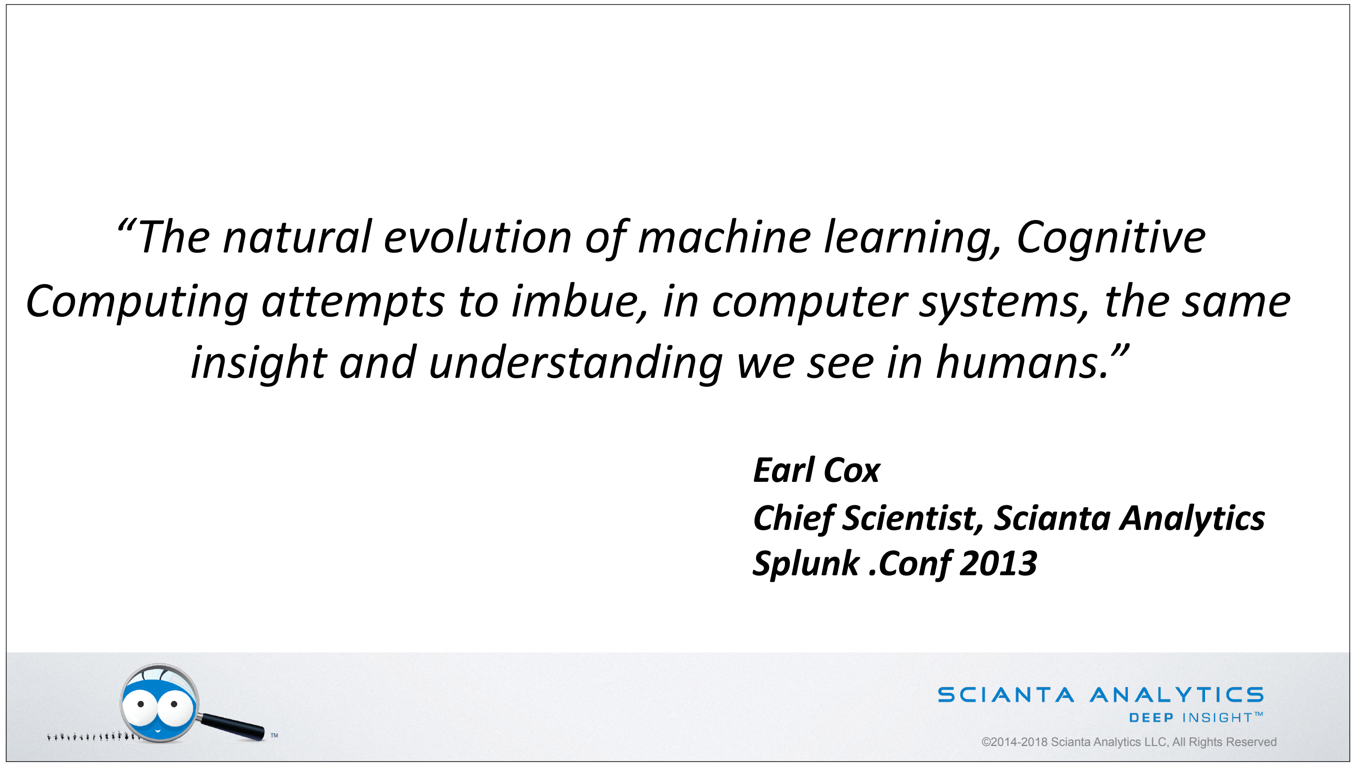
INTRODUCTION TO MACHINE INTELLIGENCE

™

SCIANTA ANALYTICS

DEEP INSIGHT™



“The natural evolution of machine learning, Cognitive Computing attempts to imbue, in computer systems, the same insight and understanding we see in humans.”

Earl Cox
Chief Scientist, Scianta Analytics
Splunk .Conf 2013

SCIANTA ANALYTICS
DEEP INSIGHT™

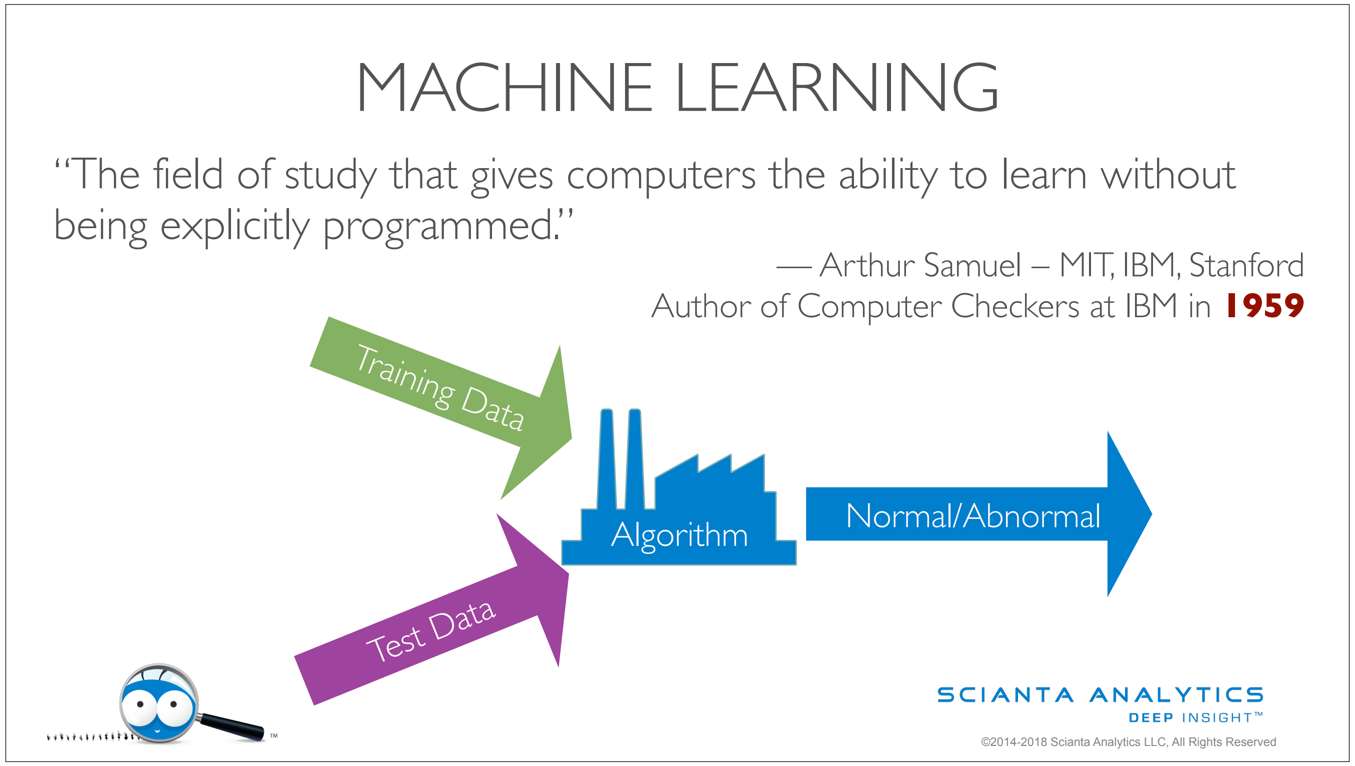
©2014-2018 Scianta Analytics LLC, All Rights Reserved

