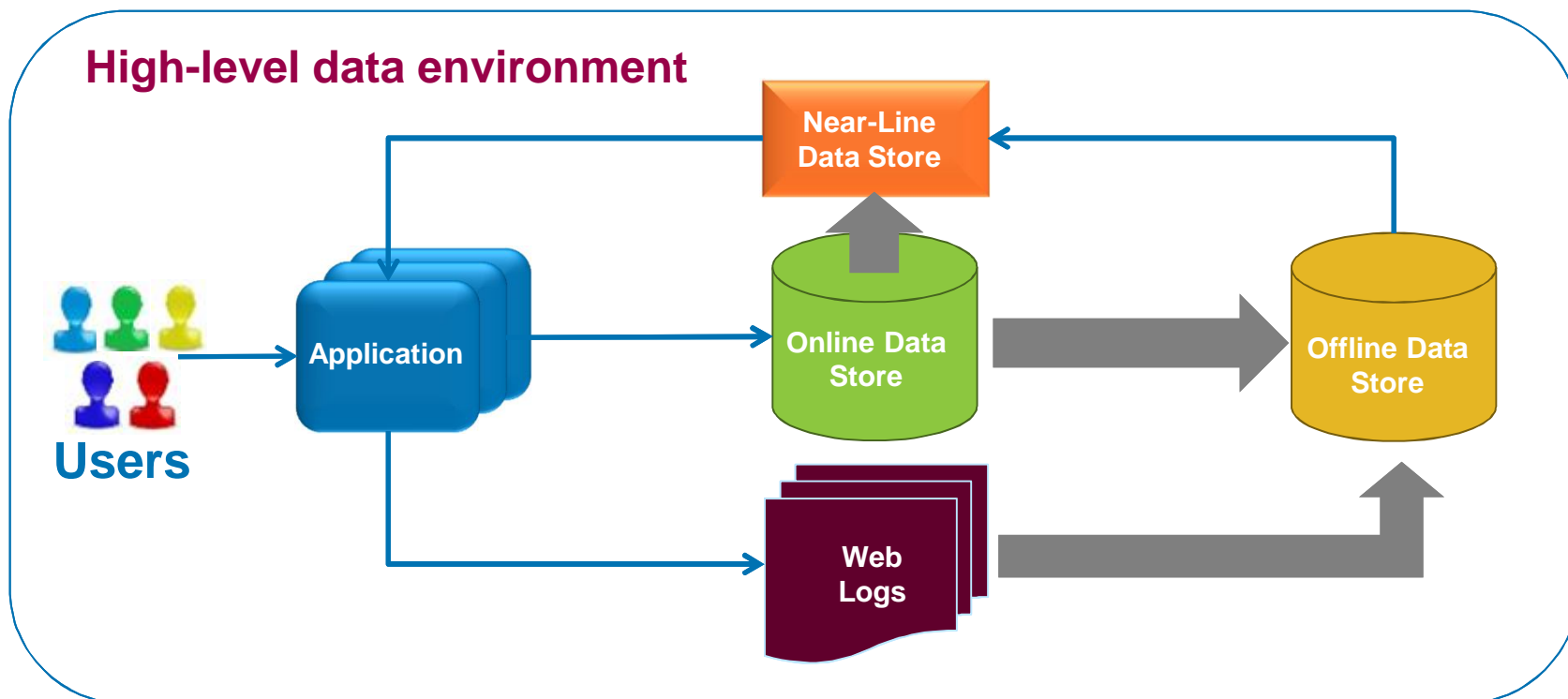
LinkedIn Reportal

Processing

 **Gephi**
makes graphs handy

prefuse

Big Data at LinkedIn



**Challenges so complex which
off-the-shelf or a few
technologies can't address**



