



ANÁLISIS Y VISUALIZACIÓN DE DATOS CON PYTHON

INTRODUCCIÓN AL ANÁLISIS DE DATOS

Felipe Restrepo Calle, PhD – ferestrepoca@unal.edu.co

Fabio A. González O., PhD - fagonzalezo@unal.edu.co

Universidad Nacional de Colombia

INTRODUCCIÓN

¿Qué es el análisis de datos?

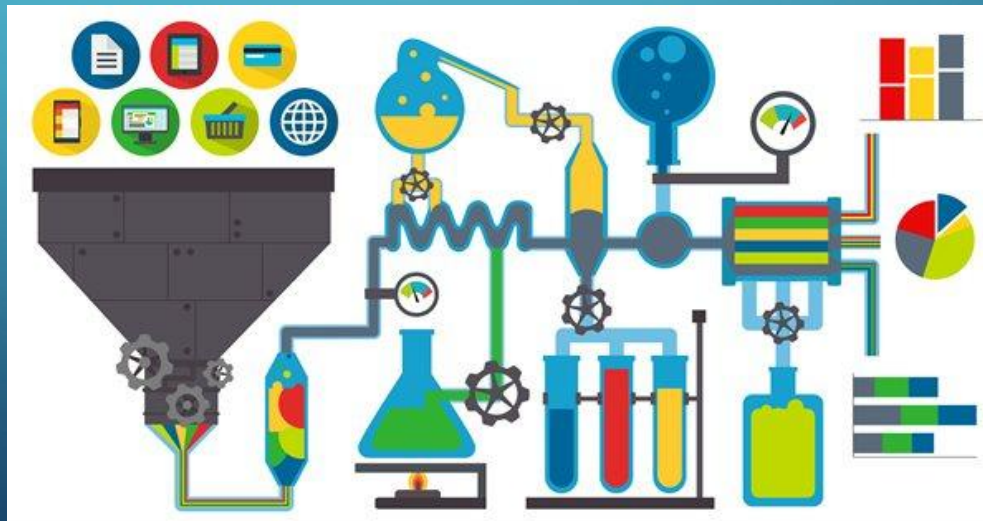
John Tukey definió el análisis de datos en 1961 así:

"Procedimientos para analizar datos, técnicas para interpretar los resultados de dichos procedimientos, formas de planear la recolecta de datos para hacer el análisis más fácil, más preciso o más exacto."

INTRODUCCIÓN

¿Qué es el análisis de datos?

- Es el proceso de: inspeccionar, limpiar y transformar datos
- Con el objetivo de resaltar y extraer información útil para apoyo en la toma de decisiones.



INTRODUCCIÓN

¿Big Data vs. Data Science?

#1. MEANING

BIGDATA



- Huge volumes of data which cannot be handled using traditional database programming
- Characterized by volume, variety, and velocity

DATA SCIENCE



- A data focused scientific activity
- Approaches to process big data
- Harnesses the potential of big data for business decisions
- Similar to data mining

INTRODUCCIÓN

¿Big Data vs. Data Science?

#2. Concept

BIGDATA



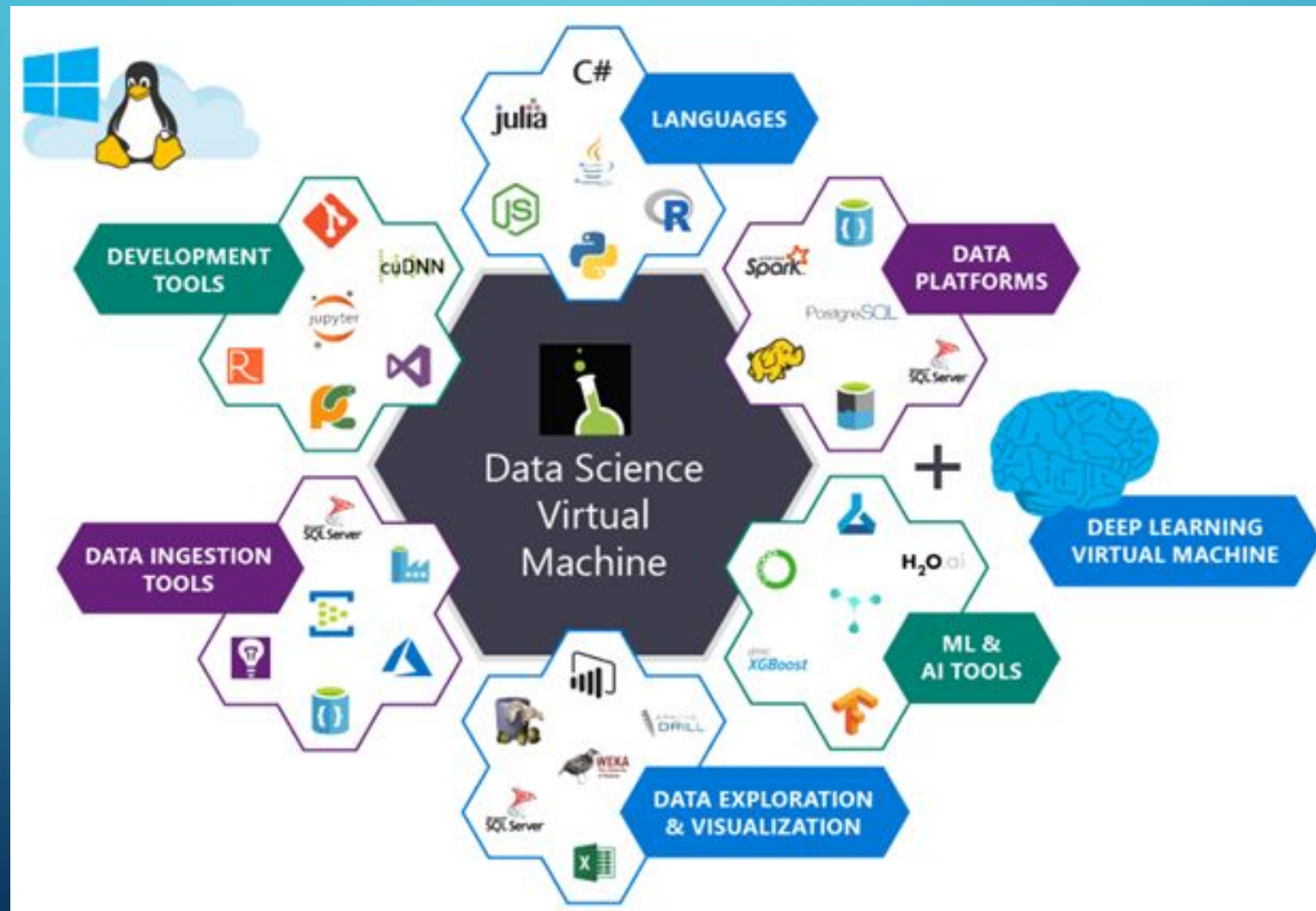
- Diverse data types generated from multiple data sources
- Includes all types and formats of data

