WK#10 Summary

- Technology trends in 2014 and beyond … 2020
- My other courses
- My website and contact information
- Independent study
- Q&A

Six Fast-Growing Careers Taking Over the U.S.

Career #2: Database Administrator

Projected Job Growth 2010-2020: 31 percent*

If you are excited by working with digital information, a rapidly-growing career as a database administrator may feed your data desire.

The U.S. Department of Labor states that database administrators are responsible for making sure that the company's computer database runs smoothly. This entails creating and managing the database, in addition to making revisions and performing tests when necessary. Database administrators also work to prevent security breaches and data loss, says the Department of Labor.

Why It's Taking Over: "With the explosion of digital information, database administrators, particularly in SQL Server and Oracle, are essential to protecting company information from digital security threats," says Merritt.

Click to Find the Right IT and Information Systems Program.

The Department adds that the rise of electronic medical records will also spur dema
administrators working in the health care industry. Additionally, the Department not
administrators are in high demand because there is simply a shortage of qualified w

Education Options: To pursue a career as a database administrator, you may nee
degree in management information systems or a computer-related field, according
Companies with large databases may prefer those with an MBA.

http://education.yahoo.net/articles/six_fast-growing_jobs.htm

Career #1: Personal Financial Advisor
Career #2: Database Administrator
Career #3: Market Research Analyst
Career #4: Medical Assistant
Career #5: Physical Therapy Assistant
Career #6: Software Developer

Cloud, Consumerization (BYOD), Big Data/Analytics, Social Media

Top 10 Strategic Technology Trends for 2014

Gartner®

Converging Forces	Mobile Device Diversity & Management
	Mobile Apps & Applications
	The Internet of Everything
	Hybrid Cloud & IT as Service Broker
Derivative Impact	Cloud/Client Architecture
	The Era of Personal Cloud
	Software Defined Anything
	Web Scale IT
Future Disruption	Smart Machines
	3-D Printing

http://www.gartner.com/newsroom/id/2603623

Gartner
SYMPOSIUM ITXPO® 2013

6 – 10 October | Orlando, Florida
gartner.com/us/symposium

The World's Most Important Gathering of CIOs and Senior IT Executives

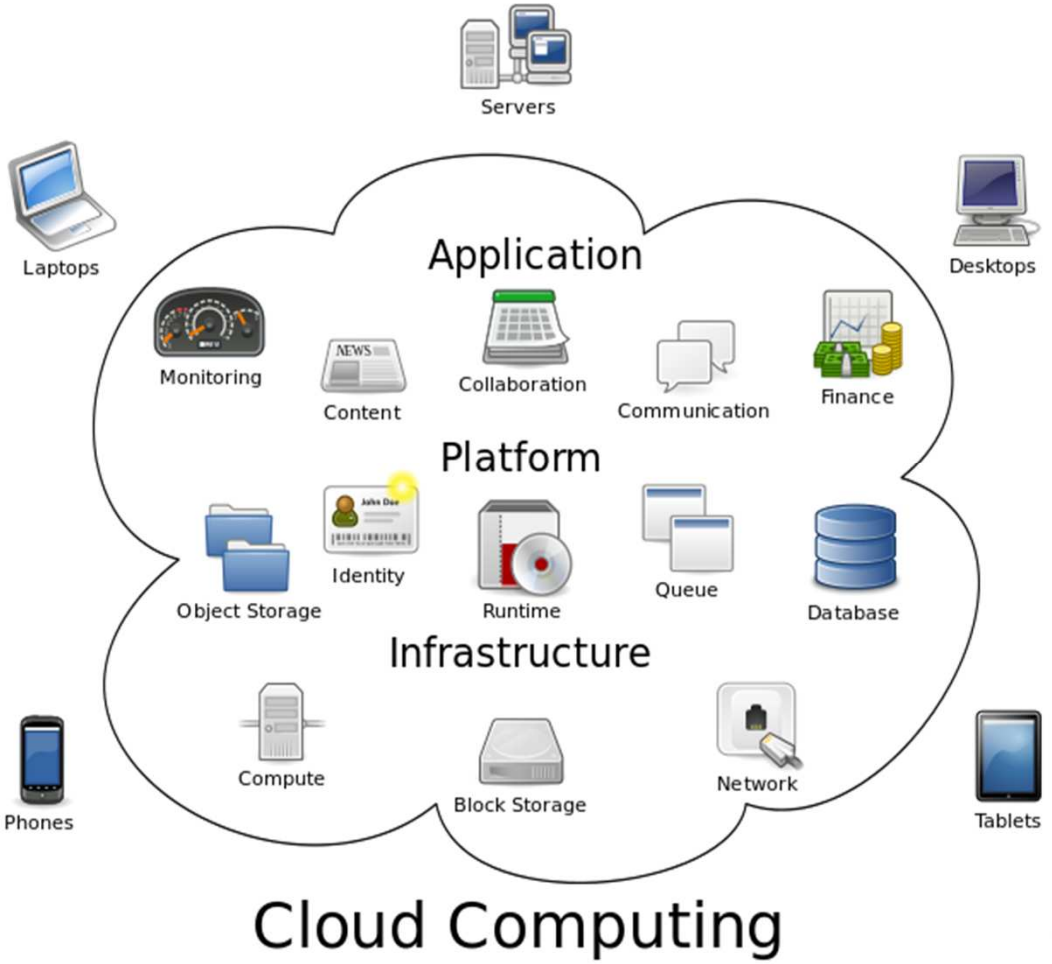
Mobile Device Diversity & Management

- Gartner expects no single vendor or platform will completely dominate the mobile industry. **Google**, **Apple**, and **Microsoft** will all be important players.
- A lot more personal area networks by **2017** to connect all their **wearable devices**, which will connect though their smartphones to the network.
- A typical knowledge worker will use **three to five devices by 2016.**

http://forwardthinking.pcmag.com/none/316750-gartner-s-top-10-strategic-technology-trends-for-2014

Hybrid Cloud & IT

- "hybrid IT" affects not only technology, but also the organization model.
- The role of the IT department changes, from that of a service provider to an advisor, broker, and provider to the business, helping lines of the business navigate through a variety of private cloud, public cloud, traditional services, and mobile apps, often creating an "enterprise app store."



