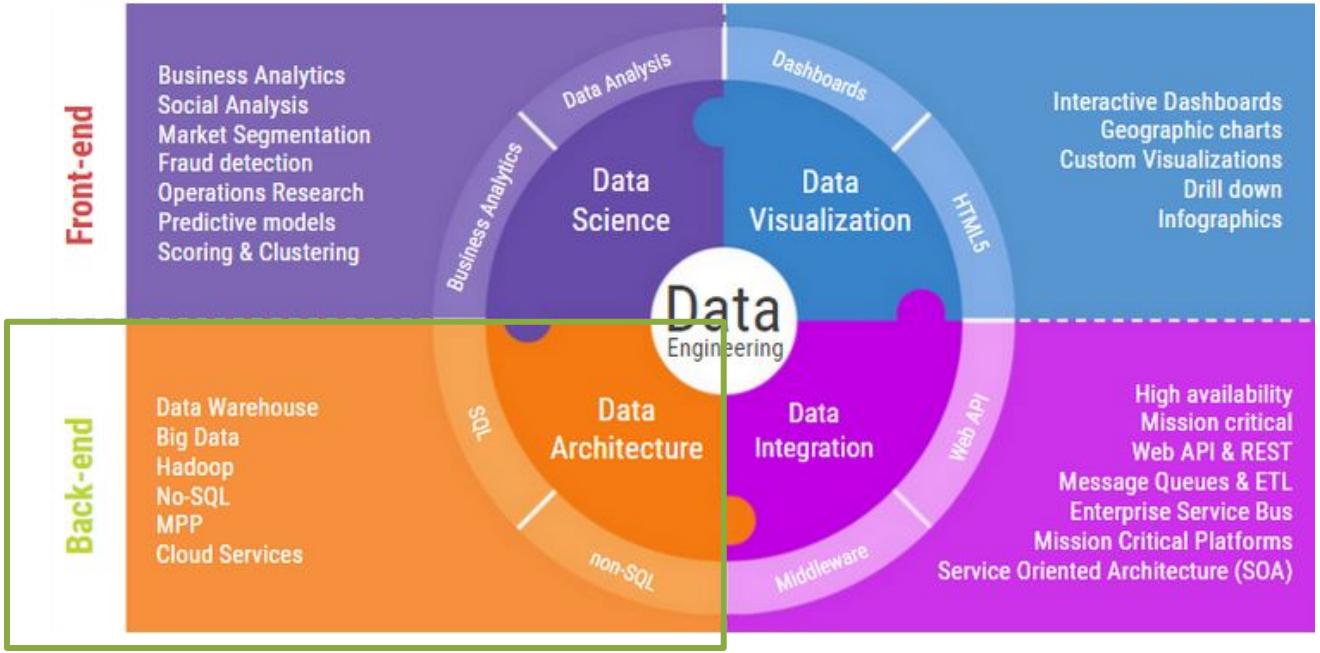
Module I:Introduction to Big Data

Globant
we are ready

Big Data - Data Engineering



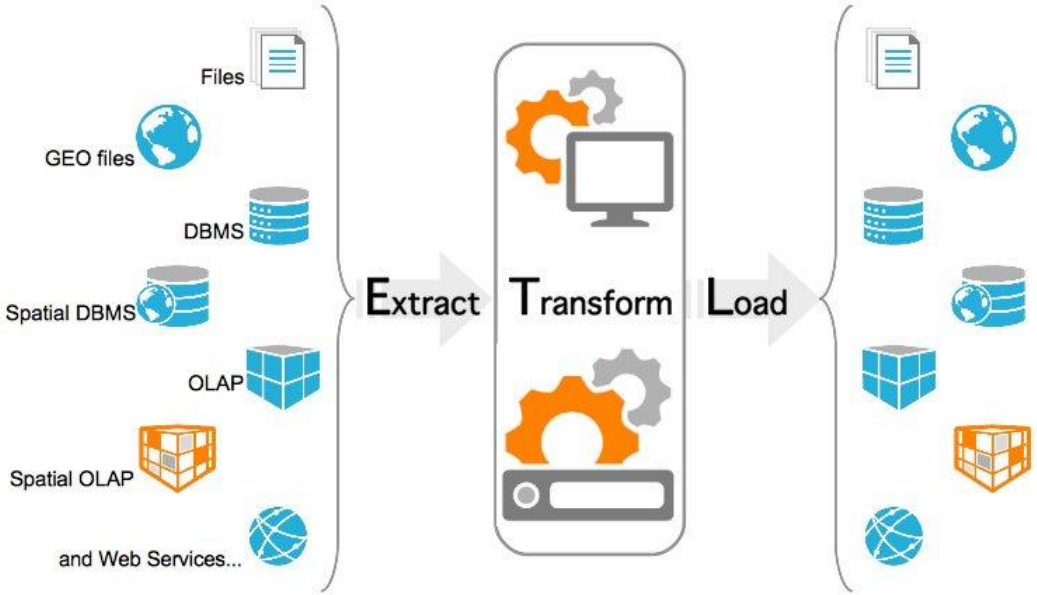
Globant
we are ready

Data Visualization.

Modern Visual Communication

