Matters of Discussion

Brief Evolution of DSA

DW-OLAP

KDD

Data mining& analytics

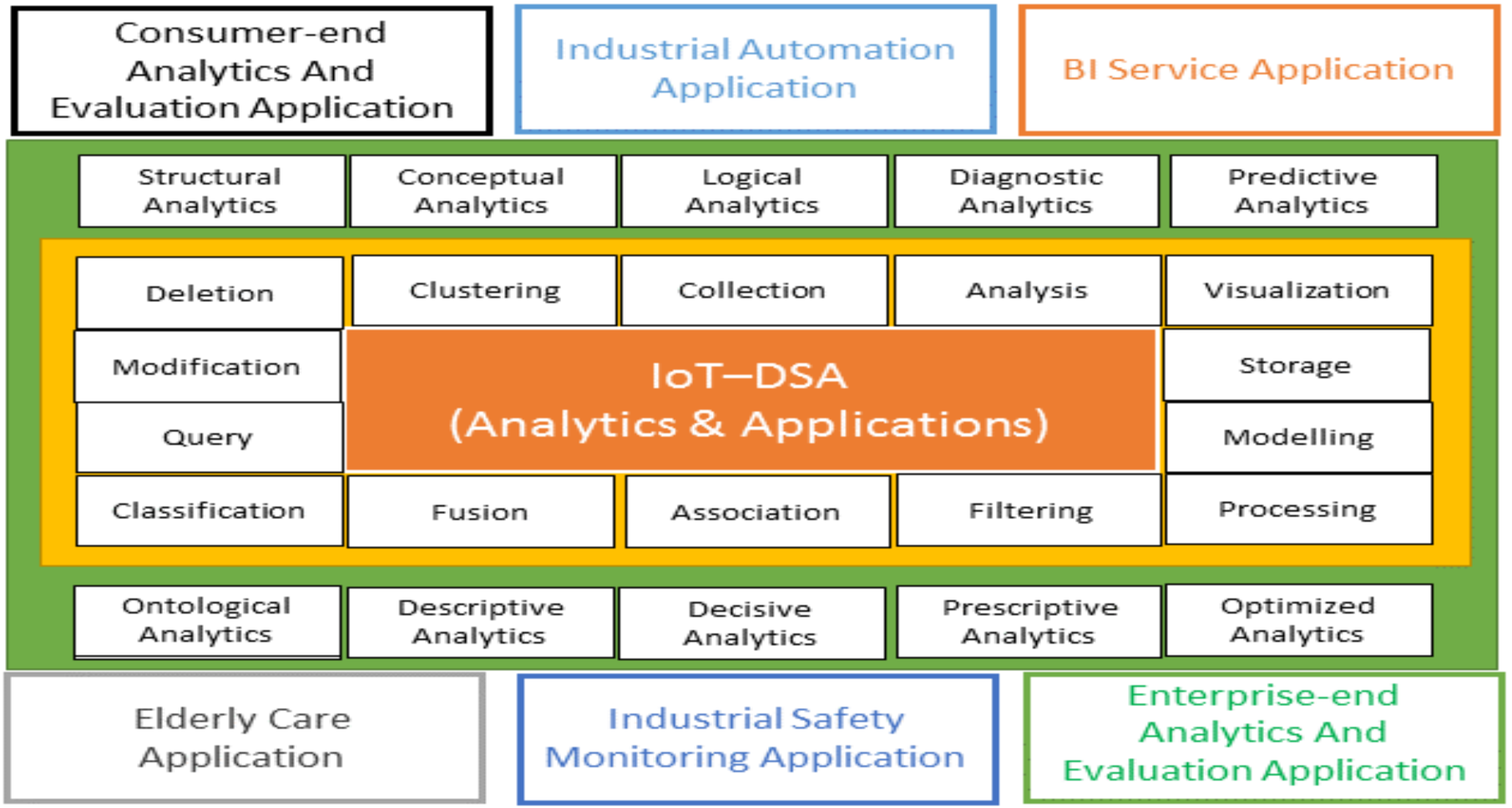
Core Tasks, Apps & Algorithms

Compiled By: Dr. Nilamadhab Mishra [(PhD- CSIE) Taiwan]

Brief Evolution of DSA

Year/Duration	Features Included in DSA
1960	Data science as a substitute of CS
1974	DS as data processing methods
1977	Exploratory data analysis
1989-1996	Data classification, mining, and knowledge discovery
1997-2001	Statistical computing, KDD
2005	Analytics and fact based decision
2010-11	Statistics & machine learning
2012 to till date	IoTA, Cognitive learning, Big data analytics.

