+ Stockage et traitement de données à grande échelle
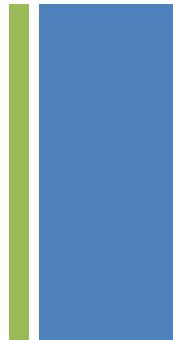
Claudia Roncancio, Ugo Comignani, **Yves Denneulin**, Sylvain Bouveret, Idir Benouaret

Prenom.Nom@grenoble-inp.fr

Distributed, ubiquitous, large scale systems

- Data everywhere!!

- Data, information and knowledge management
  - very small → big data
  - mobile, elastic, … !

- Extremely heterogeneous application domains and requirements
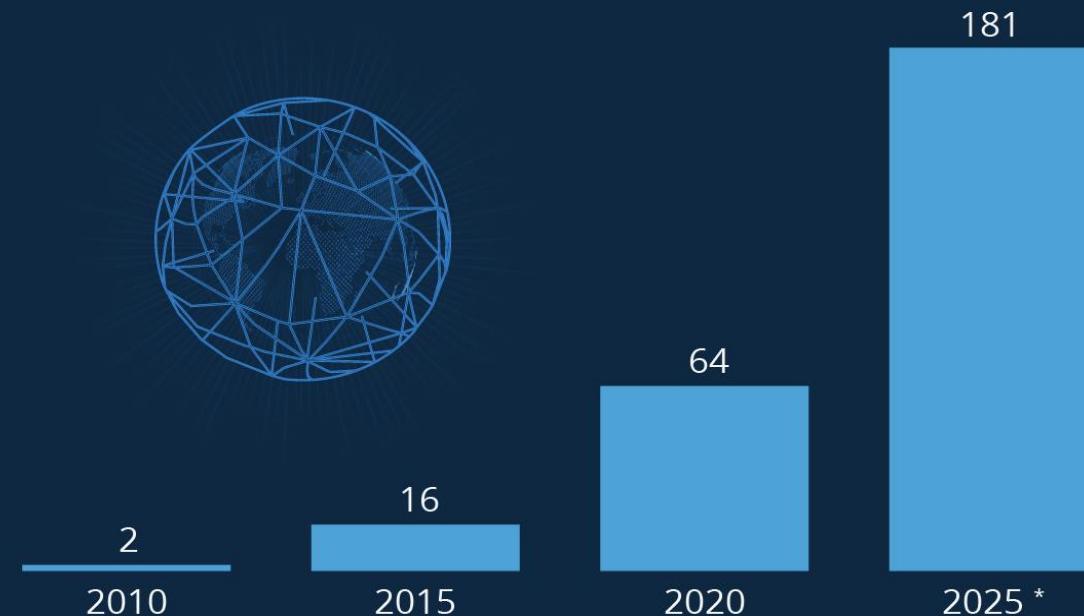
Digital world… *Big Data* !

- Data production and use as never
  - Applications and physical devices , sensors, mobiles, networks, simulations, programs, satellites, …
  - Managers, workers … everybody, pets, machines, …

- Data-driven decisions, actions, process

- Data, information & knowledge
  - **Extracting value**
  - Data analysis, prediction, *business intelligence*, data-intensive science, research, data-based services, machine learning…

- Operational and analytics

Le Big Bang du Big Data

Estimation du volume de données numériques créées ou répliquées par an dans le monde, en zettaoctets

2010	2015	2020	2025 *
2	16	64	181

Un zettaoctet équivaut à mille milliards de gigaoctets.
* Prévision en date de mars 2021.
Sources : IDC, Seagate, Statista
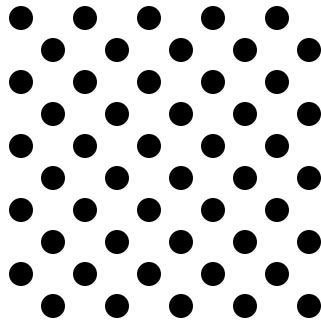
statista

...Dimensions of Big Data
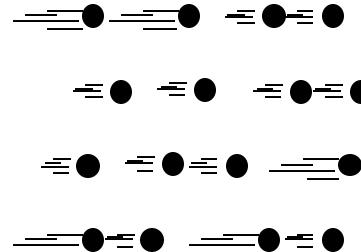
Volume	Velocity	Variety	Veracity*
Data at Rest	**Data in Motion**	**Data in Many Forms**	**Data in Doubt**
Terabytes to exabytes of existing data to process	Streaming data, milliseconds to seconds to respond	Structured, unstructured, text, multimedia	Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations
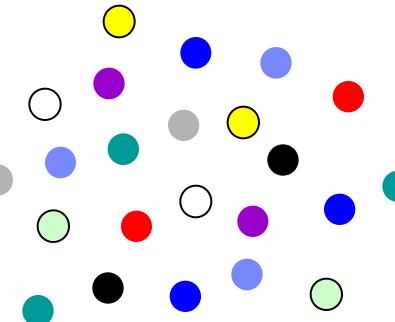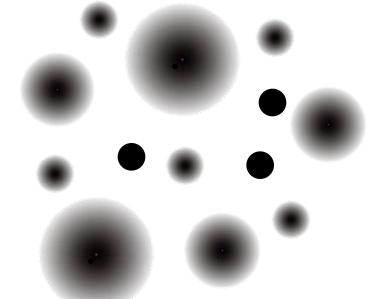
* Truthfulness, accuracy or precision, correctness

Many heterogeneous data management systems

Data Models

- Relational, Key-Value, Graphs, Documents
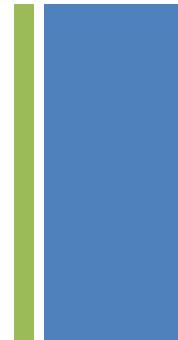
Querying & programming approaches

- Map-Reduce, spark, frameworks

- Scalability

- Memory management & storage

System aspects

- Concurrency control, consistency, transactions

Plan du cours

- Stockage et traitement des masses de données
  - MapReduce, Spark
  - Gestion répartie des données
  - Systèmes NoSQL

- Traitement /exploitation des données
  - Données géographiques
  - Systèmes de recommandation