Globant
we are ready

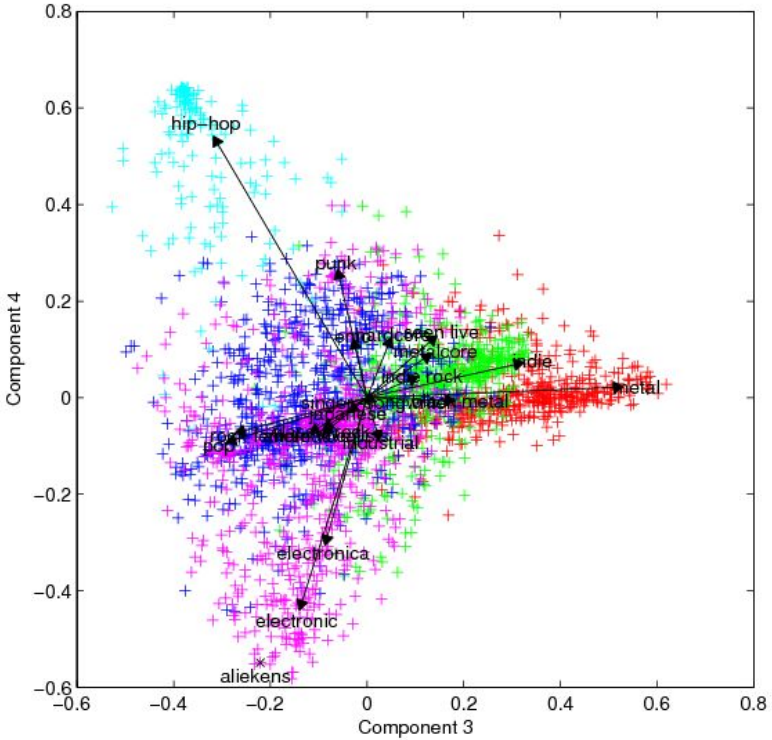
Data Integration.



The combination of technical and business processes used to combine **data** from disparate sources into meaningful and valuable information.