**Built our own combination of
toolsets/ technologies to
meet specific requirements**

LinkedIn Data Stack – Online

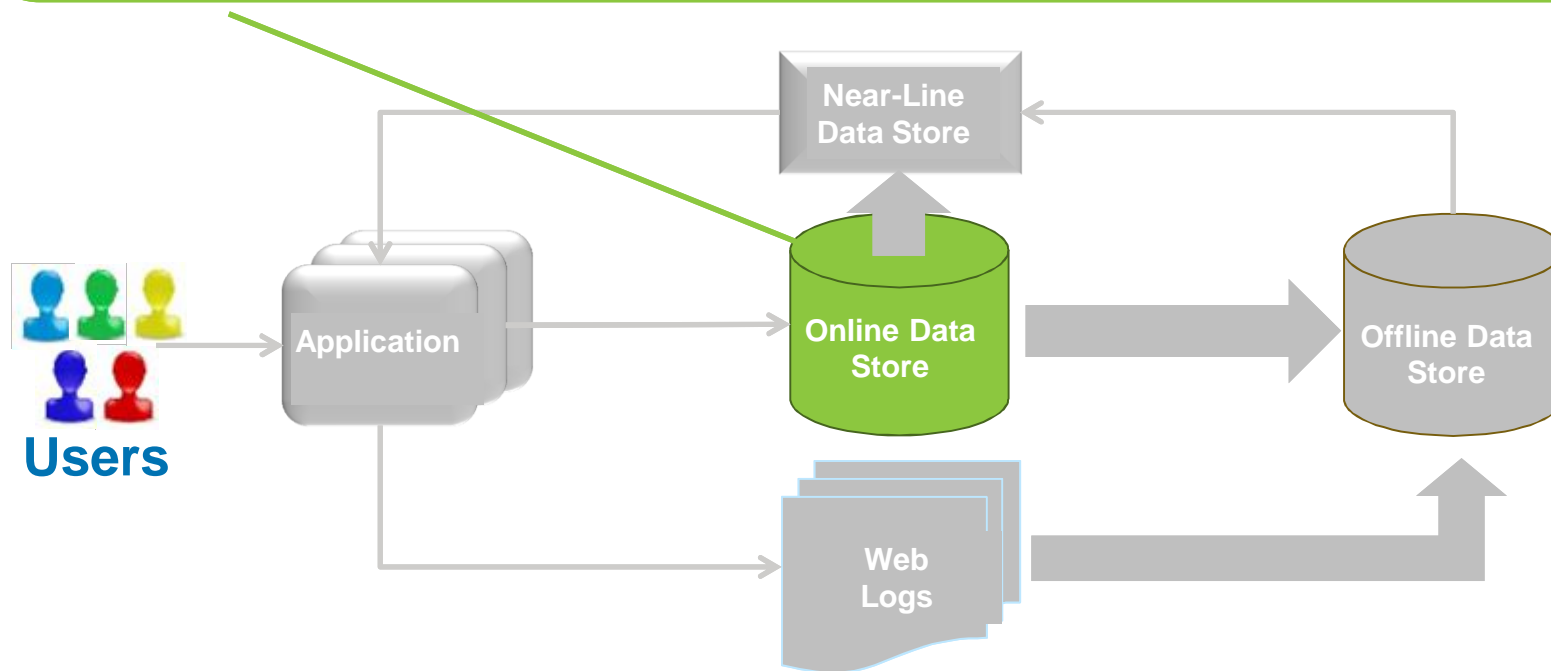
Systems

ORACLE®

ESPRESSO

Capabilities

