

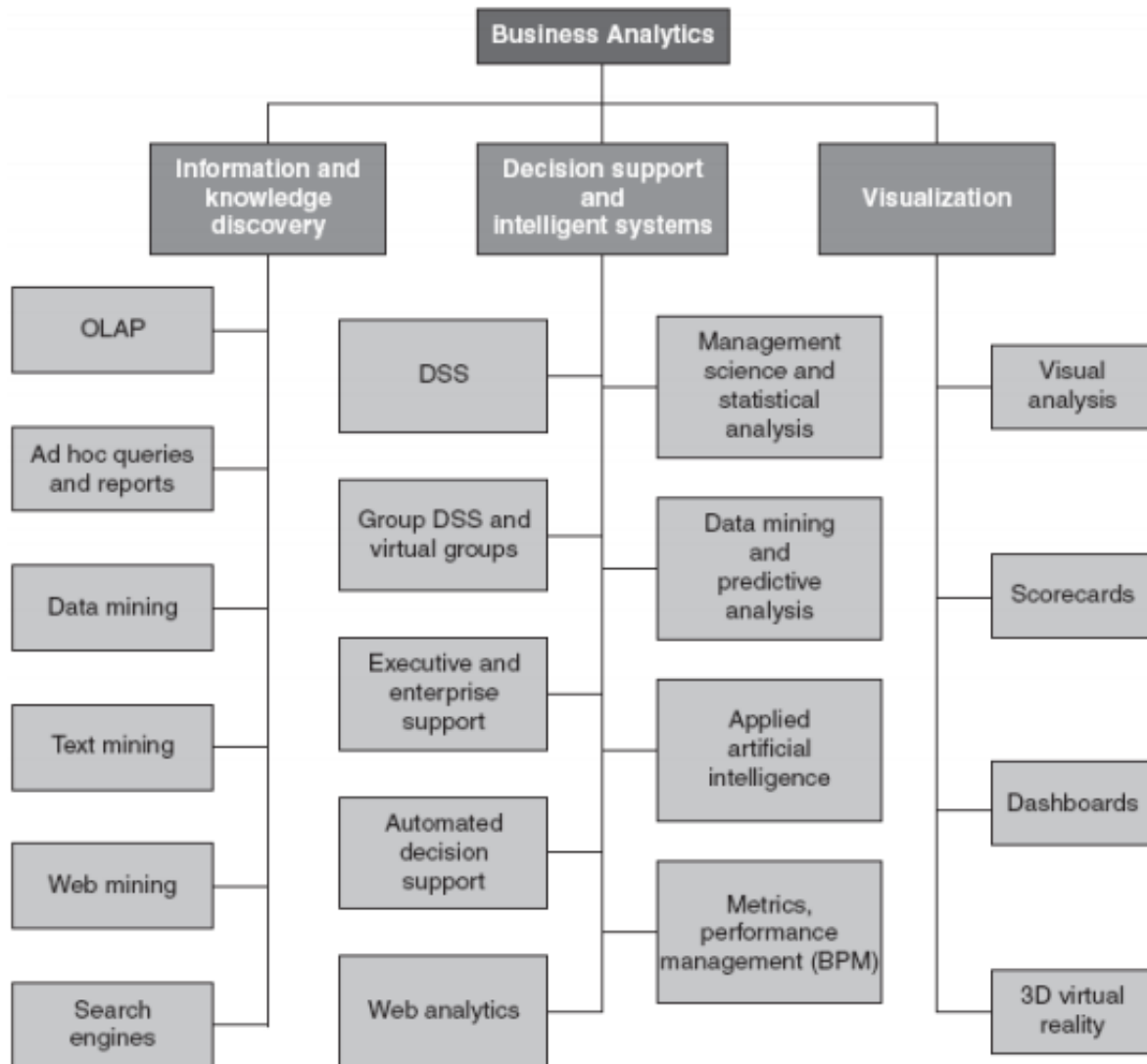
## Table of Contents

<b>Lecture 3&amp;4 Visual Analytics</b> .....	<b>3</b>
The BA field: (an overview).....	3
The Nature of Data .....	4
Structured data .....	4
Unstructured/textual data.....	4
Semi-structured data .....	4
Readying the data for analytics.....	5
Data Pre-processing Tasks and Methods.....	6
Business Reporting .....	8
Different Types of Charts and Graphs .....	8
Basic Charts and Graphs.....	8
<b>Visual Analytics</b> .....	<b>12</b>
Information visualization.....	12
Predictive analytics.....	12
Performance Dashboards .....	13
<b>Lectures 5&amp;6 Data Mining</b> .....	<b>14</b>
Data Mining at the Intersection of Many Disciplines .....	15
Data in Data Mining.....	15
Nominal .....	15
Ordinal.....	15
Interval .....	15
Ratio .....	15
A Taxonomy for Data Mining Tasks.....	16
Accuracy of Classification Models .....	16
Data Mining Process: CRISP-DM.....	17
Data Preparation .....	17
Data Mining Process: SEMMA.....	18
An Estimation Methodology for Classification .....	18
<b>Lecture 7 BPM</b> .....	<b>19</b>
A Closed-loop Process to Optimize Business Performance .....	19
Monitor: How Are We Doing?.....	19
Act and Adjust: What Do We Need to Do Differently? .....	20
BPM Methodologies Balanced Scorecard.....	20
BPM Methodologies .....	21
Strategy map .....	21
Expedia's Uses a Balanced Scorecard .....	22