Globant
we are ready

Data Science



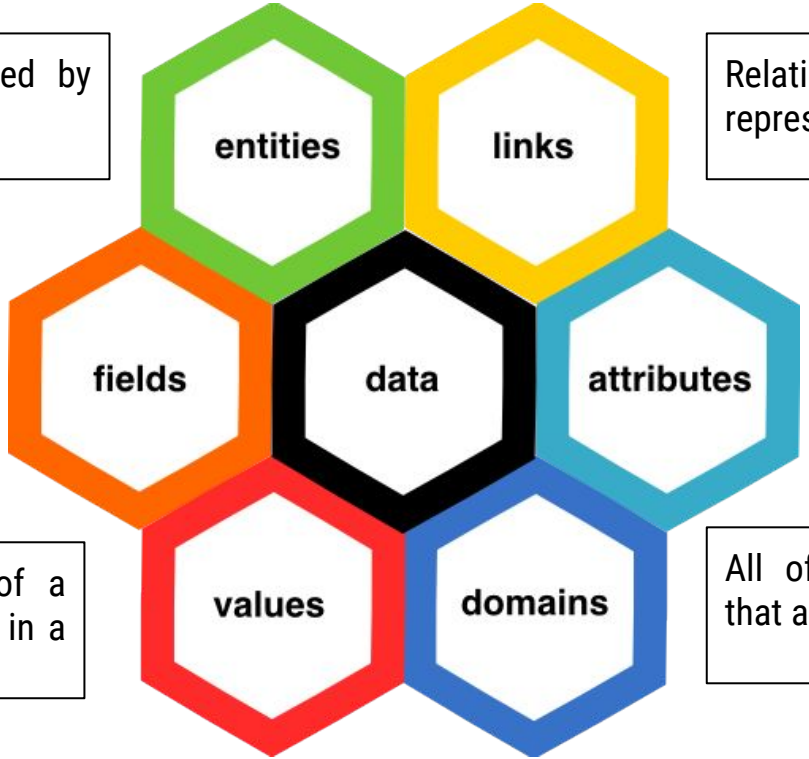
The extraction of knowledge from data

Globant
we are ready

Definitions

Data structure modeled by one or more tables.

Relationship between tables representing interaction.



Minimum division within a record defined by a table.

Particular characteristic of a given entity.

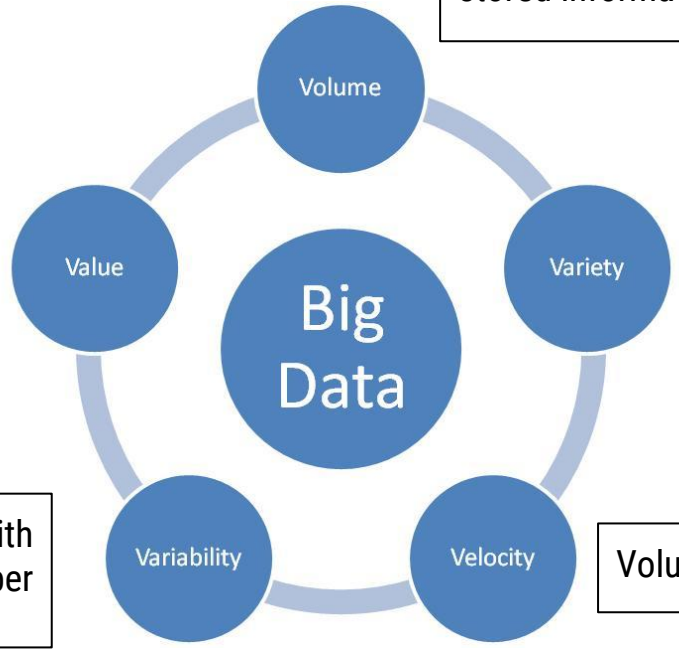
Actual representation of a characteristic of a field in a particular record.

All of the potential values that a given field could take.

Globant
we are ready

Big Data Characteristics

Accumulated size of stored information



Possibility of extracting useful information.

