Enabling Big Data with Cloud

Go faster
Reduce risk
Scale as you grow
Avoid mistakes

Dr. Phil Shelley

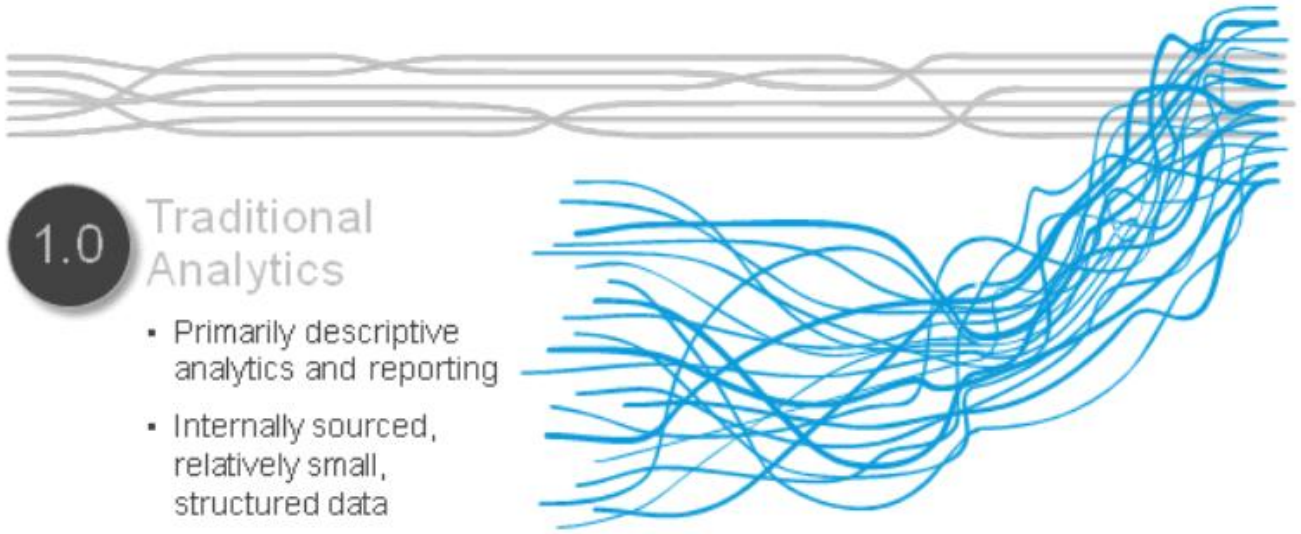
NEWTON PARK PARTNERS

Why Cloud and Big Data?

- Complexity
- Speed
- Cost
- Skills
- Support
- Technology

ANALYTICS 2.0 | THE BIG DATA ERA

INTERNATIONAL INSTITUTE FOR ANALYTICS.



1.0 Traditional Analytics

- Primarily descriptive analytics and reporting
- Internally sourced, relatively small, structured data
- "Back room" teams of analysts
- Internal decision support

2.0 Big Data

- Complex, large, unstructured data sources
- New analytical and computational capabilities
- "Data Scientists" emerge
- Online firms create data-based products and services

NPP NEWTON PARK PARTNERS

Hot Searches

Top Charts New!

