MOVE OUT OF YOUR COMFORT ZONE: BIG DATA, THE ERA OF ABUNDANCE

ERIC HUNLEY
SENIOR MANAGER, STATE AND LOCAL GOVERNMENT PRACTICE

§sas
THE POWER TO KNOW.

- What?

- Why?

- How?

- Where?

WHAT

"Big data is what happened when the cost of storing information became less than the cost of making the decision to throw it away."

George Dyson

Big Data *(Noun)*

Data of such volume, variety or velocity that an organization must move ***beyond its comfort zone*** technologically to derive intelligence for effective decisions. Unstructured data introduces more variability and complexity.

http://bcove.me/1te37l04

BIG DATA | INTERESTING FACTS

- By 2020 Emerging Markets will lead the world in growth of data

- Within 5 – 6 years there will be more than 2.8 Zetabytes of data captured

  - KB, MB, GB, TB, PB, EB, ZB

  - 50X larger than today

- By 2016 there will be enough data accumulated equivalent to a 2 lane highway reaching from Tokyo to San Francisco

- Nearly ½ of the worlds data is unprotected

- Only .5% of all data is actually analyzed
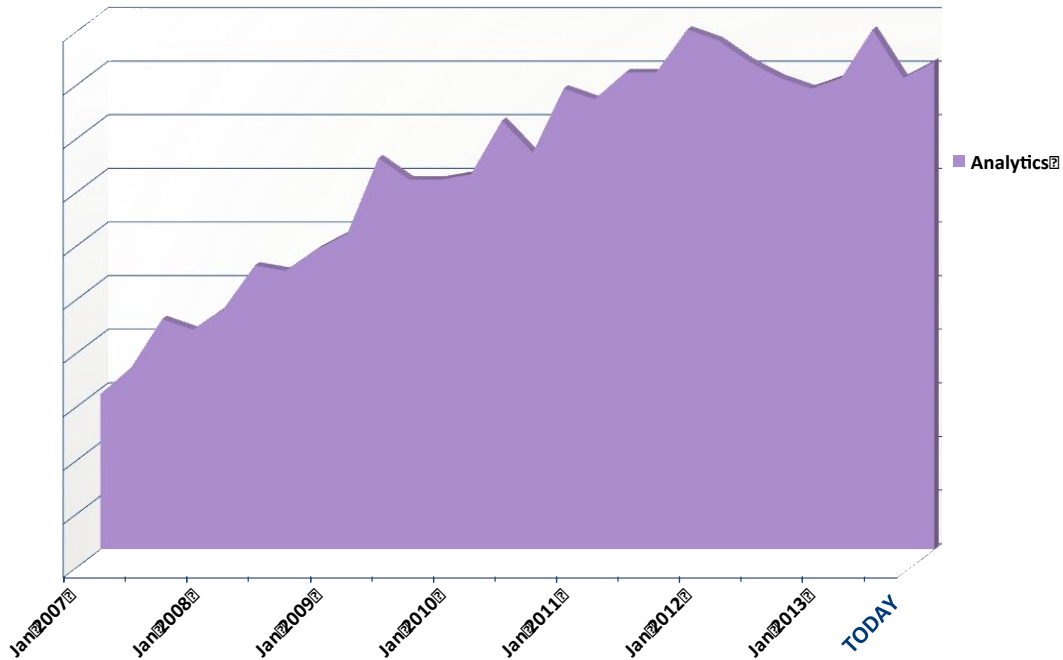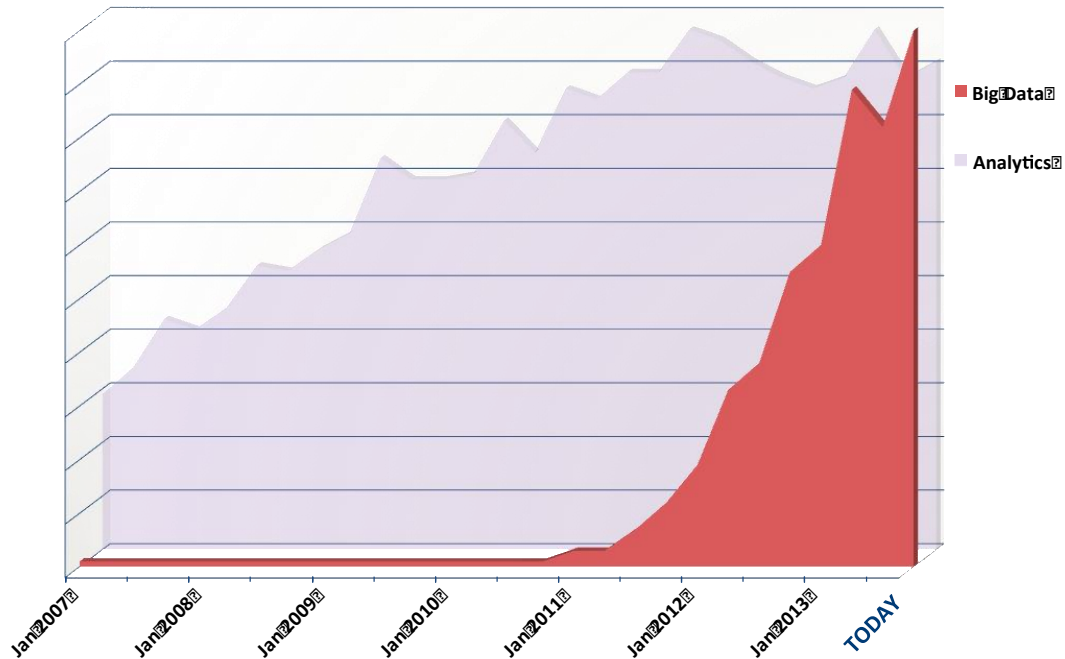
- Data → Information → Knowledge → Wisdom

§sas. | THE POWER TO KNOW.

