Big Data and Business Analytics: Accelerating Digital Transformation in Enterprises and Industries

James PANG, Ph.D.
Visiting Associate Professor
School of Computing
National University of Singapore

Jorge SANZ, Ph.D.
Director Business Analytics Center
School of Computing – Business School
National University of Singapore
Agenda

1. Business Analytics Solution Landscape
2. Big Data, Big Opportunity
3. Future Outlook
4. NUS Business Analytics Center Projects
What is Analytics?

Analytics is broadly defined as the process of deriving **insight** from **data** in order to make better **decisions**

**Use case** | **Data** | **Insight** | **Action/Decision**
--- | --- | --- | ---
Pricing optimization | Past sales, price levels, etc | Predict sales level from price | Set prices in order to maximize profit/revenue
Outcome Based Management | Social data | Social context for each case | Select best social program
Traditional Analytics Landscape

Based on: Competing on Analytics, Davenport and Harris, 2007
Dimensions of Analytics

**Data Sources**
- Internal data
- External data

**Insight**
- Trends, patterns, sentiments, etc.

**Analytics Process**
- Data driven, human driven, event driven

**Maturity Level**
- Descriptive, predictive, prescriptive, etc.

**Decision Automation**
- Decision support vs Decision automation

**Genericity**
- Industry specific vs Cross industry

**Usage**
- Inward/IT vs Outward/LOB

**Solution Granularity**
- Component, asset, product, solution

**Computation Complexity**
- Simple (e.g. reporting), complex (predictive, optimization), very complex (NLP + machine learning + scoring, etc.)

**Users**
- Business user, Data scientist, Business Analyst/Consultant, Application developer

* NLP – Natural Language Processing
* LOB – Line of Business

© 2016
Select the Right Analytics Techniques for the Use Cases

- The five common used business analytics techniques are shown below:

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimization</td>
<td>Mathematically finds the best solution to complex problems with many decision options and constraints</td>
</tr>
<tr>
<td>Simulation</td>
<td>Simulate and analyze a system, process, behaviour, or business problem</td>
</tr>
<tr>
<td>Visualization</td>
<td>Allows users to visually gain insight from data and modeling results, e.g. data exploration, dashboard, BI reports</td>
</tr>
<tr>
<td>Advanced Statistics &amp; Machine Learning</td>
<td>Transform complex data into business insight through identifying the data patterns and trends</td>
</tr>
<tr>
<td>Rule Engine</td>
<td>Creates, manages and executes business rules in robust repositories, for use across wide applications</td>
</tr>
</tbody>
</table>

Typical Analytics Use Cases:
- Process Improvement
- Resource Planning / Scheduling
- Customer Segmentation
- Forecasting
- Preventative Maintenance
- Root Cause Analysis
- Recommendation System
- Anomaly Detection
- Risk & Fraud Analytics

© 2016
Only Analytics Techniques are Not Enough

- Integration
- System Scalability
- Hardware Infrastructure
- Security
- Data Storage & Processing

Information Technologies (IT)
3 Key Elements of a Successful Business Analytics Solution

- Mathematical analysis expertise
- Transform business requirements into mathematical requirements
- Build the analytics models and algorithms

Industry Domain Experts

- Provide industry domain knowledge and experience to ensure the analytics models representing the real world requirement
- Analyze and confirm the analytics results

Analytics Experts

IT Experts

- IT Architecture design
- Applications development
- Data storage and processing
# Build the Business Analytics Solution with 3 Key Elements

## Industry Analytics Applications

<table>
<thead>
<tr>
<th>Finance / Banking</th>
<th>Retail / E-Commerce</th>
<th>Healthcare</th>
<th>Telecom</th>
<th>Manufacturing</th>
</tr>
</thead>
</table>

## Analytics Core Services

- Advance Statistics and Machine Learning
- Optimization
- Simulation
- Rule Engine
- Visualization

## Data Storage and Processing

- Hadoop System
- Stream Computing
- Enterprise Data Warehouse

## Information Integration, Security & Governance

## Hardware Optimization

(Compute-Centric Model -> Data Centric Model)
A New Big Data Era

- **90%** of the world’s data is created in the last two years
- **80%** of the world’s data today is unstructured
- **1 Trillion** connected devices generate 2.5 quintillion bytes data / day

**Volume**

- **Data at Scale**
  - Terabytes to exabytes of existing data to process (e.g. CRM, ERP data, etc.)

**Velocity**

- **Data in Motion**
  - Streaming data, milliseconds to seconds to respond (e.g. data from smart sensors, mobile device, etc.)

**Variety**

- **Data in Many Forms**
  - Structured, unstructured, text, multimedia (e.g. rational DB, images, free text, video, etc.)

**Veracity**

- **Data Uncertainty**
  - Uncertainty due to data error, inconsistency & incompleteness, ambiguities, model approximations (e.g. manual errors, device errors, models errors, etc.)

