Chris Williams
Chief Architect – Big Data and information Management
IBM UK and Ireland
24th April 2013

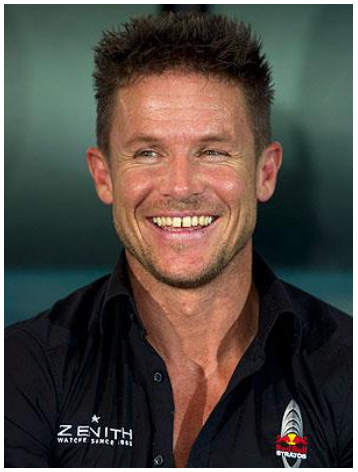
Getting Started with Big Data

*Turning **big data** into **smarter decisions***

Agenda

- Why Big Data Matters
- Entry points for Big Data
- Big Data and Governance
- IBM solutions to accelerate your journey

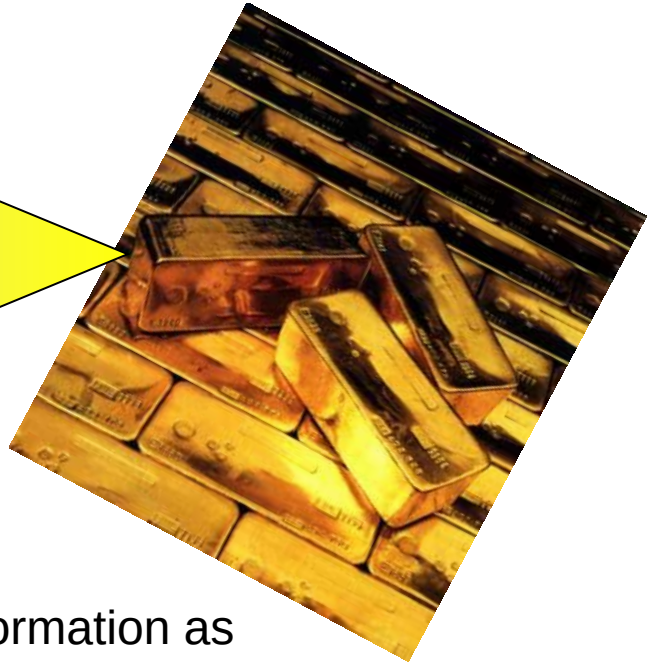
Felix Baumgartner



"I need to see the data……"

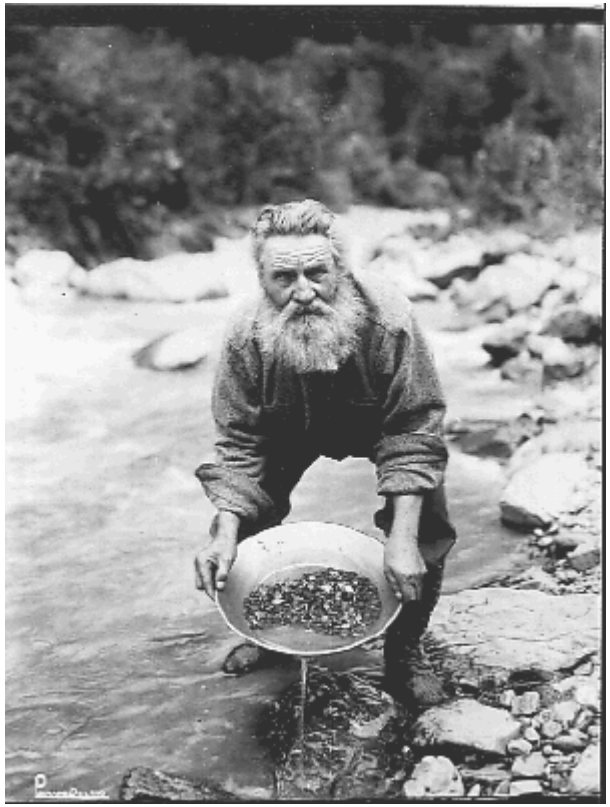
Data Alchemy

Data to run
the business



Information as
a strategic
Asset

Do we really need alchemy to find the gold?