The Internet of Everything / The Rise of Smart Machines

- From **Internet of Things (IoT)** to **Internet of Everything**
- This includes **people**, **things**, **information**, and **places** all connected.
- There will be **25 billion "things"** connected to the Internet by **2020**.
- The importance of **"big data analysis"** growing as everything is connected.
- This could end up changing **usage models**, and in turn, **business models**.
  - For instance, cars that know where and how they are driven could result in **pay-as-you-go** insurance.
- An explosion of new content coming from **Big Data** and the **Internet of Things**, **network-scale computing**, more powerful hardware, and new algorithms in areas such as **deep learning** and **natural language processing**.

IT Salary Survey by Dice.com

Tech salaries saw a nearly 3% bump last year, and IT pros with expertise in big data-related languages, databases and skills enjoyed some of the largest paychecks. Average U.S. tech salaries climbed to $87,811 in 2013, up from $85,619 the previous year, according to Dice's newly released 2013-2014 Salary Survey. Significantly, nine of the top 10 highest paying IT salaries are for skills related to big data. Here's the full list of the top 10 highest paying IT salaries:

1. R: $115,531
2. NoSQL: $114,796
3. MapReduce: $114,396
4. PMBok: $112,382
5. Cassandra: $112,382
6. Omnigraffle: $111,039
7. Pig: $109,561
8. Service Oriented Architecture: $108,997
9. Hadoop: $108,669
10. Mongo DB: $107,825

"Companies are betting big that harnessing data can play a major role in their competitive plans, and that is leading to high pay for critical skills," said Shravan Goli, president of Dice, in a statement. "Technology professionals should be volunteering for big data projects, which makes them more valuable to their current employer and more marketable to other employers."

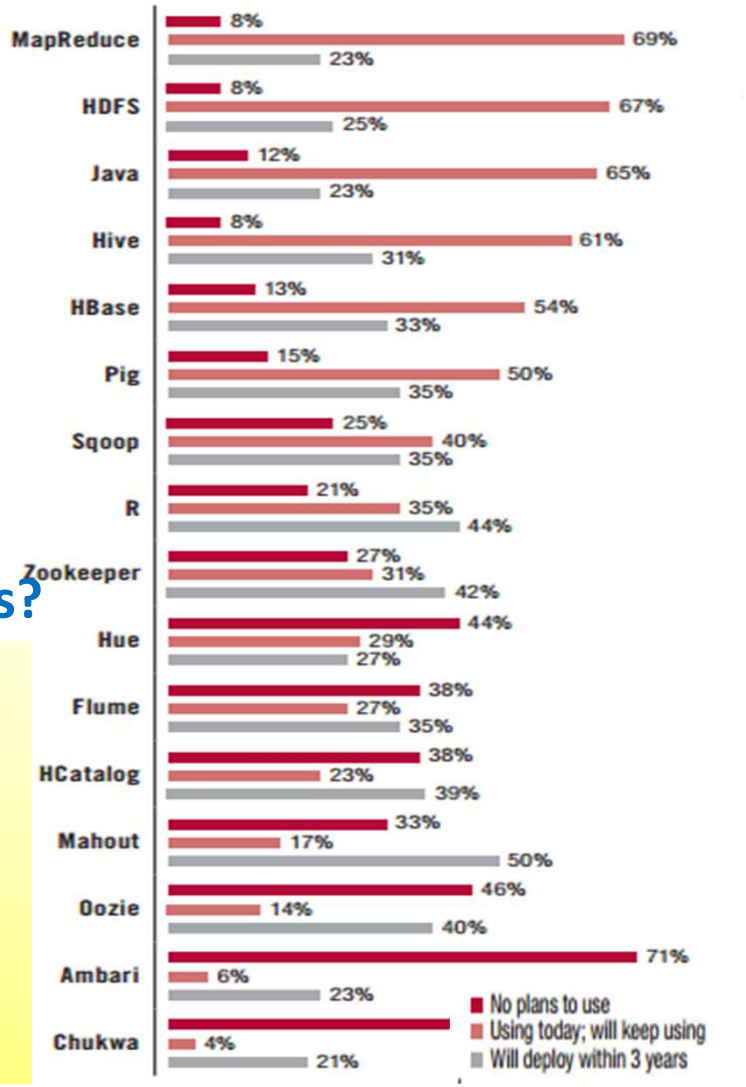
http://www.networkworld.com/news/2014/020714-big-data-dice-278557.html

TDWI BEST PRACTICES REPORT SECOND QUARTER 2013

INTEGRATING HADOOP INTO BUSINESS INTELLIGENCE AND DATA WAREHOUSING

Which of the following Hadoop and related technologies are in production in your organization today?
Which will go into production within three years?