DSA For Analytics and Applications



Information Framework for IoT-DSA.

TECHNOLOGIES & SUPPORTED TOOLS/ PROCESSES/ FRAMEOWRKS/ TASKS/ APPLICATIONS

DW-OLAP

- **Like SQL in DBMS**
- OLAP is the dynamic synthesis, and analysis of large volumes of multi-dimensional data.
- OLAP uses multi-dimensional view of aggregate data to make forecasting.
- **OLAP finds- what is happening?**

Multi-dimensional data

City	Time	Total Revenue
Glasgow	Q1	29726
Glasgow	Q2	30443
Glasgow	Q3	30582
Glasgow	Q4	31390
London	Q1	43555
London	Q2	48244
London	Q3	56222
London	Q4	45632
Aberdeen	Q1	53210
Aberdeen	Q2	34567
Aberdeen	Q3	45677
Aberdeen	Q4	50056
.....	.....	.....
.....	.....	.....

(a)

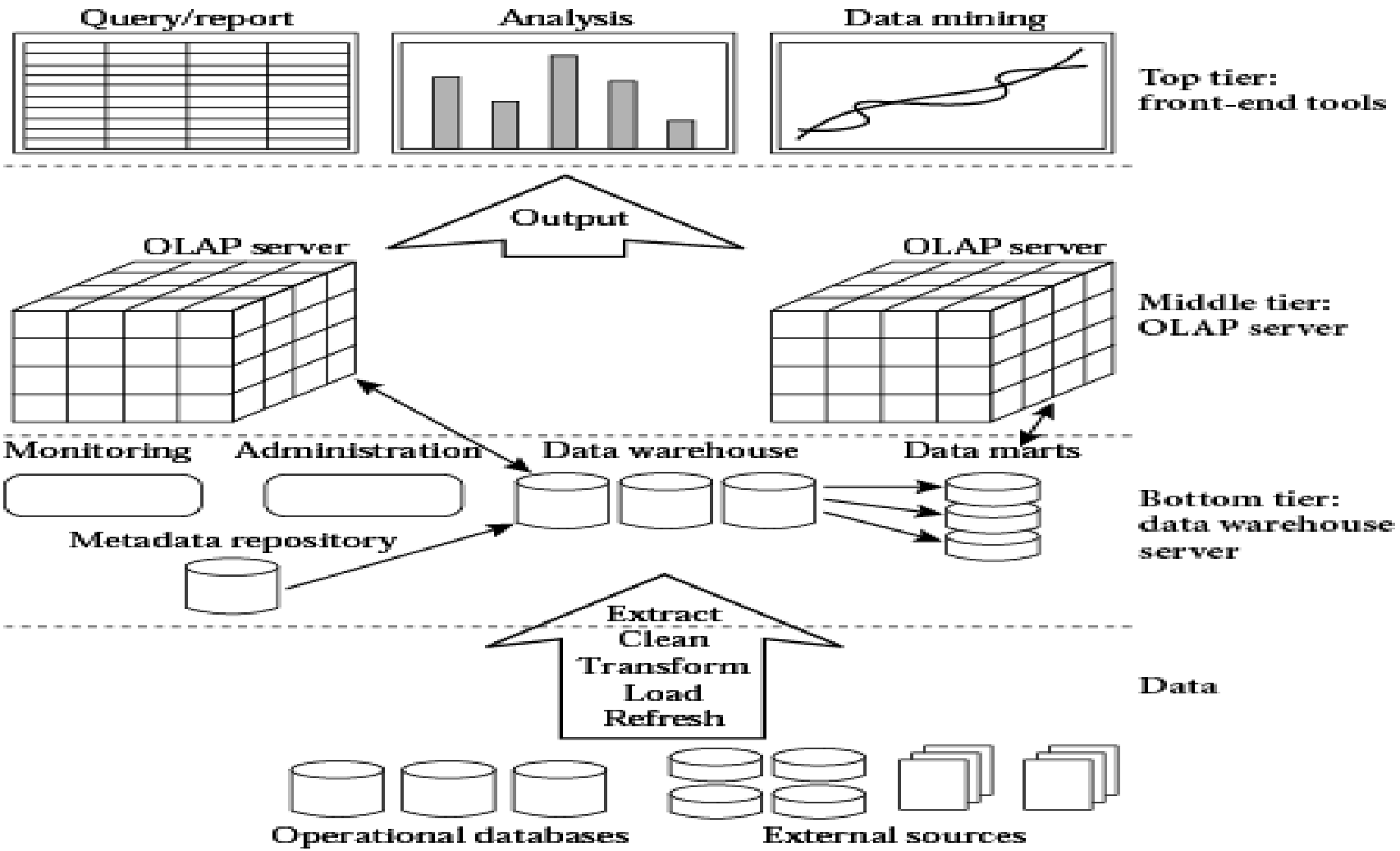
City

Time \ City	Glasgow	London	Aberdeen	.....
Q1	29726	43555	53210	.....
Q2	30443	48244	34567	.....
Q3	30582	56222	45677	.....
Q4	31390	45632	50056	.....

(b)

One-dimensional **Two-dimensional**

OLAP Architecture



OLAP Applications

1. Finance: Budgeting, activity-based costing, financial performance analysis, and financial modeling.
2. Sales: Sales analysis and sales forecasting.
3. Marketing: Market research analysis, sales forecasting, promotions analysis, customer analysis, and market/customer segmentation.
4. Manufacturing: Production planning and defect analysis.

OLAP Limitations

- Limitation:-
  - can not predict :
  - what will happen in future?
  - Why happens?
- **How to overcome this limitation--**KDD

KDD process

- knowledge discovery from database[KDD].