WHY

GOOGLE TRENDS

"ANALYTICS"

Analytics

Jan 2007   Jan 2008   Jan 2009   Jan 2010   Jan 2011   Jan 2012   Jan 2013   TODAY

GOOGLE TRENDS

"BIG DATA"

Big Data

Analytics

Jan 2007 Jan 2008 Jan 2009 Jan 2010 Jan 2011 Jan 2012 Jan 2013 TODAY
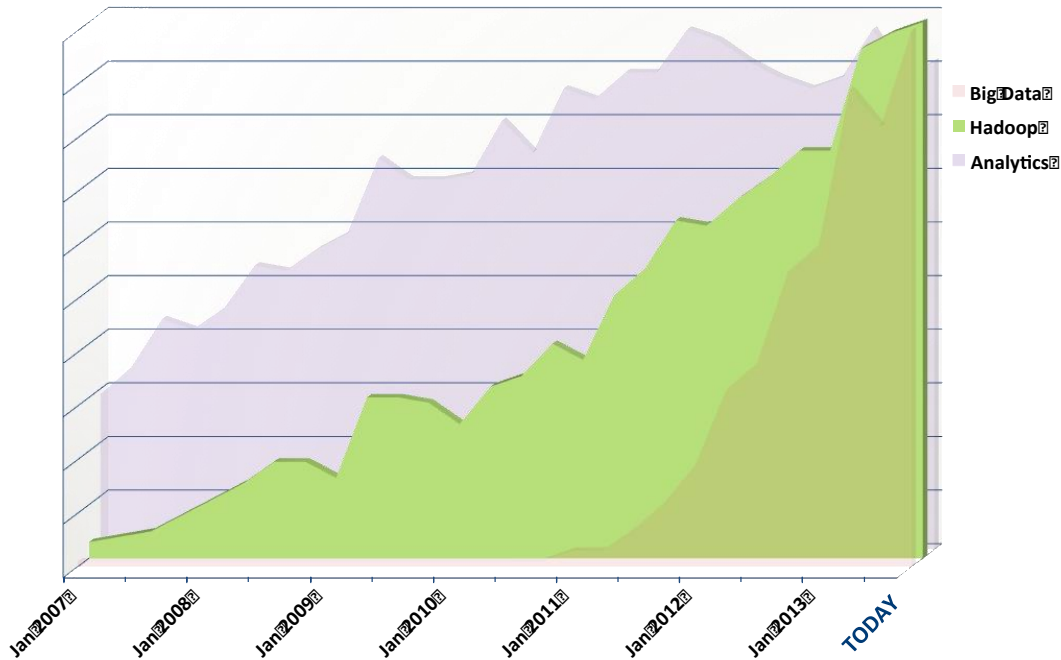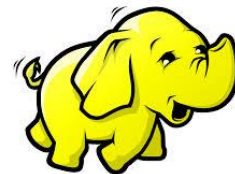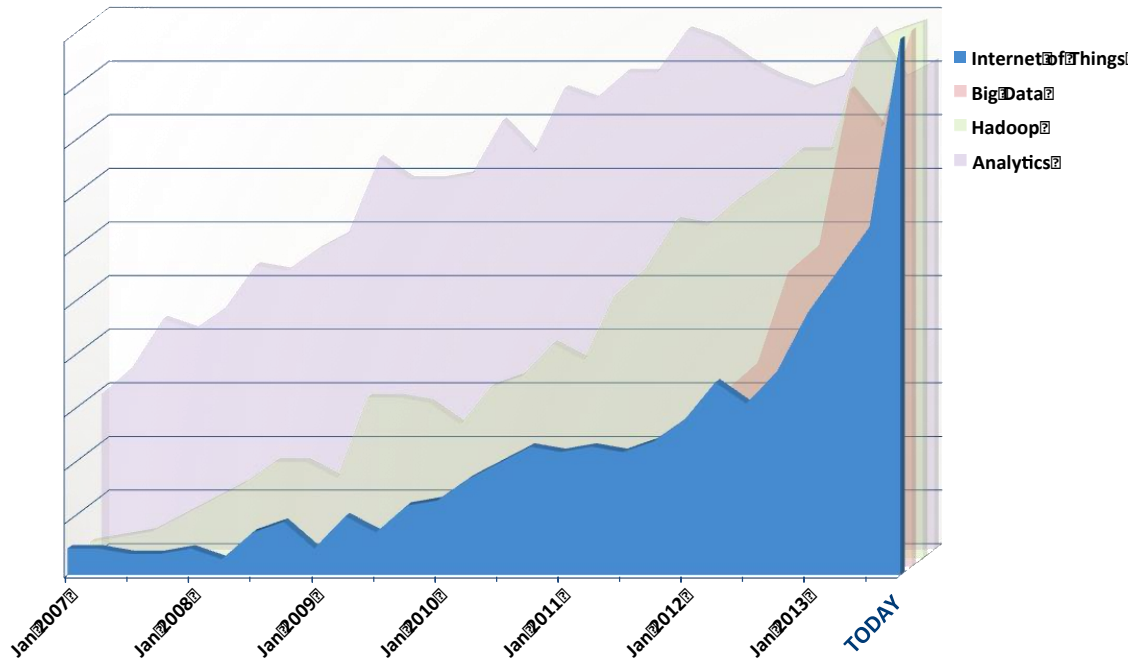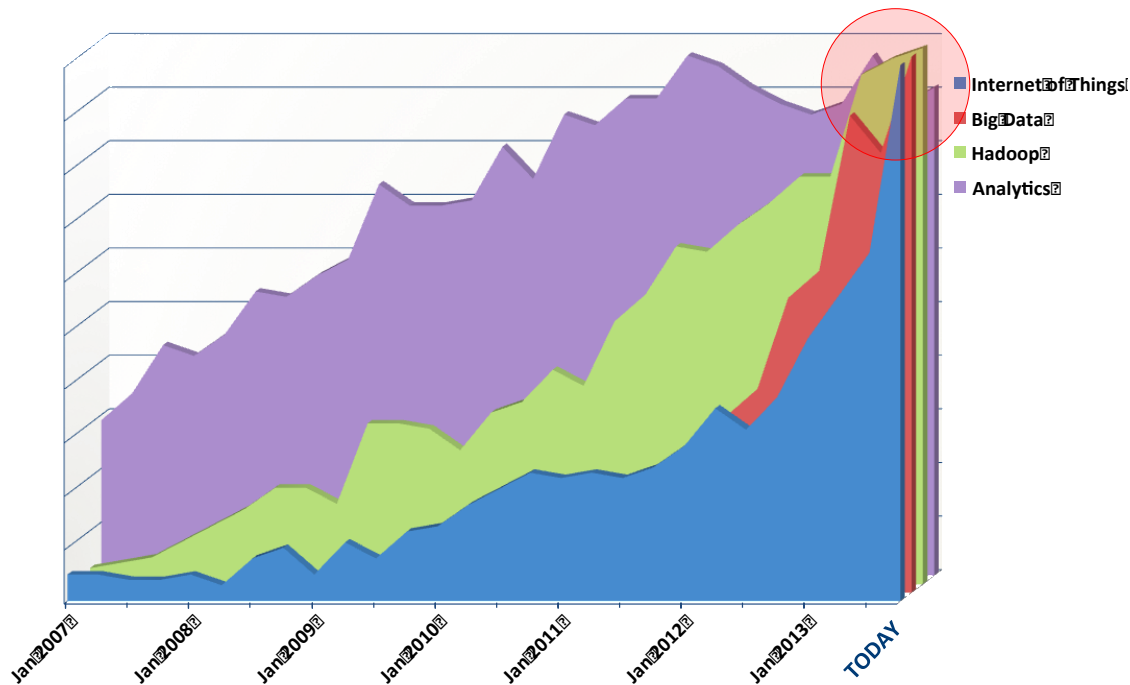
THE ERA OF ABUNDANCE ... EVOLVING MINDSETS