DATA SCIENCE

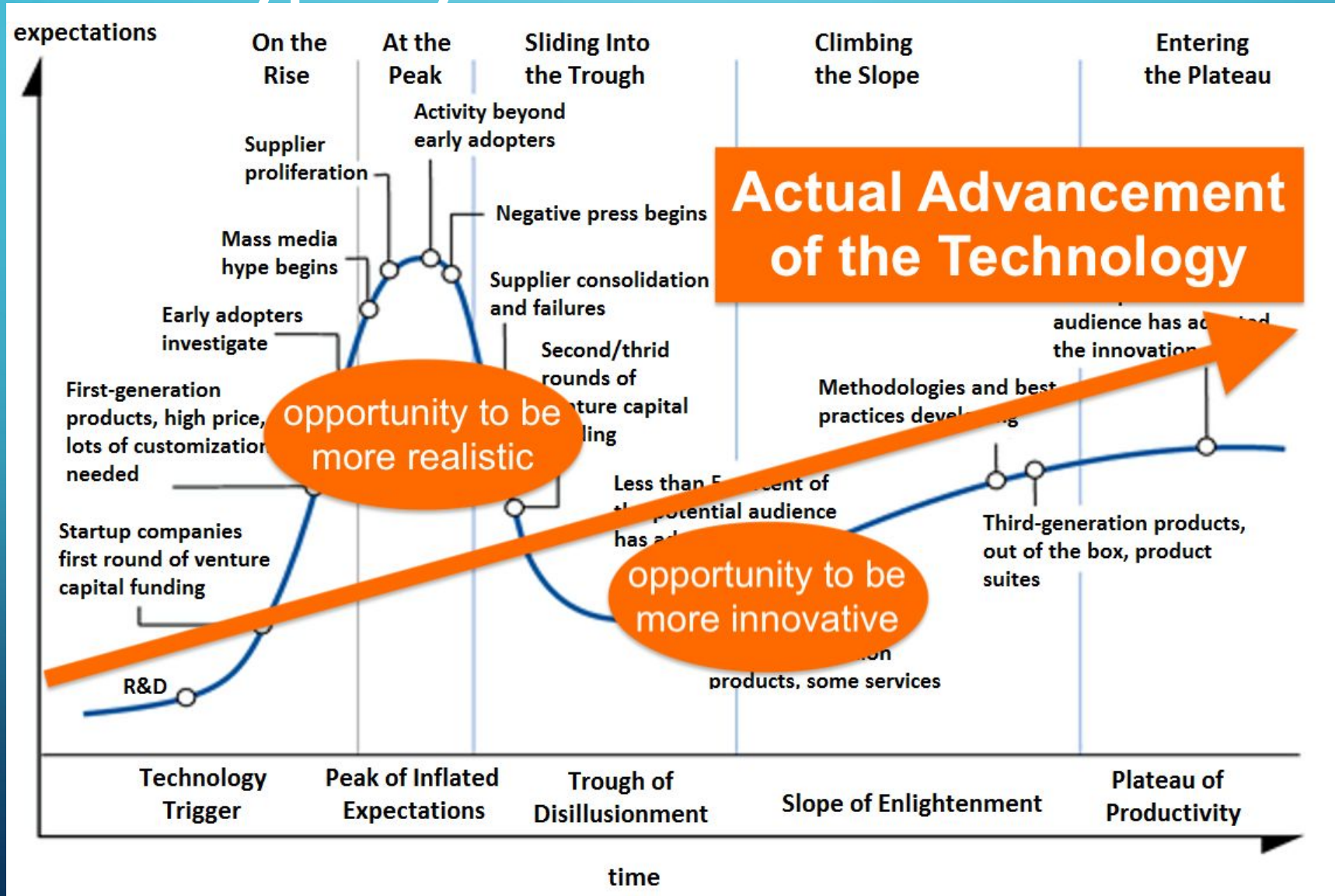


- A specialized area involving scientific programming tools, models and techniques to process big data
- Provides techniques to extract insights and information from large data sets
- Supports organizations in decision making