- Rich structures (e.g. indexes)
- Change capture capability



LinkedIn Data Stack – Nearline

Systems



Voldemort



Zoie



Bobo

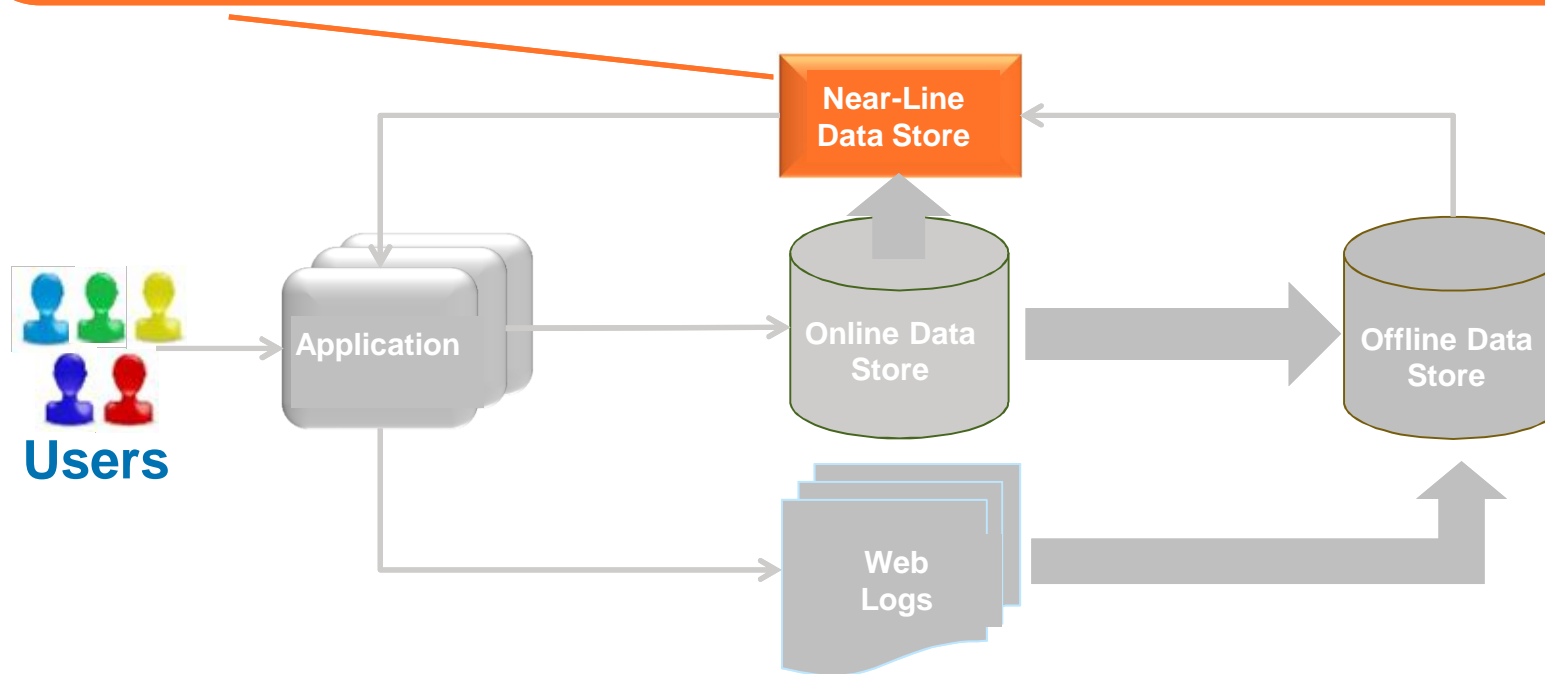


Sensei

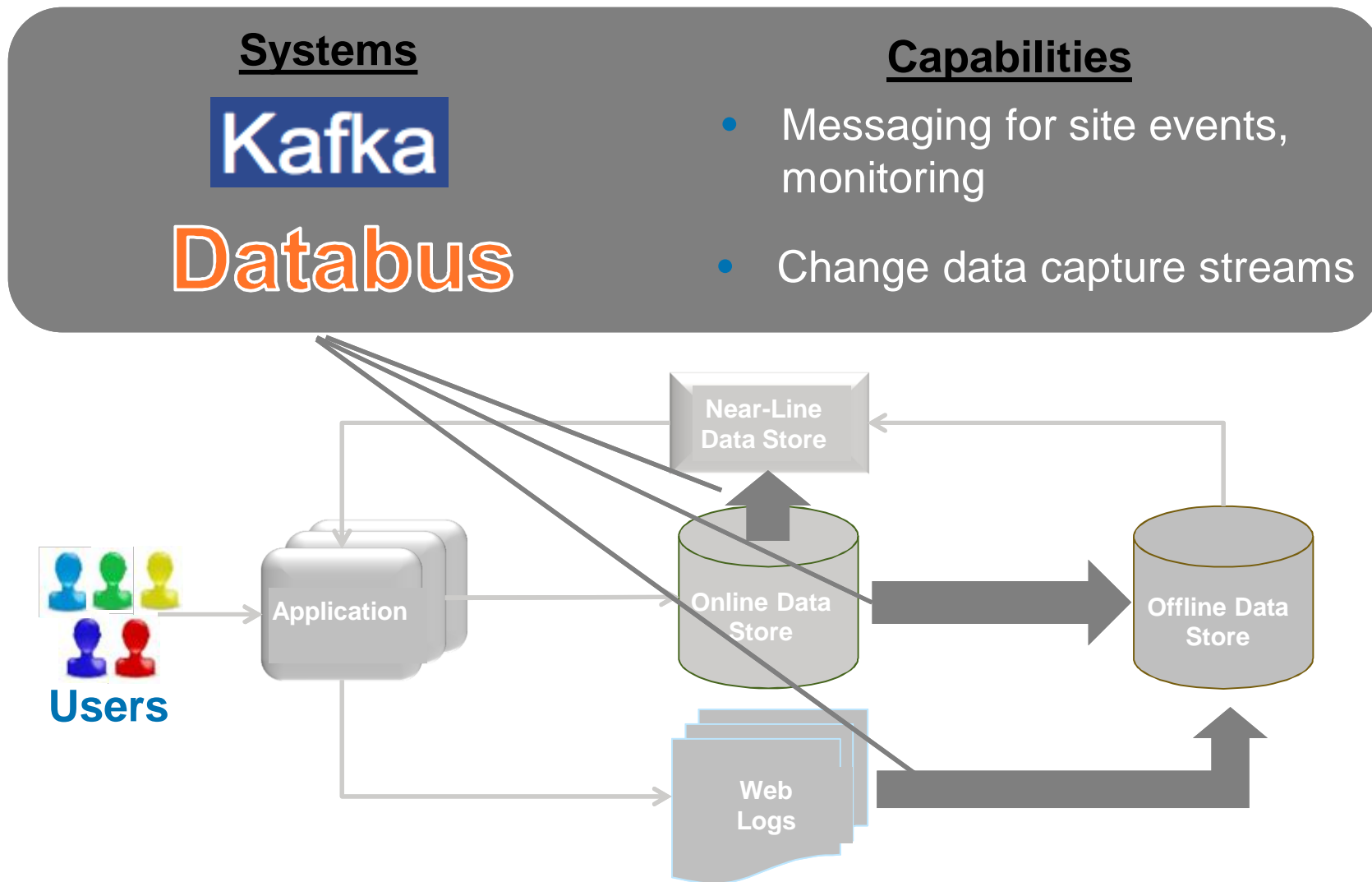
D-Graph