SCARCITY MINDSET

- Technology constrained
- Process-centric
- Focus on cost control
- **Everything is forbidden unless it is permitted**

ABUNDANT MINDSET

- Technology empowered
- Discovery-centric
- Focus on value
- **Everything is permitted unless it is forbidden**

§sas | THE POWER TO KNOW.

1. Never forgets.

2. Is a Social Animal (teamwork)

Doug Cutting, Created Hadoop
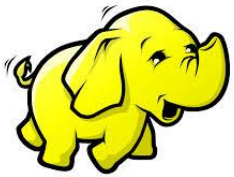
HADOOP FUNDAMENTALS
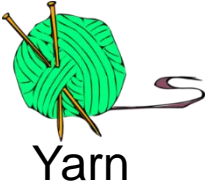
Distributed Computation.

Distributed Storage.

MapReduce
Distributed Data Processing

HDFS
Distributed, Redundant, Reliable Storage

Yarn

ZooKeeper

Head | Data Nodes
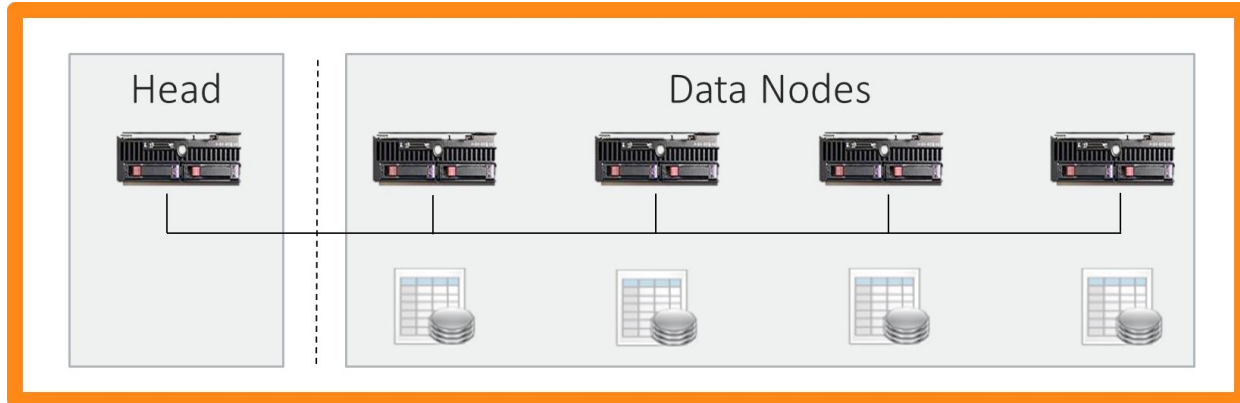
Pig

Who is the youngest person in the room?