HERRAMIENTAS

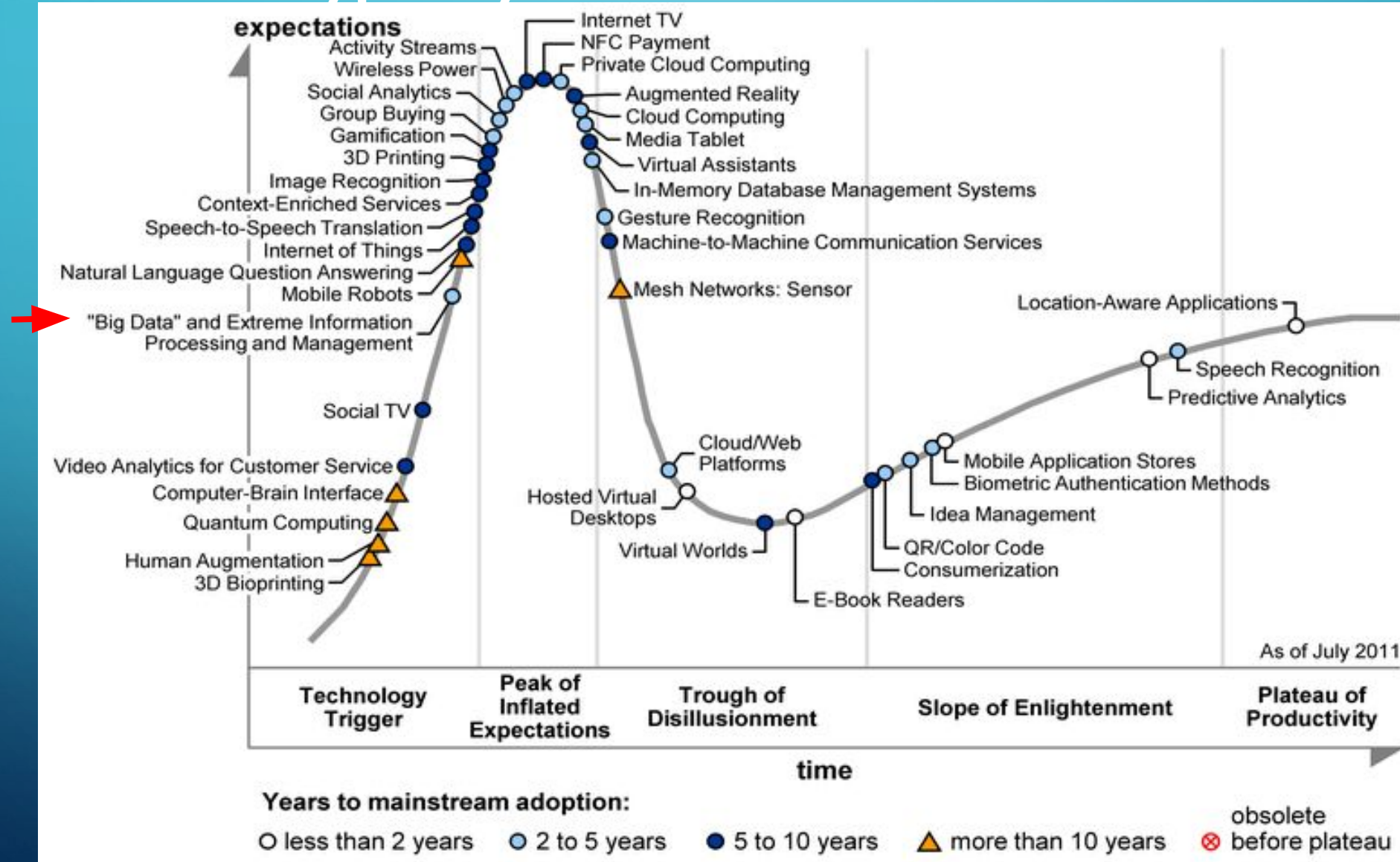


Gartner's hype cycle



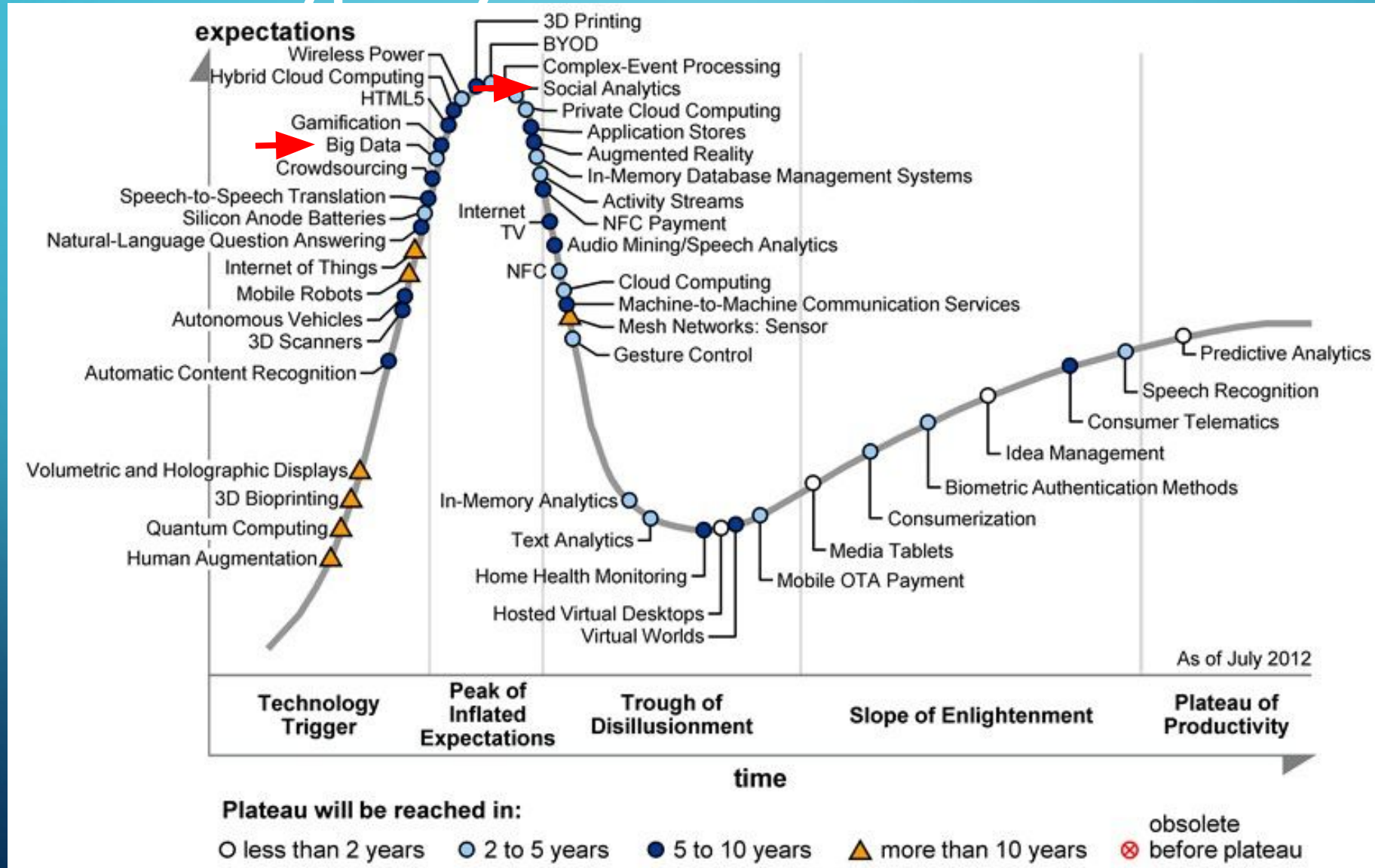
Gartner's hype cycle 2011

<https://www.gartner.com>



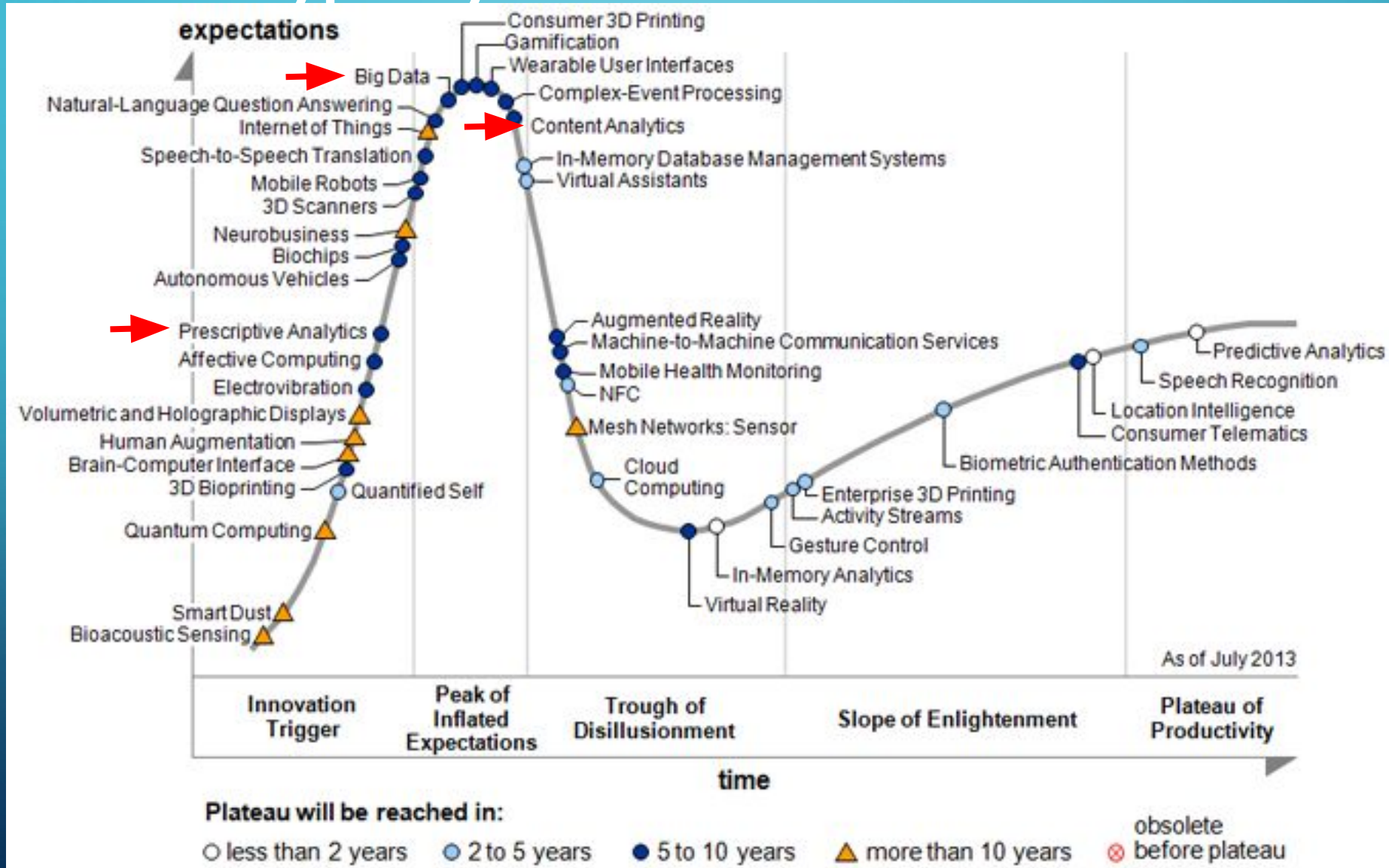
Gartner's hype cycle 2012