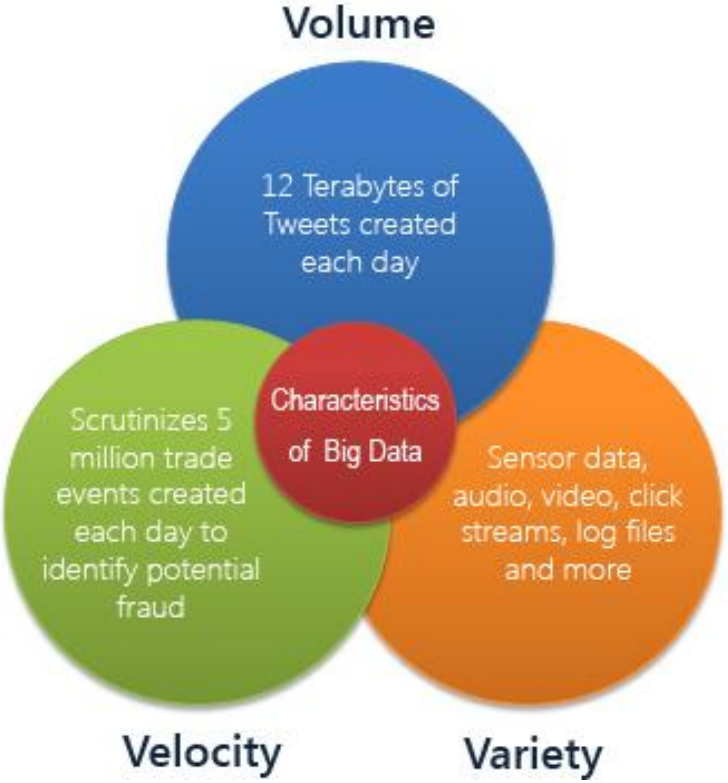
Different Data sources Generating different entities

Same entities with different attributes per row

Volume growth speed

Globant
we are ready

Big Data Characteristics



Globant
we are ready

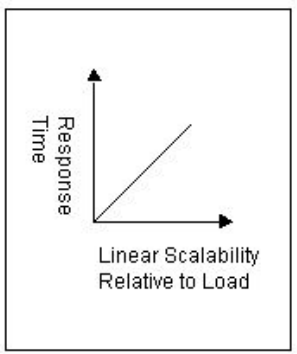
What is the Big Data practice?

It is a set of tools that provide the means to exploit datasets.

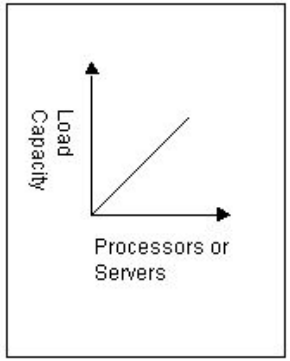
It is completely measurable.

It scales horizontally, in terms of processing, hardware requirements and costs. Linearly or sub-linearly.

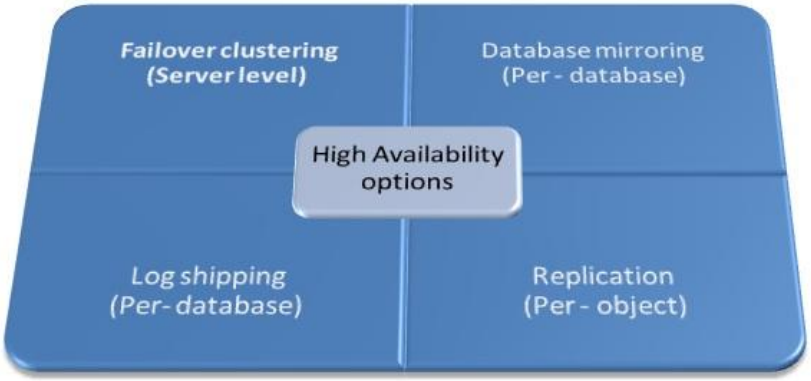
It provides persistence through high availability.