Studies show that organizations competing on analytics outperform their peers

substantially o**utperform**

220%

1.6x Revenue Growth

2.5x Stock Price Appreciation

2.0x EBITDA Growth

Big Data without
analytics is
just a lot of data.

UK Deficit Analysis (2011-12)

- Total Debt (Dec 2012) £1111000000000
- Government Borrowing £121600000000
- Government Income £550600000000
- Government Spending £694890000000
- Interest £48200000000

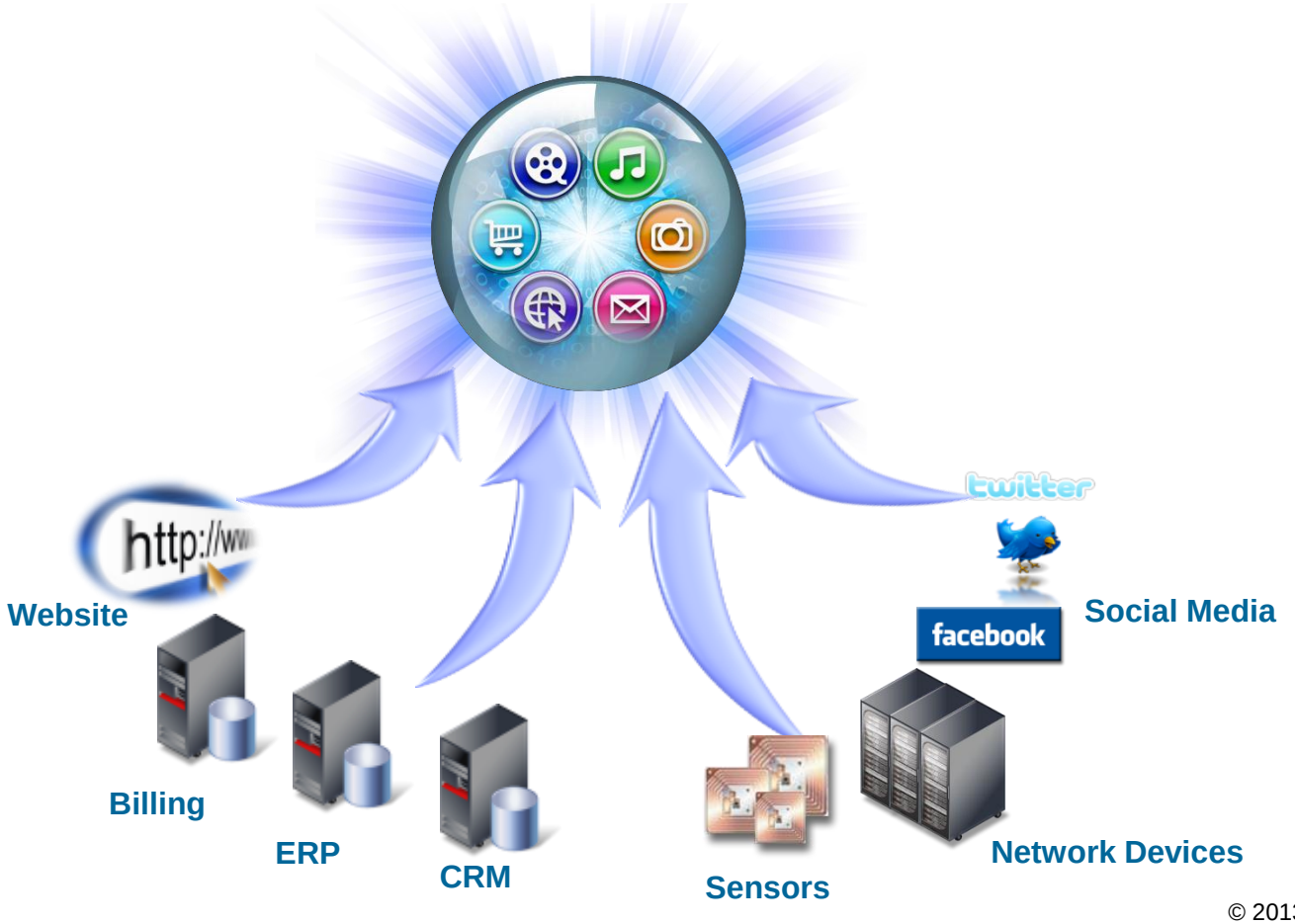
Household Budget

Outstanding Mortgage	£111,100
New debt this year	£12,160
Income	£55,060
Spending	£69,489
Mortgage Payments	£4,820

Big Data is a Hot Topic Because Technology Makes it Possible to Analyze ALL Available Data

