



# Unlocking the Intelligence in **Big Data**

---

**Ron Kasabian**

General Manager Big Data Solutions  
Intel Corporation

# What's Driving Big Data?

Volume &  
Type of Data



**10X**

Data growth by 2016  
*90% unstructured*<sup>1</sup>

Lower Cost of  
Compute & Storage



New Investments in  
Tools & Services



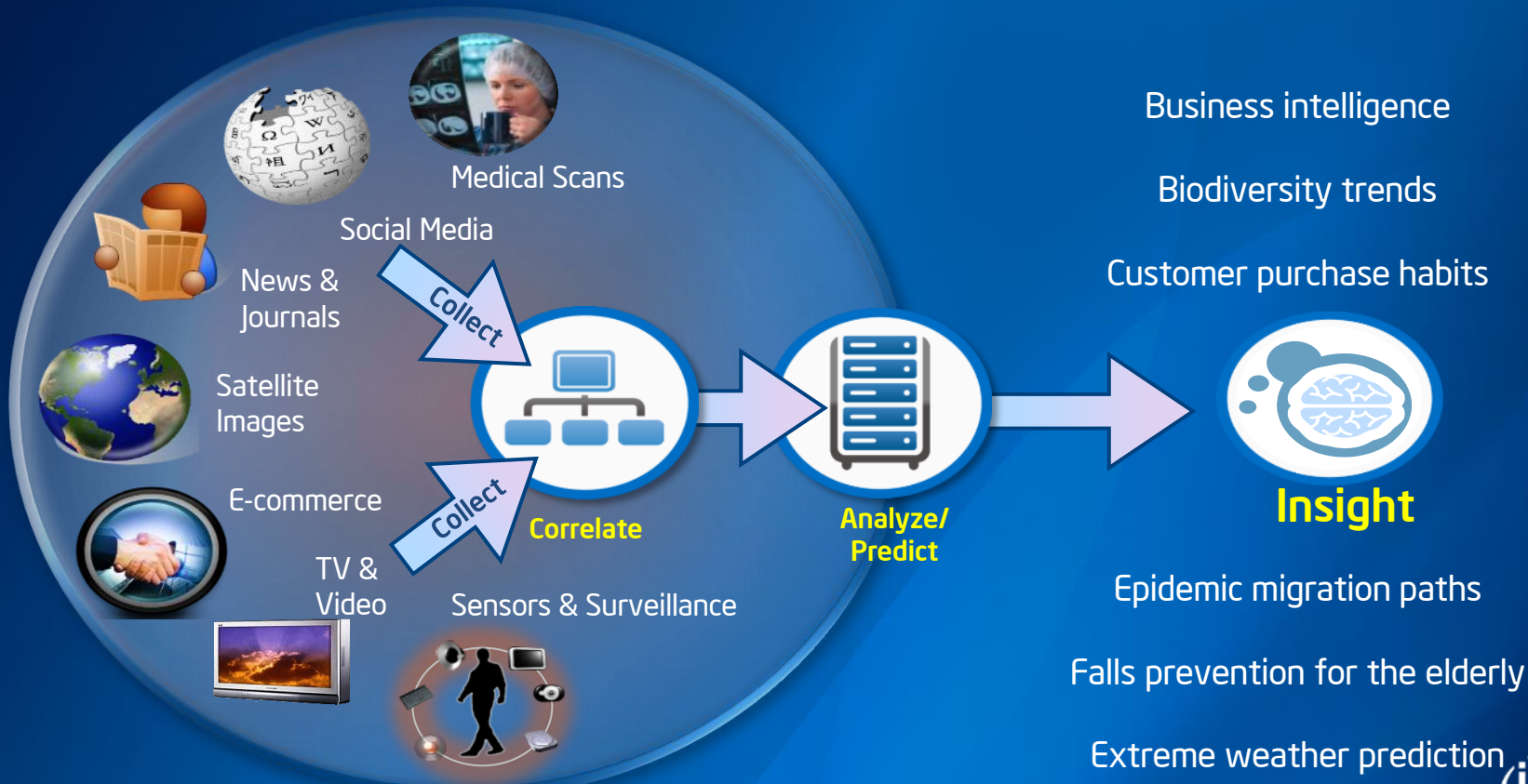
1: IDC Digital Universe Study December 2012

2: Intel Forecast

3: IDC WW Big Data Forecast



# Untapped Value of Big Data



# Today's Big Data Insights

## Operational Efficiency



Traffic Optimization



Smart Energy Grid

## Consumer Behavior



Location Aware Ad Placement



Buyer Protection Program

## Security & Risk Management



Personalized Preventive Care



Claim Fraud Reduction

# Big Data in Action at Intel



## Test Time Reduction:

Predictive analytics in manufacturing to identify failing parts quickly  
Expected to save ~\$20M in 2013



## Reseller Channel Management:

Improved reseller engagement with customer profile analytics  
Increased sales by \$5M per Qtr. while reducing costs



## Malware Detection:

Analyzing ~4B access events per day at the system, network, and application levels to discover new malware threats before they arise.