AGENDA

Introduction to Machine Intelligence	Data Handling 1	Data Handling 2	Anomaly Detection	Transactional Behavior	Impact Analysis
Academic Concepts	Collection	Retention	Anomaly Definition	Defining Transactions	Organizational Visibility
Data Systems	Storage	Format	Measuring Normality	Transaction Relationships	Types of Impact
Maturity Curve	Security	Labeling		Probability Measurement	Responsiveness

SCIANTA ANALYTICS
DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved



MACHINE LEARNING

"The field of study that gives computers the ability to learn without being explicitly programmed."

— Arthur Samuel – MIT, IBM, Stanford
Author of Computer Checkers at IBM in **1959**

Training Data

Test Data

Algorithm

Normal/Abnormal

SCIANTA ANALYTICS
DEEP INSIGHT™

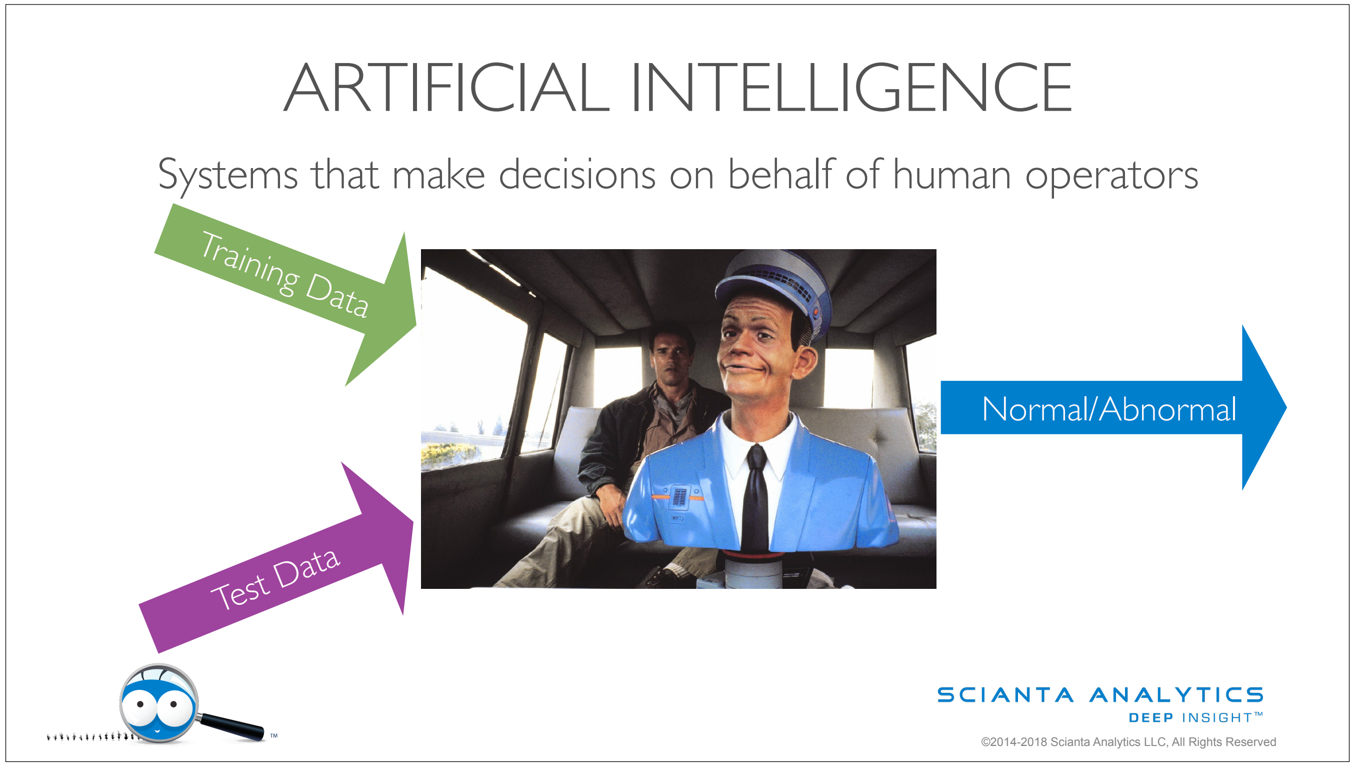
©2014-2018 Scianta Analytics LLC, All Rights Reserved

What is Machine Learning?

Train a system with a bunch of high quality labeled data, it finds the statistical outliers.

This assumes that outliers are bad, and you know what they say about assumptions.

Still, Machine Learning techniques are widely available now, and they're pretty useful.