<https://www.gartner.com>



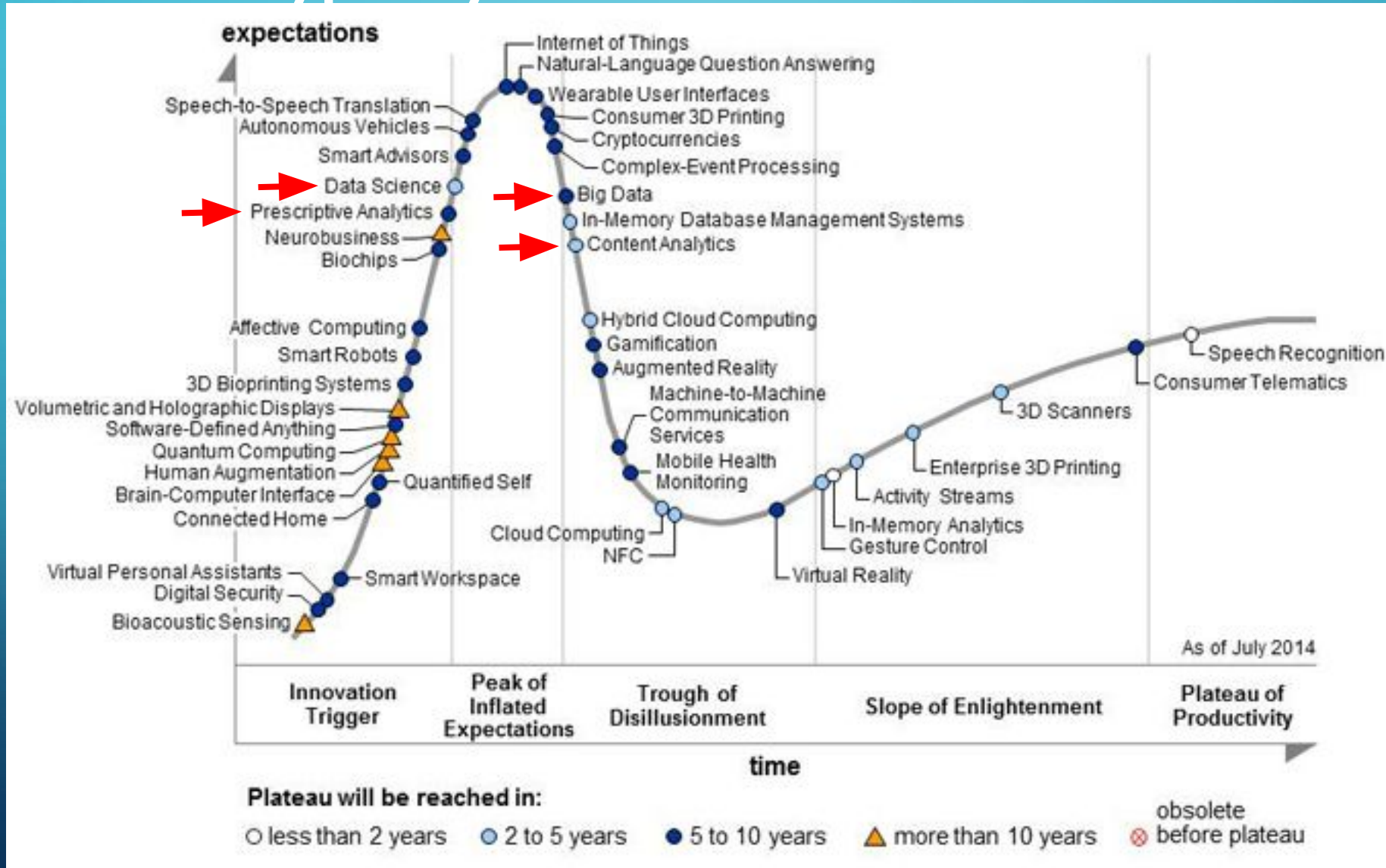
Gartner's hype cycle 2013

<https://www.gartner.com>



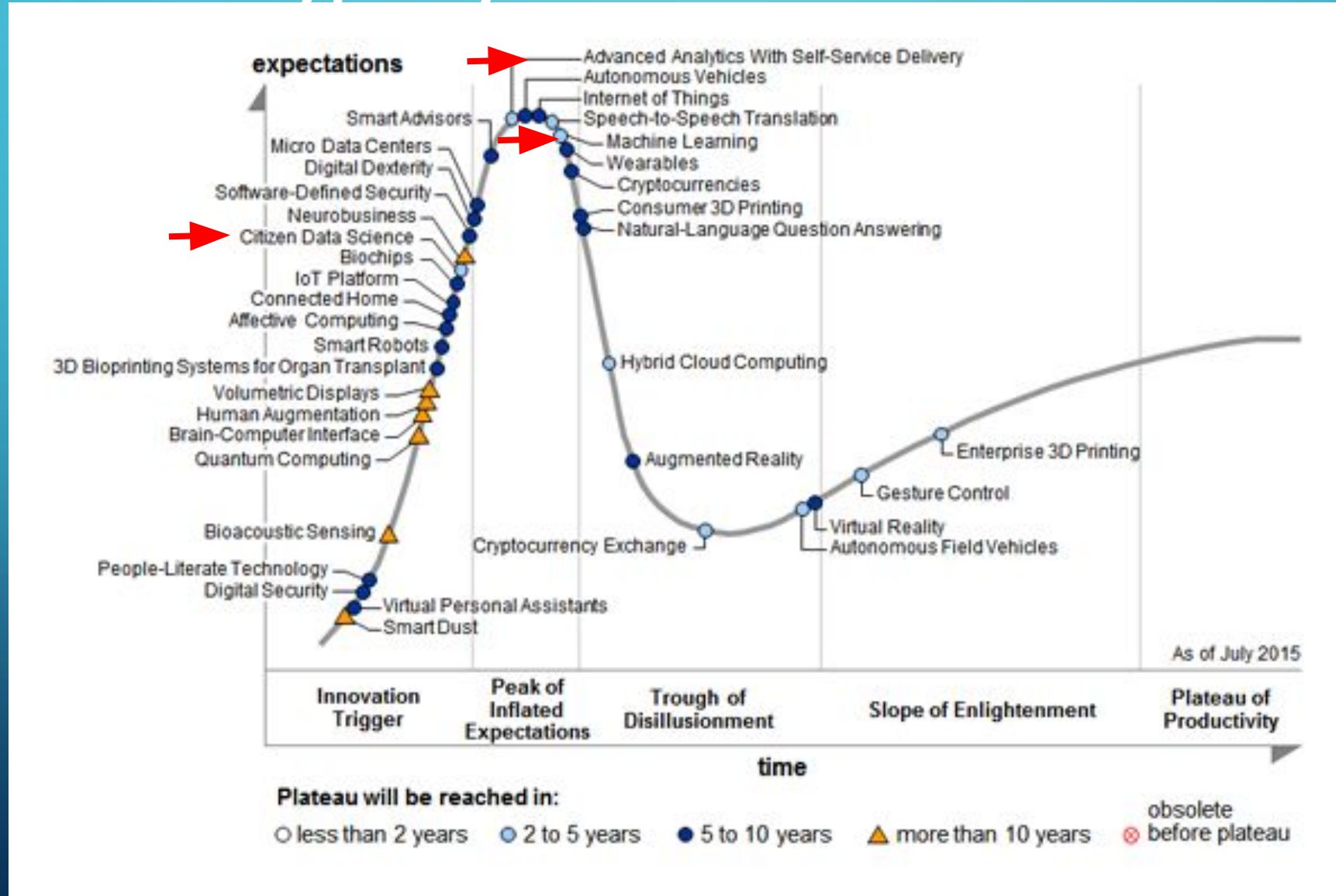
Gartner's hype cycle 2014

<https://www.gartner.com>



