Big Data Use Cases

NASPI Working Group Meeting
October 23, 2014

Siamak Tavallaei
Chief Architect, Moonshot
Distinguished Technologist, HP Server
Hewlett-Packard Company

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Outline

- Common Use Cases of Big Data
- Big Data Goal
- Fields of Study
- Big Data Tools
- Customer Use Case Examples
- Results

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

A New Era of Data-Collecting Devices and Sensors

Purpose-built, energy-efficient, lots of *little data*

2013

98,000 tweets

23,148 apps downloaded

400,710 ad requests

60 sec

2000 lyrics played on Tunewiki

1,500 pings sent on PingMe

208,333 minutes Angry Birds played

Growing Internet of Things (IoT)

Pervasive Connectivity | Smart Device Expansion | Explosion of Information

By 2020

30 Billion[1] Devices

40 Trillion GB[2] DATA

Mobile Apps **10 Million**[3]

... for 8 Billion[4]

A New Style of IT is Required for IoT Solutions

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Common use cases for Big Data

- Customer Knowledge
- Fraud Detection Risk Compliance
- Targeted Marketing
- Security
- Energy Transportation
- Optimizing Operations

Big Data Goal (accuracy, agility, quality, precision , lower cost)

Extract Business-value of Data

Source: Combining Moonshot with TI Keystone SoC

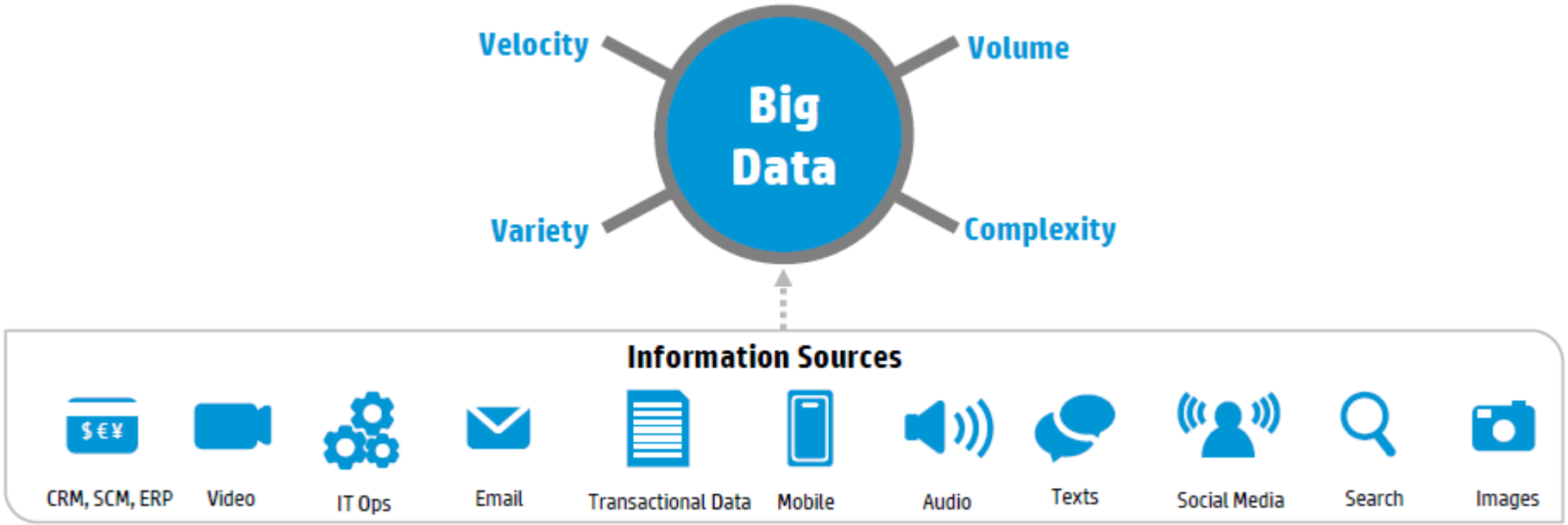
Apply a variety of

data acquisition,
data storage,
data management,
data analysis, and
presentation tools

to arrive at *insights* and
make better *predictions*

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Why do we call it *Big Data?*

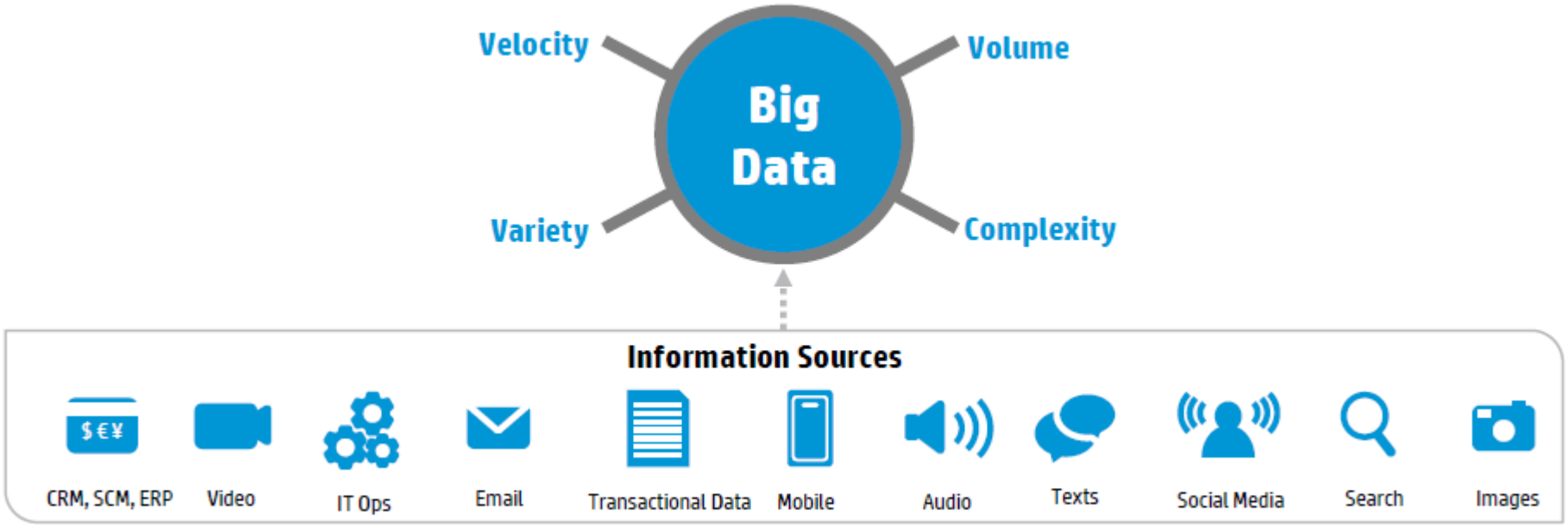


© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Why do we call it *Big Data?*