Linear Scalability Relative to Load

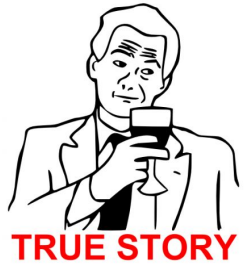


Linear Scalability Relative to System Resources (e.g. haradware)



Globant
we are ready

Big Data Is not for everyone



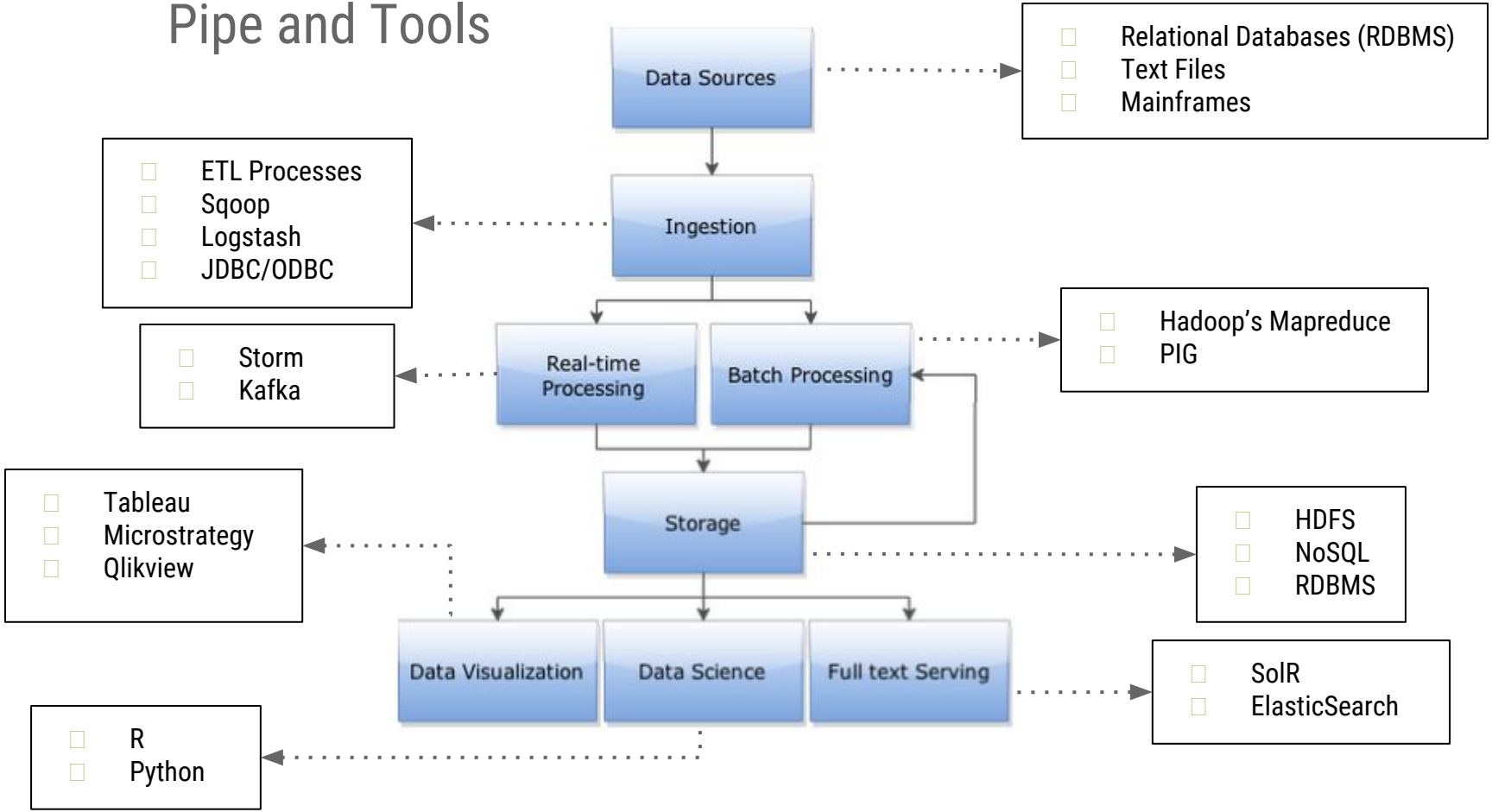
Globant
we are ready

Architecture questions

