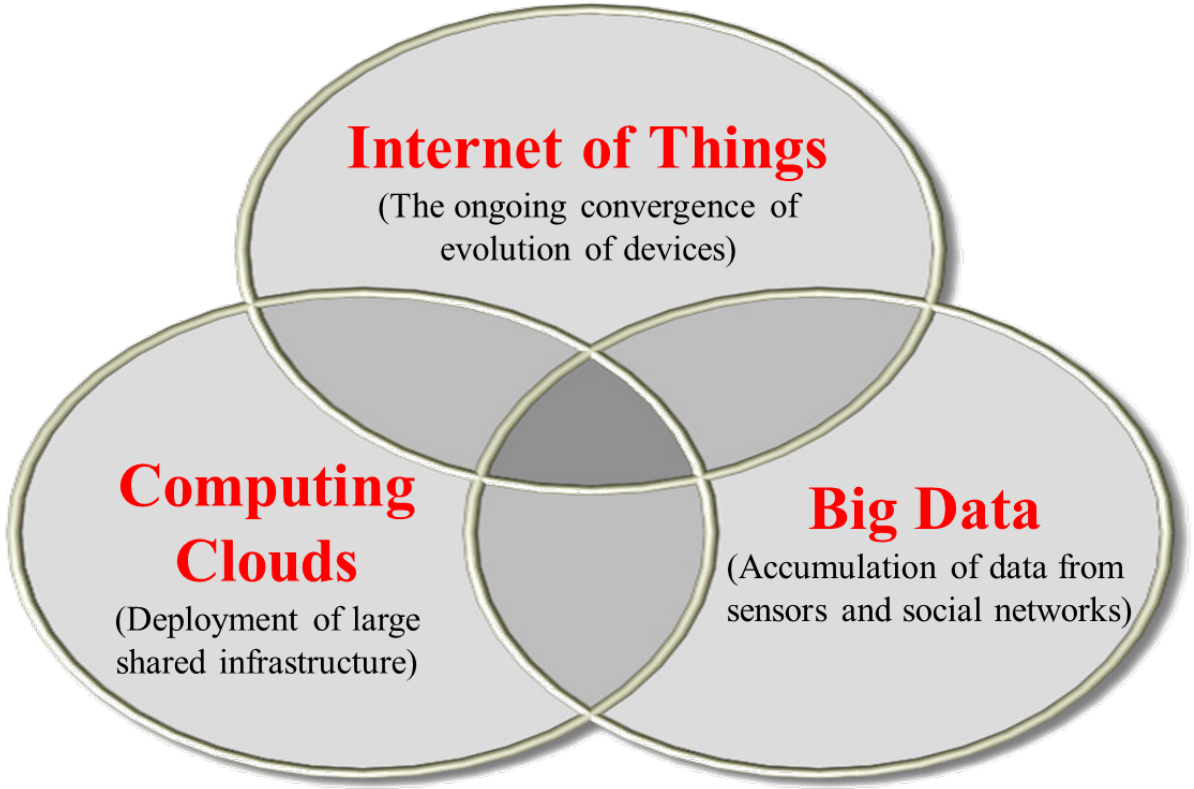
IoT Data Services and Analytics

Software Architecture and Scalability for Internet of Things

Dr Jonathan Kua

IoT, Cloud Computing and Big Data are tightly linked!



For Data, these means using the best new database technologies

NOSQL

- NoSQL is an approach to database design that can accommodate a wide variety of data models, including key-value, document, columnar and graph formats.