People-produced *Machine-produced* *Computer-produced*

Digital *Analog*



© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Big Data Tools

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Fields of Study

Study of *Objects*: Physics, Biology, Geology, …

We also need *Data Scientists* for
the analysis and interpretation of *Data*

Mathematics

- Statistic, Inference, Heuristics, Numerical Analysis

Computer Science

- Game Theory, Graph Theory, Cluster Analysis

Psychology, Sociology

- *How to interpret the results*

Art

- *How to display the results*

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change with

Understanding the interrelationships of

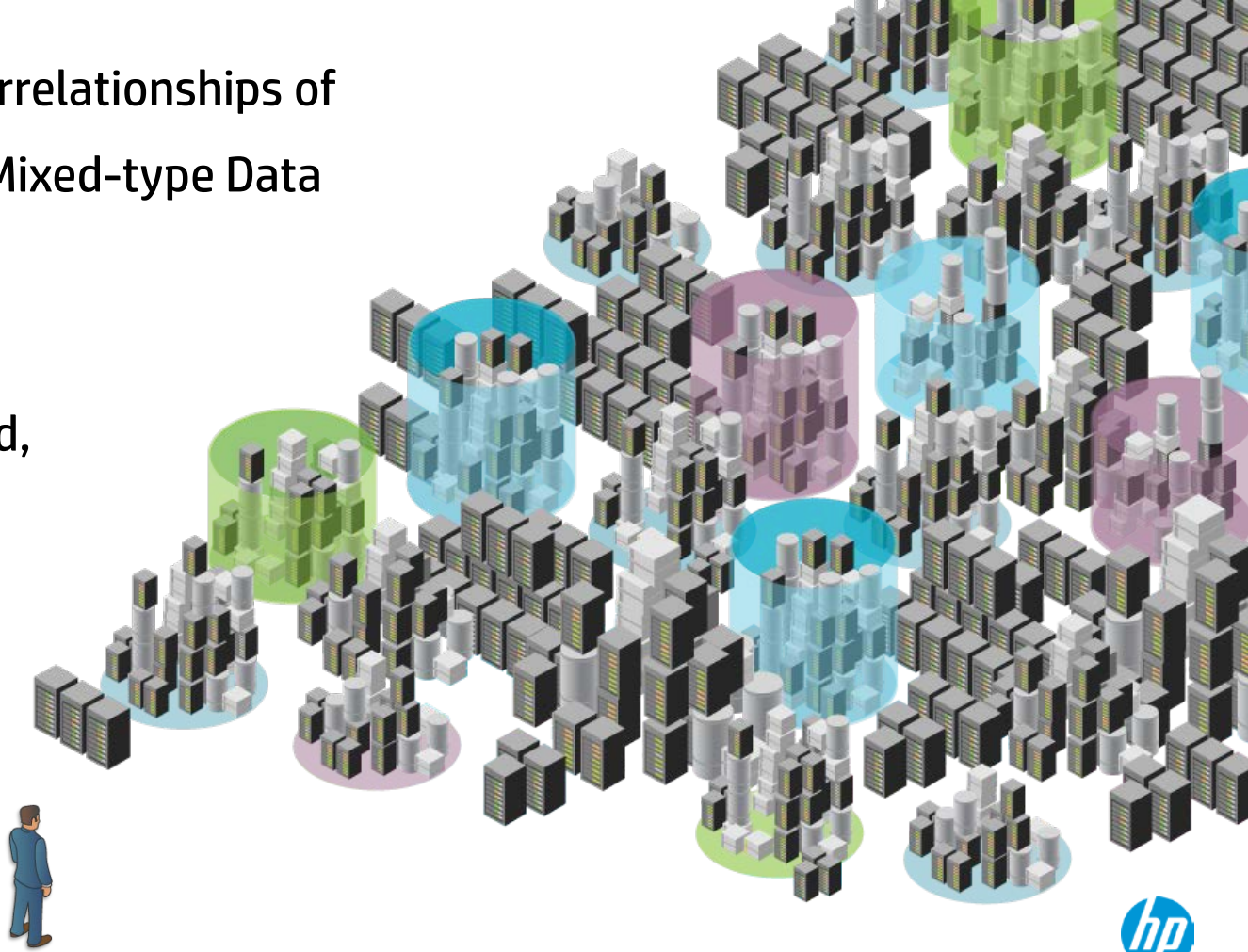
Clusters of Dispersed, Mixed-type Data

Apply:

Hierarchical, Distributed,

Heterogeneous

Computing



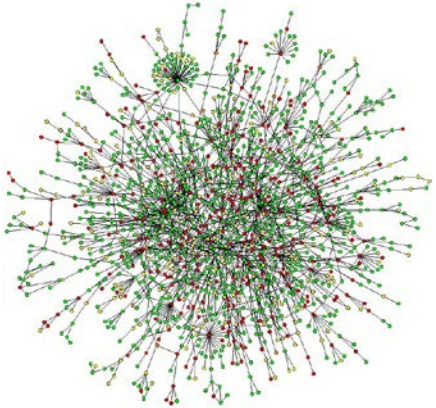
 © Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Graph Analysis

Ubiquitous Data Structures

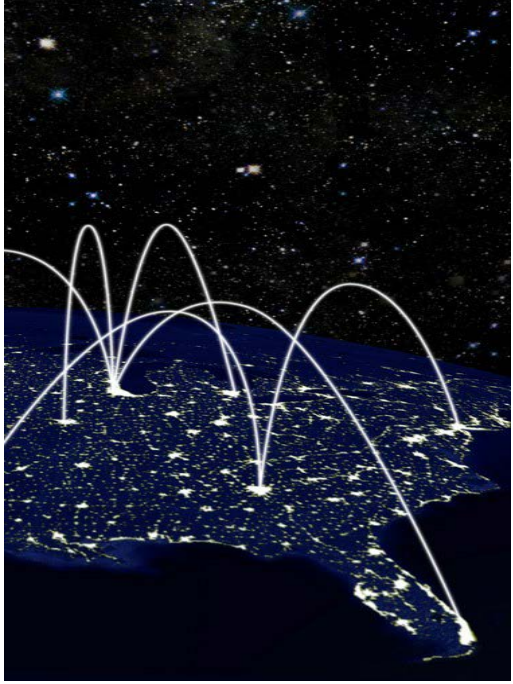
A collection of binary relationships

Networks of pairwise interactions



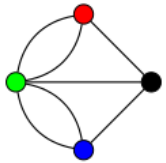
Yeast protein interactions

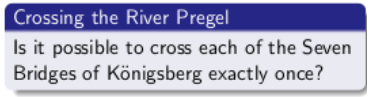
http://stat.genopole.cnrs.fr/



Examples:

- Utility grids
- Social networks
- Digital networks
- Road networks
- Internet
- Protein interactomes (molecular interactions in a particular cell)





Seven Bridges of Konigsberg problem: ~300 years ago, the first graph problem consisted of 4 vertices and 7 edges.