Gartner's hype cycle 2015

<https://www.gartner.com>



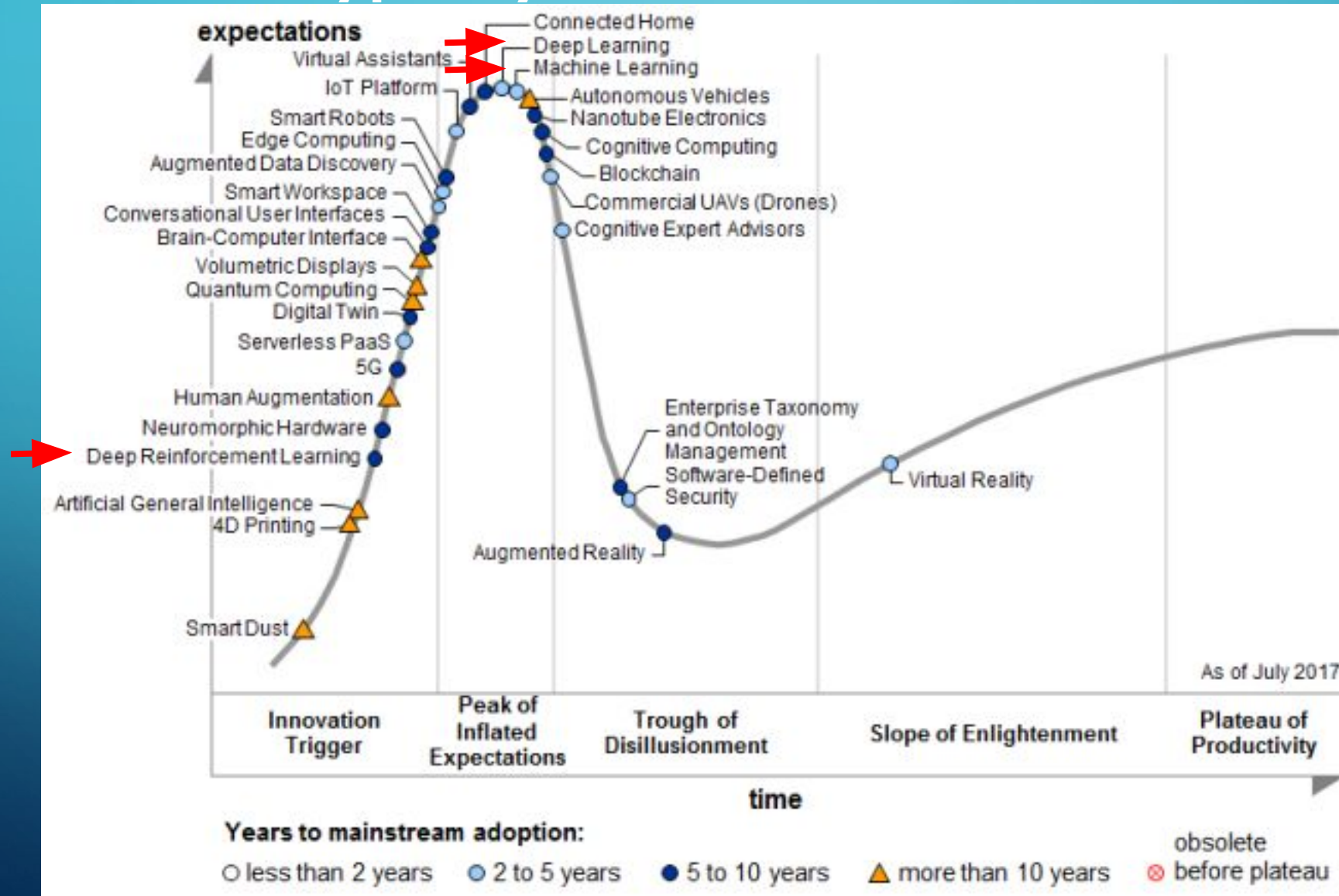
Gartner's hype cycle 2016

<https://www.gartner.com>



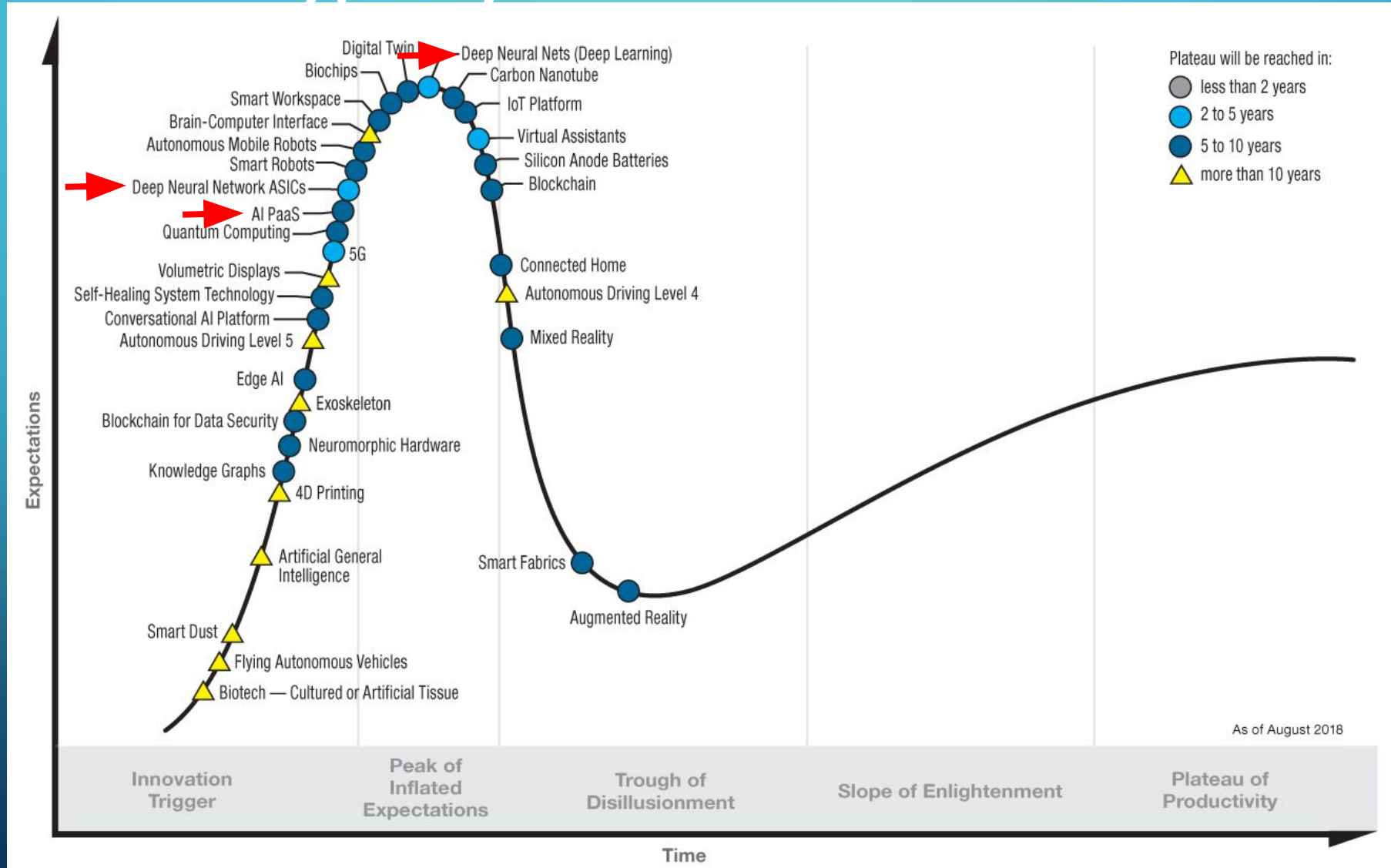
Gartner's hype cycle 2017

<https://www.gartner.com>