Cost effectively manage and analyze*all available data in its native form*

"unstructured", "structured", streaming



If you're not analysing all of the data you're not getting all of the insight.

5 Big Data Patterns

Big Data Exploration

Find, visualize, understand all big data to improve business knowledge

Enhanced 360º View of the Customer

Achieve a true unified view, incorporating internal and external sources

Security/Intelligence Extension

Lower risk, detect fraud and monitor cyber security in real-time

Operations Analysis

Analyze a variety of machine data for improved business results

Data Warehouse Augmentation

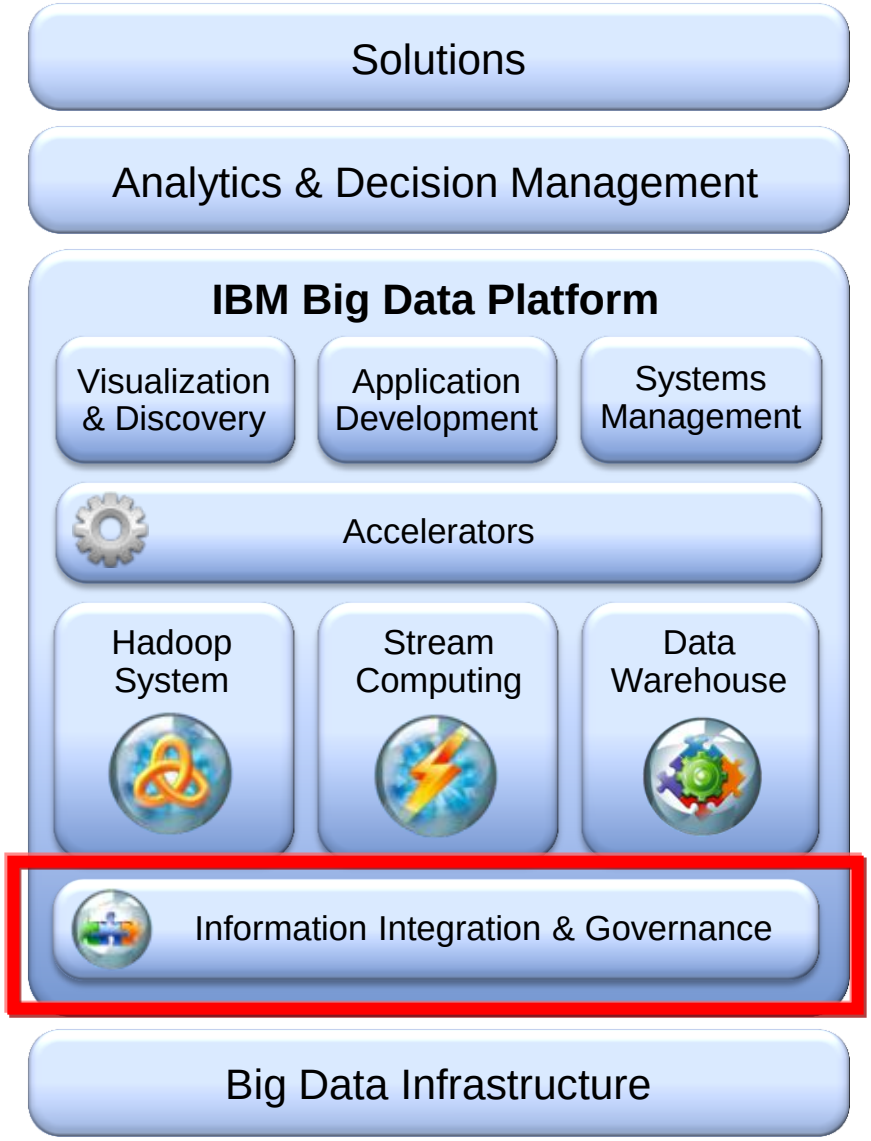
Integrate big data and data warehouse capabilities to increase operational efficiency