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Graph Databases

Tools for solving Big Data Problems

Uncovering fraud rings

- Traditional relational database techniques require modeling as a set of tables and columns
- Carrying out a series of complex joins and self-joins
- Such queries are incredibly complex to build and expensive to run

Scaling challenge

- Real-time access poses significant technical challenges
- Performance becomes exponentially worse as the size of the ring increases or as the total data set grows

Solution:

- Graph databases have emerged as an ideal tool for overcoming these hurdles
- E.g., Cypher Query Language provides a simple semantic for detecting rings in the graph
- Navigating connections in memory and in real-time

http://thenewstack.io/how-graph-databases-uncover-patterns-to-break-up-organized-crime/

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Algorithms

Tools for solving Big Data Problems

Gaming companies

- Problem: Content recommendation
- Solution: Regression algorithms (e.g., LASSO, logistic, linear)

Healthcare companies

- Problem: Patient analysis
- Solution: Boosting, Regression

Banks, E-commerce

- Problem: Customer segmentation and classification
- Solution: Clustering algorithms, Regression algorithms, Random Forest, Machine Learning

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Hardware

Tools for solving Big Data Problems

Scale-up vs. Scale-out

- Complexity, flexibility, scale, …
- Big servers with lots of cores and lots of memory vs. distributed/parallel computing
- Cost Considerations: DRAM, HDD, SSD, …

- Bear-metal, Virtualization, Physicalization, Specialization

Require Different Treatment

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

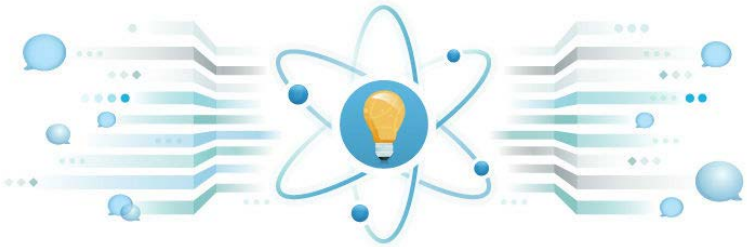
http://www8.hp.com/us/en/software-solutions/big-data-platform-haven/

Software

Tools for solving Big Data Problems

Different Programming Environments

- OpenCL, OpenMP, CUDA, MPI, OpenMPI, MathLab, LabVIEW, …



Different applications

- In-memory database: SAP HANA and ProLiant DL580/980 vs. distributed databases
- Hortonworks, MapR, Vertica, Atonomy, HAVEn
- Parallel machine learning platforms such as Vertica Distributed R

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

http://www8.hp.com/us/en/software-solutions/big-data-platform-haven/

Business Value of Data

Grows as value is added along the way in its lifecycle …

Raw Data Value?

Value of Processed Data Grows

- Collected, Transformed
- Filtered, Sorted
- Stored
- Managed
- Visualized
- Analyzed, Interpreted
- Transmitted, Presented

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained h

What do people do?

Tools for solving Big Data Problems

Typical Big Data deployments

- Collect the Data into a distributed file system such as HDFS
- Apply a NoSQL database such as HBase or Cassandra to process events
- Load data into a Hadoop for filtering, sorting, and manipulation
- Map portions of Data into
  - an in-memory analytic solution such as Apache Spark
  - a Columnar Database such as Vertica
- Apply various Search Algorithms

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Hadoop Environment

A Distributed Computing Environment

Hadoop

- Appears in the majority of server deployments for Big Data
  - Estimated 100K new servers to run Hadoop in 2014

YARN

- Enhances Hadoop with an ecosystem of applications in a container framework
  - Storing data in HDFS
  - HP Vertica database has a plug-in to run on top of Hadoop within YARN containers