Capabilities

- Key value access
- Search platform
- Distributed Graph engine



LinkedIn Data Stack – Pipeline



LinkedIn Data Stack – Offline

Systems

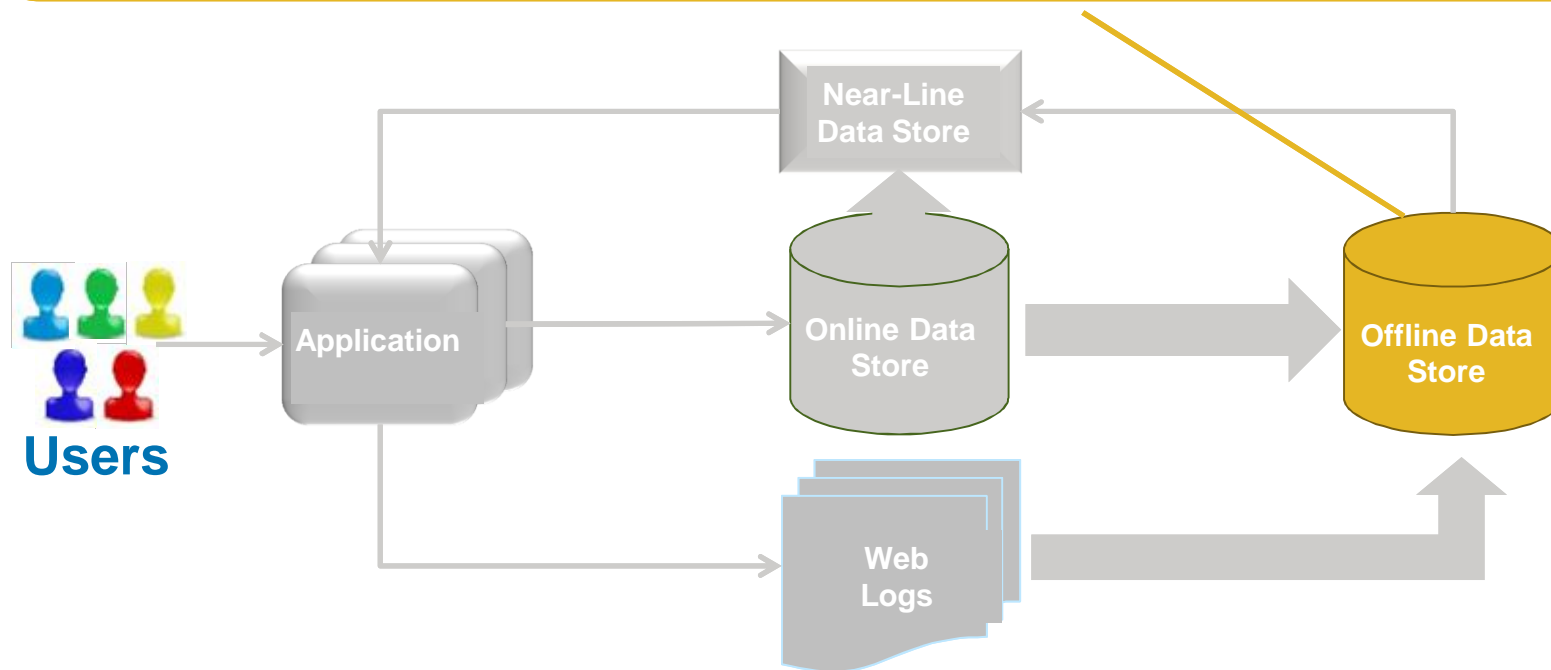


TERADATA ASTER