Big Data without governance is dangerous.

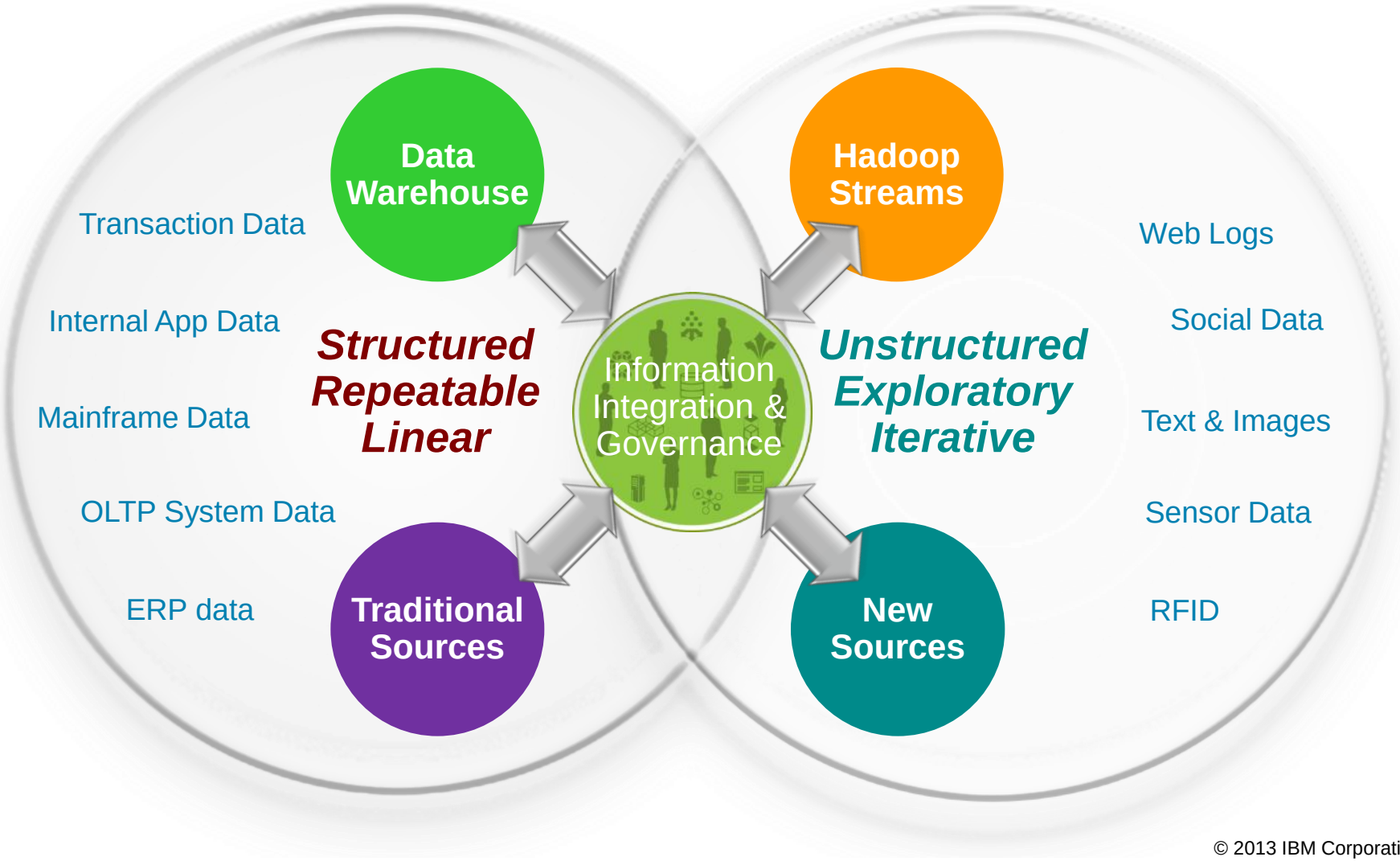
IBM's Big Data Platform is underpinned by Integration & Governance