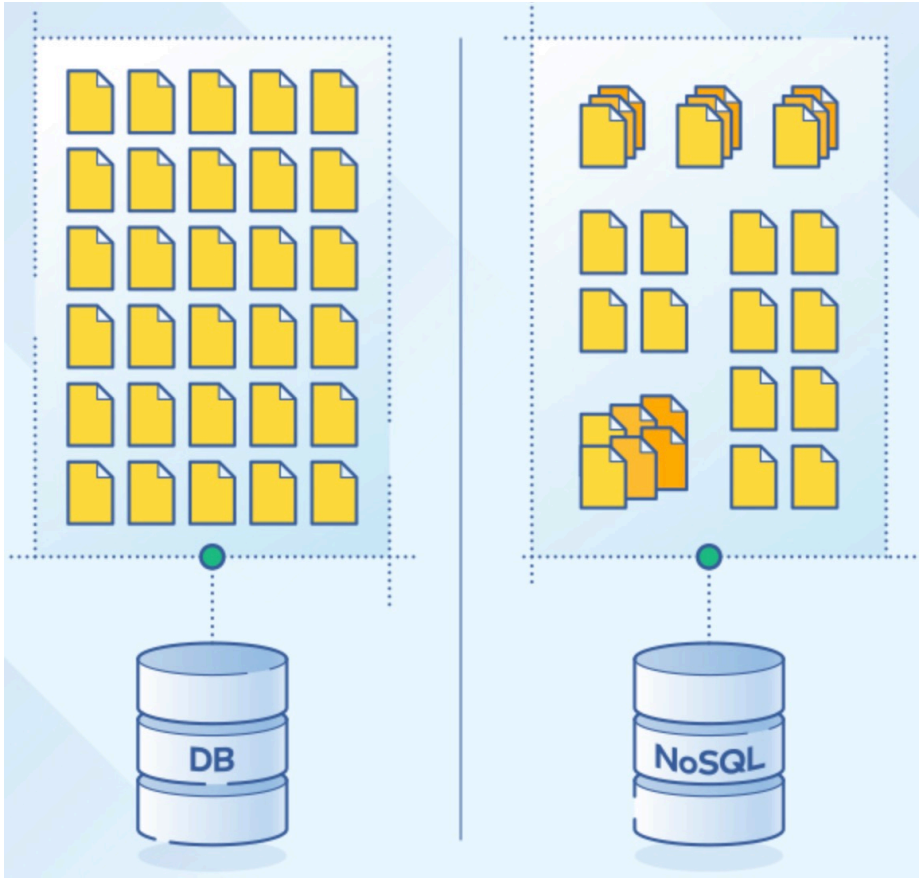
- NoSQL, which stand for "not only SQL," is an alternative to traditional relational databases in which data is placed in tables and data schema is carefully designed before the database is built. NoSQL databases are especially useful for working with large sets of distributed data.

https://searchdatamanagement.techtarget.com/definition/NoSQL-Not-Only-SQL

NOSQL

- Unlike relational databases, NoSQL databases are based on key-value pairs.

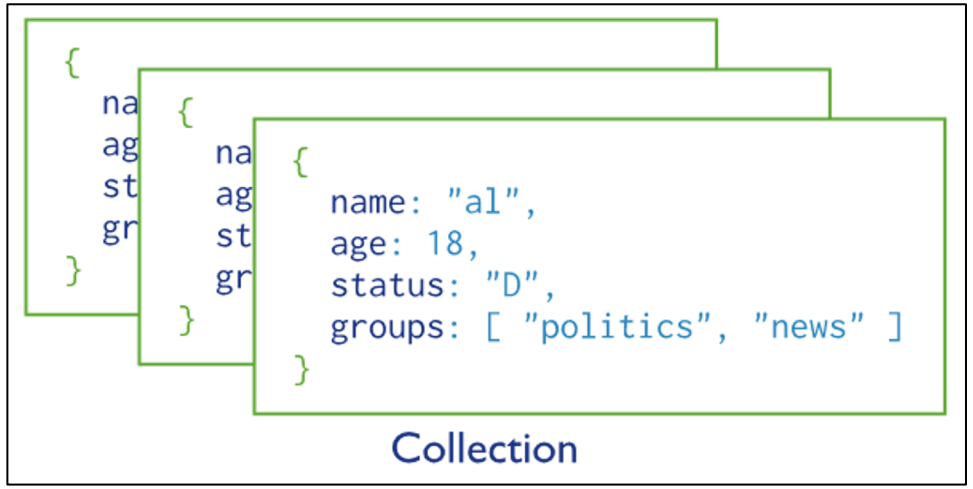
- Some store types of NoSQL databases include column store, document store, key value store, graph store, object store, XML store, and other data store modes.