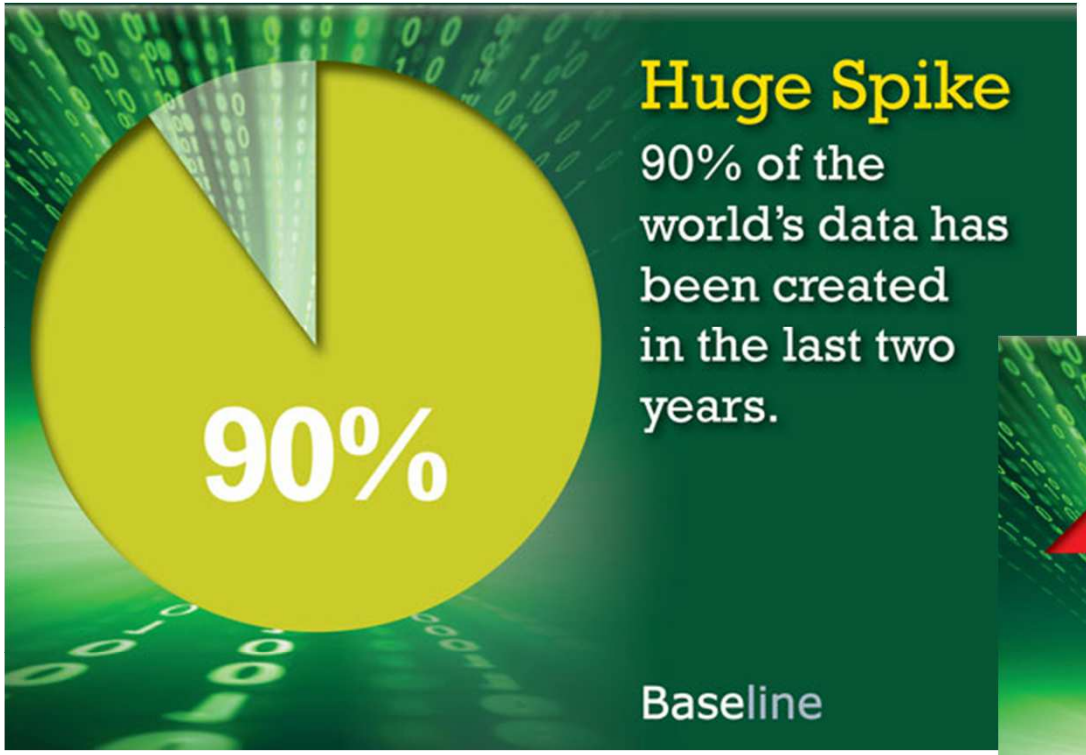
- According to this report's survey, the Hadoop products **most commonly used today are MapReduce, HDFS, Java, Hive, HBase, and Pig.**
- Only **10%** of organizations surveyed have a Hadoop implementation in production today, but a whopping **51%** say they'll have one **within three years.**



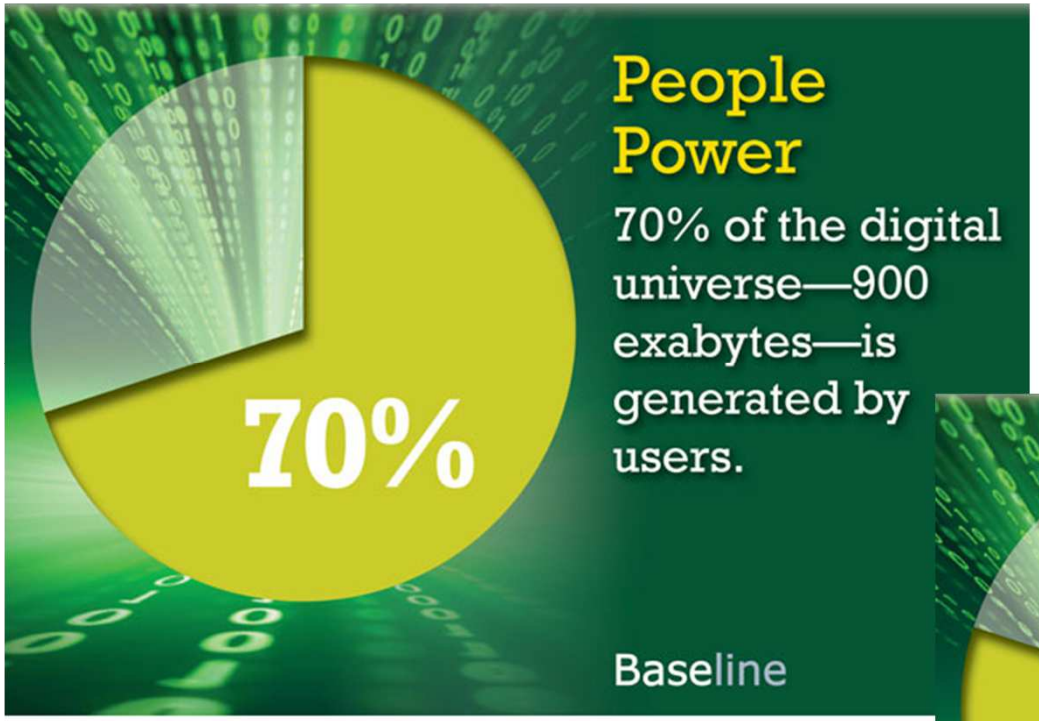
Big Data Is about to Get Even Bigger

- The growth of cloud computing and of technology in general has led to massive needs for storage and computational power, at a level that was previously reserved for only the largest of enterprises just a few years ago. In today's IT, a storage requirement of one petabyte (PB) is usually a good indicator of whether a company has joined the big data revolution.
- An InformationWeek article notes that, within 10 years, the massive storage solutions of today will not qualify as being on the scale of enterprise. In the next decade, exabytes (EBs) - each one of which accounts for just over 1,000 petabytes (PBs) - will be the new threshold for big data solutions. There are already a handful of massive organizations operating at that kind of scale, but those numbers will increase each year until the exabyte (EB) is the norm.
- In a Baseline article, an International Data Corporation (IDC) report suggests that the digital universe is doubling every 2 years and will reach 40,000 exabytes by 2020.