ARTIFICIAL INTELLIGENCE

Systems that make decisions on behalf of human operators

Training Data

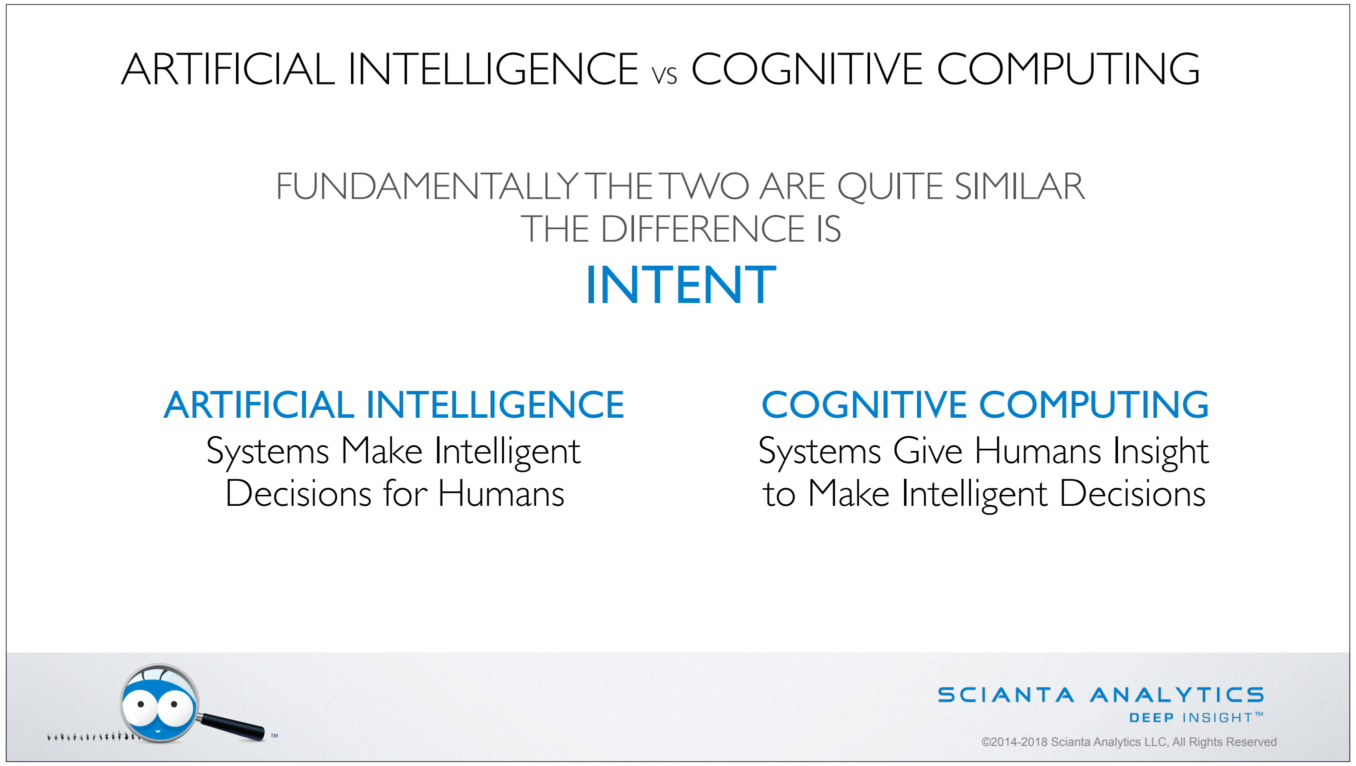
Test Data

Normal/Abnormal

SCIANTA ANALYTICS
DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved

So take a bunch of different machine learning algorithms, stack them into a self-training genetic model, recurse the data streams, think of every possible scenario, and prevent outside context problems… you've got an AI. There's some success stories in narrowly targeted use cases here, but general AI is still a long way away.



ARTIFICIAL INTELLIGENCE vs COGNITIVE COMPUTING

FUNDAMENTALLY THE TWO ARE QUITE SIMILAR
THE DIFFERENCE IS

INTENT

ARTIFICIAL INTELLIGENCE
Systems Make Intelligent
Decisions for Humans

COGNITIVE COMPUTING
Systems Give Humans Insight
to Make Intelligent Decisions

SCIANTA ANALYTICS
DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved

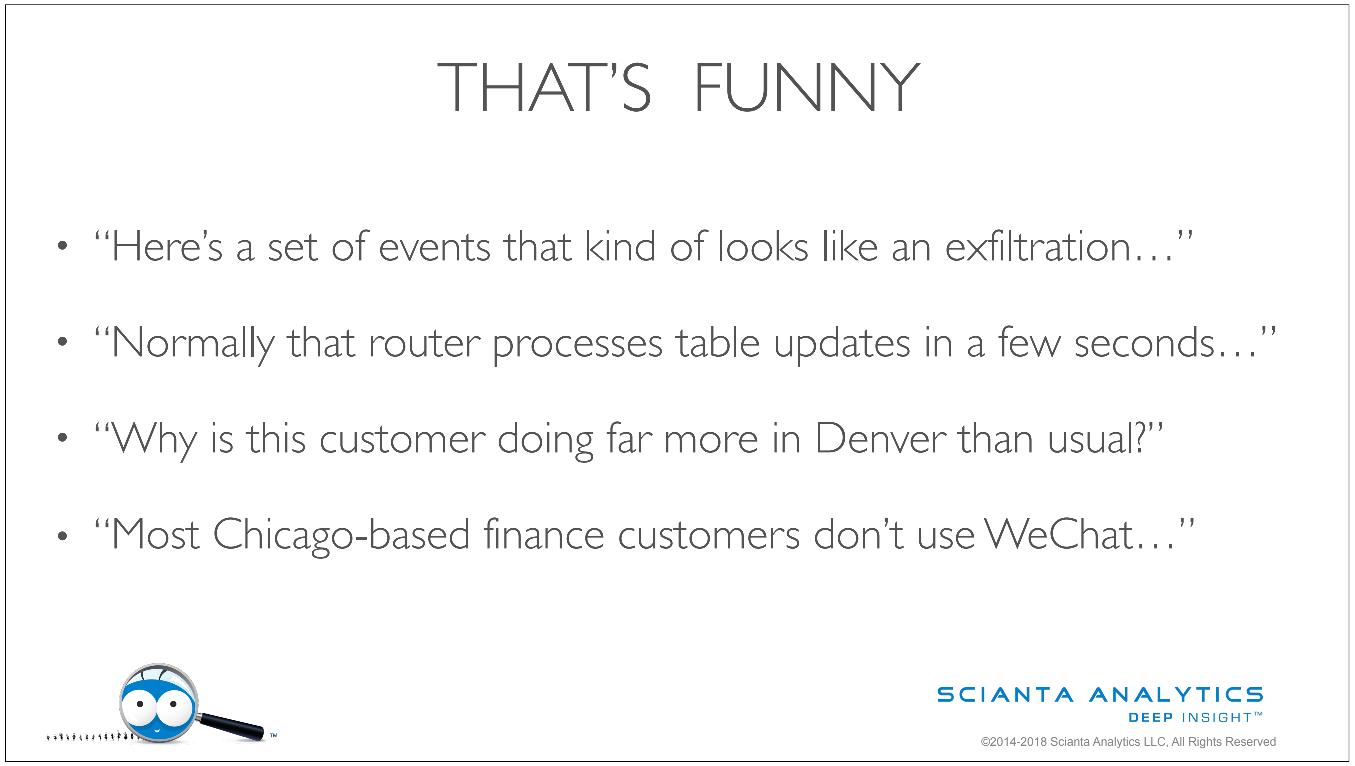
Cognitive Computing is more focused than AI and more complete than ML; it’s supporting tools for already intelligent Data Scientists. There’s a lot of brain power on this planet, and Cognitive Computing’s goal is to help focus that.