https://docs.infoworks.io/datafoundry-2.7.2/nosql-ingestion

NoSQL: MongoDB: Document Database

- In MongoDB, a **database** is a set of **collections of documents**
- A collection is analogous to a **table** in relational databases.
- A **document** is analogous to a **record/row** in relational databases
- A document is a set of **field and value pairs - similar to JSON objects**.
- The values of fields may include other documents, arrays, and arrays of documents.




```
{
  name: "sue",                    <-- field: value
  age: 26,                        <-- field: value
  status: "A",                    <-- field: value
  groups: [ "news", "sports" ]    <-- field: value
}
```

Data Modelling

- Determine the purpose of your database
- Find and organize the information required to be recorded in the database.
- Divide your information into major entities or subjects, such as Products or Orders. Each subject then becomes a table in relational database, or documents in document Database, etc).
- Turn information on each entity to columns in relational database, or field/value in document Database, etc).

Data Modelling - Relationships

- Once we structure the data in tables or documents. We find relationships between these entities. In case of relational database, these are modelled using primary key & foreign keys.
- In document databases, our target is to consolidate all the data we need and store about entities in one collection.
- We manage relationship as follows:
  a. One-to-one relationships are modelled by embedding documents
  b. One-to-many relationships are modelled by either embedding or referencing documents

MongoDB - One-to-One relationship (embedding)

User has one contact and belongs to one group

```
{
  _id: <ObjectId1>,
  username: "123xyz",
  contact: {
              phone: "123-456-7890",
              email: "xyz@example.com"
            },
  access: {
              level: 5,
              group: "dev"
            }
}
```

Embedded sub-document

Embedded sub-document

MongoDB - One-to-Many relationship (embedding)

User has multiple addresses - note that the address does not mean anything by itself - not an entity

```
{
   _id: "joe",
   name: "Joe Bookreader",
   addresses: [
                {
                  street: "123 Fake Street",
                  city: "Faketon",
                  state: "MA",
                  zip: "12345"
                },
                {
                  street: "1 Some Other Street",
                  city: "Boston",
                  state: "MA",
                  zip: "12345"
                }
              ]
 }
```