<b>Lecture 8 Other Modelling Methods</b> .....	<b>23</b>
Categories of Models.....	23
Structure of Mathematical Models for Decision Support .....	23
Examples - Components of Models .....	24
Categories of Models.....	24
Static and Dynamic Models.....	24
Static Analysis.....	24

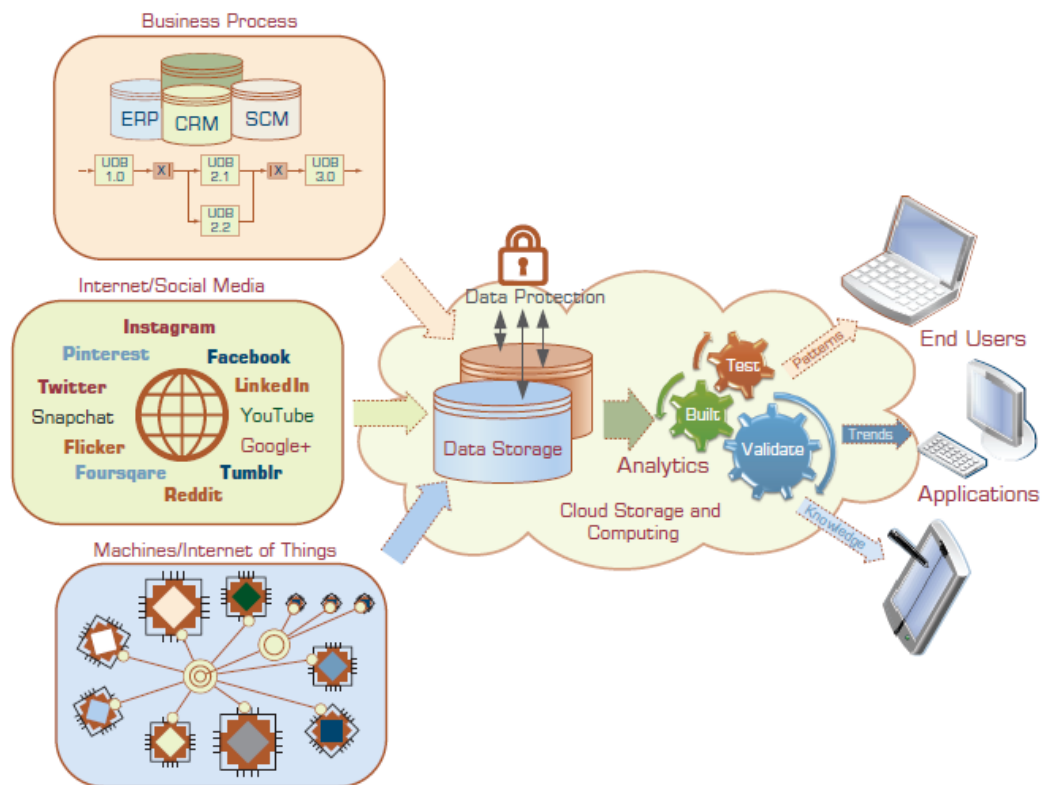
Dynamic Analysis.....	25
<b>Simulation .....</b>	<b>25</b>
<b>Simulation Methodology .....</b>	<b>25</b>
<b>Simulation types.....</b>	<b>25</b>
Probabilistic/Stochastic vs. Deterministic Simulation .....	25
Time-dependent vs. Time-independent Simulation .....	25
Discrete Event vs. Continuous Simulation.....	26
Simulation Implementation.....	26
<b>Lecture 9 Text and Web Mining .....</b>	<b>27</b>
Text mining process.....	27
Text analytics and text mining .....	27
Sentiment Analysis Process .....	28
Web Mining.....	29
Web Usage Mining (clickstream analysis) .....	29
Process View of Web Site Optimisation .....	30
Customer Strategy Framework .....	30
<b>Lecture 10 &amp; 11 Big Data Emerging Trends .....</b>	<b>31</b>
Critical Success Factors for Big Data Analytics .....	31
Skills That Define a Data Scientist .....	32
Big Data and Analytics in Politics - Application Case 13.4 .....	32
Hadoop versus Data Warehouse When to Use Which Platform.....	33
Coexistence of Hadoop and DW .....	33
Stream Analytics; A Use Case in Energy Industry.....	34
Service-Oriented DSS/BI .....	34