Rob High, CTO of IBM’s Watson team, says it well: “What it’s really about is involvement of a human in the loop… Cognitive Computing is ‘augmented intelligence’ rather than ‘artificial intelligence.’”

AGENDA

Introduction to Machine Intelligence	***Data Handling 1***	***Data Handling 2***	***Anomaly Detection***	***Transactional Behavior***	***Impact Analysis***
Academic Concepts	***Collection***	***Retention***	***Anomaly Definition***	***Defining Transactions***	***Organizational Visibility***
Data Systems	***Storage***	***Format***	***Measuring Normality***	***Transaction Relationships***	***Types of Impact***
Maturity Curve	***Security***	***Labeling***		***Probability Measurement***	***Responsiveness***

©2014-2018 Scianta Analytics LLC, All Rights Reserved



THAT’S FUNNY

- “Here’s a set of events that kind of looks like an exfiltration…”
- “Normally that router processes table updates in a few seconds…”
- “Why is this customer doing far more in Denver than usual?”
- “Most Chicago-based finance customers don’t use WeChat…”

SCIANTA ANALYTICS
DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved

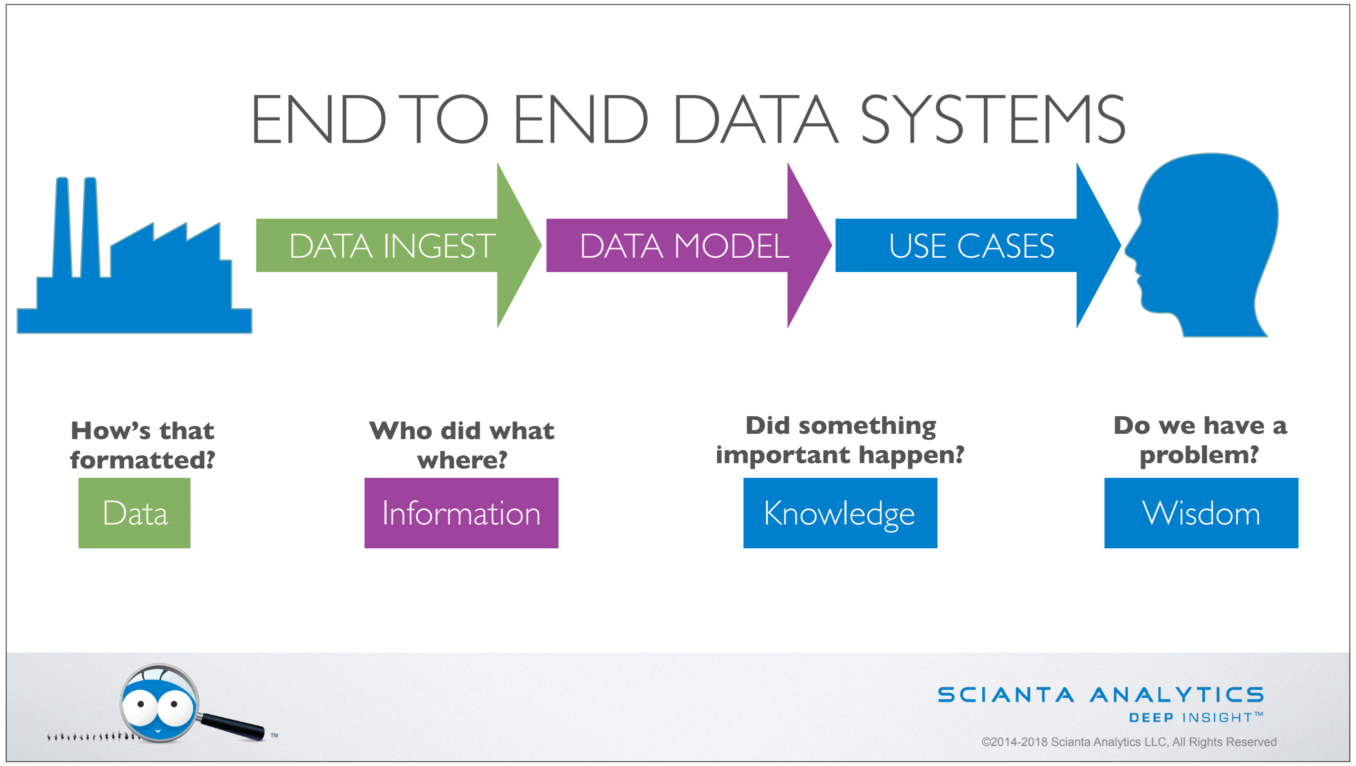
So, what kind of use cases can we solve with Cognitive Computing? The sky is the limit! In fact, we can do anything with sufficient time and money. More seriously, Cognitive Computing techniques are useful when you want to understand transactional behaviors, find anomalies in complex data systems, and help your analysts find unknown problems.

Is this set of actions similar to a known transactional pattern?

Is this transaction different from other transactions?

Does this actor normally act like they are acting now?

Are this actor’s actions strange compared with their peer group?



END TO END DATA SYSTEMS

DATA INGEST

DATA MODEL

USE CASES

How's that formatted?

Data

Who did what where?

Information

Did something important happen?

Knowledge

Do we have a problem?