Source: IBM GTO 2012
Big Data Brings New Opportunity and Value

The value of analytics grows by incorporating **new sources of data**, composing a variety of **analytic techniques**, spanning organizational silos, and enabling iterative, user-driven interaction.

![Diagram showing sources and types of data](image)

- **Structured or standardized**
  - Sales-based demand forecasting
  - Segmentation-based market impact estimates
- **New format or usage of data**
  - Intent-to-buy trends
  - Multi-modal demand forecasting
  - Price-based demand forecasting (own & competitors)

**Scope of decision**
- **Low**
- **High**

Source: IBM GTO 2012
### Big Data Requires to Expand the Analytics Landscape

<table>
<thead>
<tr>
<th>New Data</th>
<th>Traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adaptive Analysis</strong></td>
<td><strong>Continual Analysis</strong></td>
</tr>
<tr>
<td><strong>Optimization under Uncertainty</strong></td>
<td><strong>Optimization</strong></td>
</tr>
<tr>
<td><strong>Predictive Modeling</strong></td>
<td><strong>Simulation</strong></td>
</tr>
<tr>
<td><strong>Forecasting</strong></td>
<td><strong>Alerts</strong></td>
</tr>
<tr>
<td><strong>Query/Drill Down</strong></td>
<td><strong>Ad hoc Reporting</strong></td>
</tr>
<tr>
<td><strong>Standard Reporting</strong></td>
<td><strong>Collect and Ingest/Interpret</strong></td>
</tr>
</tbody>
</table>

#### New Methods
- **Adaptive Analysis**: Responding to context
- **Continual Analysis**: Responding to local change/feedback
- **Optimization under Uncertainty**: Quantifying or mitigating risk
- **Optimization**: Decision complexity, solution speed
- **Predictive Modeling**: Causality, probabilistic, confidence levels
- **Simulation**: High fidelity, games, data farming
- **Forecasting**: Larger data sets, regression
- **Alerts**: Rules/triggers, context sensitive, complex events
- **Query/Drill Down**: In memory data, geo spatial
- **Ad hoc Reporting**: Query by example, user defined reports
- **Standard Reporting**: Real time, visualizations, user interaction
- **Entity Resolution**: People, roles, locations, things
- **Relationship, Feature Extraction**: Rules, semantic inferencing, matching
- **Annotation and Tokenization**: Automated, crowd sourced

### Cognitive (Learn)
- **Prescriptive (Decide and Act)**
- **Predictive (Understand and Predict)**
- **Descriptive (Report)**

Source: IBM GTO 2012
Analytics Paradigm Shifts Enabled by Big Data
Leverage more of the data being captured

TRADITIONAL APPROACH

All available data
Analyze small subsets of data

BIG DATA APPROACH

All available data
Analyzed “all” data

google.org  Flu Trends

United States flu activity: Low  Past years

© 2016
Analytics Paradigm Shifts Enabled by Big Data
Leverage data as it is captured

TRADITIONAL APPROACH

Data → Repository → Analysis → Insight

Analyze data after it’s been processed and landed in a data warehouse or data mart

BIG DATA APPROACH

Data → Analysis → Insight

Analyze data in motion as it’s generated, in real-time

Big Data enabled doctors from University of Ontario to apply neonatal infant monitoring to predict infection in ICU 24 hours in advance

IBM Data Baby
youtube.com
Analytics Paradigm Shifts Enabled by Big Data
Reduce effort required to leverage data

TRADITIONAL APPROACH
Carefully cleanse data before any analysis

BIG DATA APPROACH
Analyze data as is, cleanse as needed

Small amount of carefully organized data
Large amount of messy data
Organizational Adoption of Big Data and Business Analytics Technologies Has Picked Up Pace …

Big data adoption over time

Educate: Learning about big data capabilities
Explore: Exploring internal use cases and developing a strategy
Engage: Implementing infrastructure and running pilot activities
Execute: Using big data and analytics pervasively across the enterprise

Source: Analytics: The upside of disruption. IBM Institute for Business Value 2015 Analytics research study
… and Once Implemented, Big Data and Business Analytics Technologies Are Paying Off

Return on investment from big data technologies

Source: Analytics: The upside of disruption. IBM Institute for Business Value 2015 Analytics research study
Trend 1: Faster and Deeper Analytics

Smarter Analytics
- Faster Decisions
- Real-time Awareness

Adaptive Analytics
- Deep Analytics
- Predictive Models

Data in Motion
- Deeper Insights

Data at Rest
- Big Data
Future Shift of Analytics Frontier

- Large-scale Analytics Platform
  - Not possible today
  - Moving to new frontier of feasibility