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Hadoop Environment

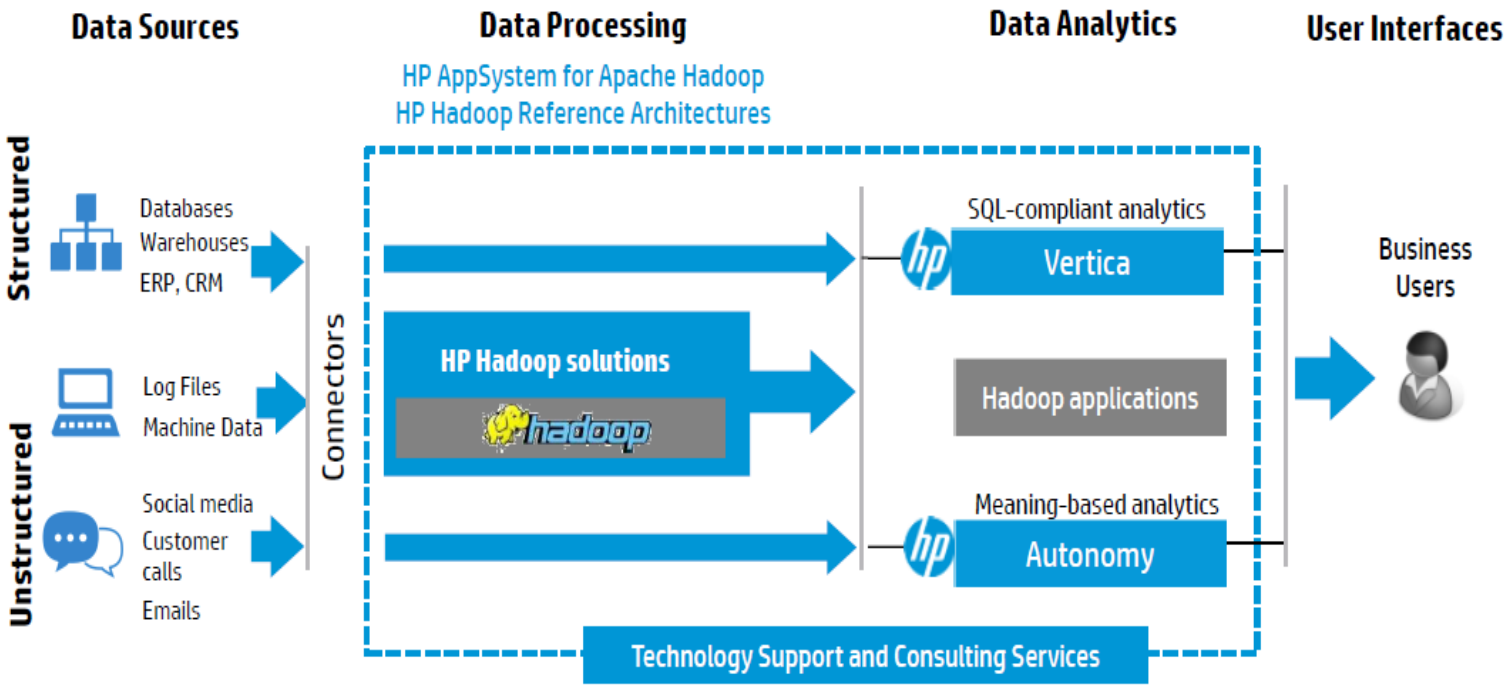
Tools for solving Big Data Problems

MapReduce

- MapReduce is the most common Hadoop workload
  - Map/Distribute
  - Filter/Sort
  - Peer-to-peer Shuffle
  - Reduce
  - Repeat as necessary
  - Arrive at Results
  - Present/Display Results
  - Store/Transmit Results

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

A complete Big Data analytics platform



© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained h

Hadoop Environment

Tools for solving Big Data Problems

Hardware for Hadoop

- Big Data is suited for Distributed Computing running Open Source Software (low cost)
- Cluster of Ethernet-connected servers and storage
  - Variety of local storage and network-attached storage
- Each environment requires different type or size of compute and storage
  - Customers often build multiple compute/storage clusters and
  - Move large datasets through the interconnecting fabric to solve their business problems
- Deploying a server per application is costly if the server is oversized
  - NAS, SAN: shared, converged infrastructure are costly and require expert skillset
  - Distributed, right-sized servers/storage are more suitable

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Hadoop Environment

Tools for solving Big Data Problems

Optimization

- Need efficiency at every level (map function, network, computation, storage, presentation)
- Special-purpose computing helps reduce execution time
- A modular, distributed computing system allows a variety of tools to work collaboratively:
  - Hadoop, Vertica, Casandra, HBase, ArcSight,
  - Trafodion (transactional SQL on HBase), …

HP Moonshot 1500

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Tools for Visualizing Data

Complex datasets require finesse!

Collect Data

- Myriad of Sensors
- Capture and transfer Data

Analyze Data

- Filter, Sort, Rearrange, Combine Data
- Produce Insightful Results

Present Insights

- Extract Business-value of Data
- Make Predictions

3D Graphics Representation & some Artistic Talent Required!

2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

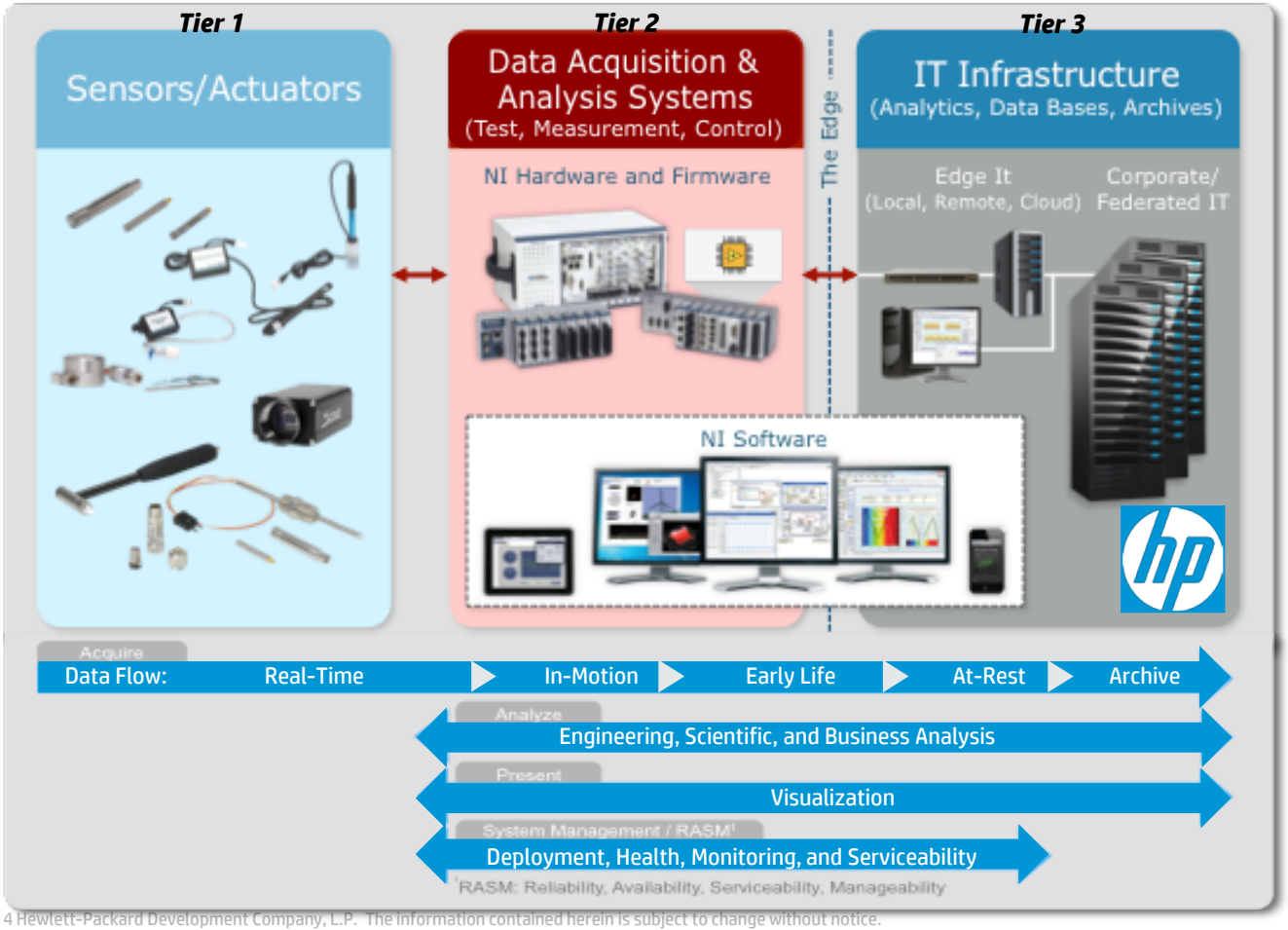
Needle in a Haystack?