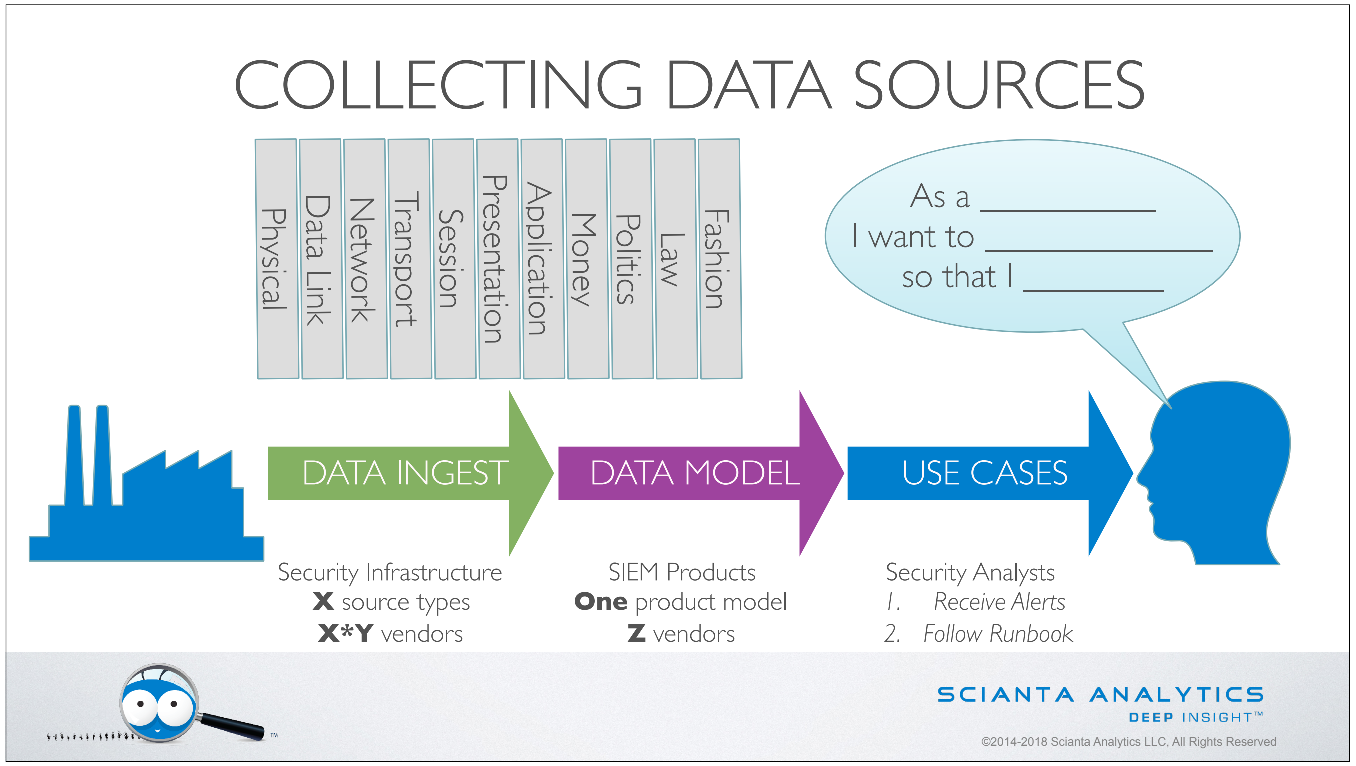
Wisdom

SCIANTA ANALYTICS

DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved

To do those things, a Cognitive Computing approach needs a data system, which is a processing path from raw data to something a human can use. Data has to be collected, modeled, and used before it can turn into information, knowledge, or wisdom. Maybe you can do it with a notebook and a pen, or maybe you need to process data at Internet scale…. But someone needs to do the work to enable another person to get value from the data.



COLLECTING DATA SOURCES

Physical
Data Link
Network
Transport
Session
Presentation
Application
Money
Politics
Law
Fashion

As a ___________
I want to ______________
so that I __________

DATA INGEST

DATA MODEL

USE CASES

Security Infrastructure
X source types
X*Y vendors

SIEM Products
One product model
Z vendors

Security Analysts

1. *Receive Alerts*
2. *Follow Runbook*

SCIANTA ANALYTICS
DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved

What questions do you want to answer? What data sources could answer those questions? How easy or hard will it be to get those data sources?

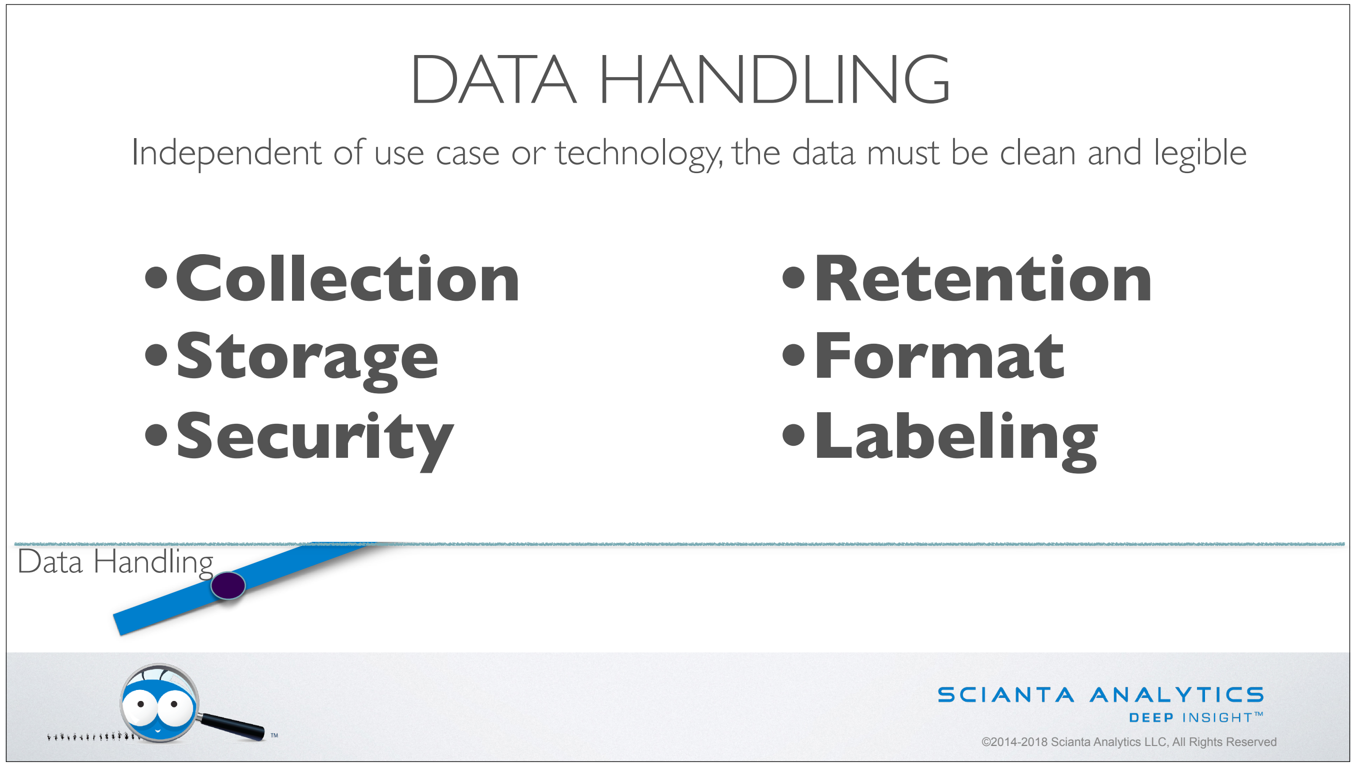
In some data systems these three questions are straightforward to answer; for instance in security, the SIEM product category has established common use cases and common data sources. In other data systems, you may have less structure to work from out of the box.