The desire for trusted information drives the requirements for an information integration and governance platform that is foundational to big data

- Ensure the highest quality information
- Master data into a single view
- Govern data throughout its lifecycle
- Protect and secure all information
- Integrate all data for a common view
- Ensure a single understanding and set of knowledge



Vision for information integration & governance



Big Data Governance Check List

- Compliance & Regulation
- Security
- Data cleansing
- Reference Data
- Master Data
- Stewardship
- Meta Data
- Auditing
- Policy
- Retention and Life Cycle Management
- Test Data Generation
- Privacy
- Integration and Consistency

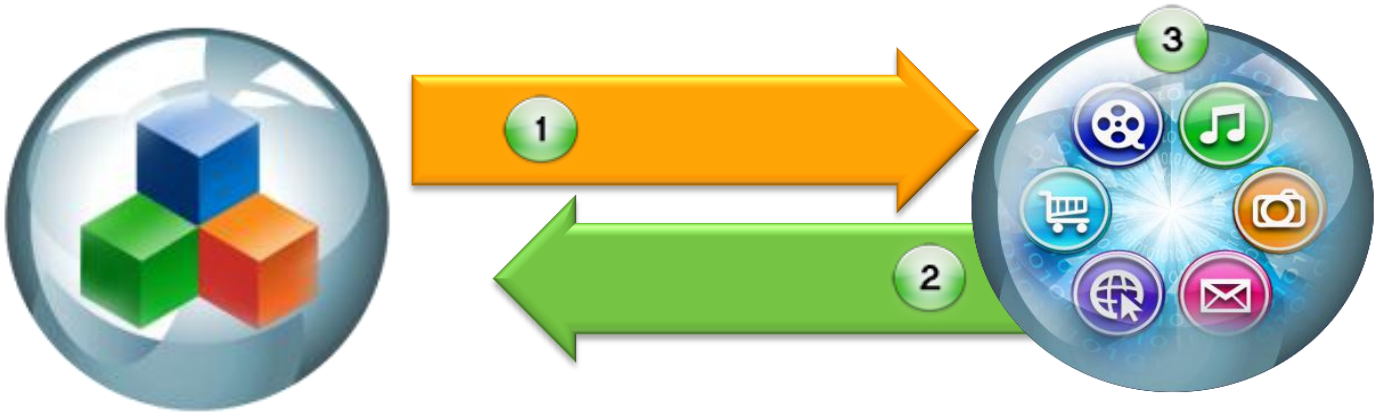
Master data meets Big Data

Governed, Mastered Data

- Governs data creation & links
- Supports enterprise use of insight
- Shared metrics

Big Data

- Leverages data from big data sources
- Drives into enterprise systems
- Leverages master data with big data insights



Level 1
MDM system holds social media profile (social media contacts, preferences, etc.)

Level 2
MDM system links appropriate inputs from Big Data (product sentiment, etc)

Level 3
Big Data applications combining trusted & "big data" data (location offers, social media analytics, micro-campaigns, etc.)