Order from Chaos.

Examples of using Big Data

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Generalized 3-Tier Big Analog Data™ Solution (National Instruments)



Acquire:

Analyze:

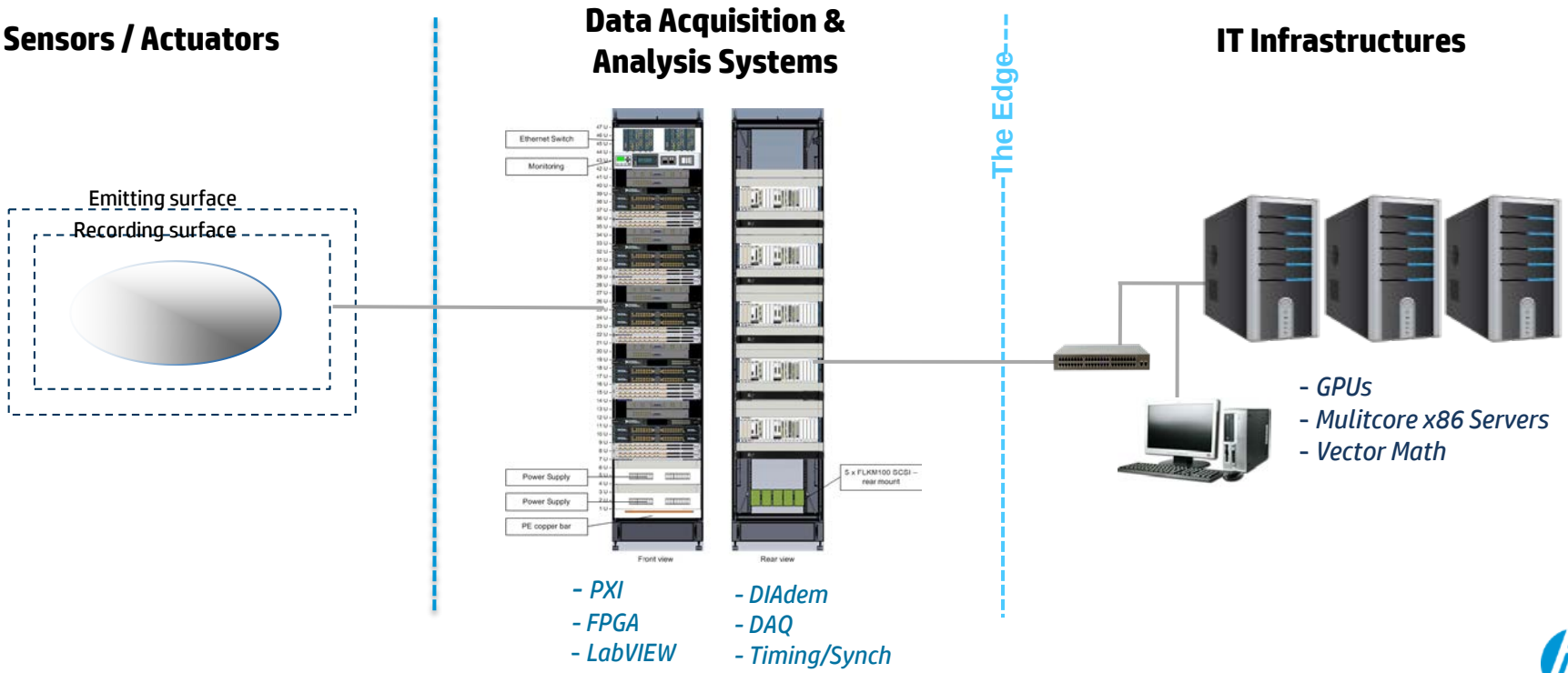
Present:

RASM:

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Example: 3-Tier Big Analog Data™ Solution

Scientific Research (Seismic)



© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Other Examples of using Big Data

HP

Flight Recorder:

- Consider HP: one server every 10 seconds
- Each server with a sea of sensors
- Different data types, values, and ranges, …
- Many servers, …
- Performance data, diagnostics data, catching anomalies, outliers, …
- Feedback loop to service, design, …

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Customer Use Case Examples

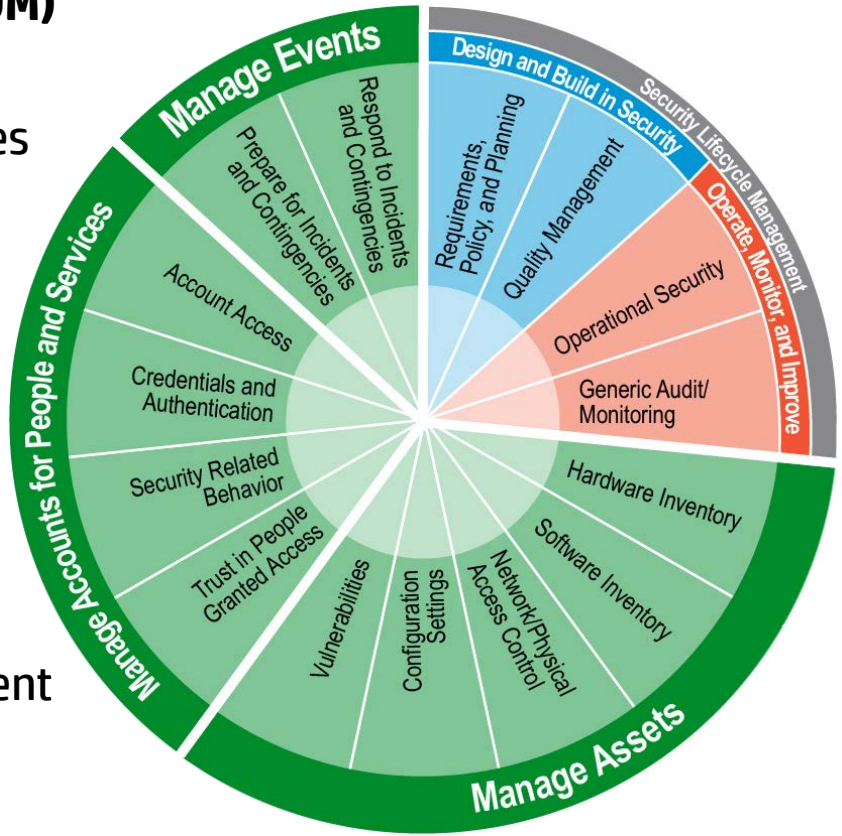
© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Fraud, risk, compliance