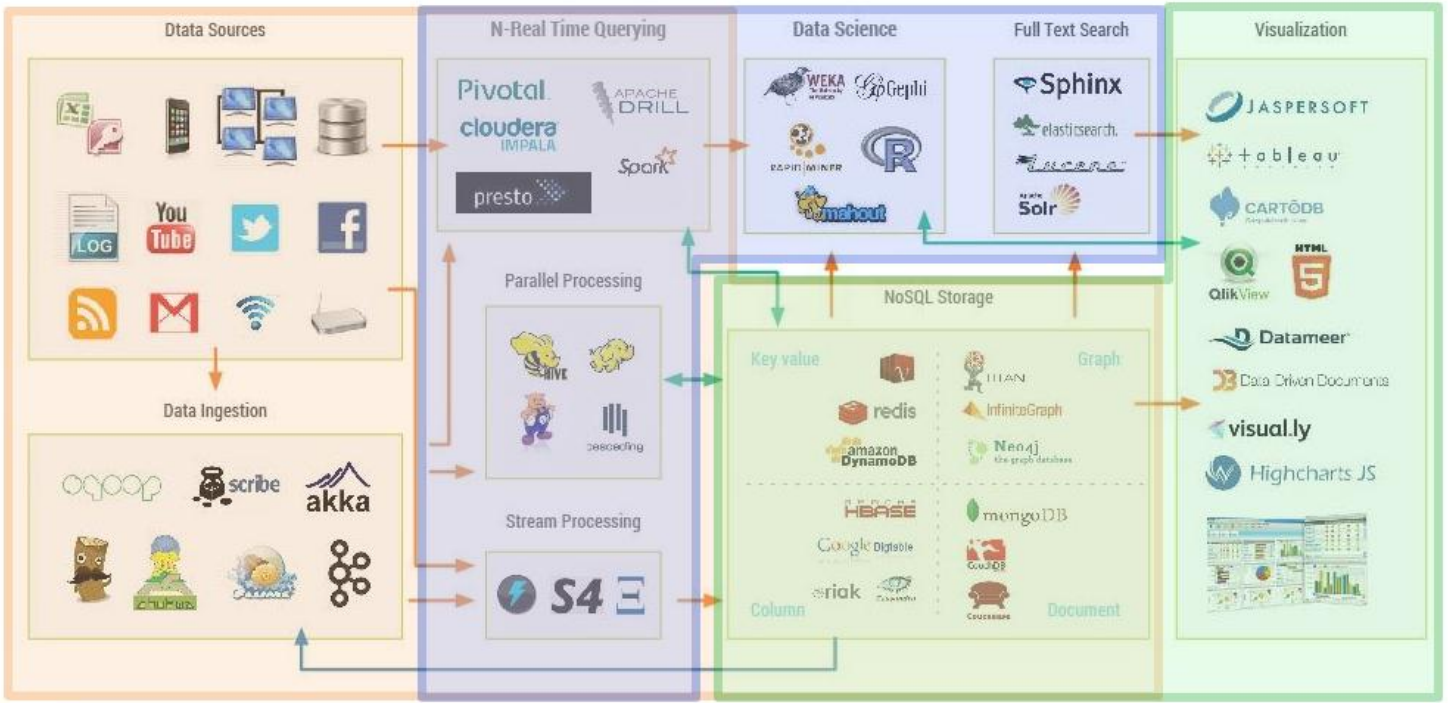
- What does the customer want to do with his data?
- Where is the Data now?
- How much data is it?
- How fast does it growth?
- How fast do we need to process the results?
- What technology does the customer already have?
- What sort of end user will consume the results and how?