Current Analytics Sweet Spot

- Low-latency Analytics Platform

Data Scale:
- PB
- 100TB
- 10TB
- 1TB
- 100GB
- 10GB
- GB
- MB
- KB
- B

Decision Frequency:
- Occasional
- Frequent
- Real-time
Next Generation Analytics Systems (NGAS)

System Capacity (capability)

Device Clusters

High

Med

Low

Single Device

Low

Med

High

Scale-out

Scale-up

Scale-down

Scale-in

Cloud Computing

Exascale

Physical Limits

System Density (1/Latency_{end-to-end})

Scale-down

Maximize feature density

Atom Transistor

Atom Storage

Scale-up

Maximize device capacity

Gigabyte HDD

Terabyte HDD

Scale-out

Maximize system capacity

Blade Server

NAS

Scale-in

Maximize system density

Minimize end-to-end latency

Manycore

FPGA

3D Chips

BPRAM/SCM

FLASH SSD

Interconnect

In-mem DB

DAS

Source: IBM GTO Study

© 2016
Context is what gives Meaning to Data.
Data without Context is Meaningless.
Meaningless Data is Worthless.
The more Context, the higher the Value.
What does Context Mean for the Future of the Enterprise?

A simple analogy for a complex concept

**Leaves** convert solar energy into sugar to power the organism, driving growth and productivity.

**Roots** provide water and nutrients required for operation and growth as well as a stable foundation.

**Trunks** connect and integrate the entire organism.

**People** power the enterprise by converting business situations into communications, decisions and actions, resulting in **products** and **services**.

**Decision Makers**

**Information Technology** connects and integrates the entire enterprise.

**Data** provides the raw materials for insight, and ‘grounding’ in the reality of the enterprise’s environment.
The Future Contextual Enterprise

The ability to dynamically build and accumulate context at scale from data sources to deliver new business value

**Traditional Approach**
Structured, analytical, logical
- Systems of Record

**New Approach**
Creative, holistic thought, intuition
- Systems Of Engagement

- Big Data and NoSQL / NewSQL tools help to realize business value without the limitation of fixed relational schemas that make existing infrastructures too brittle.
- Big Data enables extracting value and information from things in unanticipated ways.
- Capture relationships fit to purpose for industry domains

**Structured**
- Repeatable
- Linear

**Unstructured**
- Exploratory
- Dynamic

- Data Warehouse
- Hadoop and Streams
- Multimedia
- Web Logs
- Social Data
- Text Data: emails
- Sensor data: images
- RFID

- Transaction Data
- Internal App Data
- Mainframe Data
- OLTP System Data
- ERP data

**Traditional Sources**
- Traditional
- Repeatable
- Linear

**New Sources**
- New
- Unstructured
- Exploratory
- Dynamic

- ERP data
- Transaction Data
- Internal App Data
- Mainframe Data
- OLTP System Data

**Enterprise Integration and Context Accumulation**

**Systems Of Record**
- ERP data
- Transaction Data
- Internal App Data
- Mainframe Data
- OLTP System Data

**Systems Of Engagement**
- Multimedia
- Web Logs
- Social Data
- Text Data: emails
- Sensor data: images
- RFID

**Data Sources**
- Traditional
- New
Context at Scale brings Technical Challenges

- Fully contextualized information will require at least 10x the storage of raw data.

- Continual ingestion and curation will require continual deep analytics to discover new insights.

- Dynamic schema requirements and temporality will drive new database requirements.

- High dimensional data with dynamic content and relations lead to irregular graphs that are notoriously hard to partition and favor large in-memory systems.

- Context never rests but data grows continuously with streaming high velocity input. Highly dynamic graphs will require ultra scalable data-structures, that support local graph traversal and at the same time representation/querying of global properties.

- Context size, dynamics, and access patterns will require data-centric, scale-in, highly integrated systems.

- Hybrid memory cubes and other disruptive technologies will enable large-scale, real-time, contextual processing.
Trend 3: Analytics Ecosystem Transformation

**Traditional Analytics Ecosystem**
A siloed, linear Model

Clients | Industry | Communities
---|---|---
Communications | Financial Services | Distribution
Industrial | Public | General Business

**Emerging Analytics Ecosystem**
A collaborative, co-developing, and value co-creation model

Clients | Industry | Communities
---|---|---
Raw Data Providers | Data Aggregators | Analytics Providers

**Analytics as a Service**
- Leverage cross domain data accesses
- Deliver broad and deep analytics services via Web
- Offer customization services for industries
- Partner with industries and technology providers

**Data as a Service**
- Partner with data providers / aggregators to gain access to a wide range of public and private data
- Leverage Cloud infrastructure for scale
- Offer data access, processing, and mgmt services
- Amortize common costs