- KDD- find useful information or knowledge & pattern from data.
- Data mining uses algorithms to extract information & pattern derived by KDD process.

Cont..

- **ANN/ machine learning**:- transform database into a knowledge base system. **Part of data mining technique.**
- **Data mining is a part of KDD.**
- KDD process- **selection**(obtain data from source), **preprocessing**(data cleaning), **transformation**(into desired data format), **data mining**(obtain desired result), **interpretation**(present result to user meaningfully).

- Data Mining:- computational process of discovering patterns in large data sets.
- Integration of artificial intelligence, machine learning, statistics, and database systems.
- Knowledge Discovery in Databases (KDD) process:-

1. Data Selection
2. Pre-processing (attribute extraction & Normalization)
3. Transformation- transform data into desired format.
4. Data Mining-- discovering patterns.
5. Interpretation/Evaluation

Data mining Core Tasks, Apps & Algorithms

1. Classification task :- Identifying to which category an object belongs to.

Applications: e-mail Spam detection, Image recognition.

Algorithms: SVM, nearest neighbors, random forest.

2. Regression task :- Predicting a attribute value associated with an object.

Applications: Drug response, Stock prices.

Algorithms: SVR, ridge regression.

3. Clustering task:- Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes

Algorithms: k-Means, spectral clustering.