Continuous Monitoring – the Federal Model

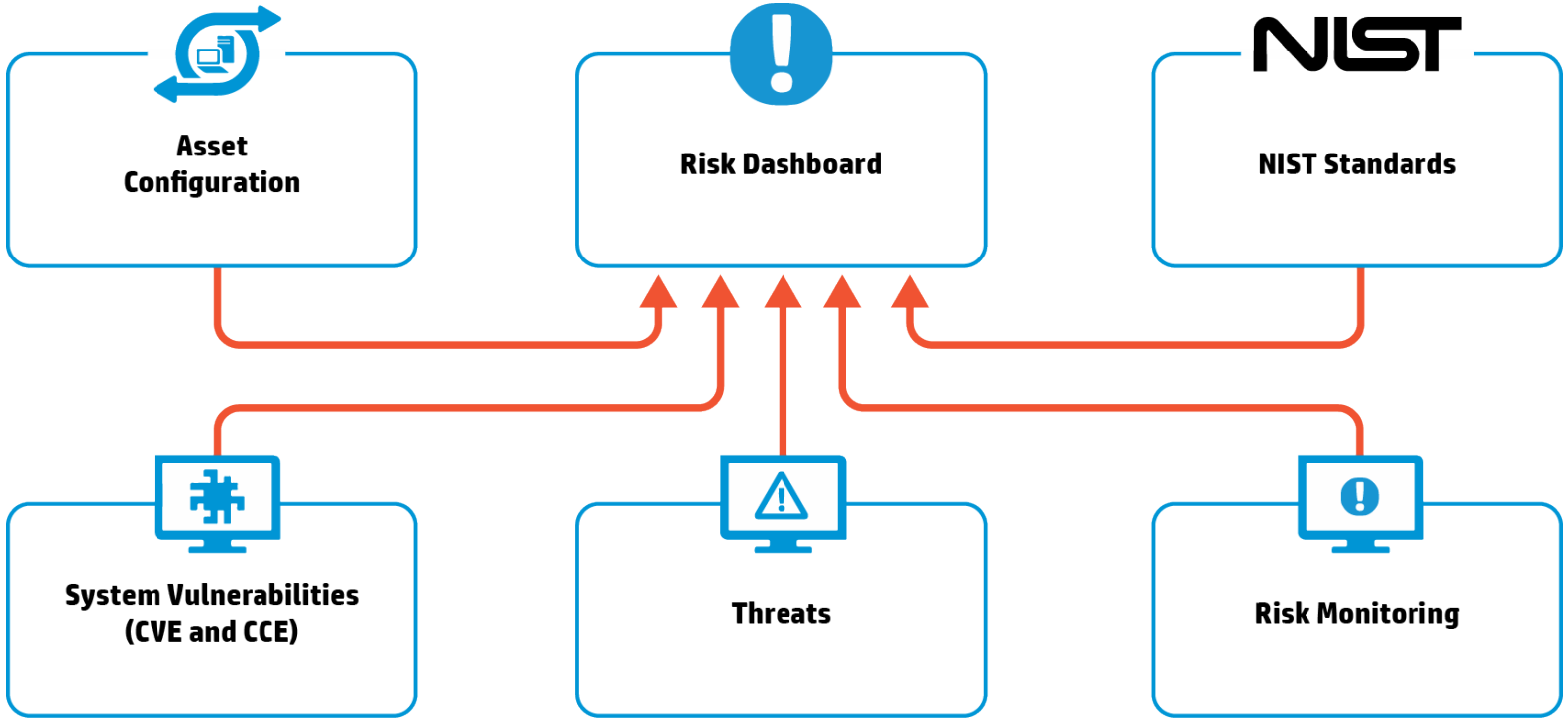
Continuous Mitigation & Diagnostics (CDM)

- Manage Accounts for People and Services
- Manage Events
- Manage Assets
- Implement Security Lifecycle Management



© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained here

Continuous Monitoring Framework



© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Iberdrola

Security analytics in the utilities industry

Challenge

- Manage 15+ million security events generated daily by the myriad of devices across the network
- Consolidate security-related information into an easily understood and manageable format
- Assess and reduce threat exposure
- Control and measure efficiency of security devices and policies

Solution

- HP ArcSight ESM (HAVEn single engine: ArcSight Logger)

Results

- Fast, effective response to attacks and abnormal situations across the entire organization
- Clear view of threat exposure to reduce corporate and IT risk
- Manage 15+ million security events generated daily