Surprising Statistics About Big Data



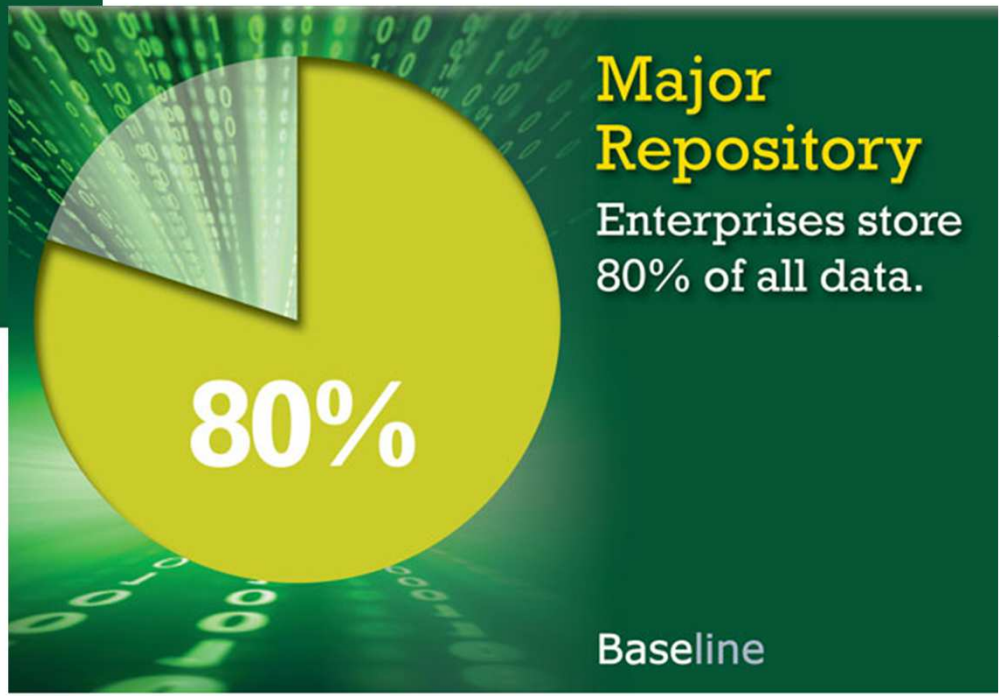
http://www.baselinemag.com/analytics-big-data/slideshows/surprising-statistics-about-big-data.html/



People Power

70% of the digital universe—900 exabytes—is generated by users.