Explore

Search terms

× big data

× hadoop

+ Add term

Limit to

Web Search

Worldwide

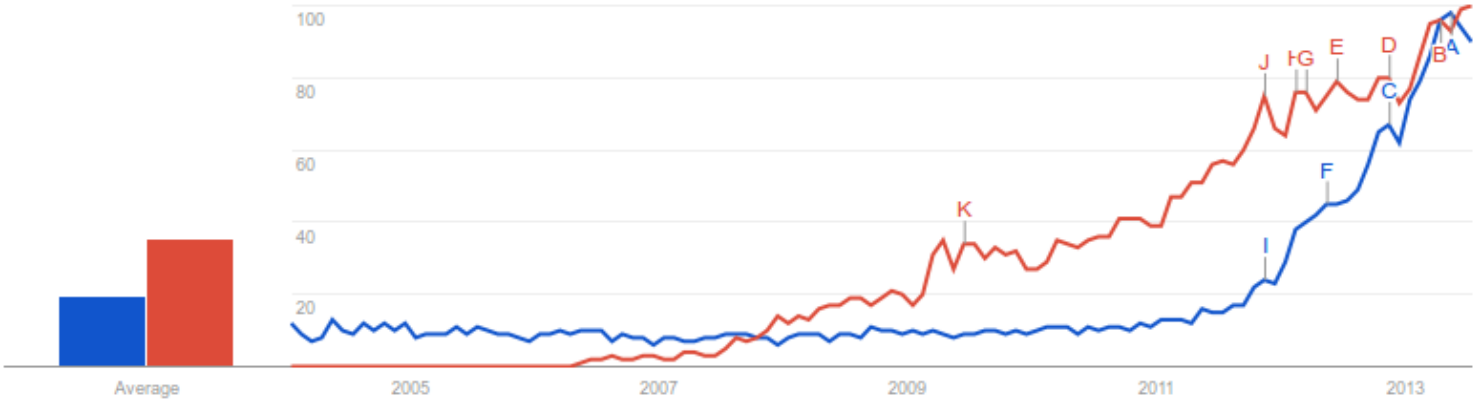
2004 - present

All categories

Interest over time

The number 100 represents the peak search interest

News headlines Forecast



Embed

big data | hadoop

Regional interest

0 100

Region | City

View change over time

Embed

Related terms

Top | Rising

data analytics	100
big data analytics	100
big data hadoop	75
hadoop	75
google big data	45
big data ibm	35
oracle big data	30
big data database	30
big data analysis	25
wiki big data	25

Embed

NEWTON PARK PARTNERS

Industry Trends

IT prepare the data → Business Self-Service
Reporting → Drive business value from data
Copy and Use → Source Once & Re-Use
Linear → Parallel Processing
Proprietary → Open Source
Heavy Iron → Commodity
Capital → Cloud Expense
Computing Power → Up
In-House → Cloud
Batch → Real Time
Costs → Down

NEWTON PARK PARTNERS

Why Big Data – From an IT Perspective

- **Responsiveness**
  - Business user no-longer have to wait for an ETL setup or change
  - Data is available without IT involvement
  - No concept of building a schema or cube before the business can use the data

- **ETL complexity is needed no-longer – DATA HUB**
  - Source Once – Re-Use many times
  - ETL changes to ELTTTTTT

- **Latency in data is a thing of the past**
  - Analysis is routinely possible within minutes of data creation

- **Long-running workload**
  - Can be eliminated and executed at any time
  - Run times are a fraction of the original clock-time

- **Batch processing on mainframes or Data Warehouse workload**
  - Moved to Hadoop
  - Run 10, 50, even 100 times faster

  - **Intelligent Archive**
  - Put your archives/tape data on Hadoop and make it Intelligent
  - Archive with the ability to run analytics or join it with other data

- **Modernize Legacy**
  - Mainframe MIPs reduction has very attractive ROI
  - Move Data Warehouse workload – Reduce Cost – Go Faster – Avoid traditional warehouses

Why Big Data – From a Business Perspective

- **Focus on question that bring value from the data...**
  - Not finding the data
  - Not the schema of the data
  - Not waiting for IT whenever a question comes along
  - …. but test hypothesis without data or system restrictions

- **Focus on Discovery**
  - No-longer reporting
  - Now the inquisitive mind is needed (data scientist)
  - Move to ... I wonder if, or I wonder what …… then test it
  - Ask the previously impossible questions, without:
    - System restrictions
    - Inadequate history or granularity