❑ Two ways to get the answer:

1. Ask one person at a time (sequentially)
2. Work as a team and share the work (distributed)

INTERNET OF THINGS

LOYALTY CARDS …

INTERNET OF THINGS

AUTOMOBILES

- Over 700,000 accident free miles
- > 90% of all accident Human Error

§sas. | **THE POWER TO KNOW.**

The Internet of Things (IoT)

- The total economic value-add from IoT across industries will reach $1.9 trillion worldwide in 2020 (Gartner).

- Fifty billion devices will be connected to the Internet by 2020 (Cisco)

- Wide deployment of IoT technologies in the auto industry could save $100 billion annually in accident reductions

SOCIAL MEDIA IS DRIVING BIG DATA TO NEW LIMITS
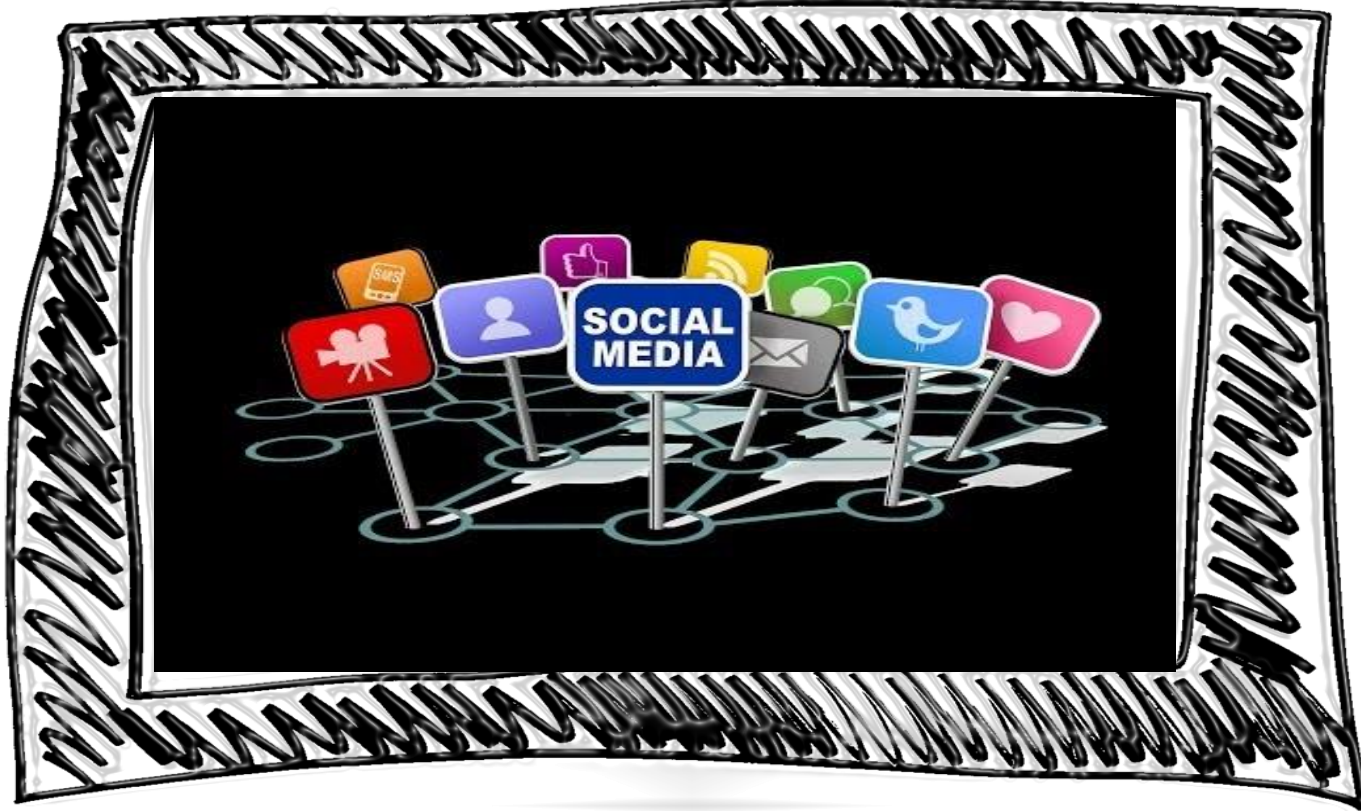
https://www.youtube.com/watch?feature=player_embedded&v=1pahLo5TTy4

SOCIAL MEDIA | INTERESTING FACTS

- Years to reach 50 Million users
    - Radio – 38 years
    - TV – 13 years
    - Internet – 4 years
    - iPod – 3 years
    - Facebook – 9 months to reach 100M

- Facebook would be the 4th largest country

- 55 – 65 years old fastest growing segment on Facebook

- YouTube 2nd largest search engine

- Social Media is not a fad … fundamental shift in how we communicate

HOW DO WE MANAGE DATA IN THE PHYSICAL WORLD?

1. Receive

2. Analyze

3. Prioritize, Act and Store
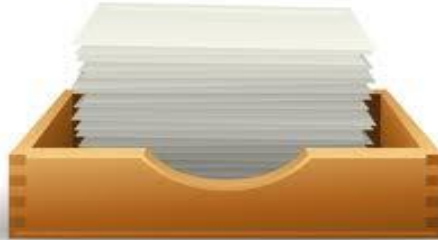
Trash

Cache

Storage

§sas | THE POWER TO KNOW.