## Lecture 3&4 Visual Analytics

The BA field: (an overview)



## The Nature of Data



### Structured data

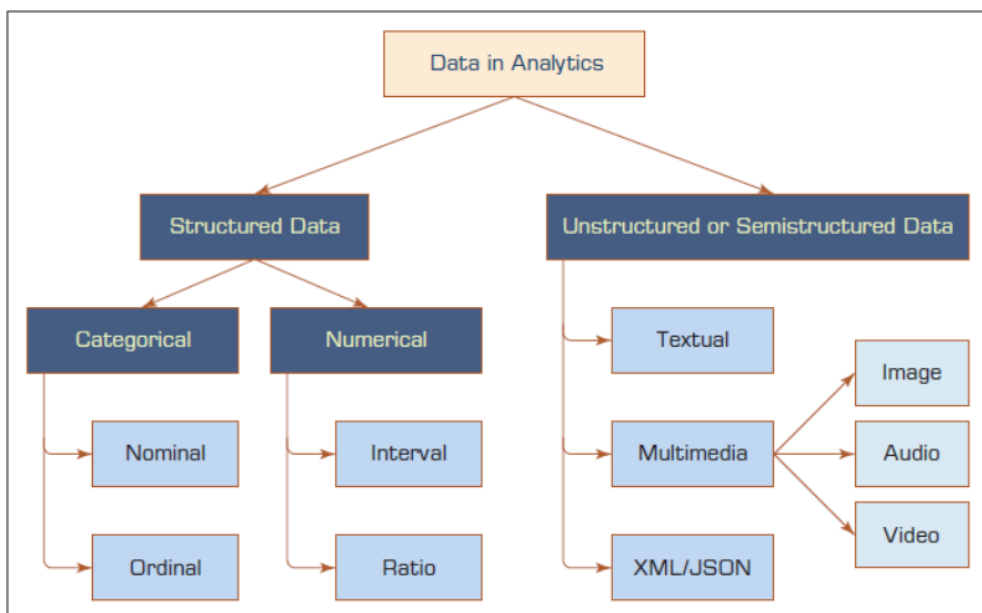
- Targeted for computers to process
- Numeric versus nominal