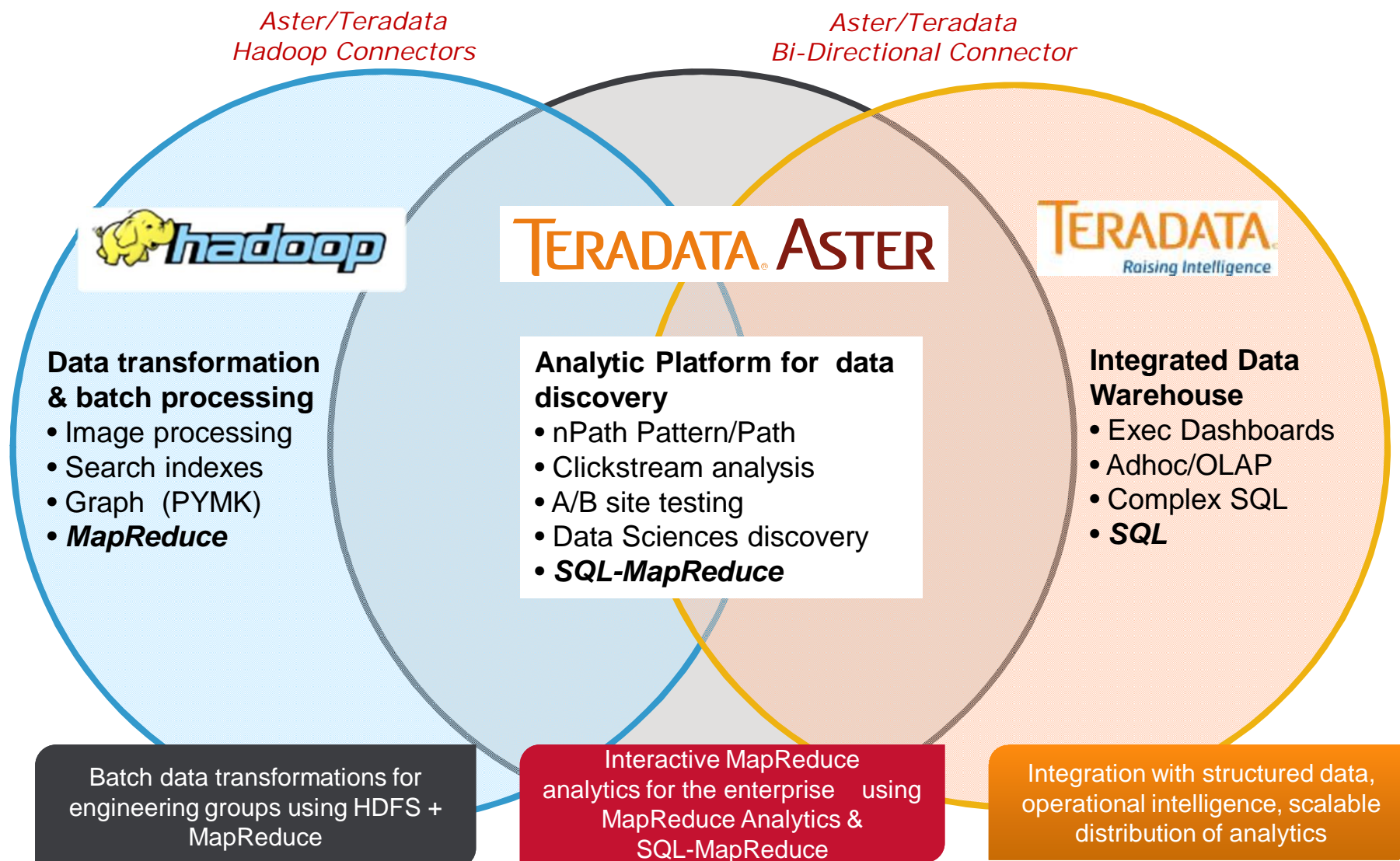
TERADATA

Capabilities

- Machine learning, ranking, relevance
- Warehouse and analytics



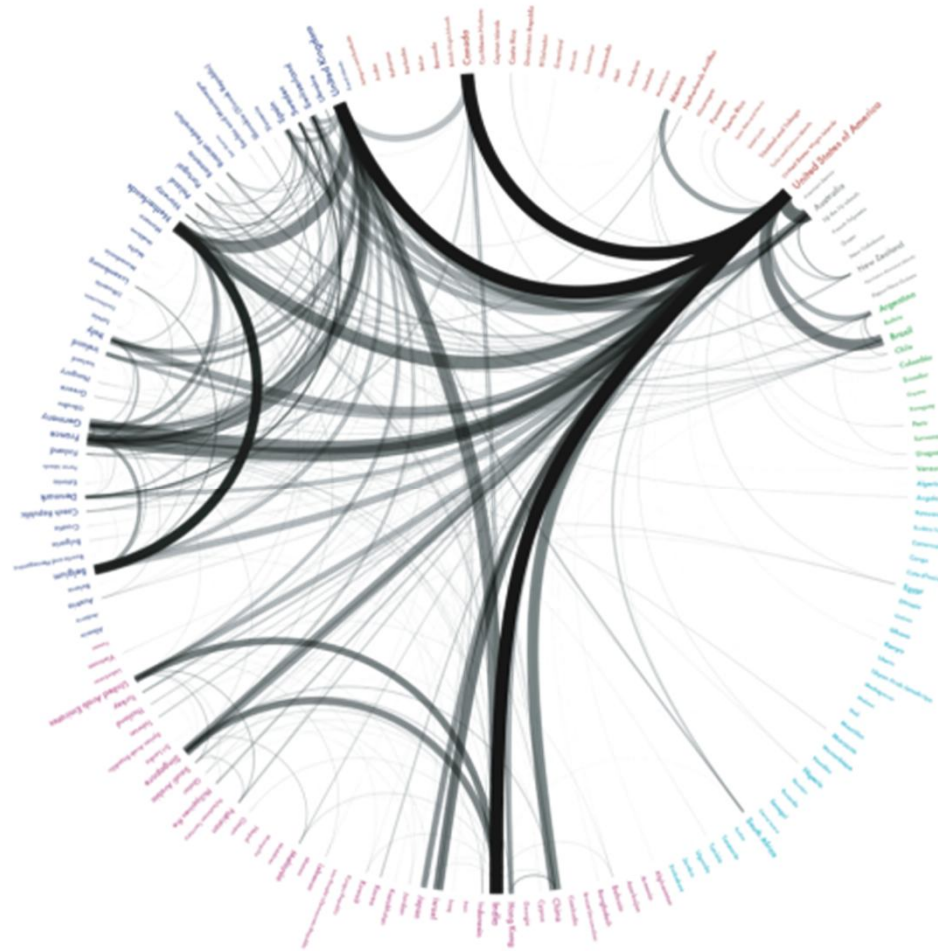
LinkedIn with Hadoop, Aster, and Teradata