HOW TO MANAGE DATA IN THE DIGITAL WORLD?

1. Stream It "Data Management"
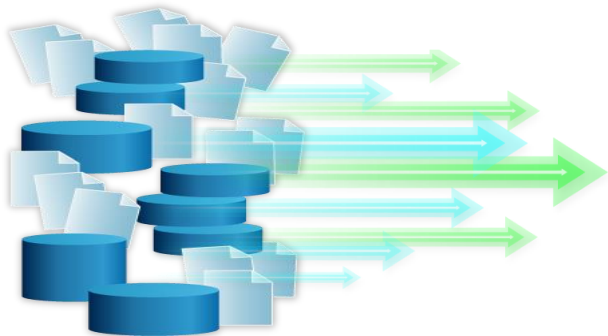
2. Score It "Analytics"

ANALYTICS & MACHINE LEARNING

3. Prioritize, Act and Store

Trash

IN-MEMORY

Cache

Storage

§sas | THE POWER TO KNOW.

https://www.youtube.com/watch?feature=player_embedded&v=WNICBy07z08

ANALYTICS | INTERESTING FACTS

- A new way of thinking

- Look beyond traditional metrics

- Out of the box thinking, be creative

- Look for new knowledge, uncover inefficiencies

- Find Value & Innovate out of necessity

§sas | THE POWER TO KNOW.

THE SCARCE APPROACH

HEALTHCARE LENS

Episodes of Care

Outcomes Analysis

Payment Models

Eligibility

Food Stamps

Healthcare Analytics

Mental Health

Children's Services

§sas | THE POWER TO KNOW.

THE SCARCE APPROACH

EDUCATION LENS

Value-added Measures

Student Growth

Student Projections

College Readiness

Earning Potential

School Performance

§sas | THE POWER TO KNOW.

THE SCARCE APPROACH

PUBLIC SAFETY LENS

Intelligent Policing

Risk Assessment

Community Corrections

Recidivism

Prison Population

Criminal History

THE ABUNDANT APPROACH

CITIZEN LENS

Know Your Citizens!!

Episodes of Care
College Readiness
Intelligent Policing
Payment Models
Risk Assessment
Earning Potential
Community Corrections
Recidivism
Healthcare Analytics
Children's Services

Value-added Measures
Student Growth
Outcomes Analysis
Student Projections
Eligibility
Prison Population
Criminal History
School Performance
Food Stamps
Mental Health

§sas | THE POWER TO KNOW.

WHERE

LOS ANGELES COUNTY DSS

BUSINESS ISSUES
- Needed to uncover systemic fraud in delivery of social services to economically disadvantaged citizens.

EXPECTED RESULTS
- Investigators can spot more cases of fraud sooner, resulting in fewer losses, lower investigative costs and greater public confidence.
- Total cost avoidance projected from the pilot results is expected to exceed $6.8 million.

US STATE AND LOCAL GOVERNMENT

"The system analyzes social networks to determine if individuals are likely to commit fraud. Social network analysis also helps identify collusive fraud rings companion cases.

The bottom line: more than $6.8 million in savings."

§sas | THE POWER TO KNOW.

NEW YORK STATE DEPARTMENT OF HEALTH BUREAU OF NARCOTIC ENFORCEMENT (BNE)

BUSINESS ISSUES

- More effectively detect and investigate suspicious activities around narcotics prescriptions in the state.
- Lacked data driven decisioning for effective prescription drug monitoring, including:
  - **Prescribers.** Which prescribers have suspicious or unusual behavior patterns that warrant further investigation? What are common characteristics of prescribers who engage in inappropriate and/or unlawful prescribing patterns?
  - **Patients.** Which patients have suspicious or unusual prescription patterns that warrant further investigation or intervention?
  - **Dispensers.** Which dispensers (e.g., pharmacists) are not reporting filled prescriptions accurately and/or timely?

EXPECTED RESULTS

- Increase the value of its data by using it to reduce illegal use and abuse of prescription narcotics.
- Will be able to:
  - Use its limited resources more effectively and better focus its investigation efforts.
  - Empower non-analysts to benefit from the data that can aid in investigations.
  - Identify network individuals who are conducting illegal activity.

US STATE AND LOCAL GOVERNMENT

"The abuse of prescription medications has reached epidemic proportions and the costs to society are enormous."

Thomas DiNapoli
New York State Comptroller

§sas | THE POWER TO KNOW.

COMMONWEALTH OF KENTUCKY

BUSINESS ISSUES

- Leveraging federal funds with ever-tightening federal and state budgets.
- Kentucky sought to save taxpayer dollars by tackling fraud within its government programs.
- Take a more holistic approach to confronting fraud, thereby reducing improper payments and lessening the impact of fraud, waste and abuse.

EXPECTED RESULTS

- By approaching fraud, waste and abuse at an enterprise level, Kentucky is able to identify misspent funds that can be put back into state coffers.
- Key sponsors within Kentucky state government see this as an opportunity to use analytics more widely across all state government agencies

UNITED STATES STATE & LOCAL GOVERNMENT

"SAS Fraud Framework became a part of Kentucky's Health Benefits Exchange in late 2012. This success led to an enterprise license of the SAS Fraud Framework to the Commonwealth of Kentucky for use beyond the Cabinet for Health and Family Services. This was all made possible by Kentucky's early recognition of the value of addressing fraud, waste and abuse on an enterprise level rather than the typical, siloed approach."

Commonwealth of Kentucky

§sas | THE POWER TO KNOW.