Providing a view into the Data

Data Explorer can provide a **Business User interface to trusted master data combined with related content** from other structured and unstructured data sources

Business User sees 360° view of entity including structured and unstructured data, from sources both inside and outside the enterprise

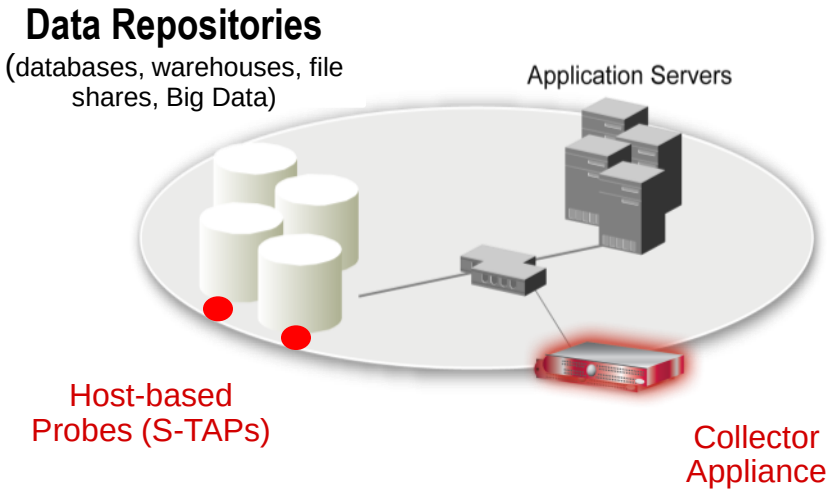
Master Data Management

Data Explorer

© 2013 IBM Corporation

IBM Guardium provides real-time data activity monitoring for security & compliance

- ✓ **Continuous, policy-based, real-time monitoring of all data traffic activities, including actions by privileged users**
- ✓ **Database infrastructure scanning for missing patches, mis-configured privileges and other vulnerabilities**
- ✓ **Data protection compliance automation**



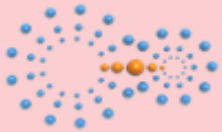
Key Characteristics

- Single Integrated Appliance
- Non-invasive/disruptive, cross-platform architecture
- Dynamically scalable
- SOD enforcement for DBA access
- Auto discover sensitive resources and data
- Detect or block unauthorized & suspicious activity
- Granular, real-time policies
  - *Who, what, when, how*
- 100% visibility including local DBA access
- Minimal performance impact
- Does not rely on resident logs that can easily be erased by attackers, rogue insiders
- No environment changes
- Prepackaged vulnerability knowledge base and compliance reports for SOX, PCI, etc.
- Growing integration with broader security and compliance management vision

IBM delivers a governable, consumable Big Data platform that's steeped in analytics for data in-motion and data at-rest.