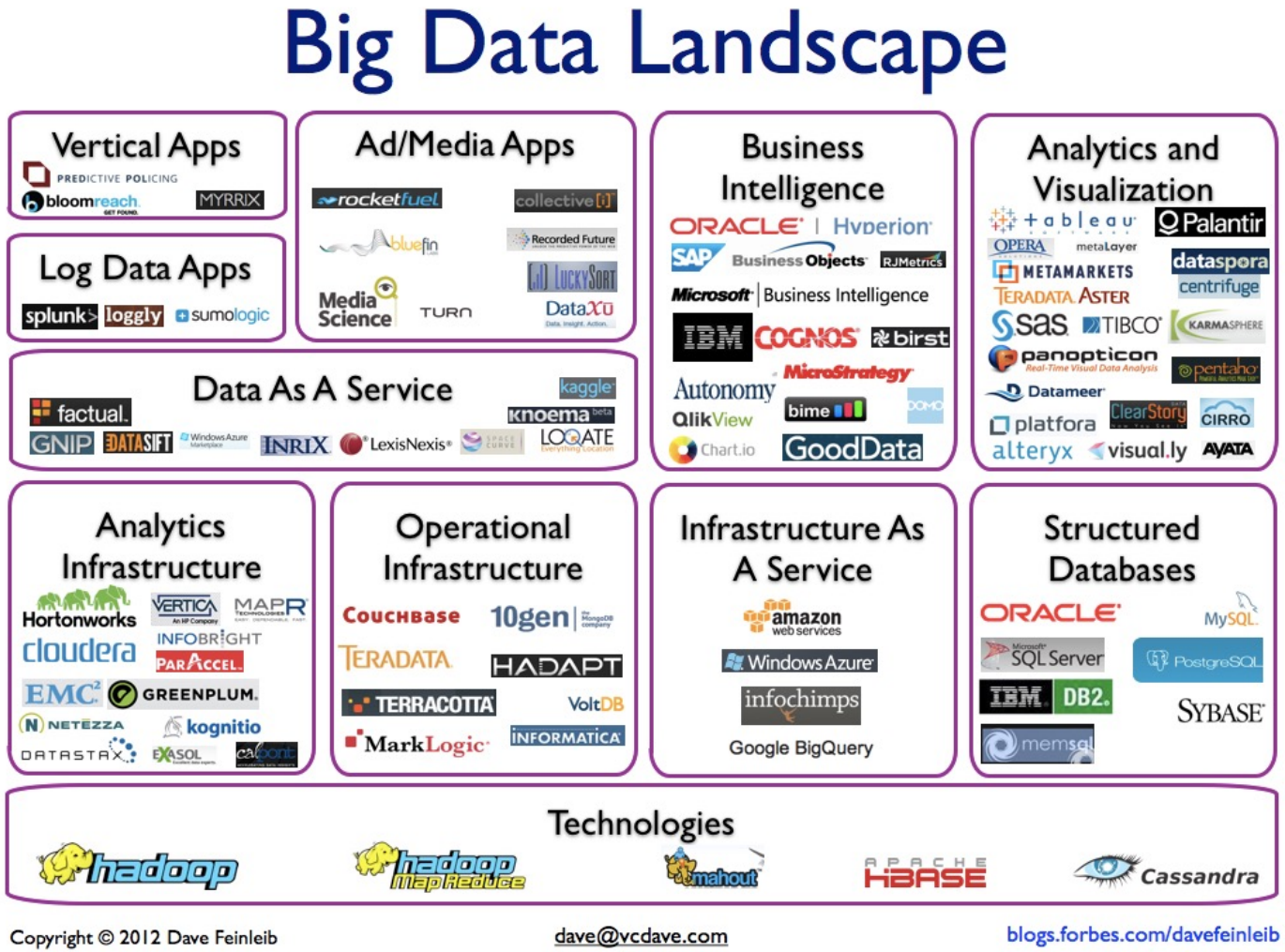
MongoDB - One-to-Many relationship (referencing)

- A publisher has multiple books - note that both need to exist as standalone entities
- Add reference to book in the publisher - you will have to modify for every new book!
- Add reference to publisher in the book