CITY OF MARKHAM

BUSINESS ISSUES

- Growing need for data integration and synchronization – over 60 different sources
- Desire for a consistent view of data across major business applications

RESULTS

- Standard platform that spans multiple data management areas
- Production time reduced by up to 40% for applications requiring data integration
- Development of standard templates and enforced data standards
- Ability to have a unified and consistent view of addresses across major applications

GOVERNMENT

"Having an accurate view of our organization, our customers and our business environment allows us to optimize profitability, mitigate risks and reduce costly operational inefficiencies."

Teodor Tecsa
Manager of Applications and GIS

§sas | THE POWER TO KNOW.

"SAS is committed to ensuring that we continue to have a leading-edge anti-fraud solution. We are very pleased with the results."

Head of Group Fraud Risk, Global Security and Fraud Risk

CREDIT CARD COMPANIES

Business Issue

- Reduce incidence of fraud.
- Lower costs.
- Increase customer confidence.

Results/Benefits

- Achieved significantly lower incidence of fraud across tens of millions of debit and credit card accounts.

GOVERNMENT DATA ANALYTICS CENTER (GDAC)

Government Data Analytics Center (GDAC)

The GDAC is an enterprise program to promote the use of North Carolina's **data as an asset** to support business decisions. The GDAC fosters interagency **collaboration** among and between the branches of governments and their sub-units to establish statewide business intelligence (BI) **standards**, to **improve data sharing, quality and consistency** and facilitate the identification, development and support of BI solutions for the State.

Government Data Analytics Center (GDAC)

- Compliance, Waste, and Abuse
- Performance and Outcome Monitoring
- Improved Citizen Experience & Relationship
- Manage Once, Use Many!!!
    - Governance
    - Quality
    - Share & Collaborate

§sas. | THE POWER TO KNOW.
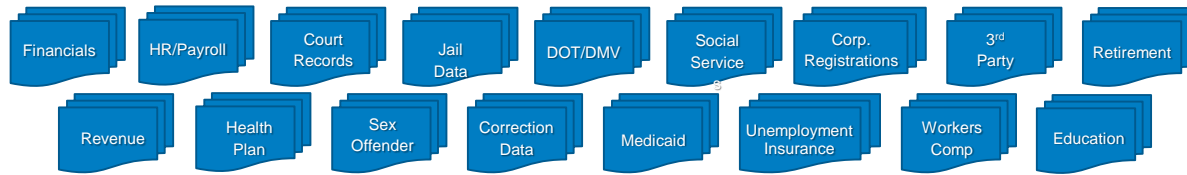
Government Analytics Center of Excellence

Data Sources (top):
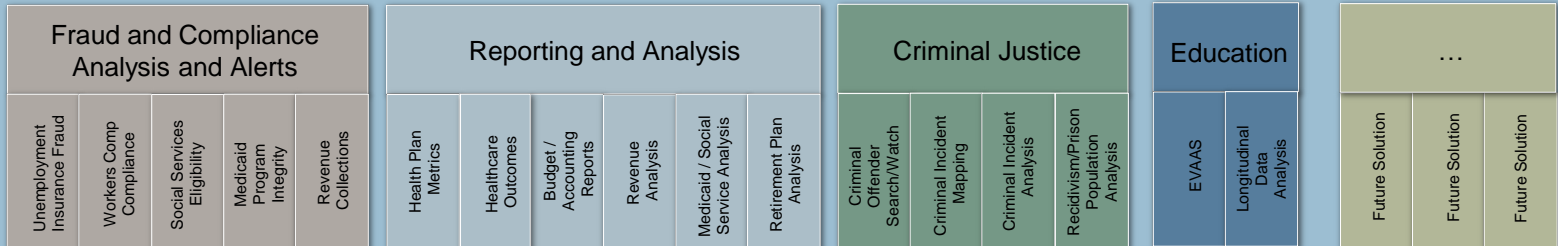Financials, HR/Payroll, Court Records, Jail Data, DOT/DMV, Social Service, Corp. Registrations, 3rd Party, Retirement
Revenue, Health Plan, Sex Offender, Correction Data, Medicaid, Unemployment Insurance, Workers Comp, Education

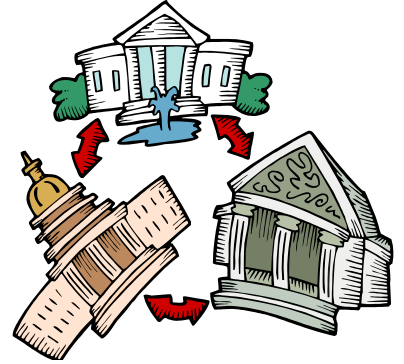**Extract, Transform and Load**

Data Governance (left column):
- Identity Mgt. and Authorization
- Data/System Monitoring
- Business Intelligence
- Policy and Procedure
- Data Standards
- Data Mgmt. and Admin
- Security

Security and Auditing

Governance

Data stores: 3rd Party Data, DOT/DMV, HR/Payroll, Education, Healthcare, Unemployment Insurance, Medicaid, Revenue, Death Master File, Criminal Justice, Financials, Wage and Labor, Social Services

Fraud and Compliance Analysis and Alerts
- Unemployment Insurance Fraud
- Workers Comp Compliance
- Social Services Eligibility
- Medicaid Program Integrity
- Revenue Collections

Reporting and Analysis
- Health Plan Metrics
- Healthcare Outcomes
- Budget / Accounting Reports
- Revenue Analysis
- Medicaid / Social Service Analysis
- Retirement Plan Analysis

Criminal Justice
- Criminal Offender Search/Watch
- Criminal Incident Mapping
- Criminal Incident Analysis
- Recidivism/Prison Population Analysis

Education
- EVAAS
- Longitudinal Data Analysis

…
- Future Solution
- Future Solution
- Future Solution

INFORMATION DELIVERY PORTAL

§sas | THE POWER TO KNOW.