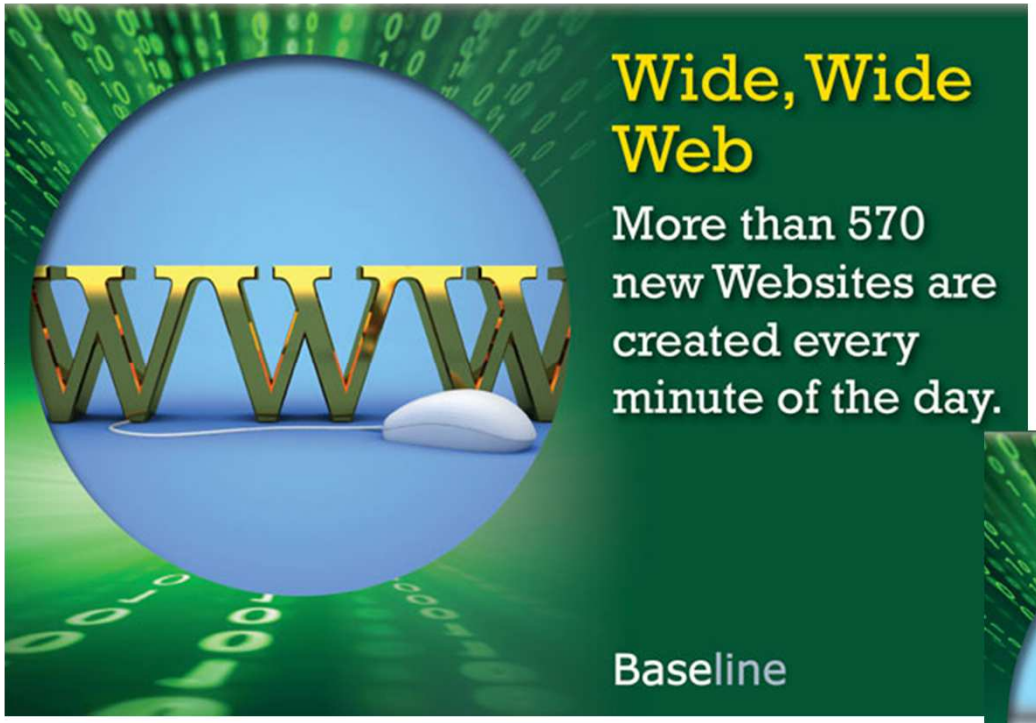
Baseline

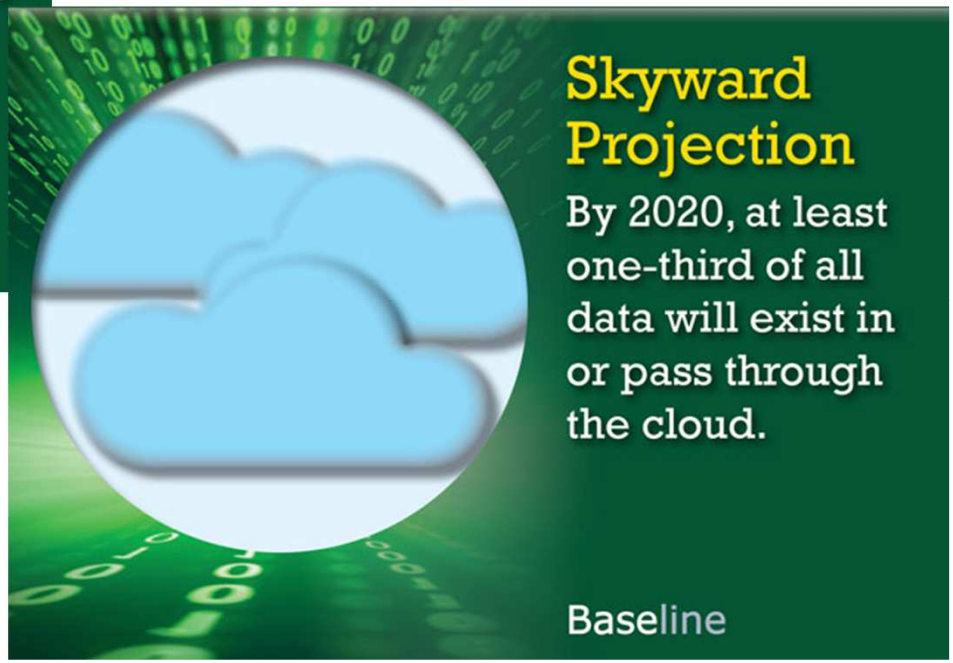


Major Repository

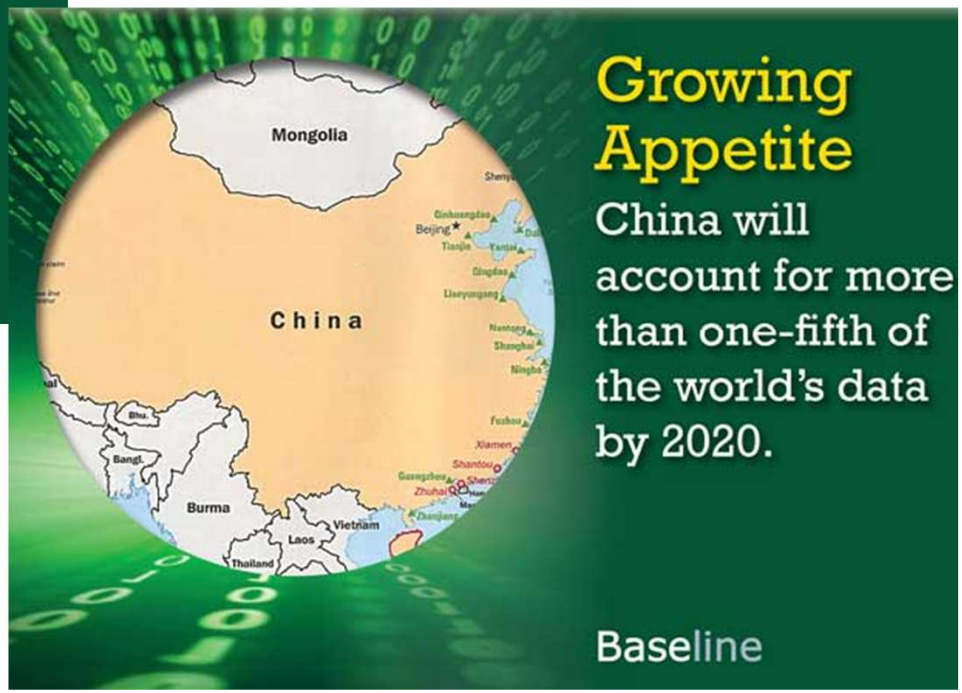
Enterprises store 80% of all data.

Baseline

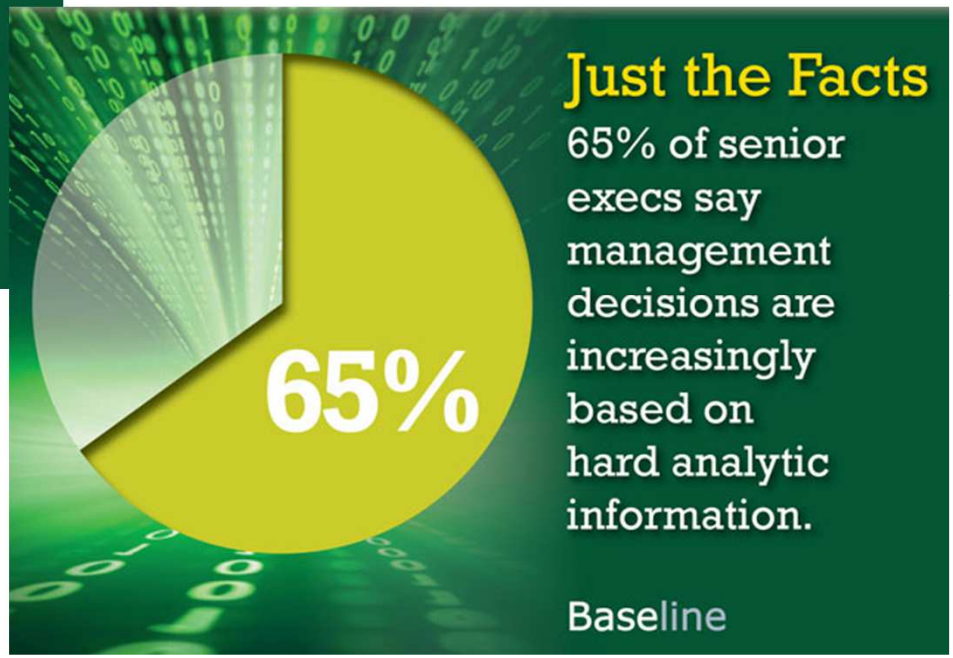

Wide, Wide Web
More than 570 new Websites are created every minute of the day.
Baseline


Skyward Projection
By 2020, at least one-third of all data will exist in or pass through the cloud.
Baseline

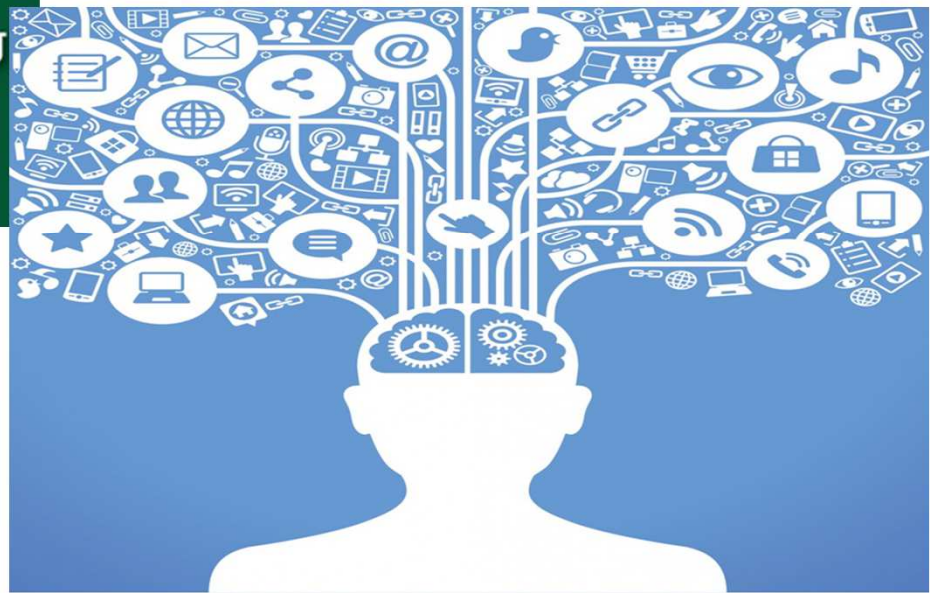

$200
MILLION
Executive Order
The White House administration is investing $200 million into big data research projects.
Baseline