### Unstructured/textual data

- Targeted for humans to process/digest

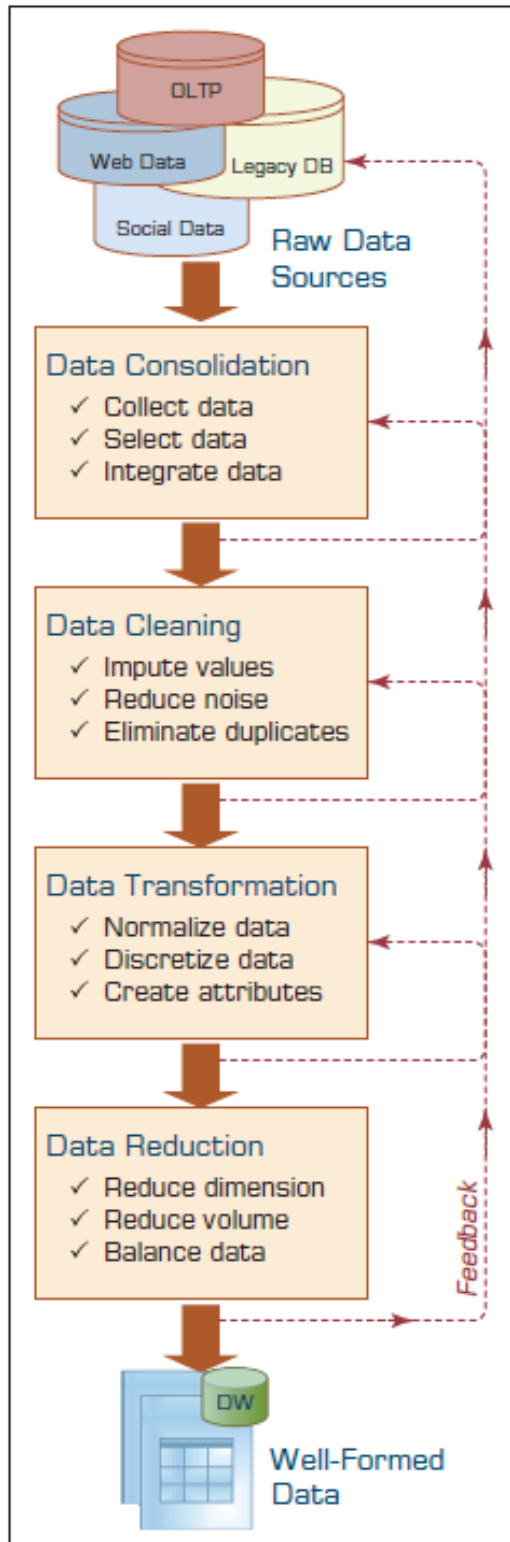
### Semi-structured data

- XML, HTML, Log files, etc



## Readying the data for analytics

- Data consolidation
- Data cleaning
- Data transformation
- Data reduction



## Data Pre-processing Tasks and Methods

A Summary of Data Pre-processing Tasks and Potential Methods		
Main Task	Subtasks	Popular Methods
Data consolidation	Access and collect the data Select and filter the data Integrate and unify the data	SQL queries, software agents, Web services Domain expertise, SQL queries, statistical tests SQL queries, domain expertise, ontology-driven data mapping
Data cleaning	Handle missing values in the data	Fill in missing values (imputations) with most appropriate values (mean, median, min/ max, mode, etc.); recode the missing values with a constant such as "ML"; remove the record of the missing value; do nothing
	Identify and reduce noise in the data	Identify the outliers in data with simple statistical techniques (such as averages and standard deviations) or with cluster analysis; once identified, either remove the outliers or smooth them by using binning, regression, or simple average
	Find and eliminate erroneous data	Identify the erroneous values in data (other than outliers), such as odd values, inconsistent class labels, odd distributions; once identified, use domain

		expertise to correct the values or remove the records holding the erroneous values
Data transformation	Normalize the data	Reduce the range of values in each numerically valued variable to a standard range (e.g., 0 to 1 or -1 to +1) by using a variety of normalization or scaling techniques
	Discretize or aggregate the data	If needed, convert the numeric variables into discrete representations using range or frequency-based binning techniques; for categorical variables, reduce the number of values by applying proper concept hierarchies
	Construct new attributes	Derive new and more informative variables from the existing ones using a wide range of mathematical functions (as simple as addition and multiplication or as complex as a hybrid combination of log transformations)
Data reduction	Reduce number of attributes	Principal component analysis, independent component analysis, chi-square testing, correlation analysis, and decision tree induction
	Reduce number of records	Random sampling, stratified sampling, expert-knowledge-driven purposeful sampling
	Balance skewed data	Oversample the less represented or under sample the more represented classes