**Cloud Infrastructure**

**Data Engineer**
- Developer
- Partners
- Communities

**Business Analyst**
- Data Scientist
- Analytics Providers

**Raw Data Providers**
- Data Aggregators

**Analytics Providers**
- SPSS
- SAS
- COGNOS
- SAP BusinessObjects

**Data Aggregators**
- Experian
- LexisNexis
- AXIOM

**Raw Data Providers**
- Private
- Public

**Data Aggregators**
- Facebook
- Google
- YouTube
- Twitter
IBM DataWorks Platform for Analytics Ecosystem Transformation

User Experiences
- Find
- Share
- Collaborate

Solution Blueprints
- Self-Service Analytics
- Internet of Things
- Data Lake
- Mobile Applications

Individual Services
- Data Access
- Data Recognition
- Advanced Analytics

Governance
- Access & Ingest
  - IOT
  - Streaming
  - ETL
- Store
  - Hadoop
  - NoSQL/SQL
  - Object Store
- Analyze & Build
  - Descriptive
  - Predictive
  - Prescriptive
  - Dev environment
- Deploy
  - Apps/APIs
  - Reports
  - Models

Powered by Spark

© 2016
For Users in the Ecosystem to Collaborate Together

OUTPUT

INPUT

Data Engineer
Architects how data is organized & ensures operability
DataWorks Forge

Data Scientist
Gets deep into the data to draw hidden insights for the business
Data Science Experience

Business Analyst
Works with data to apply insights to the business strategy
Watson Analytics

App Developer
Plugs into data and models & writes code to build apps
Bluemix

ANALYSIS

Ingest data

Explore and understand data

Transform:
shape

Transform:
clean

Evaluate

Create and build model

Communicate results

Deliver and deploy model

Understand problem and domain

OUTPUT

INPUT

ANALYSIS

INPUT
Other Important Business Analytics Trends

- Cognitive Computing, Machine Learning, Artificial Intelligence
- Open source analytics platform
- Self-service analytics
NUS Business Analytics Center: Vision and Mission

- Support the goal of making Singapore the regional hub for Business Analytics
  - Build long-term, world-class capabilities to support Singapore’s strategic economic focus on business analytics
  - Develop a center of excellence that equips local companies and organizations with talented skills and leaders in business analytics
    - Deep relevance to different industries through direct engagements with companies and enterprises
    - Hands-on practical experience and PoCs with external organizations

- Build a Center with world-class academic and professional reputation
  - Differentiated study program that combines business and computing into analytics curriculum and projects
  - Bring academics and practitioners from different NUS departments and industry to work together with students
Developing an accurate model to provide estimates on how long a job should take given the characteristics of the job

---

**Finance**

- ALM Roll-Tagging Prediction
- Risk Classification Model
- Analysing High Risk Segments in Auto Loan Portfolio

- OCBC Bank

- Customer-Money Life Cycle
- Marketplace analysis

- PayPal

- PnL Analytics
- Cyber Security
- Anti Money Laundering
- Fraud analytics
- Risk assessment model for investigation program

- DBS

- Developing an accurate model to provide estimates on how long a job should take given the characteristics of the job

- UBS

- Analysis of Customer Queuing Time & Headcount planning

- Swiss Re

- Swiss Re

- Motor Pricing KPIs
- Travel Pricing KPIs and exploratory analysis
- Cross-sell and up-sell in insurance

- Allianz

- Social Network and Geospatial Analytics in context of Insurance

- MUFG

- Optimising Endowment Portfolio Performance

- WAREES
NUS Business Analytics Center Projects, 2016 (II)

Healthcare

- Frequent Attenders to the Emergency Department
- Deriving Insights from NEHR (National Electronic Health Record)
- Emergency Medical Service (EMS) Ambulance Demand Analytics & Prediction
- Breakout detection for Hep C patients.

Information Technology

- Sales Management Analytics
- HR Analytics
- Pricing Assessment Tool based on Analytics
- BlueMix & Watson Analytics
- Online Analytics
- An Analytics Approach to Improve Subscription Rate for Nursing Course (prelim title)
- Global logistics cost optimisation
- Global project - Social Media
- Predicting High Risk Churn Segments Via Product Usage Data
NUS Business Analytics Center Projects, 2016 (III)

Logistics & Manufacturing

- POS transactional data
- Analysis on overtime cost
- Customer credit risk analysis
- IoT / Event Analytics in Manufacturing
- Data Lake architecture to deliver a virtualization layer for disparate data sources.

Retail

- Supply Chain Network Optimization
- News recommendation engine for high net worth customers
- Economic Scenario Stress Testing
- Determining optimal level of markdowns through customer segmentation for revenue maximization
- Market Research
- Social Media /Digital Marketing/PR

Government

- Optimising maintenance schedule for fleet management
- Automatic Rostering System
- Market Research
- Social Media /Digital Marketing/PR
Thank You!

Dr. PANG Yan, James
Email: jamespang@nus.edu.sg

Dr. Jorge SANZ
Email: jorges@nus.edu.sg