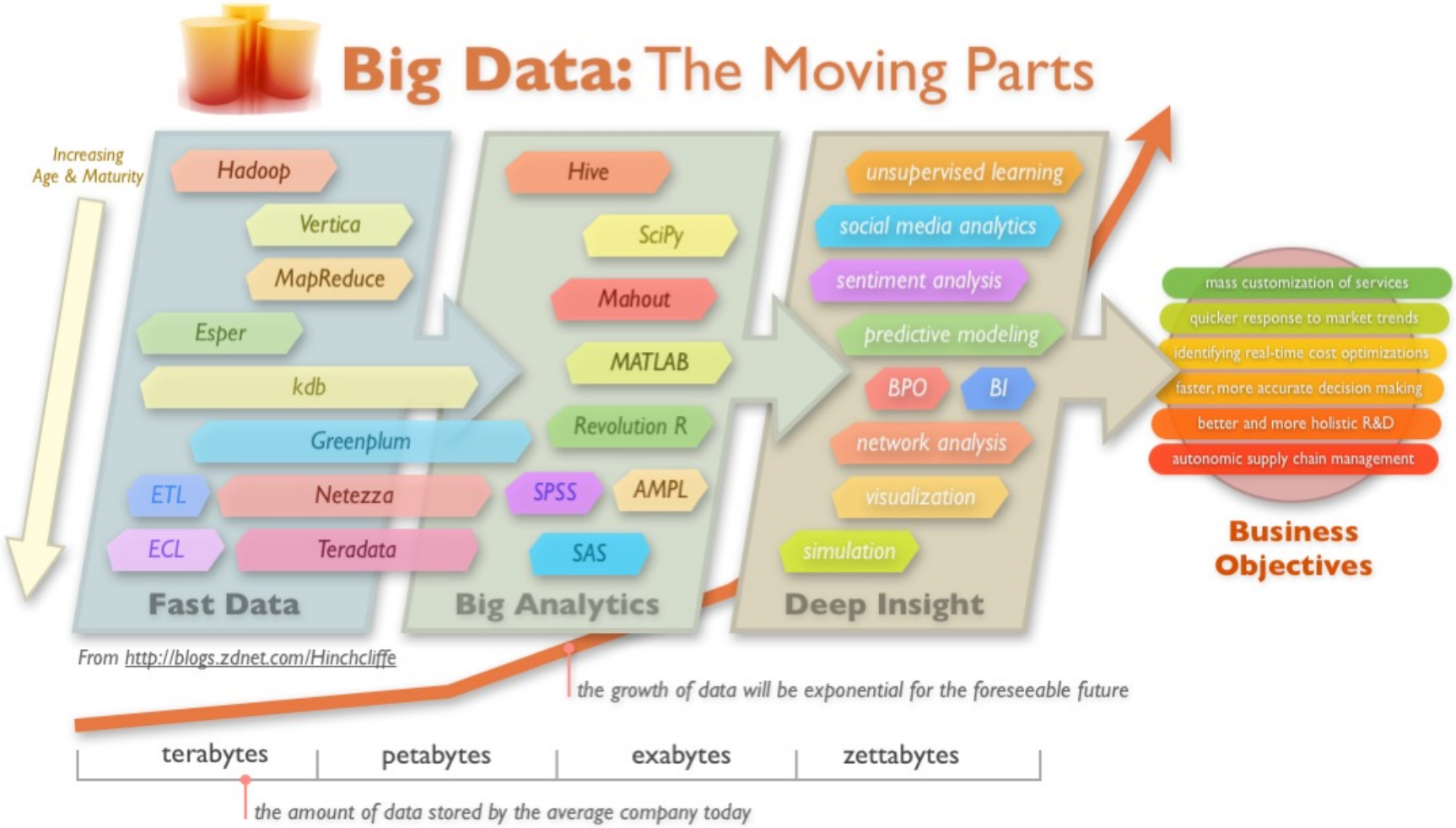
```
{
   name: "O'Reilly Media",
   founded: 1980,
   location: "CA",
   books: [123456789, 234567890, ...]
}

{
    _id: 123456789,
    title: "MongoDB: The Definitive Guide",
    author: [ "Kristina Chodorow", "Mike Dirolf" ],
    published_date: ISODate("2010-09-24"),
    pages: 216,
    language: "English"
}

{
   _id: 234567890,
   title: "50 Tips and Tricks for MongoDB Developer",
   author: "Kristina Chodorow",
   published_date: ISODate("2011-05-06"),
   pages: 68,
   language: "English"
}
```

```
{
   _id: "oreilly",
   name: "O'Reilly Media",
   founded: 1980,
   location: "CA"
}

{
   _id: 123456789,
   title: "MongoDB: The Definitive Guide",
   author: [ "Kristina Chodorow", "Mike Dirolf" ],
   published_date: ISODate("2010-09-24"),
   pages: 216,
   language: "English",
   publisher_id: "oreilly"
}

{
   _id: 234567890,
   title: "50 Tips and Tricks for MongoDB Developer",
   author: "Kristina Chodorow",
   published_date: ISODate("2011-05-06"),
   pages: 68,
   language: "English",
   publisher_id: "oreilly"
}
```