INSIGHT



It's a global economy



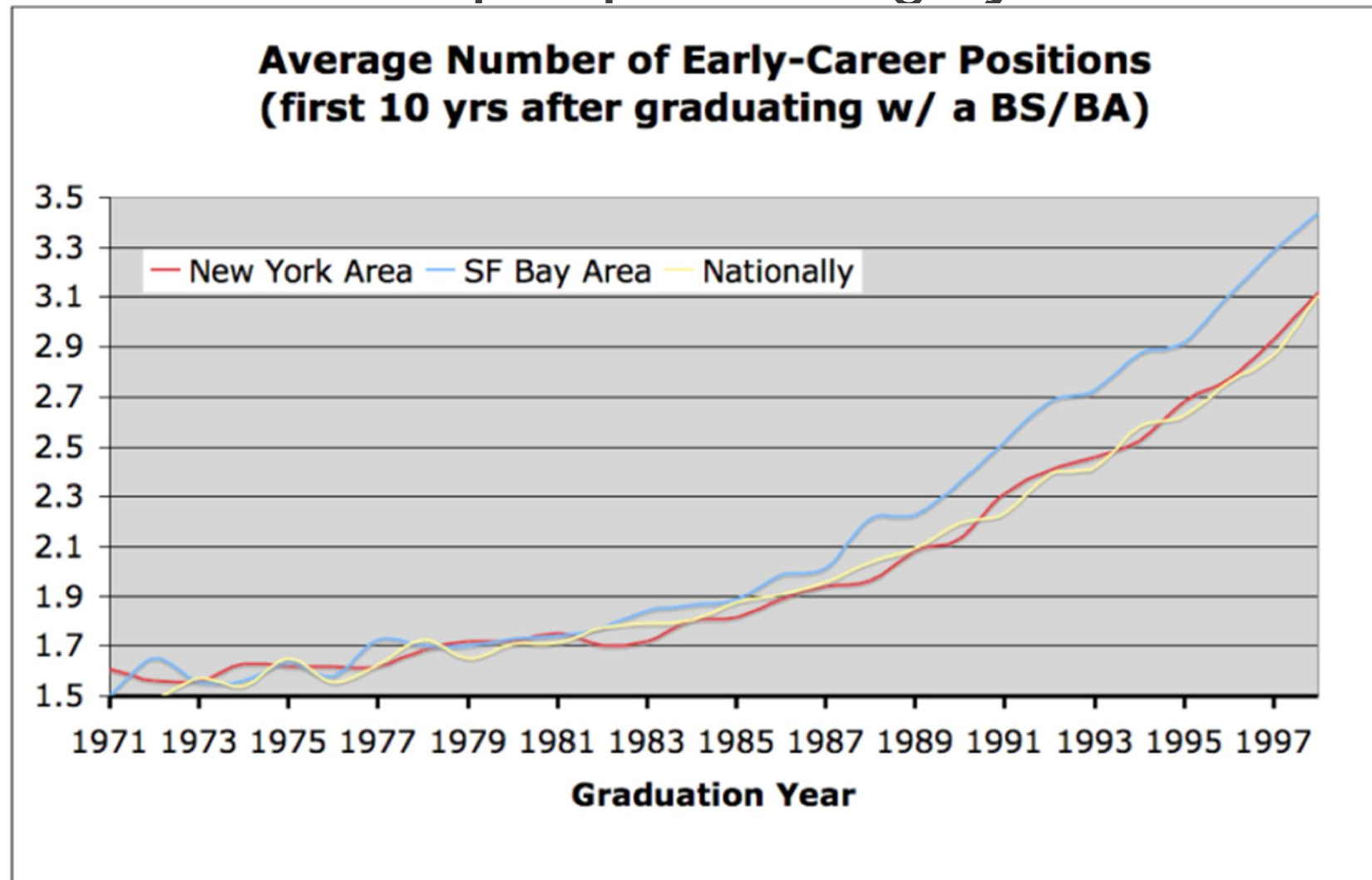
Country connectedness on LinkedIn

Data deep dives



Job migration after financial collapse

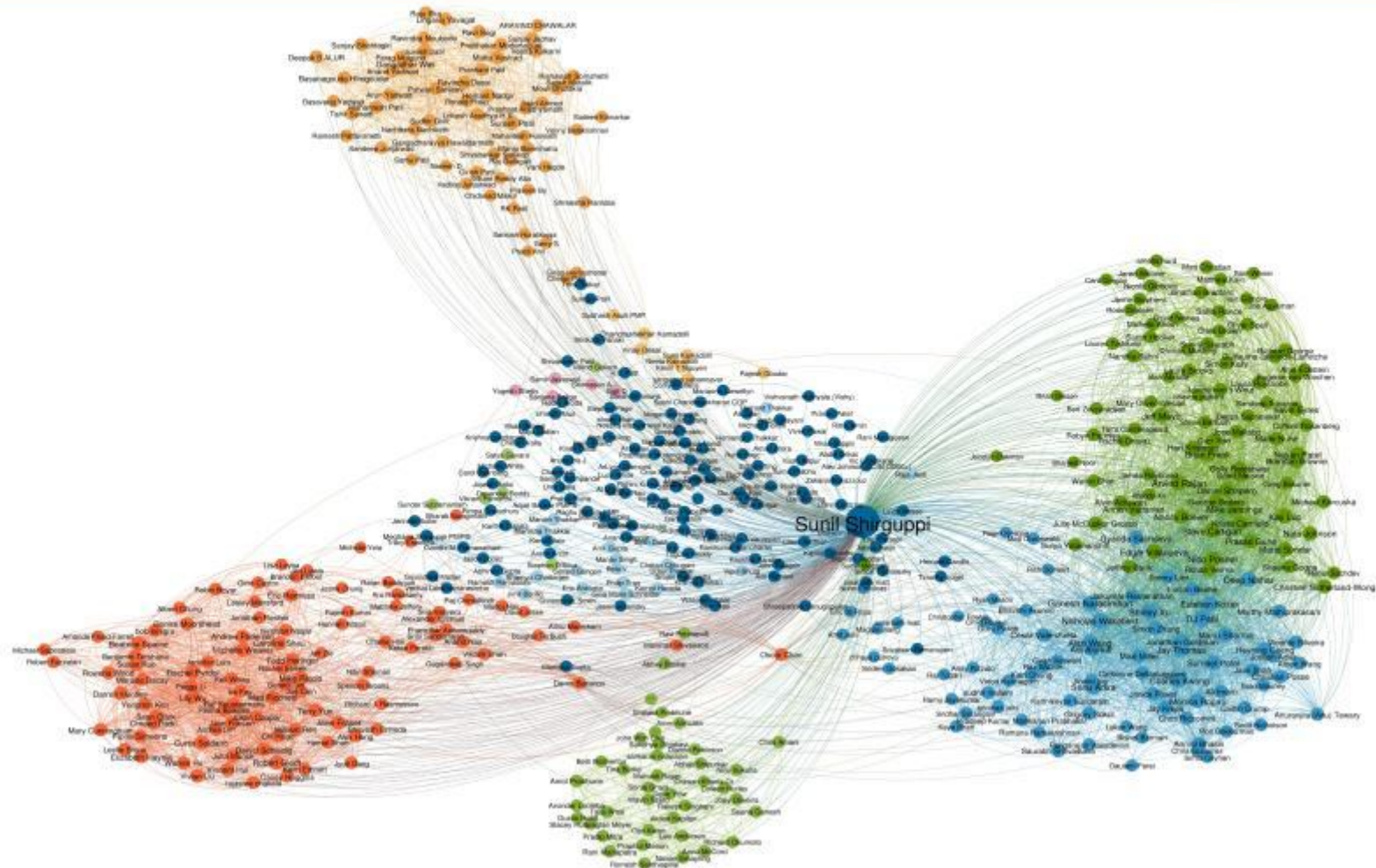
How Often do people change jobs?