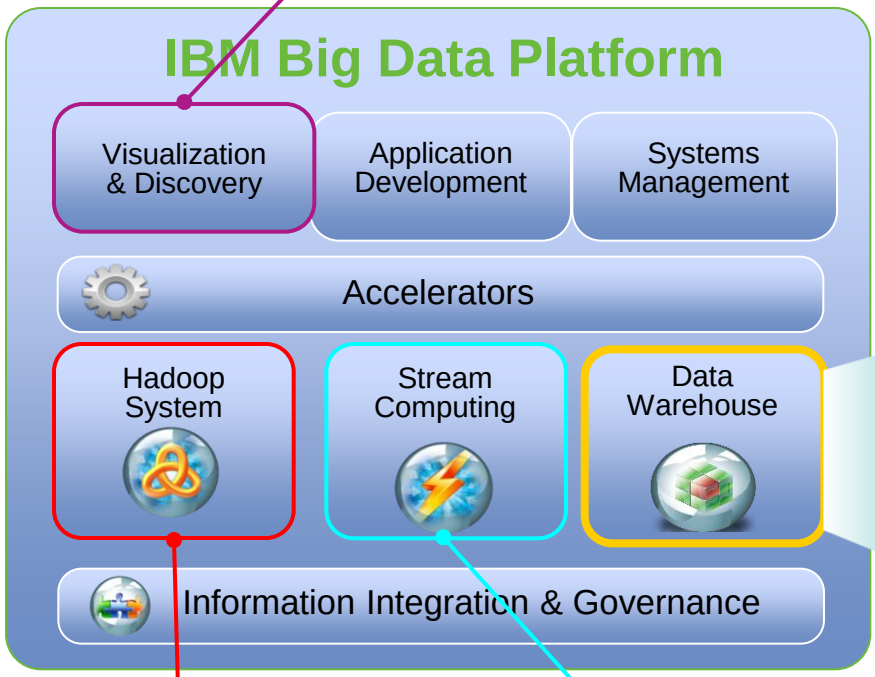
A Big Data Platform Manifesto

CONSUMABLE
(What not How/Patterns/Expert Systems, +++)

Understand and Navigate Federated Big Data Sources		Federated Discovery and Navigation
Manage and Store Huge Volume of any Data		Hadoop File System MapReduce
Structure and Control Data		Data Warehousing
Manage Streaming Data		Stream Computing
Analyze Unstructured Data		Text Analytics Engine
Integrate and Govern all Data Sources		Integration, Data Quality, Security, ILM, MDM

Unlock Big Data

InfoSphere Data Explorer



FORRESTER®

"IBM has the deepest Hadoop platform and application portfolio. IBM, an established EDW vendor, has its own Hadoop distribution; an extensive professional services force working on Hadoop projects; extensive R&D programs developing Hadoop technologies; connections to Hadoop from its EDW." –*The Forrester Wave™: 1Q12*

PureData

Optimized systems delivering data services for analytics & reporting

Analyze Raw Data & reduce costs

InfoSphere BigInsights

Analyze Streaming Data

InfoSphere Streams

Let's simplify Big Data …

From custom and complex …To organized simplicity

Designed to…

- Simplify the building, deploying and management of a Hadoop cluster
- Speed the time-to-value for Hadoop and unstructured data
- Maximize the overall analytic ecosystem
- Provide enterprise security and platform management

[1]*Based on IBM internal testing and customer feedback. "Custom built clusters" refer to clusters that are not professionally pre-built, pre-tested and optimized. Individual results may vary.*

Announcing the new PureData System for Hadoop

Simplifying Big Data for the enterprise

- Accelerate time to value
- Accelerate time to insight
- Simplify big data adoption and consumption
- Extend the value of the data warehouse
- Implement enterprise class big data
- Minimize system setup and administration

PureData
System for Hadoop

Consumer Intelligence

Personal Attributes
- **Identifiers**: name, address, age, gender, occupation…
- **Interests**: sports, pets, cuisine…
- **Life Cycle Status**: marital, parental

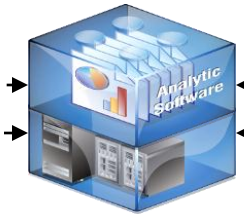
Social Media based 360-degree Consumer Profiles

Timely Insights
- **Intent to buy** various products
- **Current Location**

Products Interests
- **Personal preferences** of products
- **Product Purchase history**

Life Events
- **Life-changing events**: relocation, having a baby, getting married, getting divorced, buying a house…

Relationships
- **Personal relationships**: family, friends and roommates…
- **Business relationships:** co-workers and work/interest network…



Monetizable intent to buy products

I need a new digital camera for my food pictures, any recommendations around 300?

What should I buy?? A mini laptop with Windows 7 OR a Apple MacBook!??!

Location announcements

I'm at Starbucks Parque Tezontle http://4sq.com/fYReSj

Life events

College: Off to Stanford for my MBA! Bbye chicago!