4. Dimensionality reduction task:- How to choose a good set of attributes.

Applications: Visualization, Increased efficiency

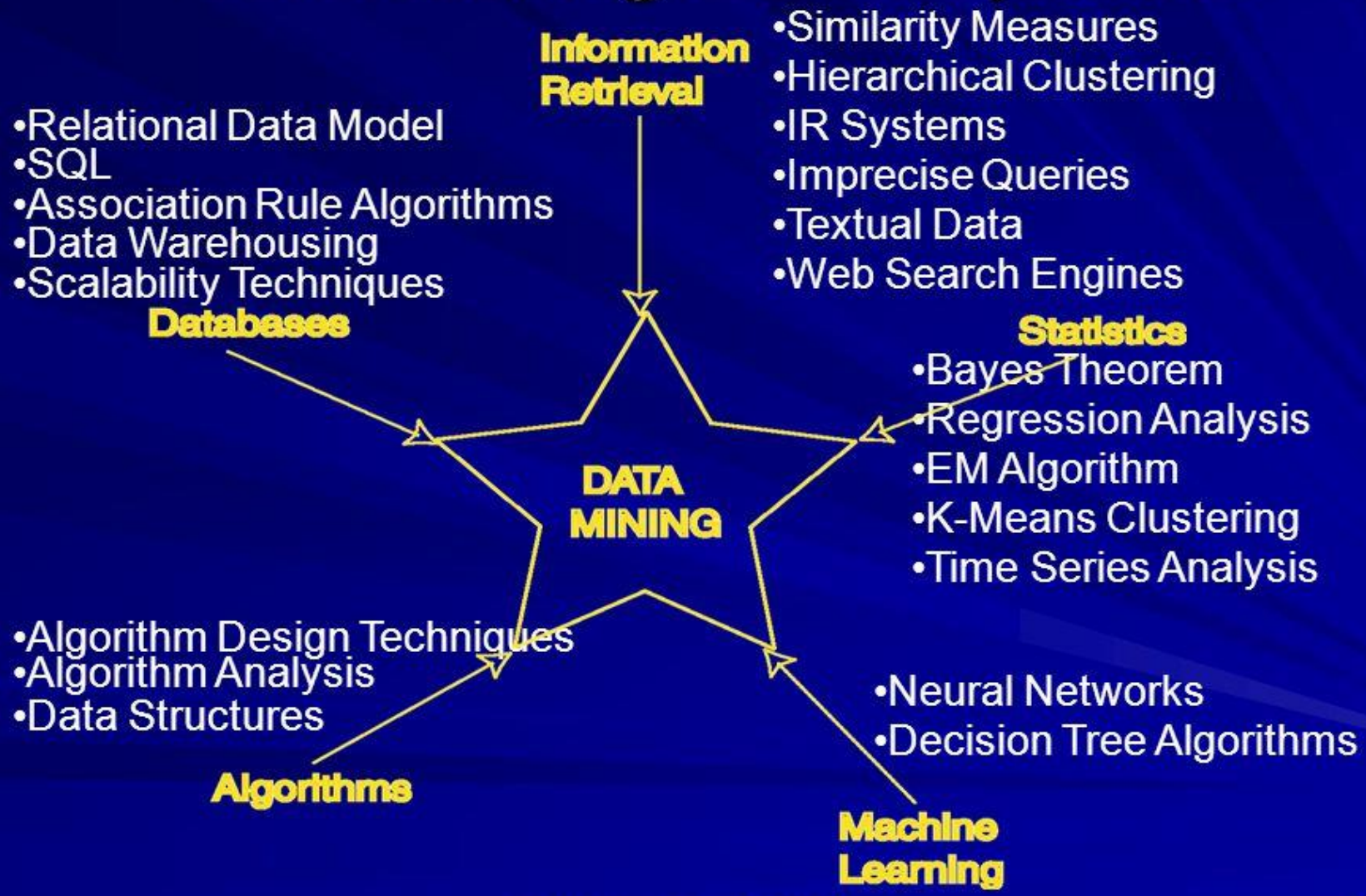
Algorithms: PCA, feature selection.

5. **Anomaly detection** (Outlier/change/deviation detection) – The identification of unusual data records, that might be interesting or data errors that require further investigation.

6. **Association rule learning** (Dependency modeling) – Searches for relationships between variables.

For example, a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis.

Data Mining Development



pcai.com

- Knowledge Based Systems, AI Languages, Neural Networks, Machine Learning, Genetic Algorithms, Evolutionary Software, Expert Systems, Fuzzy Logic, Data Mining, Intelligent Agents, Business Rules, Case-Based Reasoning, Common Sense, Data Visualization, Inferencing, Forecasting, Pattern Matching, Speech, Rule-Based Systems, Text Mining, Vision, Robotics.