Mongolia
China
Beijing
Tianjin
Shanghai
Burma
Vietnam
Laos
Thailand
Growing Appetite
China will account for more than one-fifth of the world's data by 2020.
Baseline


10%
Data Is Money
A 10% increase in data accessibility translates into an additional $65.7 million in net income for a typical Fortune 1000 company.
Baseline


65%
Just the Facts
65% of senior execs say management decisions are increasingly based on hard analytic information.
Baseline

Open your mind to all data

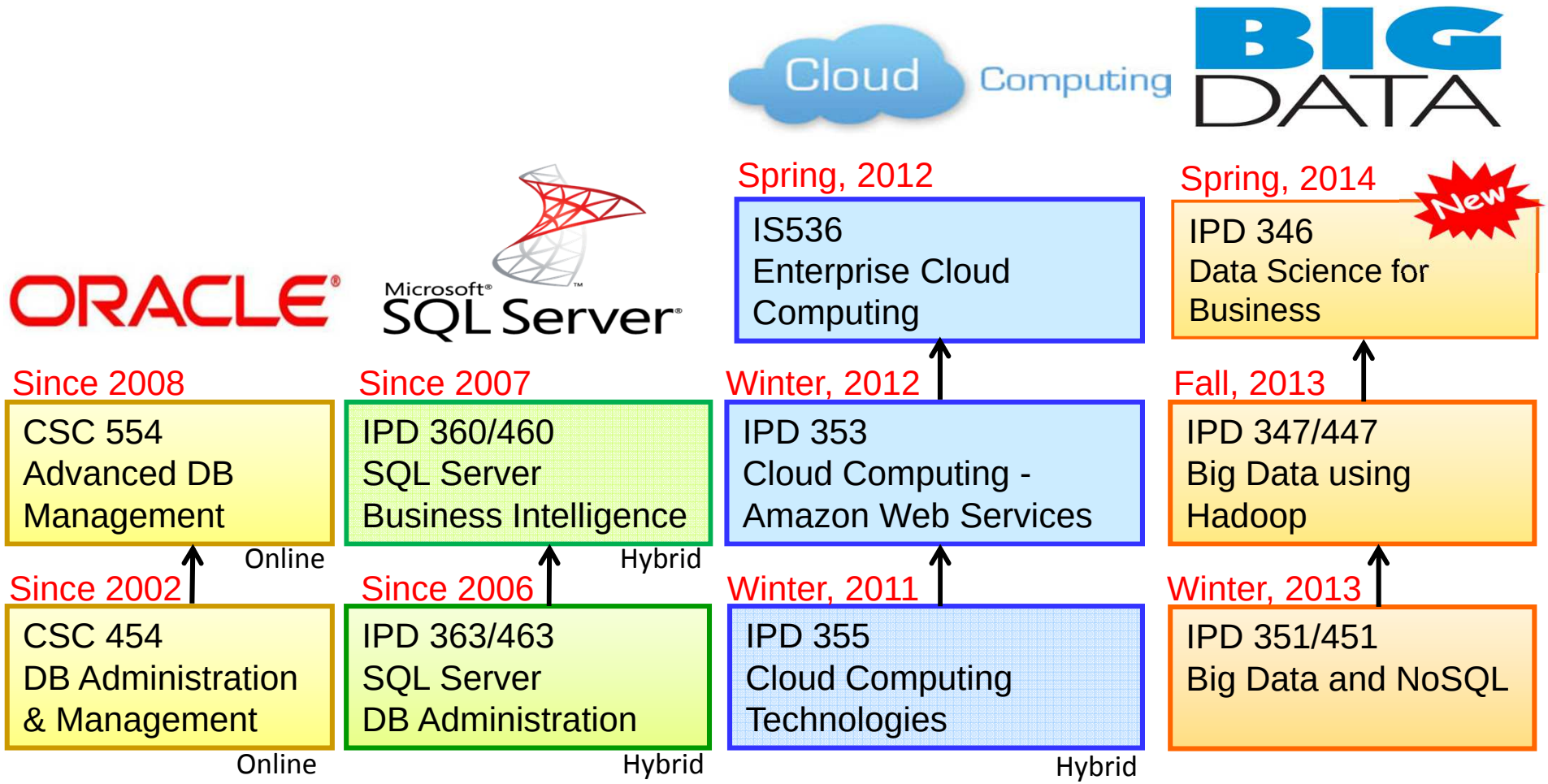


Top 12 funded Big Data Startup companies

Company	Funding (in $US mil.)	Services
Palantir	$650 million	Analytics applications
MongoDB	$231	Document-oriented database
Mu Sigma	$208	Data-Science-as-a-Service
Cloudera	$141	Hadoop-based software, services, training
Domo	$123	Software-as-a-Service business intelligence platform
Opera Solutions	$114	Data-Science-as-a-Service
Talend	$102	Application and business process integration platform
Hortonworks	$98	Hadoop-based software, services, training
Guavus	$87	Big data analytics solution
DataStax	$83.7	Cassandra-based big data platform
Opera Solutions	$84	Data-Science-as-a-Service
10gen	$73.4	MongoDB (open-source, document database)