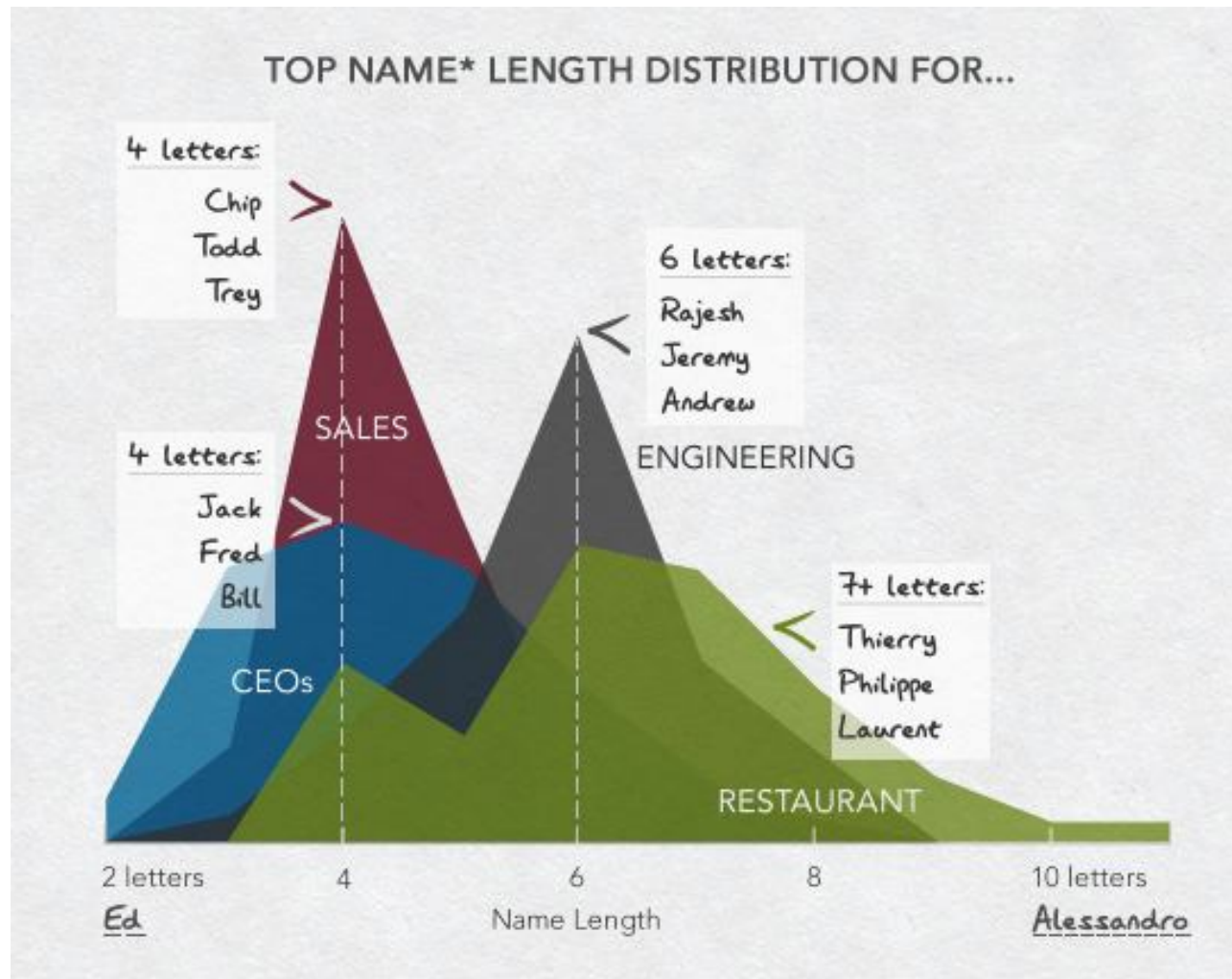
Visualization is important

as of September 28, 2010

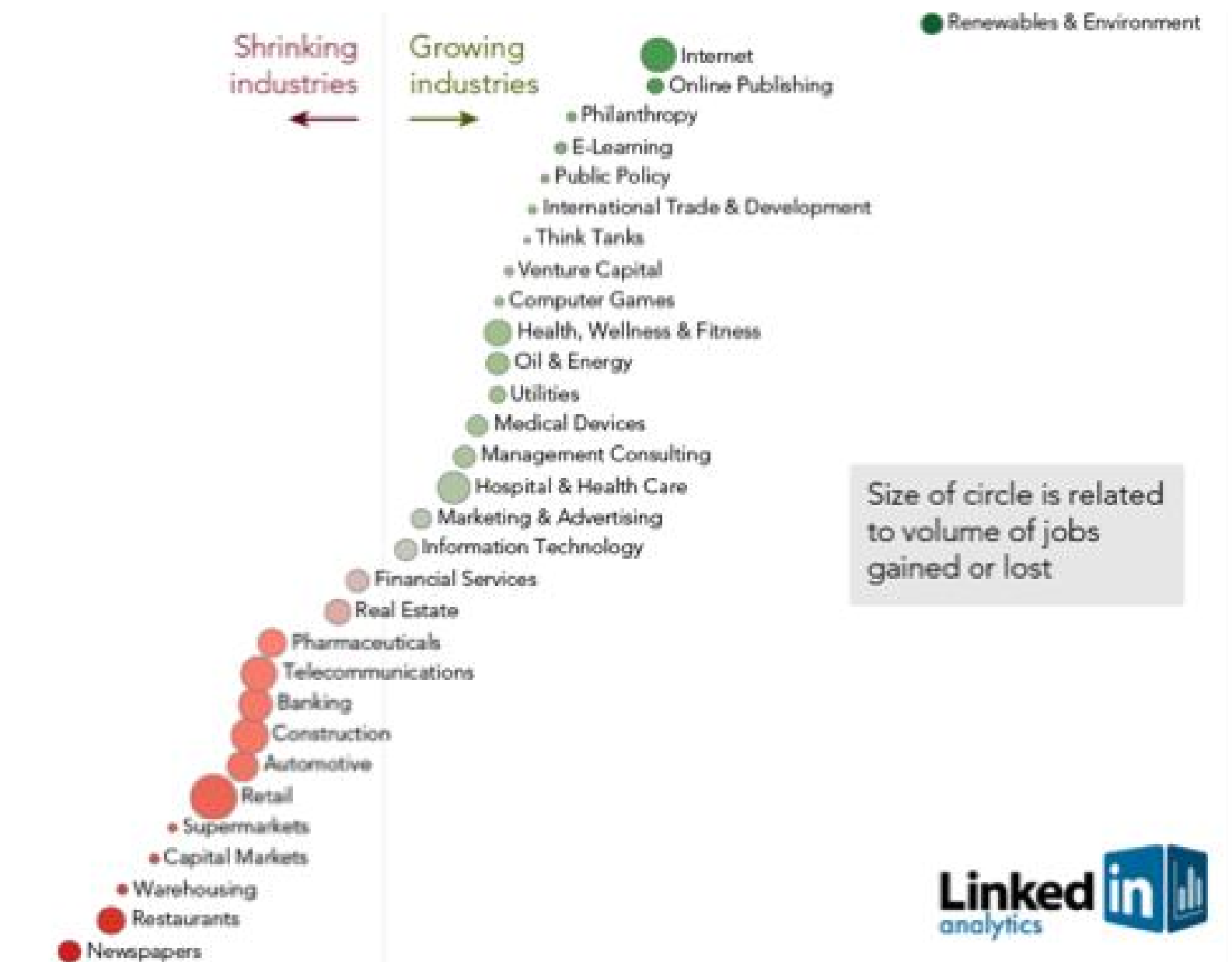
Sunil Shirguppi's Professional Network



If your name is Chip, you are likely in sales!



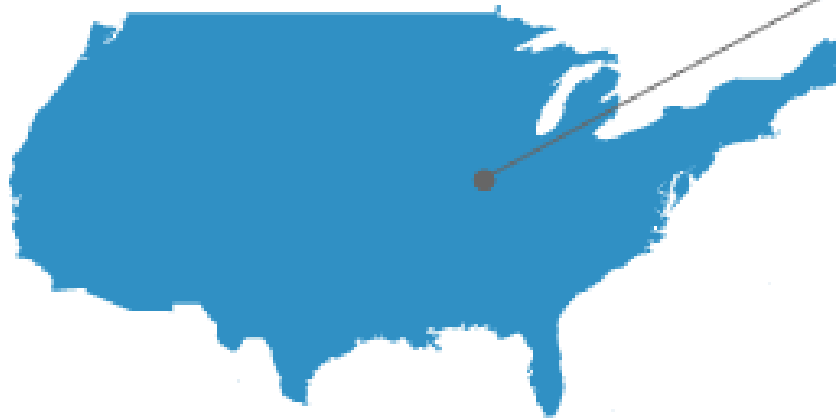
Industry Growth



Buzzwords



Top 10 overused buzzwords in LinkedIn Profiles in the USA




1. **Extensive Experience**
2. Innovative
3. Motivated
4. Results-oriented
5. Dynamic
6. Proven track record
7. Team player
8. Fast-paced
9. Problem solver
10. Entrepreneurial

Linked ®

What next?

- Self service analytics
- Metadata framework
- Integrate reporting solutions
- Go Mobile!
- Scalability and Data Quality

Challenges

- **Data volumes and availability**
 - Billion+ rows every day
 - Users in Global locations need data
 - **Data Quality**
 - User input data
 - Data standardization
 - **Multiple platforms**
 - Agile development
 - Data Integration
- 
- The IBM Canada logo is located in the bottom right corner of the slide. It features the word "IBM" in its characteristic eight-striped blue font, followed by "Canada" in a smaller, blue, sans-serif font. Below "Canada" is the word "Ltd." in an even smaller font. To the right of the main logo, there is a smaller, partially visible version of the same logo.



Key Learnings

- **Self Service**

- Making data accessible to key stakeholders in a timely manner creates tremendous value.
- Viz is more important than we think

- **Measuring your future investments**

- Performance is not the only measure
- Company fundamentals matter

- **As an Data team, be in control of your destiny**

- Identify what to measure and lead by metrics
- Become the Think-tank

Web 3.0 – It's all about data!!

ULTIMATELY...

It is all about the people!



Thank You!