CJLEADS CRIMINAL JUSTICE & PUBLIC SAFETY

- Over **50M records** and 14M unique offenders
- **25+ disparate data sources**
- **Over 548 Local, State & Federal Agencies**
  - Department of Homeland Security – ICE
  - US Postal Service – Inspector General
  - US Forest Service
  - US Department of Agriculture – Office of the Inspector General
  - USDOJ
- **Support for ~30,000** court, correction and law enforcement professionals
- **Over 11,300 users** access CJLEADS weekly
- In the last 12 months, criminal justice professionals have **run over 18.2 million searches** and accessed 15.2 million offender and DMV records

§sas | THE POWER TO KNOW.

Offender Summary Screen

Offender info for: JOHN JULIAN PORTER ✕

JOHN JULIAN PORTER 📋 *AWC* OFA SOR GNG CHP 👓   Lookup | Report | Watch | Merge

SSN: XXX-XX-1111 📋   FBI ID: 122233JB1 📋   SID: NC9988776A 📋   Age: 34   DOC ID: 0327401   DL: NC-16400000 📋

| Summary | Images | Timeline | Court Records | Outstanding | Jail | Prison | DV | ComCorr | SOR | Gangs | Marks | Watchers |

JOHN JULIAN PORTER ...   *AWC* OFA SOR GNG CHP 👓

Most Recent Update: 03/01/2011

Address (03/01/2011): 801 N OAK STREET, DURHAM, NC 27707 📄

Birth Date:	01/01/1976 📋	FBI ID:	122233JB1
Age:	34	SID:	NC9988776A
Race:	WHITE	SSN:	XXX-XX-1111
Gender:	MALE	DL:	NC-16400000
Height:	5'7"	DOC ID:	0327401
Weight:	219		
Eyes:	BROWN	SRN ID:	34567S1
Hair:	BLACK	Occupation:	

More...

DOC 03/30/2009 ➡

Outstanding Processes	NC Court Records	Jail
Unserved Processes - 0	Pending - 2	Current - No
Felony - 0	Disposed - 64	Previous - Yes
Misdemeanor - 0		Escaped - No
Infraction - 0		
Traffic - 2		

Prison	Community Corrections	Sex Offender Registry
Current - Yes	Current - No	Pending Registration - No
Previous - Yes	Previous - Yes	Active Registration - Yes
Escaped - No	Absconded - No	Prior Registration - Yes
		Registration Type - AGGRAVATED OFFENDER
		Registration Contact Agency - Wake

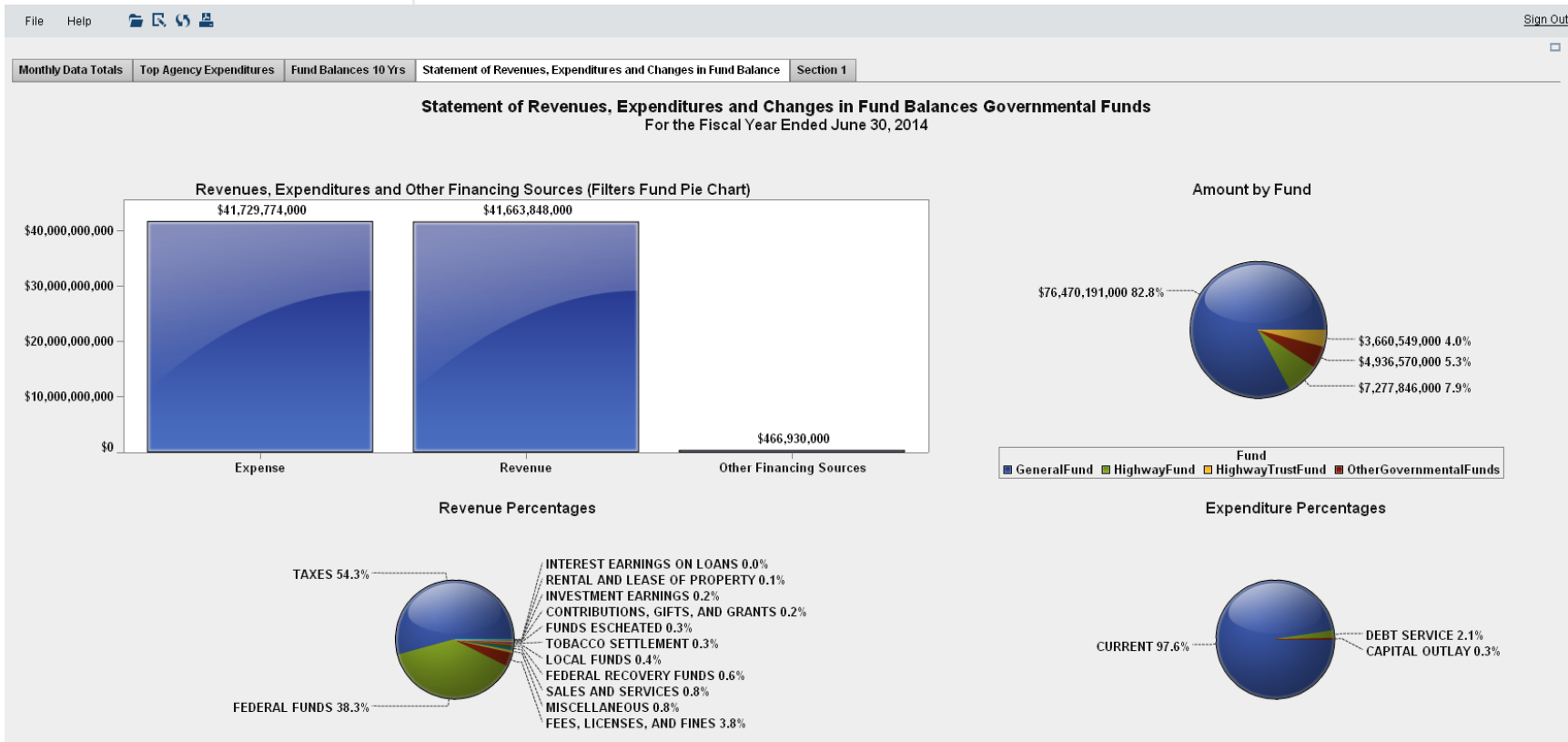| Marks / Scars	Gang Affiliations	Domestic Violence
Yes	Yes	Case(s) - 3