http://bigdataanalyticsnews.com/top-12-funded-big-data-startup-companies/

Current/Future Database/Cloud Computing/Big Data Courses



Who attended IPD Big Data classes

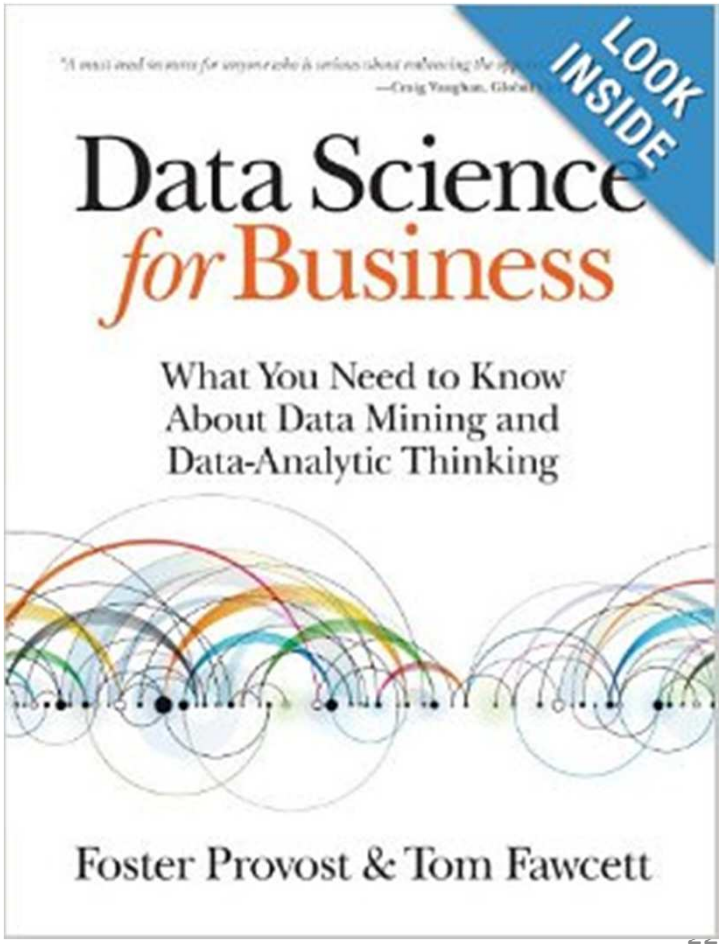
IPD 347/447 Big Data using Hadoop (Monday)

Cloudera Academic Partnership (CAP)

Week	Module	Topic
1	1	Introduction
	2	Big Data Overview
2	3	Introduction to Hadoop, Ecosystem
		Review and discuss HW#1
	4	CDH Installation
3	5	HDFS
	6	Sqoop
4	7	MapReduce
5	8	Hive
6	9	Pig
7	10	Cloudera Manager
	11	Hue
	12	Impala
8	13	Flume
	14	Hbase
9	15	Microsoft HDInsight
10	16	Planning Your Hadoop Cluster
	17	Hadoop Cluster Installation
11	18	Review, Certification Exam, Final project presentation

New class: Data Science for Business

- For decision makers, business owners
- Focus on case studies, demo, discussion
- Fewer lab/homework assignments
- Partner with Chicago local Data Scientists
- Partner with Big Data vendors
- Topics:
  - Machine Learning
  - Sentiment Analytics
  - Recommender System
  - Data Visualization
  - R, Mahout, etc.



IPD346 Data Science for Business (Tuesday)

Date	Class	Topic
4/8	1	Introduction, Technology trends, and Data Science Overview
4/15	2	Machine Learning Introduction
4/22	3	Predictive Modeling: From Correlation to Supervised Segmentation
		Clustering, Classification, Text mining
		Supervised and Unsupervised Learning
4/29	4	Introduction to R and basic data analysis
5/6	5	Building Recommender Systems
5/13	6	Business Problems and Data Science Solutions
		Sentiment Analysis, Market Segmentation, Fraud Detection, Risk Analysis
5/20	7	Data Analytics as a Service
		Monetizing Data, Internet of Things
5/27	8	Case Study: Social Media and Big Data
		Mahout
6/3	9	Data Quality and Governance
		Data Visualization
6/10	10	Data Science and Business Strategy
		Final project presentation

IPD355 Cloud Computing Technologies (Tuesday - Hybrid)

Week	Topic	Assignments
1	Cloud Computing Overview	AS#1 Workload
2	IT as a Service (IaaS, PaaS, SaaS)	AS#2 Discussion
3	Application development using Microsoft Windows Azure	AS#3 Windows Azure
4	Virtual Server in the Cloud	AS#4 VM in Azure
5	SQL Azure	AS#5 SQL Azure
6	Amazon Web Services (AWS)	AS#6 AWS
7	Cloud computing migration strategies, programming, management and monitoring, case studies	
8	Application development using the Force.com	AS#7 Salesforce
9	Cloud standards, security, compliance, and best practices	
10	Summary and final project presentation	AS#9 Final Project

IPD351/451 Big Data and NoSQL (Wednesday)

Week	Topics
1	Introduction, Big Data Overview
2	NoSQL Overview, MongoDB Concepts, Architecture and Installation
3	MongoDB Commands, Import/Export, Aggregation
4	Index, Geospatial index, Full Text Search
5	MongoDB Schema Design, MongoDB Replication
6	MongoDB Sharding
7	Cassandra Overview, CLI, CQL
8	Cassandra Data Model and DB Design
9	Cassandra Architecture and Replication
10	Graph Database
11	Review/Summary, Final Project Presentation

IPD 360/460 SQL Server Business Intelligence (Thursday - Hybrid)