Globant
we are ready

Pipe and Tools



Tools Framework



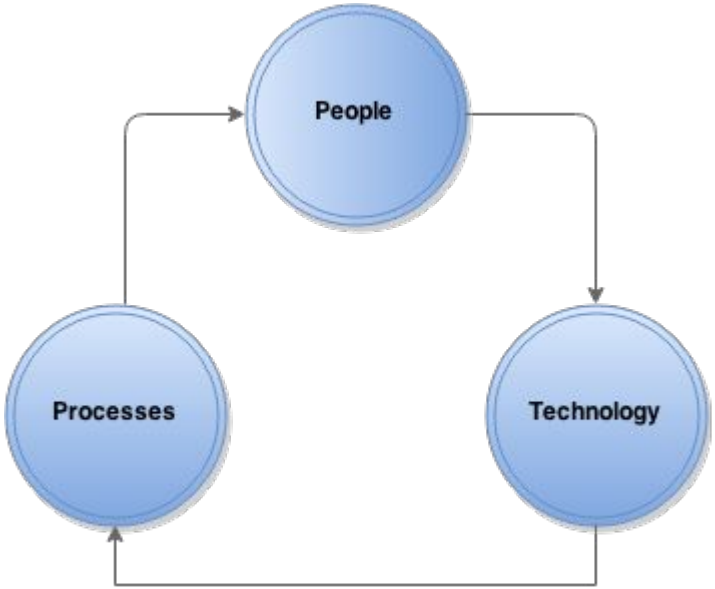
Globant
we are ready

Implementation Process

Globant
we are ready

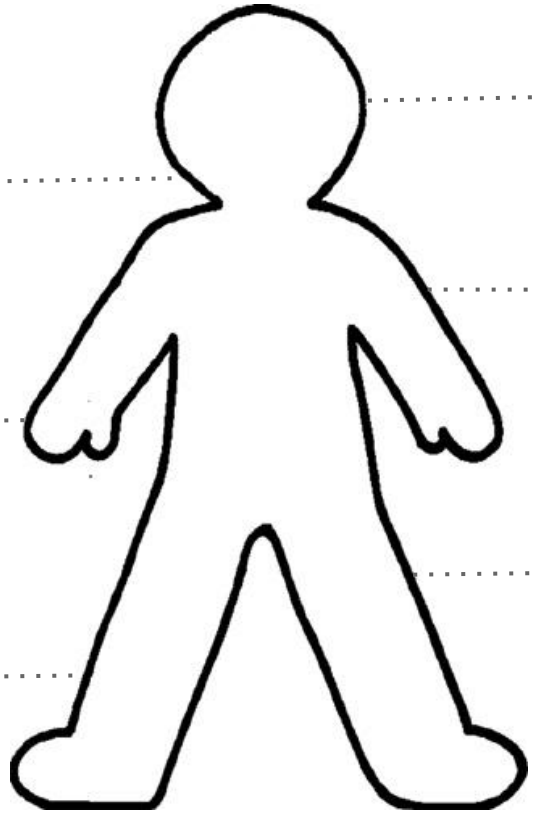
Data Excellence

- Making data consistent
- Improving data quality
- Making data accurate and complete
- Maximizing the use of data to make decisions
- Improving business planning
- Provide data access management and accountability



Globant
we are ready

Data Architecture profile



Communicative: Able to explain problems, findings and solutions.

Technologically comprehensive: Able to spend enough time reading, learning and sharing feasible documentation on new technologies

Business Aware: Able to understand the end game of the implementations and the relevant business constraints.

Analytical: Able to decompose and transform **BIG** problems into manageable solutions.

Rigorous: Able to affirm with certainty results, implementations and tests.

Production ready: Able to develop and architect with the required business Service Level Agreements (SLAs) in mind.

Globant
we are ready

This coursework

Week 1

- Big Data Introduction
- NoSQL Introduction
- Apache Hadoop Introduction and architecture
- Hadoop Installation

Week 2

- Development in Python
- MapReduce in Python
- Amazon Infrastructure I
- Amazon Infrastructure II

Week 3

- MapReduce in Java
- Apache Sqoop
- Pig Querying
- Apache Hive

Training Team

- Leandro Mora
- Renato Carelli
- Martin Cigorraga
- Ignacio Soubelet
- Juan Gaviria
- Gonzalo Zarza
- Alejandro de la Viña

Globant
we are ready