AGENDA

Introduction to Machine Intelligence	***Data Handling 1***	***Data Handling 2***	***Anomaly Detection***	***Transactional Behavior***	***Impact Analysis***
Academic Concepts	***Collection***	***Retention***	***Anomaly Definition***	***Defining Transactions***	***Organizational Visibility***
Data Systems	***Storage***	***Format***	***Measuring Normality***	***Transaction Relationships***	***Types of Impact***
Maturity Curve	***Security***	***Labeling***		***Probability Measurement***	***Responsiveness***

SCIANTA ANALYTICS
DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved



DATA HANDLING

Independent of use case or technology, the data must be clean and legible

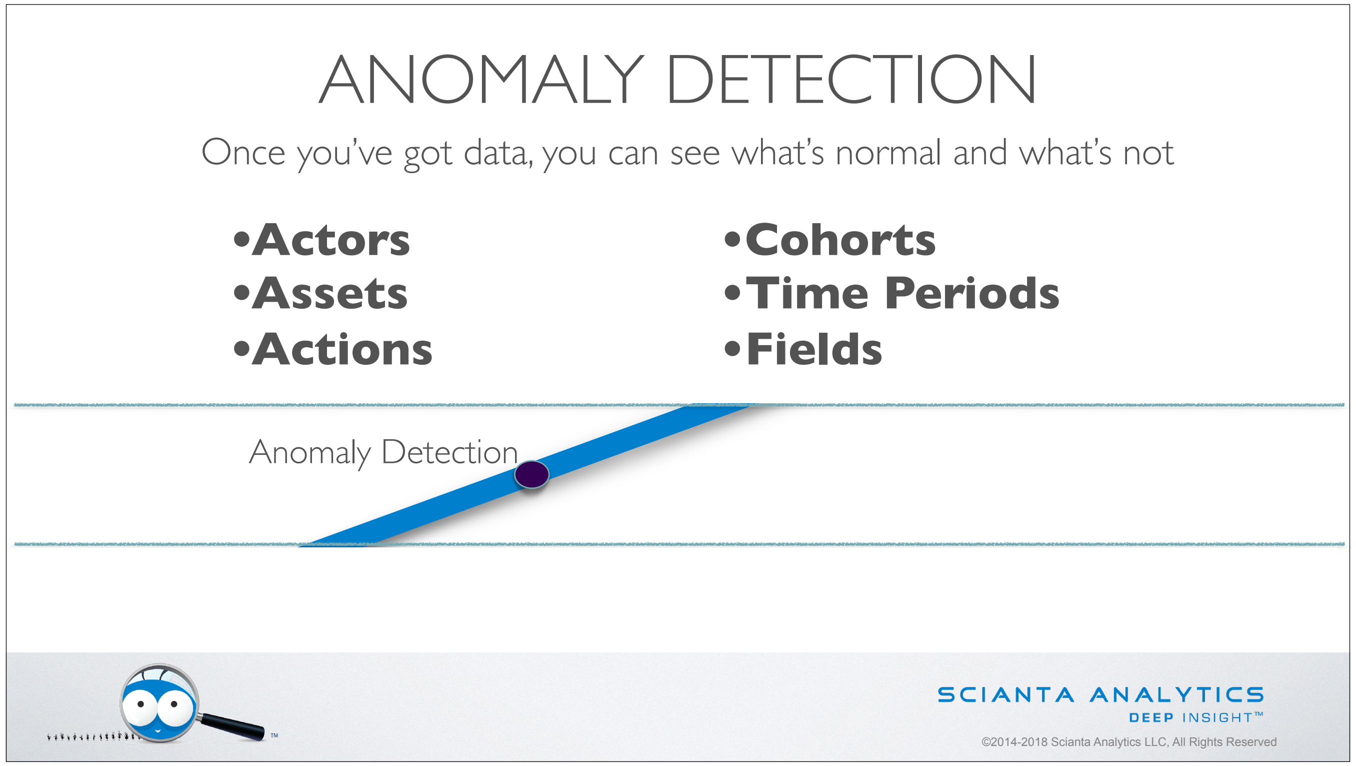
- **Collection**
- **Storage**
- **Security**
- **Retention**
- **Format**
- **Labeling**

Data Handling

SCIANTA ANALYTICS
DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved

How are you going to collect the data? Does the raw data have to be stored, or just metadata? For how long? In what format? Who gets to see it? And most importantly, how will it be labeled? Data without labels can’t be searched, analyzed, or understood. We’ll dig into all these areas in greater detail in later sessions; for now, it’s enough to start asking these questions.