# Eric Dishman

Intel Fellow,  
General Manager, Health Strategy  
and Solutions Group  
Datacenter and Connected  
Systems Group



# End To End Big Data Vision



E2E Analytics



Device Gateways



Sensors and Devices  
Generating Data



# Internet of Things Analytics Evolution

*From Devices to Systems, Monitoring to Automation*

## Analytics Today



### Single Machine or Device Focus

- Monitor a single device
- Requires Human Intervention
- Preventative Maintenance and Monitoring

## Analytics Future



### Aggregated Systems Focus

- Many devices and systems
- Autonomous decisions & real-time optimization
- Transformative business opportunities

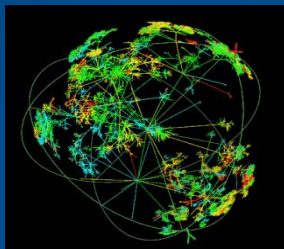


# Barriers impeding Big Data Adoption

Cost



Complexity



Confidentiality



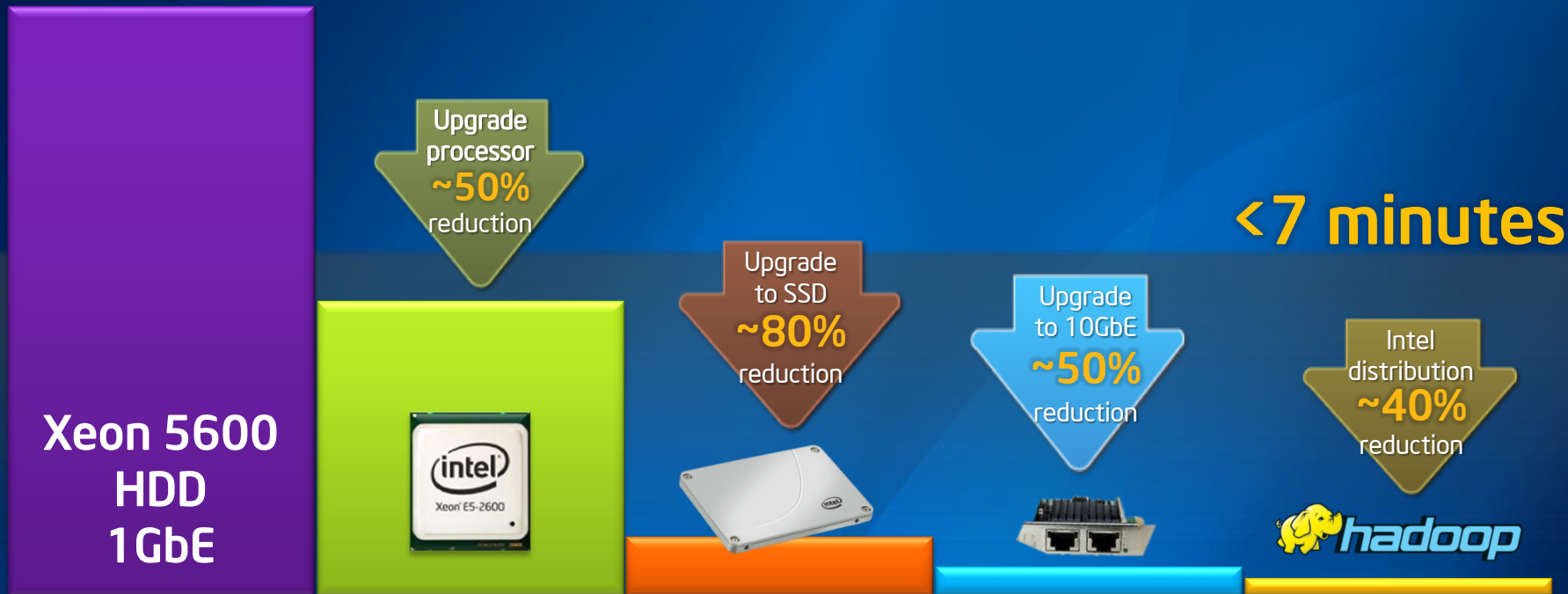
# Intel's Role in Big Data

*Lower the Time and Cost to deliver Big Data Solutions*



# The Power of Platform Solutions

TeraSort for 1TB sort:  
>4 hour process time



# Intel® Distribution for Apache Hadoop



Granular access control in HBase

Up to 20X faster decryption with AES-NI\*