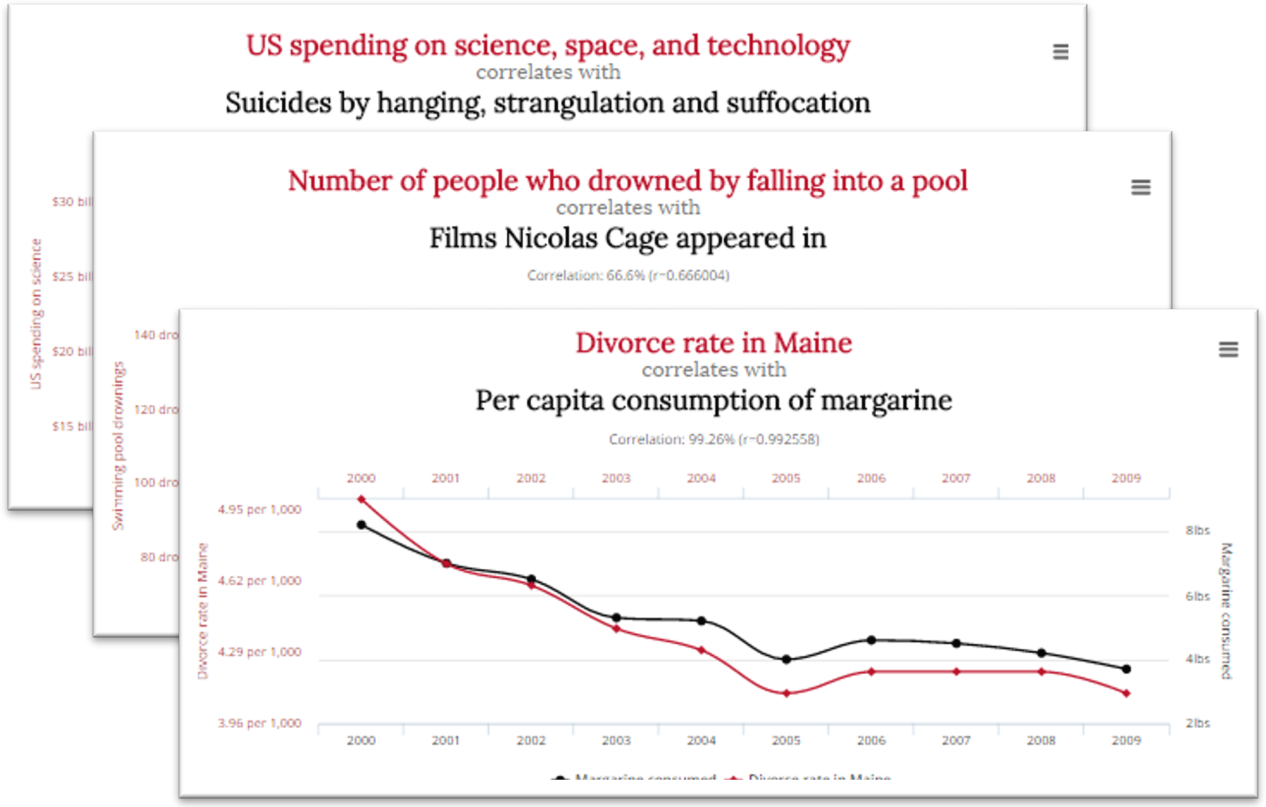
IoT Data Analytics: Big Data Landscape



Big Data Technologies



Correlations can just be coincidences!



Source: *http://www.datasciencecentral.com/profiles/blogs/spurious-correlations-15-examples*

Big Data Challenges

- **Most companies are:**
  - data rich BUT information poor
  - want (and need!) to turn their data into knowledge that informs decision making
  - Lack skill set required
- **BIG Challenges:**
  - How to identify and focus efforts on the data most relevant to the business goals
  - How to perform the most appropriate data analytics to gain useful information
  - How to interpret the value of its data analytics to gain leverage and/or competitive advantage