ANOMALY DETECTION

Once you've got data, you can see what's normal and what's not

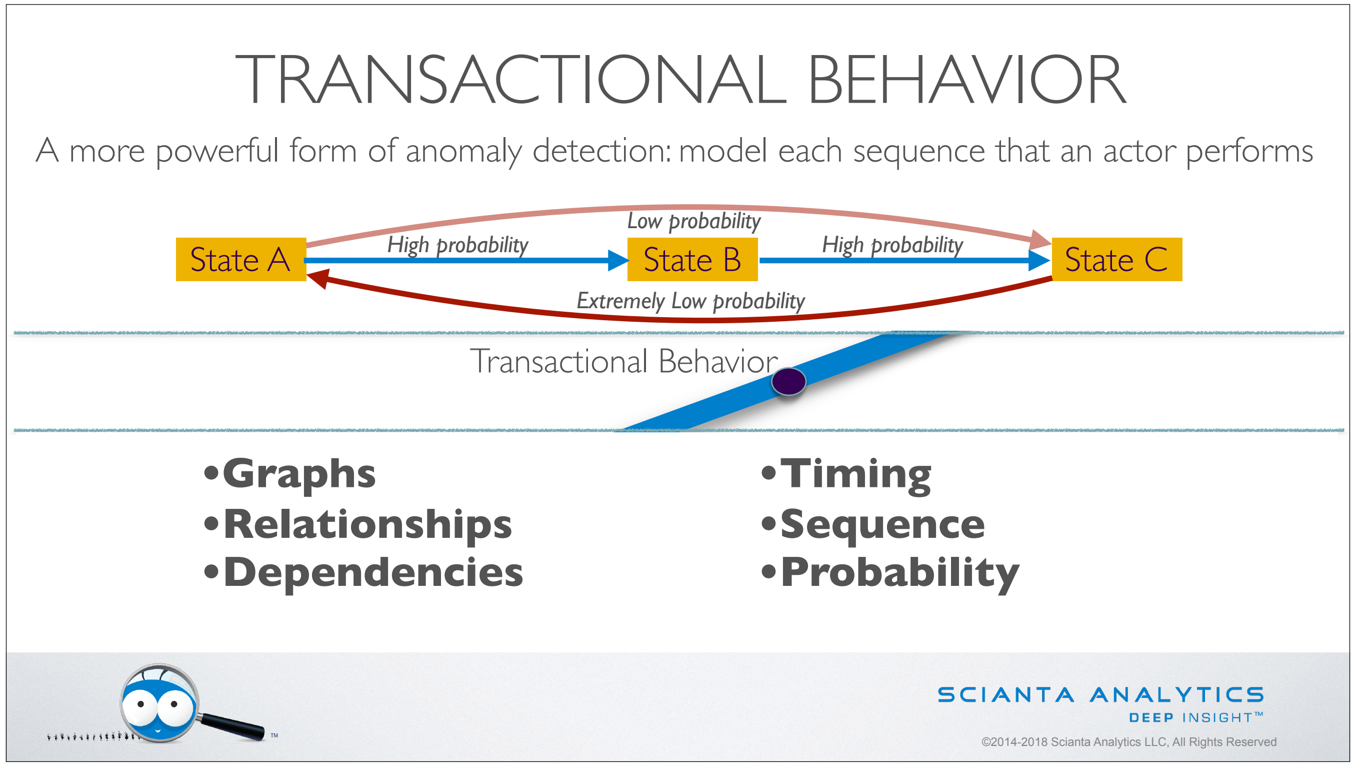
- Actors
- Assets
- Actions
- Cohorts
- Time Periods
- Fields

Anomaly Detection

SCIANTA ANALYTICS
DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved

Anomaly Detection is the 101 level of machine learning, for data systems where it works. If the past accurately predicts the future for the system you're modeling, anomaly detection is a very powerful tool. When that isn't true, anomaly detection techniques may still be useful for describing current state, but will not be able to say if that state is abnormal or not.



TRANSACTIONAL BEHAVIOR

A more powerful form of anomaly detection: model each sequence that an actor performs