Common authentication,  
access control, auditing



30X faster Terasort on Intel® Xeon  
processors, Intel 10GbE, and SSD

Up to 8.5X faster queries in Hive\*



Bringing MapReduce to  
data on Lustre FS



Up to 30% gain in job performance,  
automated by Intel® Active Tuner



Optimizing Hadoop for  
virtualization & cloud

*Drive down complexity , Increase performance on IA,  
Secure the data*



# Challenge: Extreme Complexity

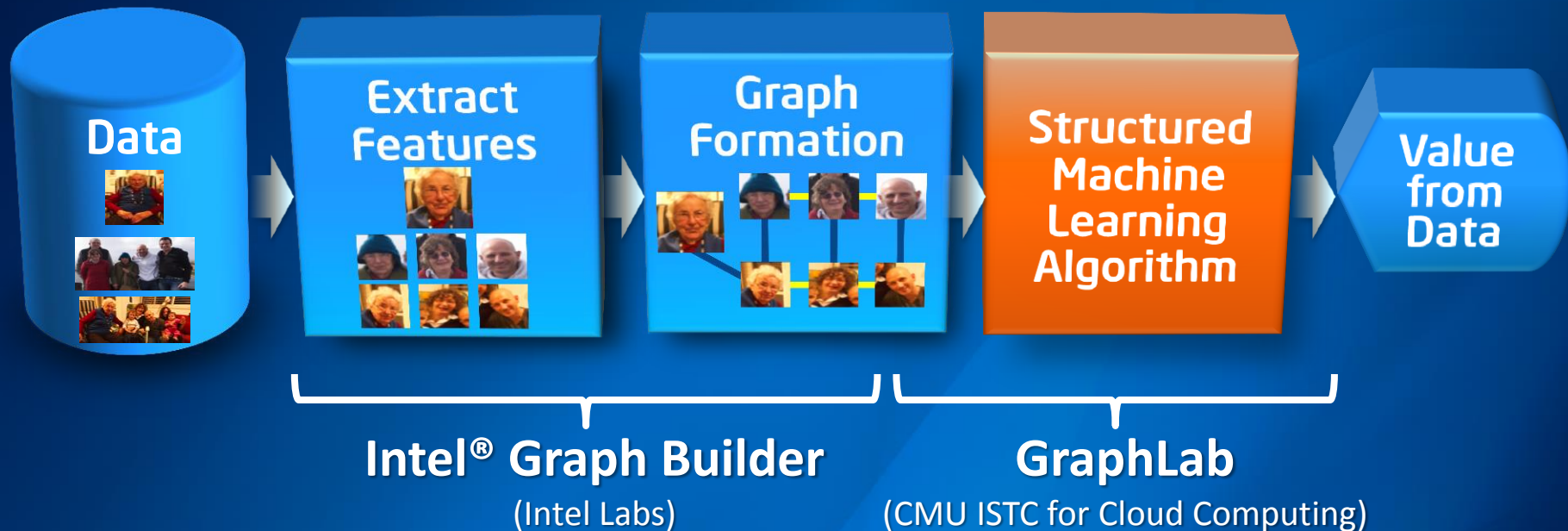
How do you  
effectively  
create and  
then analyze  
something like  
this?

Vertex view

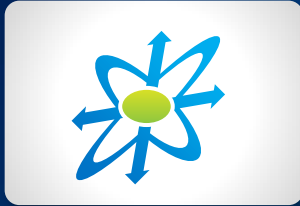
Don't have regularized structure  
Overlapping communities

# Open Source Relationship Computing

*Tools To Build & Analyze Graphs For Big Data Mining, Machine Learning*



# Intel's Big Data Research



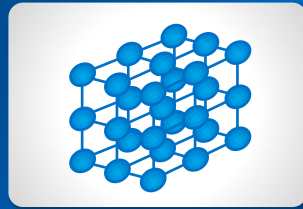
Distributed  
Data Computing



Future Data  
Experiences



Machine Learning  
Algorithms



Data Ecosystem  
Infrastructure



Analytical  
Applications

*INVESTING IN 1000'S OF SCIENTISTS ACROSS  
INTEL, ACADEMIA, INDUSTRY & GOVERNMENT LABS*