© Copyright 2014 Hewlett-Packard Development Company

GSN Games

Customer analytics in the gaming industry

Challenge

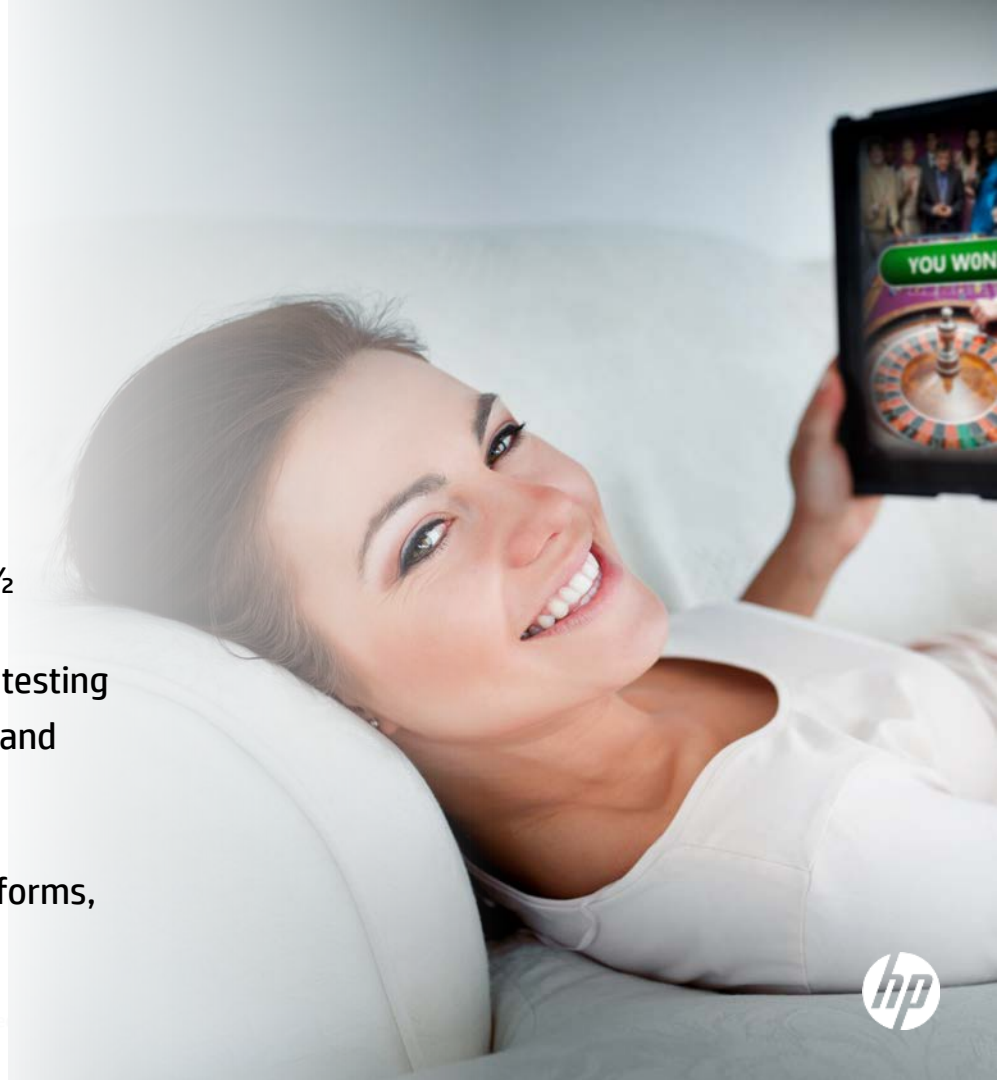
- Rapid data analysis to power strategic growth

Solution

- HP Vertica Analytics Platform

Result

- Reduced A/B test time from up to 36 hours to under ½ second; analyze trillions of data points in real time
- Improved game development through rapid, iterative testing
- Insights into whether new features will engage users and monetize well
- Increase user engagement and re-engagement
- Improved visibility into ad spend across different platforms, from Facebook to mobile



© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is s

Cerner Corporation

Customer analytics, operations analytics in the medical solutions industry

Challenge

- Improve efficiency and quality of patient care by improving the productivity of clinician users

Solution

- *Cerner Millennium* health care platform
- HP HAVEn engines: HP Vertica Analytics Platform, Hadoop

Result

- 6,000% faster analysis of timers helps Cerner gain insight into how physicians and other users use Millennium and make suggestions about using it more efficiently so the users become more efficient physicians
- Rapid analysis of 2 million alerts daily enables Cerner to know what will happen, then head off problems before they happen

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information c

Orange

Communications service provider industry

Business need

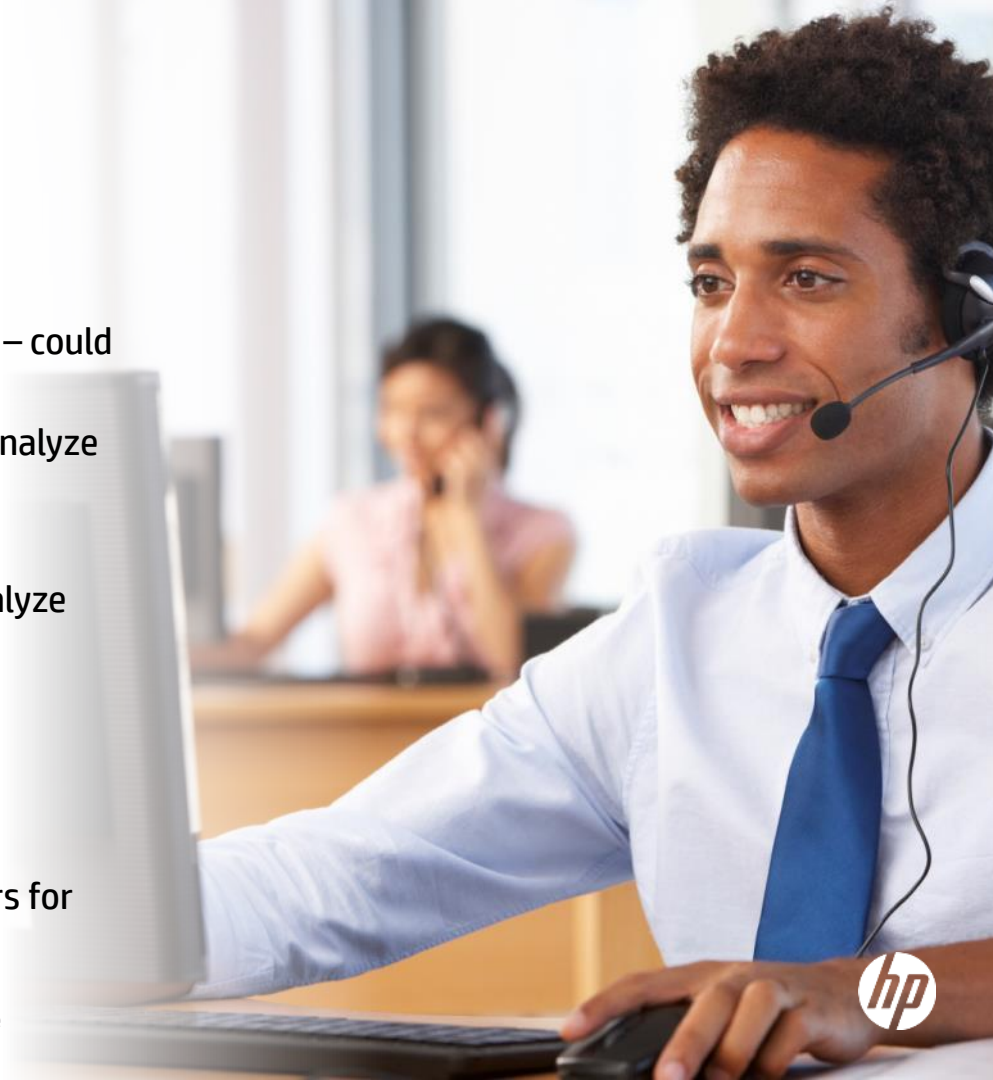
- Gather accurate and relevant information for analysis – could only analyze up to 8% of calls
- Automate the manual processes to track trends and analyze recorded customer conversations