State A

State B

State C

High probability

High probability

Low probability

Extremely Low probability

Transactional Behavior

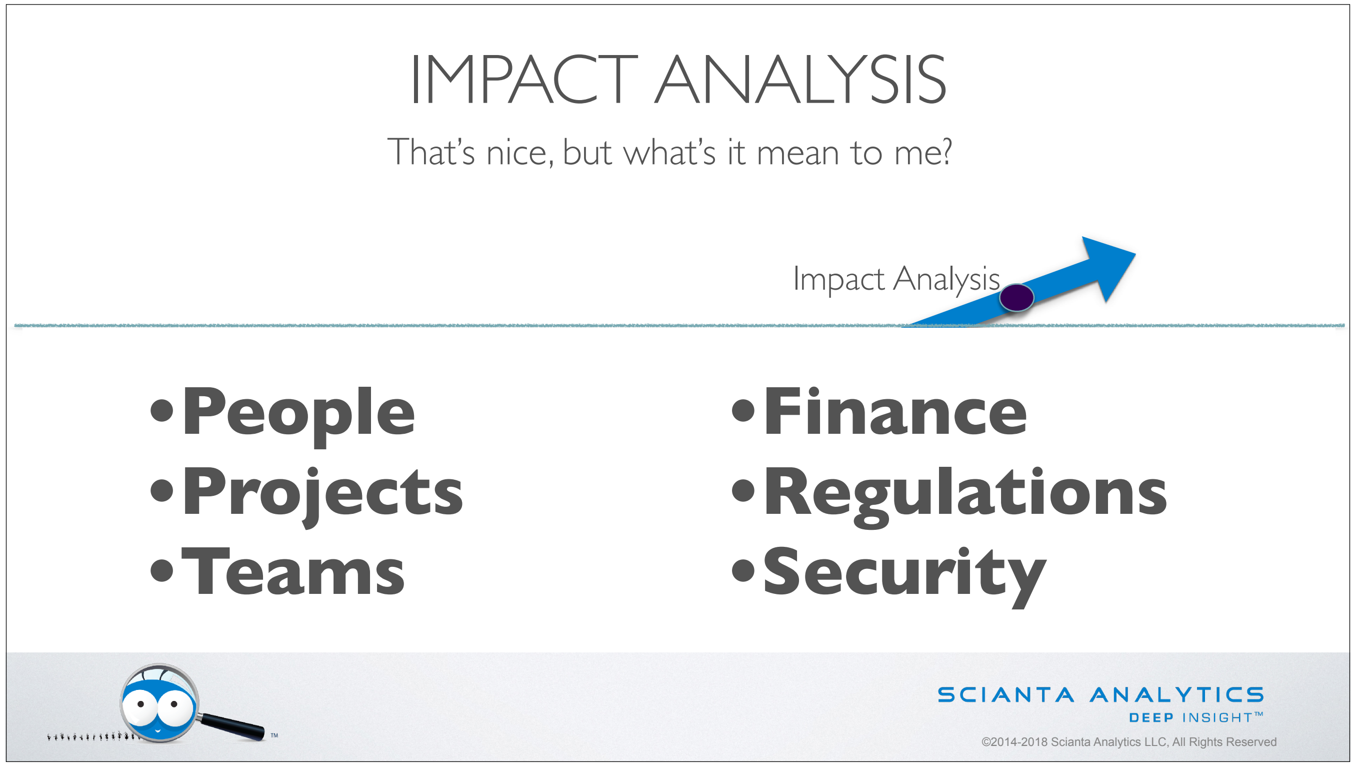
- Graphs
- Relationships
- Dependencies

- Timing
- Sequence
- Probability

SCIANTA ANALYTICS
DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved

Typically, Anomaly Detection techniques review single events and states; but what if we model the steps in a transaction, or even the entire transaction at once? By drawing a graph of each sequence committed by an actor, we can establish how likely a given sequence is, and alert on the strange ones. A slight tweak to this approach is to also evaluate the time taken per step; for instance, to distinguish human actors from scripted actors by the regularity and speed of the actions taken. Finally, each transaction can be treated as a step in a larger transactional graph, allowing dependency relationships to be understood. A good example of this is logistics; a store manager might interact with their inventory and ordering system, which is a transaction. The resulting order is processed at a distribution facility, which is a transaction. The ordered produce is then delivered to the store on a truck, which is a third transaction. All three transactions must be reviewed together to properly understand the system’s health.



IMPACT ANALYSIS

That's nice, but what's it mean to me?

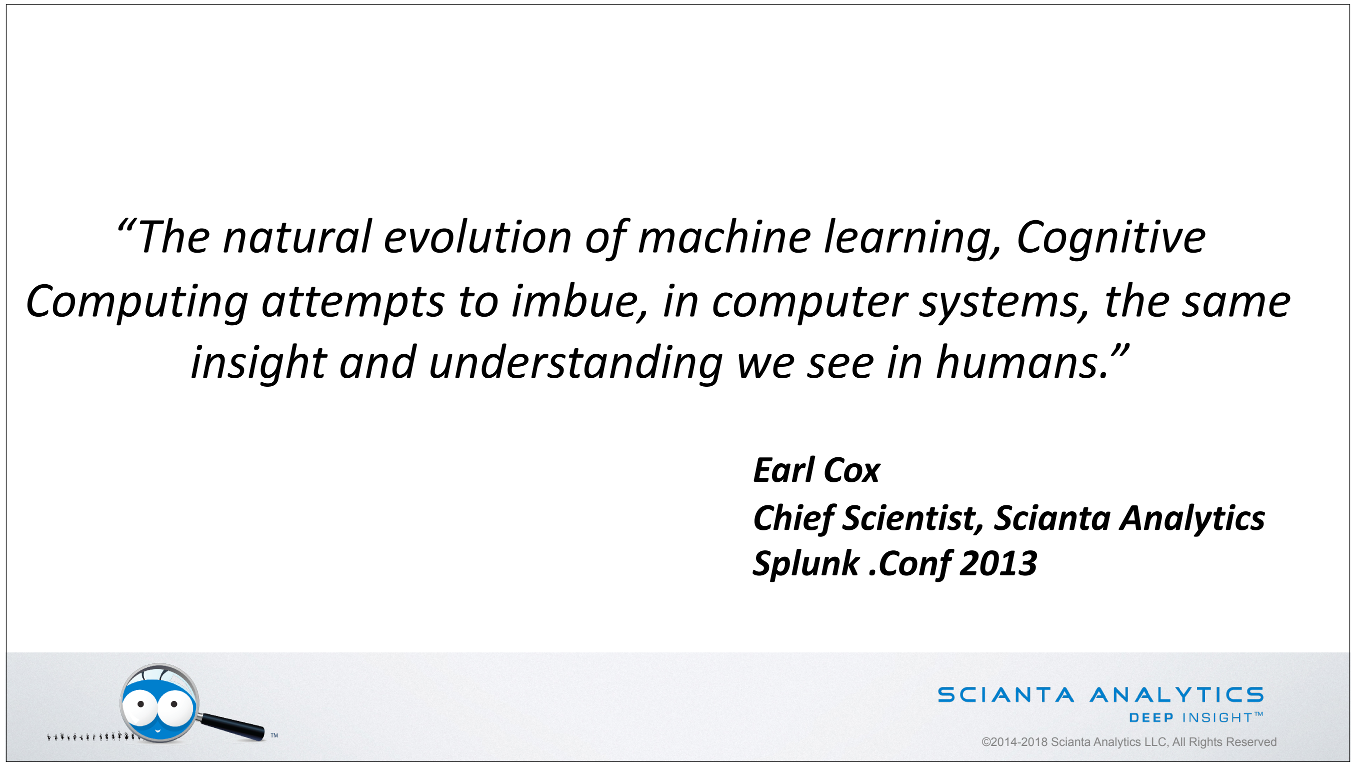
Impact Analysis

- **People**
- **Projects**
- **Teams**
- **Finance**
- **Regulations**
- **Security**

SCIANTA ANALYTICS
DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved

To continue that example, if our logistics system is failing and ordered material doesn't make it into the delivery for the store, there is a clear impact. Customers are unhappy, revenues will drop, and people will lose their jobs. If the organization's mission is not being met, impact is certain and swift. Cognitive Computing systems help to uncover systemic problems, but the analyst is still responsible for understanding and communicating the impact of discovered problems appropriately.



“The natural evolution of machine learning, Cognitive Computing attempts to imbue, in computer systems, the same insight and understanding we see in humans.”

Earl Cox
Chief Scientist, Scianta Analytics
Splunk .Conf 2013

SCIANTA ANALYTICS
DEEP INSIGHT™

©2014-2018 Scianta Analytics LLC, All Rights Reserved

Computers are force multipliers, not analysts; but they can help an analyst be more productive and successful. I hope this has been a useful introduction; next, we will go deeper into data handling techniques. Thank you!

Thank you!