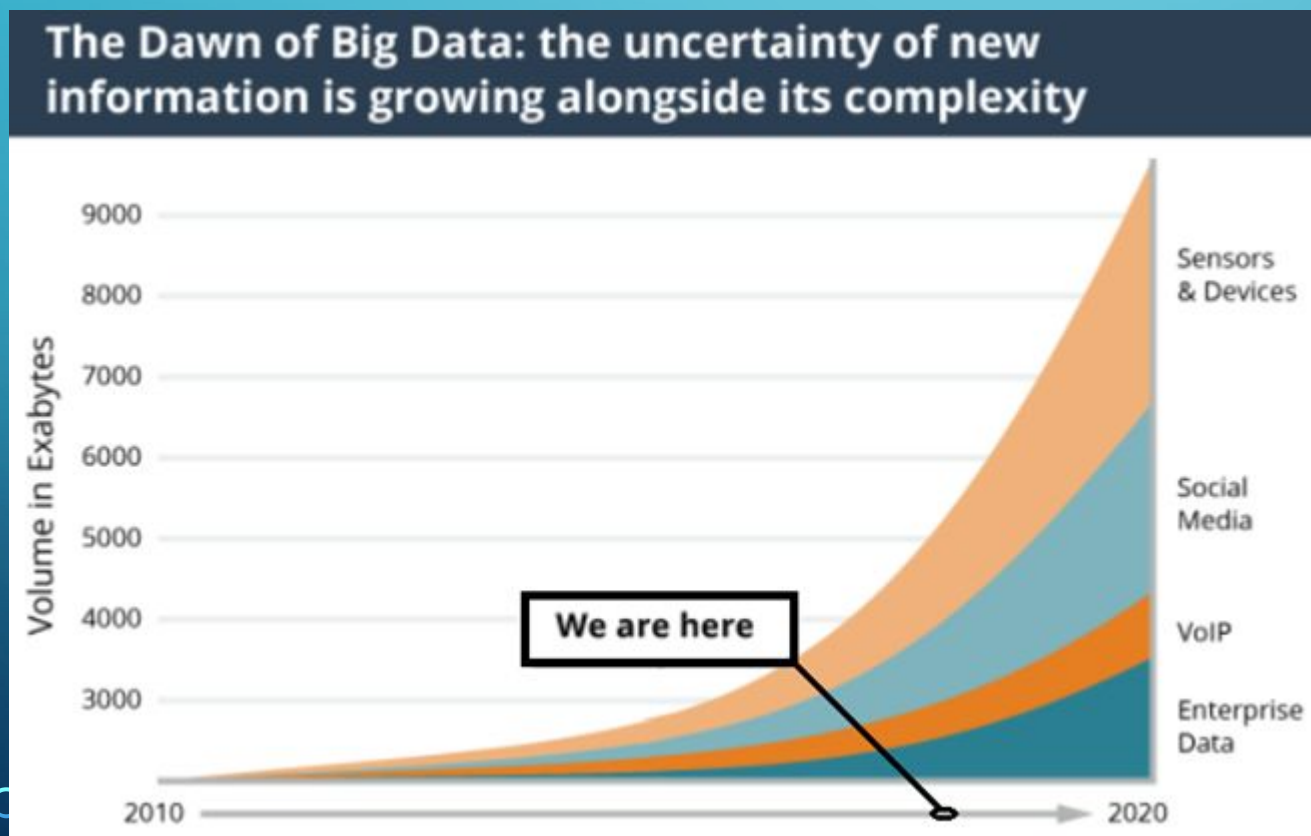
Gartner's hype cycle 2018

<https://www.gartner.com>



INTRODUCCIÓN

- Retos: Crecimiento de los datos



1 EB = 10^3 PB = 10^6 TB = 10^9 GB

Según este artículo de [Forbes](#), los datos están creciendo más rápido que nunca antes. Para el año 2020, se crearán aproximadamente 1,7 MB de nueva información por segundo para cada ser humano en el planeta.

INTRODUCCIÓN

Retos

- Almacenar
- Procesar
- Acceder
- Entender
- Seguridad y privacidad

INTRODUCCIÓN

Retos

- Almacenar
- Procesar
- Acceder
- **Entender**
- Seguridad y privacidad



- Analizar
- Cuantificar
- Comparar
- Visualizar
- Predecir

APLICACIONES



Tomada de: data-flair.training