Solution

- HP Qfiniti Workforce Optimization, including Qfiniti Analyze

Result

- 30% reduction in contact center costs
- 25% increase in quality management productivity
- Improved campaign effectiveness by 10%
- Improved root cause understanding of agent behaviors for more effective coaching and training



© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subje

Cardlytics

Consumer analytics in the financial services industry

Challenge

- Migrate to a new, scalable analytics platform to accommodate a data-intensive company undergoing rapid growth

Solution

- HP Vertica Analytics Platform

Result

- Capacity to quadruple the amount of data records added on a weekly basis – from 200 million to 800 million per week
- Typical queries reduced from up to 40 minutes to only one-half or one minute on average, up to 40-to-80X faster
- 100% reliable/stable – eliminated weekly back-ups and maintenance indexing, reducing operational support time by 90%
- Increased avg. customer pipeline 10x: 200 new merchant prospects on a weekly basis, up from 20 per week on legacy platform, with additional scalability

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject t

ANA

Transportation industry

Challenge

- Create and deliver an engaging and compelling online experience
- Replace trial and error approach with powerful multivariate testing capabilities

Solution

- HP Optimost and HP TeamSite, HP LiveSite, powered by HP IDOL, Autonomy consulting services

Result

- 30% benchmark uplift in click-thrus to purchases of domestic air tickets in just one month
- Better customer experience
- Increased online revenue

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject t

Mannheimer Swartling

Legal

Challenge

- Enable its geographically dispersed lawyers to find the right information quickly and easily across all of the firm's data repositories and systems

Solution

- HP Universal Search, powered by Intelligent Data Operating Layer (IDOL)

Result

- Significant TCO savings for both its knowledge management and document management systems
- Supports strategic initiatives such as changing business models around billing, optimizing matter management and business development
- Consolidated repositories empower lawyers to search across all from one intuitive interface to access all the firm's knowledge assets
- Users can search irrespective of locale / language preferences

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

Sample outcomes applying Big Data Analytics

Customer knowledge

15,000 conversations/min

Fraud, risk & compliance

80% drop in resolution time

Targeted marketing

30% more click-thrus

Security

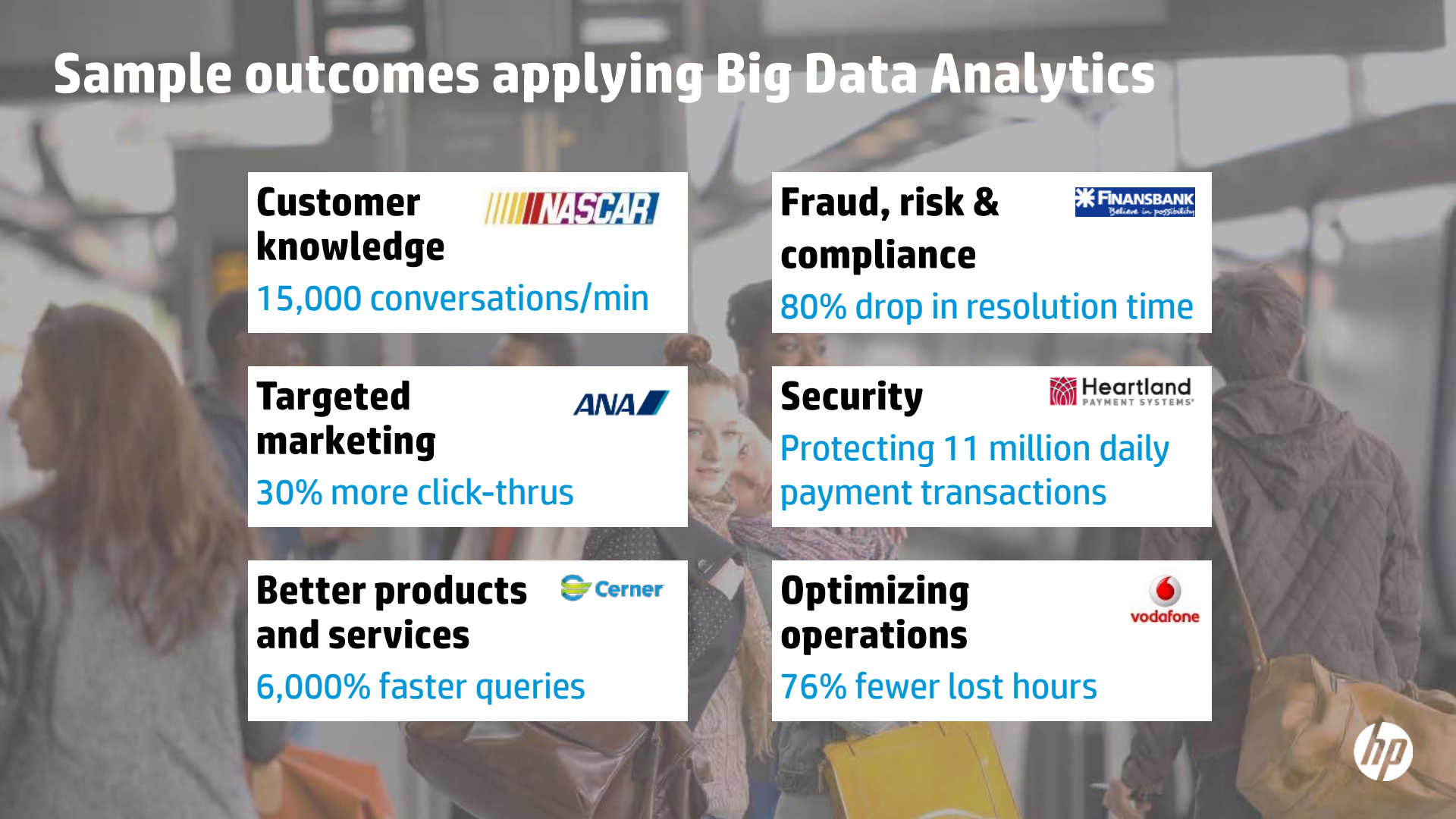
Protecting 11 million daily payment transactions

Better products and services

6,000% faster queries

Optimizing operations

76% fewer lost hours



Summary

- There are many use cases for Big Data
- Business-value of Data grows as it gets processed
- There are a number of modern algorithms, software, and hardware tools for analyzing Big Data
- By applying Big Data Analytic methods, people and businesses benefit from produced insights and predictions

© Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

 © Copyright 2014 Hewlett-Packard Development Company, L.P. The information contained h