Week	Instructor	Topic	Lecture / Lab	Assignment
1	Marco	Business Intelligence (BI) Overview Microsoft BI Platform Introduction	BI Concepts SQLServer Analysis Services (SSAS), Integration Services (SSIS), Reporting Services (SSRS) Management Studio and Business Intelligent Development Studio (BIDS)	DB Crossword
2	Marco	Integration Services I	Data Transfer - BCP, Import/Export	Building a package
3	Marco	Integration Services II	Introduction to SSIS packages – Data Flow Control Flow, Data sources, Destinations and basic transformations	Integration Services
4	Marco	Integration Services III & Power Pivot	Advanced transformations, Populating Data warehouse, Debugging Introduction to PowerPivot	Integration Services Power Pivot
5	Marco	Reporting Services I	Report Designer, Architecture Overview Building basic reports with Report Builder Report Deployment	Report development and deployment
6	Marco	Reporting Services II	Advanced Reports – Charts, Linked Reports, Cascading Parameters	Reporting Services
7	Raj	Analysis Services I	Dimensional Modeling – Dimensions, Attributes, Relationships, Hierarchies Using BIDS to develop a simple cube Building a simple cube Lab – Browsing the sample cube	Quiz (Dimensional Modeling)
8	Raj	Analysis Services II	Measure Groups, Perspectives, Calculations, Key Performance Indicators Hierarchies / Attributes drilldown Building advanced cube	Quiz – Introduction to Analysis Services
9	Raj	Analysis Services III	Advanced Dimensional Modeling	Analysis
10	Raj	Analysis Services Query Dand ata Mining	MDX fundamentals, Data mining scenarios with Excel	Project review (for graduate students)
11	Raj	Client Tools and Review	Performance Point Server, Excel	

IPD 363/463 SQL Server DBA Program – Saturday (Hybrid)

IPD 363 SQL Server Database Administration

6/7/2013

Weekday (Course Online)

SATURDAY (On-Campus)

Week #1 September 14

1. Introduction — AS#1 (6) Don't Submit
- SQL Server Overview
- SQL Server Certifications
- Database Administration

2. Installation — AS#2 (5) Don't Submit
- Planning
- Installation Options
- Post Installation Verification

Week #2

3. SQL Server Architecture — AS#3 (5)
- SQL Server Architecture
- Instance
- Upgrade vs. Migration

Week #3 September 28

4. Creating Databases and Data — AS#4 (5)
- Creating Databases
- Database Configuration

5. Data Types and Database — AS#5 (5)
- Schema
- Tables
- File Stream
- Column Store Index

Week #4

6. Data Types and Database Objects — AS#6 (5)
- Data Integrity
- Views and Synonyms
- Change Data Capture
- Change Tracking

Week #6 October 12

7. Database Backup — AS#7 (10)
- Backup Recovery Overview
- Backup Planning
- Backup Implementation Scenarios

8. Database Recovery — AS#8 (10)
- Recovery Models
- Recovery Process

IPD 363 SQL Server Database Administration

6/7/2013

Week #8

9. Large DB Features AS#9 (10)

* Stored Procedures, Functions, Triggers
* Compression, Partition

Week #7 October 28

10. Security *(Microsoft-R. Loforic)* AS#10 (5)

* Security Overview
* Security Model
* Managing Permissions
* Transparent Data Encryption
* User-Defined Server Roles
* Enhanced Auditing Features

11. Performance and Tuning

* Environment Configuration
* Index
* Locking
* Database Tuning Advisor
* Resource Governor

Week #8

12. Management Studio AS#11 (10)

* SQLCMD
* Managed Studio
* Scheduling Administrative Tasks
* Policy-based Management
* PowerShell Support

Week #9 November 8

13. DB High Availability-1 AS#12 (25) *Project

* High AvailabilityOverview
* Failover Clustering
* Log Shipping

14. DB High Availability-2

* Replication
* Database Mirroring
* Database Snapshot

Week #10

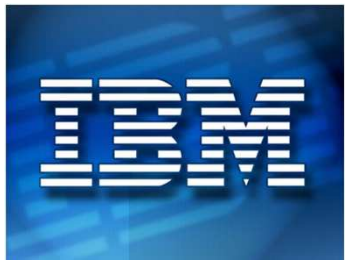
15. Other new features, Review

* AlwaysOn Availability Groups
* Distributed Replay
* Spatial Data
* Cloud DB

IPD353 Cloud Computing with Amazon Web Services (Saturday)

Week	Class	Topic
1	1	Cloud Computing Introduction, AWS Overview
	2	EC2 (Elastic Compute Cloud), S3 Overview and Setup
2	3	EC2, Load balancing, CloudWatch
	4	Route53
3	5	EMR / Hadoop
	6	SQS, SNS
4	7	Glacier, Security (IAM), CloudFormation
	8	ELB, EBS (Elastic Block Storage),
5	9	S3, CloudFront
	10	RDS, DynamoDB

Partnership



cloudera

Microsoft

DATASTAX

amazon
web services™

DataMartz
Business Analytics for Data Science

salesforce

GTR
CREATING POSSIBILITIES

Global Tech and Resources, Inc. (GTR)

Consulting/Partnership Opportunities (For You)

Goals:

- ❖ Helping students gain hands-on real-world experience
- ❖ Bridging IT, business and academia

To Do:

1. E-mail your resume to me
2. Tell me what you can do, what you want to do, and your availability

Options:

- ➢ Independent Study: Earn credits along with real-world projects
- ➢ Volunteer Work: Gain real-world experience
- ➢ Fee-based Work: Get paid

I will provide the following services to you: Free

- ➢ Matching skills and opportunities
- ➢ Coaching and mentoring
- ➢ I will NOT share your information without your consent

Consulting, Training, and Partnership (Me)

- **Provide consulting services**
  - SQL Server, Cloud Computing, Big Data (NoSQL, Hadoop)
  - Information Lifecycle Management
- **Provide corporate on-site training**
  - SQL Server, Cloud Computing, Big Data (NoSQL, Hadoop)
  - Can be customized
- **Build business partnership**
  - Consulting
  - Training
  - Sourcing

ProAct Consulting, Inc.

Marco Chou
mchou@cdm.depaul.edu
http://www.linkedin.com/in/marcochou
C: 847-226-2263

ProAct Consulting, Inc. (beta)

http://ec2-54-200-144-150.us-west-2.compute.amazonaws.com

Independent Study

- Choose a mutually agreed topic
  - Example: Big Data
    - » Social Media
    - » Sensor
    - » Cloud
    - » Mobile
- Get approval by your academic advisor
- One quarter *(summer is possible)*
- Delivery results (presentation, paper, program/app)
- Earn graduate-level 4 credits


thank
you!