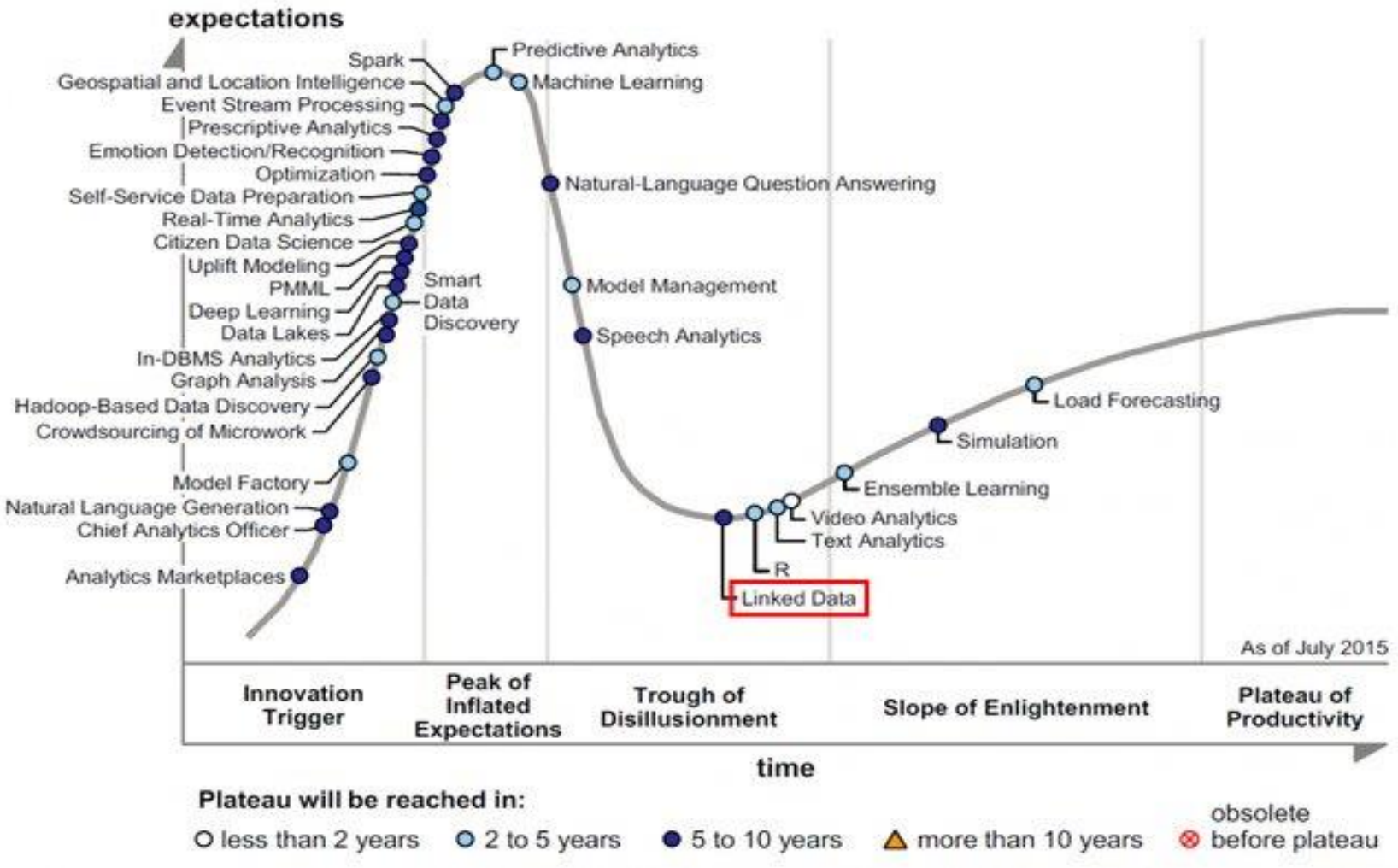
Analytic is a never ending process.

Analytic is the Major part of Data science,

Analytic is a never ending process because of progressive technological change requirements as well as the business change requirements.