CONCLUSION | **Food for Thought**
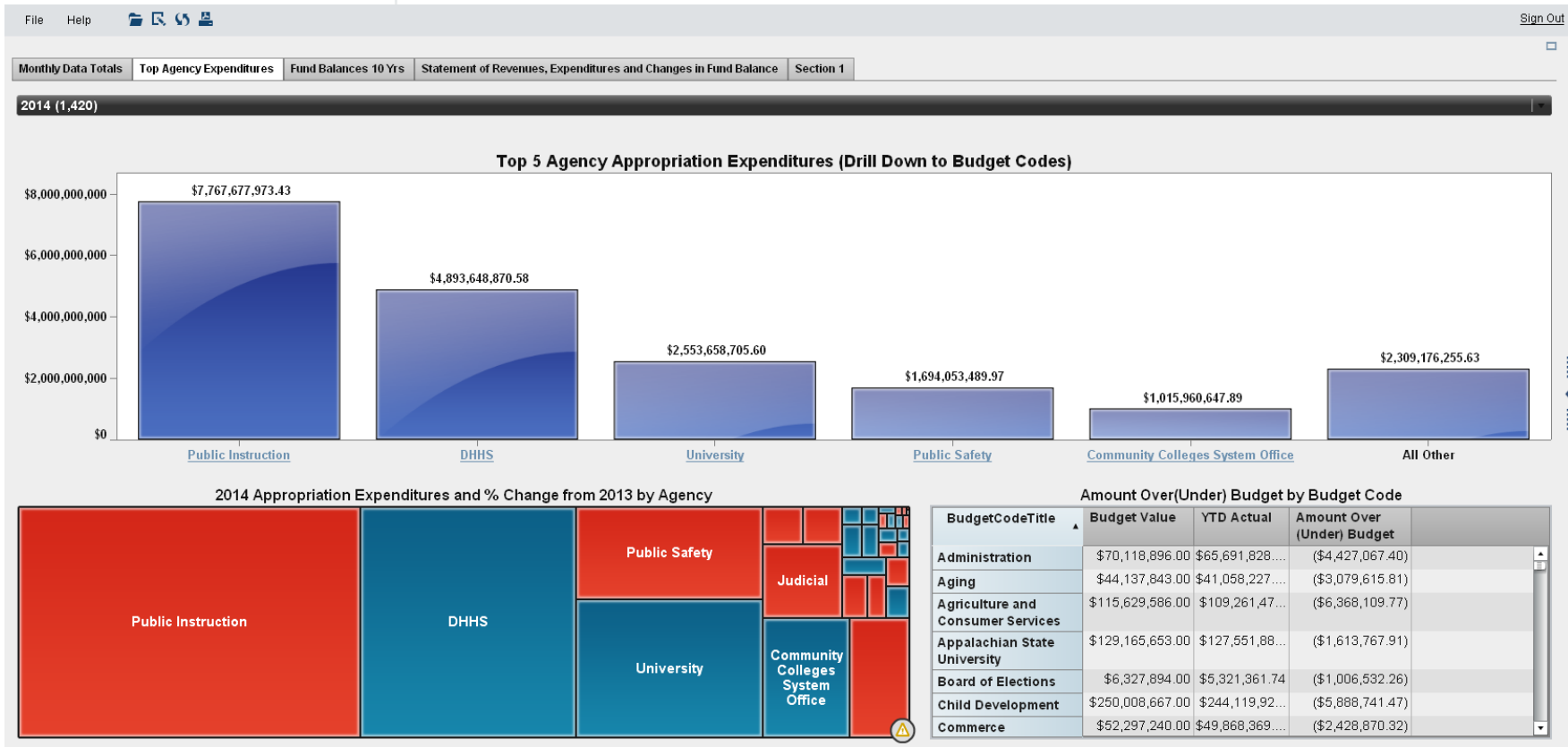
- Are you operating with a mindset of scarcity?
  - Do you have **better, faster, and more creative** ways to use your existing data?
  - Can you capture new data and **find value** from them?
- Modernize your analytics infrastructure to participate in the Era of Abundance
  - Look for existing problems to tackle – but with more data
  - Establish an Analytics Center of Excellence
- Business and IT **partnership** is critical to success