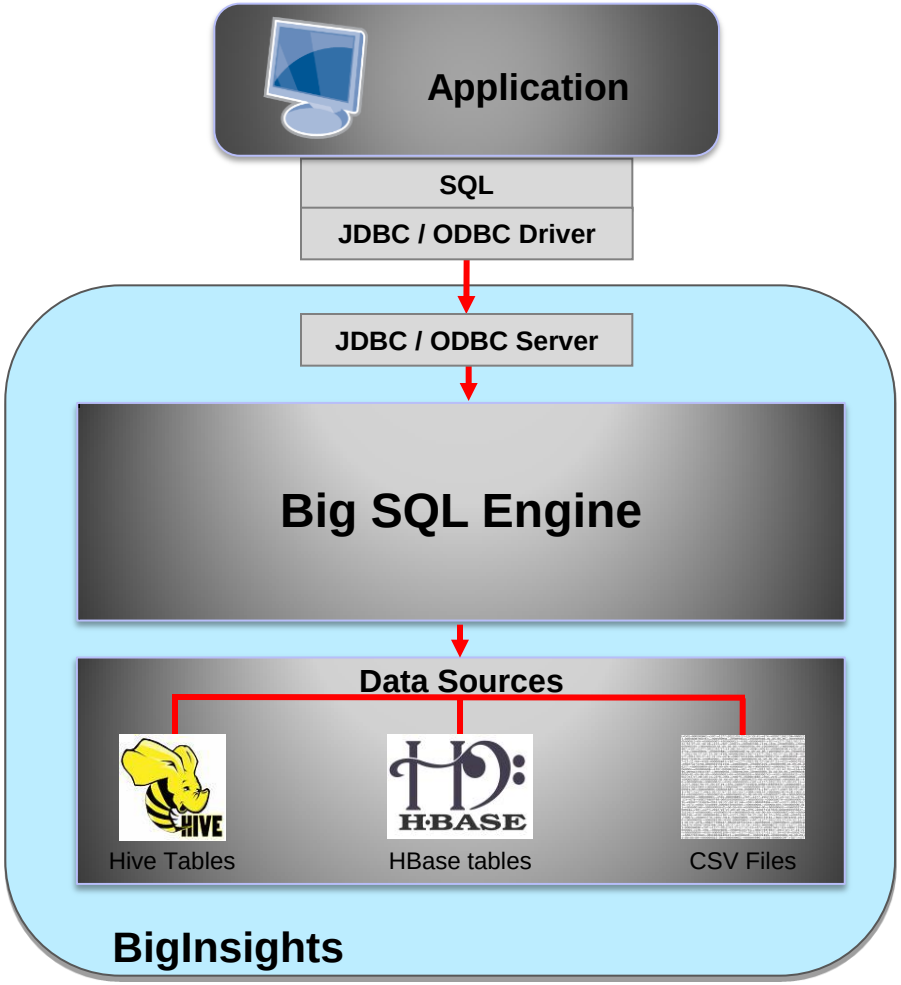
Looks like we'll be moving to New Orleans sooner than I thought.

Intent to buy a house

I'm thinking about buying a home in Buckingham Estates per a recommendation. Anyone have advice on that area? #atx #austinrealestate #austin

Big SQL: Native SQL Query Access for Hadoop

- Native SQL access to data stored in BigInsights
  - ANSI SQL 92+
  - Standard syntax support (joins, data types, …)
- Real JDBC/ODBC drivers
  - Prepared statements
  - Cancel support
  - Database metadata API support
  - Secure socket connections (SSL)
- Optimization
  - Leveraging MapReduce parallelism or…
  - Direct access for low-latency queries
- Varied data sources
  - HBase (including secondary indexes)
  - CSV, Delimited files, Sequence files
  - JSON
  - Hive tables



© 2013 IBM Corporation

Conclusions

- Big data matters – if you don't analyse all of the data you won't get all of the insight
- The big 5 – high value entry points for your big data journey
- The importance of best practice and governance
- IBM's Big Data Platform is the most complete and simplest to deploy

THINK

BIG

@chrispawilliams
chris.williams@uk.ibm.com