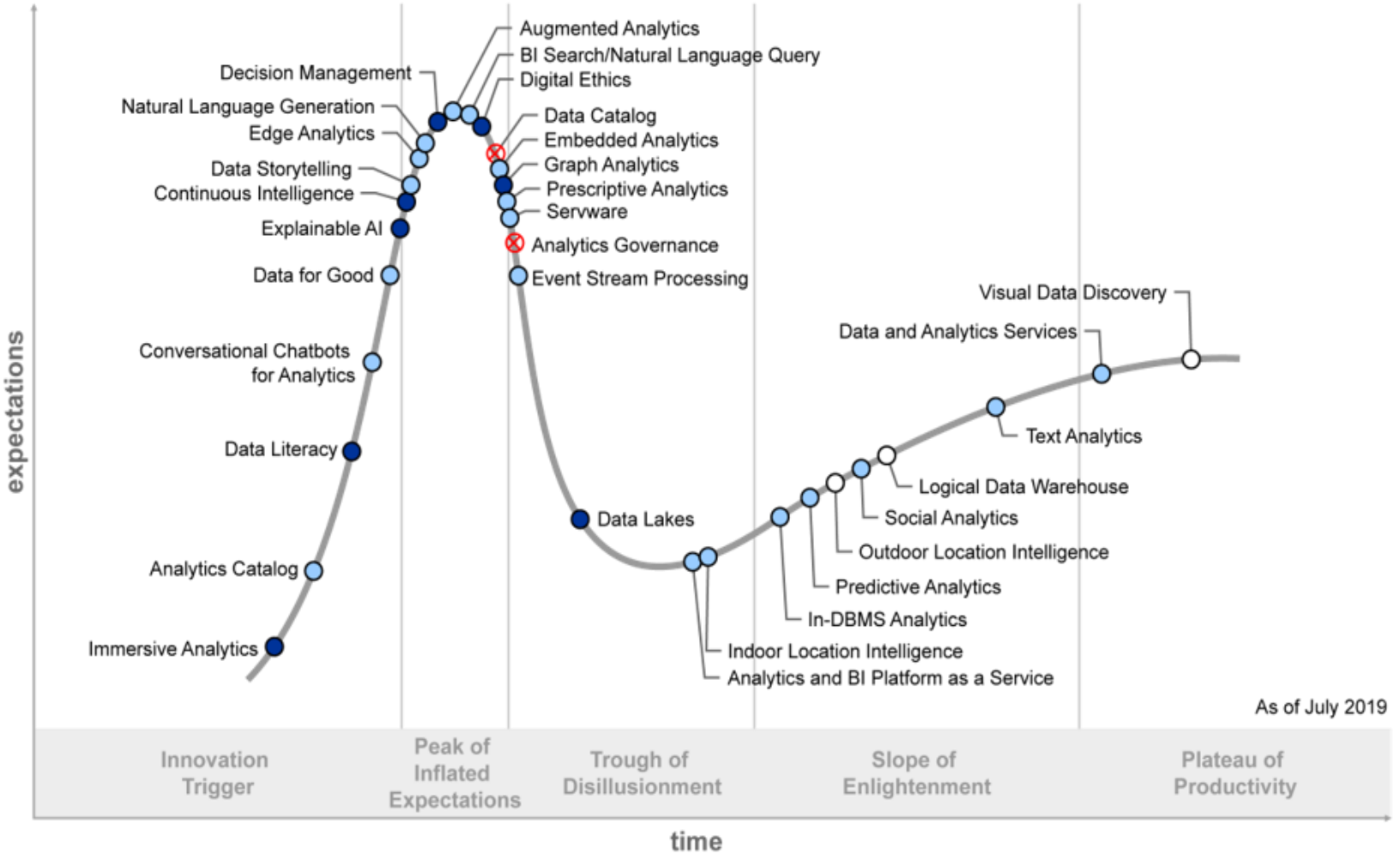
The beauty of Analytics is that two data scientist with same problem may come up with two different new solutions.

Gartner's Hype Cycle for Advanced Analytics and Data Science -2015



Compiled By: Dr. Nilamadhab Mishra [(PhD- CSIE) Taiwan]

Hype Cycle for Analytics and Business Intelligence, 2019



Source: Gartner

Compiled By: Dr. Nilamadhab Mishra [(PhD- CSIE) Taiwan]

Time to explore[Activity-01]

Investigate the numerous Data mining and analytic Core Tasks, Applications & Algorithms and prepare your investigation report.

RBT – Revised Bloom's Taxonomy
KL1 – Remember,
KL2-Understand,
KL3-Apply,
KL4-Analyse,
KL5-Evaluate,
KL6-Create
CO – Course Outcome

Cheers For the Great Patience!

Query Please?

Compiled By: Dr. Nilamadhab Mishra [(PhD- CSIE) Taiwan]