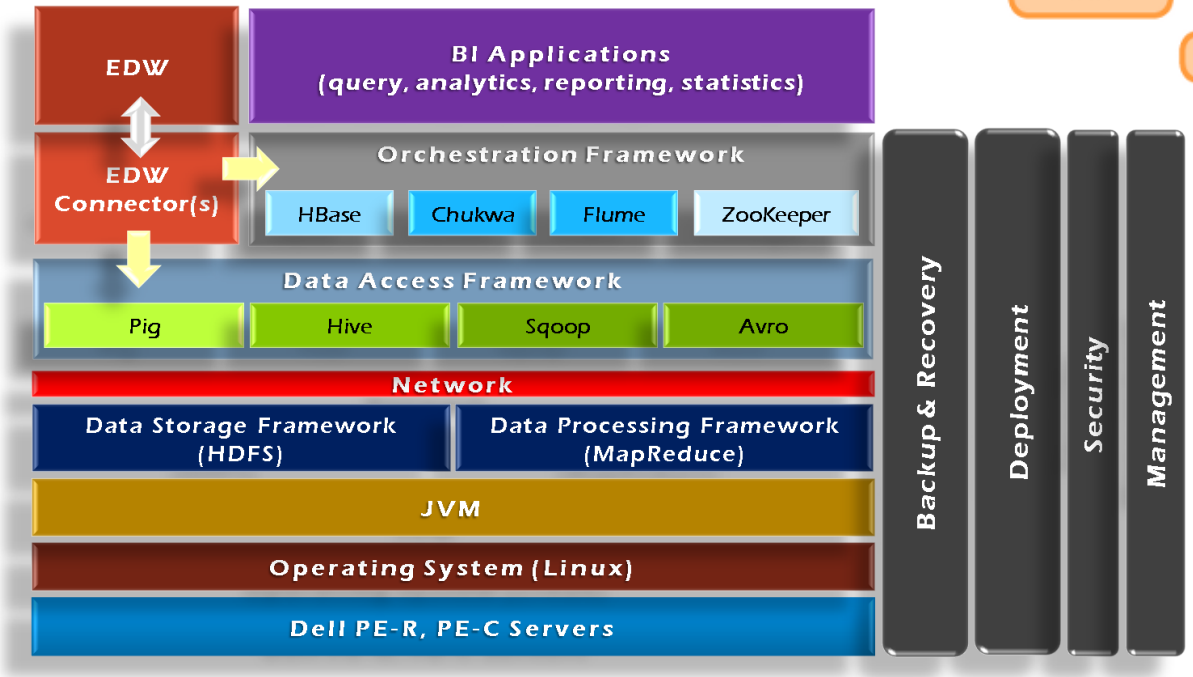
- **What about Real-Time?**

- **What about probability-based decisions and machine learning?**

Why Cloud and Big Data?

- Complexity
- Speed
- Cost
- Skills
- Support
- Technology

Complexity



NEWTON PARK PARTNERS

Speed

In-House

- Specify system and agree design
- Gain approval for the funding
- Order equipment
- Receive and install equipment
- Building the environment
- Burn-in
- Setup support, backup, monitoring and maintenance
- …Begin development

Cloud

- Spin-up the cluster
- Scale and re-size as needed
- Pay the bill as you go

Cost

Pay-per-Drink cloud model

- Avoid up-front capital and risk
- Tools included
- Ability to start small and grow
- Ability to burst to larger environments

Skills

Building, operating and maintaining Big Data Environments

- Hard to hire the needed skills
- Long time-delay to establish in-house skills
- Risk of turn-over derailing initiatives
- Have to provide challenging assignments to retain staff
- Many skills needed, ineffective if some are missing
  - Linux
  - JAVA
  - Networking
  - Security
  - Big Data Modeling
  - Big Data Integration
  - Map Reduce and noSQL development
  - …. more

Support

Keeping the environment up and running

- Monitoring and system alerting
- Hardware failures
- Blacklisted nodes
  - Repair
  - Re-introduce
- OS security patches
- JAVA patches
- Hadoop, noSQL database or other tool upgrades
- Job failure monitoring
- Long-running job diagnostics
- Data backup and restore
- Security profile and access control
- …. more

Technology

- Selection of standard and supported systems
  - What components will work together
- Maintaining currency
  - Platform as a service guarantees a working environment
  - Patching and maintenance can be handled as a service
  - Upgrades can be managed by the cloud vendor
  - Test and deploy the latest versions in the cloud

When Cloud May Not Be The Choice

High-volume environments

Rapidly changing data that needs low-latency consistency

Highly regulated industries (although that may change)

High-speed environments where microsecond latency matters

Conservative company management

Use-Case Examples

- Performance
- Modernization
- Batch processing
- ETL replacement
- Cost reductions

Use-Case Examples

- Performance
  - Long running batch jobs
  - Slow ETL
  - Slow queries due to system load
  - Analytics conflicting with transactional systems
  - Data latency
  - System unable to meet production schedules
  - Job dependencies cause schedule unreliability

Use-Case Examples

- Modernization
  - Retire obsolete systems
  - Language of code is hard to support
  - Skill sets are becoming harder to find
  - Applications hard to maintain
  - Older systems will not scale with the business

Use-Case Examples

- Batch processing
  - Faster
  - Easier to maintain
  - Cheaper
    - Mainframe COBOL
    - HP Unix
    - AIX
    - Sun
    - DB2
    - Oracle

Use-Case Examples

- ETL replacement
  - A sweet spot for Hadoop
  - Massive performance improvement
  - Shrink ETL footprint
  - Reduce licensing costs
  - Reduce the number of data copies
  - Reduce latency from data creation to data use
  - Build a data integration hub
  - Move from ETL to ELttttt...

Use-Case Examples

- Cost reductions
  - Hadoop can be 100% open-source
  - Very large cost reduction opportunities
    - Mainframe MIPS
    - DB2
    - Oracle
    - ETL tools
    - EDW – Teradata, Netezza

Retail Use Cases

Campaign Management
Customer Segmentation and Offers Management
Cross-channel consolidation
Social Networking and Sentiment Analysis
Customer Churn and Retention programs
Loyalty, Incentive and Rewards programs
Fraud and Loss Prevention Altering
Competitive Pricing and Pricing Elasticity
Supply Chain Optimization
Supplier Performance Management
Sales Planning and Forecasting
Procurement Analytics
Customer Basket Analysis and Offers
Cross Selling Optimization
Membership Initiatives

Point of Sale Alerting, Metrics and Optimization
Membership Initiatives
Survey and Customer Feedback Management
Marketing Effectiveness Analysis
Labor Prediction and Management
Competitive Intelligence
Market Sizing and Capacity Planning
Market Mix Optimization
Lifetime Value of Customers
Behavior Analysis
Pricing Analysis and Elasticity
Spend Analysis
Call Center Analytics
Customer Survey Analytics

Financial

Fraud Detection
Risk-based Pricing
Score card existing and potential customers
Customer Retention
Customer Segmentation
Risk Profiling Promotions and Offers
Collections and Recovery Analysis
Customer Lifecycle and behavioral management
Cross Selling and Up Selling
Spend Analysis
Life-time Value Analysis and Optimization
At-Risk Modeling and Mitigation
Payment and Default Predictive Analytics
Account and Application Fraud
Marketing Campaign Effectiveness and Optimization

Insurance

Loss Modeling
Hedging Strategies
Asset and Liability Analysis
Analysis of Offers and Predictive Assessment
Optimization of Re-Insurance
Claims Analysis and Resource Management
Performance and Incentive Predication
Customer Churn and preemptive retention campaigns
Price and offer Optimization
Life-time Customer Analysis
Fraud Detection and Alerting
Claims Forecasting
Underwriting Analysis and Prediction
Competitive Analysis

Health Care

Utilization and Treatment Analysis
Treatment Effectiveness
Diagnosis and Treatment correlation
Managed Care Optimization
Diagnosis Treatment and Trends and Predictions
Drug Utilization and Expense Prediction
Treatment and Outcomes Analysis and Optimization
Demand Forecasting
Price Analysis and Determination
Epidemiology Research
Provider Ratings and Benchmarking
Patient History and Digital Records Archiving and Analysis
Contract Optimization

Media and Marketing

Campaign effectiveness
Loyalty Analytics
Sentiment Analysis
Competitive Intelligence
Market Sizing and Capacity Planning
Market Segmentation
Market Mix Optimization
Lifetime Value of Customers
Behavior Analysis
Pricing Analysis and Elasticity
Spend Analysis
Call Center Analytics
Customer Survey Analytics

Big Data and Cloud

Consider you Options

… but at least start !