



Regulatory reporting à la française : the Banque de France Data Lake

Jacques FOURNIER

Director General Statistics

Banque de France

EIFR – 20/09/2016



THE GLOBAL CRISIS HAS LET TO A CONSIDERABLE GROWTH OF REQUESTS FOR DATA

- **All national stakeholders, the media, the general public as well as researchers ask for more and more data, in particular granular data**
 - The same applies to international organizations (ECB, BIS, IMF, OECD, ethical agencies, ...)
 - Our own staff and researchers have also growing requests
- **Our reflection at the Banque de France:**
 - **If we do not address those requests, ‘bad data will throw out good ones’,**
 - Garbage In-Garbage Out: especially true for data
 - The NCB has the means and the knowledge to deliver reliable intelligence at the micro-economic level
 - **Manage, manipulate and leverage on terabytes of data: paradigm shift : we cannot work as before, otherwise we will be snowed under with data of mediocre quality**
 - DG Statistics is where to answer the challenge: DQM is its job, IT use for statistics also
 - Has to be done in one place: cost-saving, effectiveness
 - **Innovation** is necessary to develop not the today but tomorrow system :

 **the DATALAKE is the new frontier**



THE BANQUE DE FRANCE IS THE “HUB” FOR ALL NATIONAL MONETARY AND FINANCIAL DATA

- The Banque de France DGS is in charge of all steps in the value chain of data processing:
 - collecting,
 - monitoring,
 - ordering,
 - analyzing
 - disseminating financial and monetary data for all national and international applicants.
- The BDF DGS is in particular in charge of:
 - balance of payments,
 - monthly economic forecasts,
 - financial inclusion,
 - savings assessments,
 - monetary and financing statistics (participate to the monetary and financial stability committees chaired by the Governor)



CURRENT AND FUTURE CHALLENGES CALL FOR INNOVATIVE SOLUTIONS

- **The observation:** the multiplication of statistical requests in the future is not hypothetical but certain
- For obvious reasons of effectiveness, innovative solutions, usable by every needs are suitable.
- **The integration and mutualisation of data management tools** provides a pragmatic approach able to meet short-term challenges
 - Ensure the high level of quality of a wide range of data collected from large cross-border banking groups
- **Through a series of projects achieved by DG-Statistics, Banque de France has already acquired a wide-ranging expertise in the statistical processing and analysis of granular data**



Capacity to collect and process large quantities of data : the Banque de France solution

- **ONEGATE**
 - Since 2010, a single point of entry (**portal One Gate**) for all data collected by the Banque de France and NSA (national supervisory authority) from financial institutions, non financial corporates, households, insurance companies,...
 - A dedicated platform for data collection :
 - Various formats accepted (XML, XBRL, CSV)
 - Management of high volume (size limit > 2 Gb per file)
 - Ability to manage up to 20 000 users and 1500 files / day
 - **200 000 files received and processed / year**



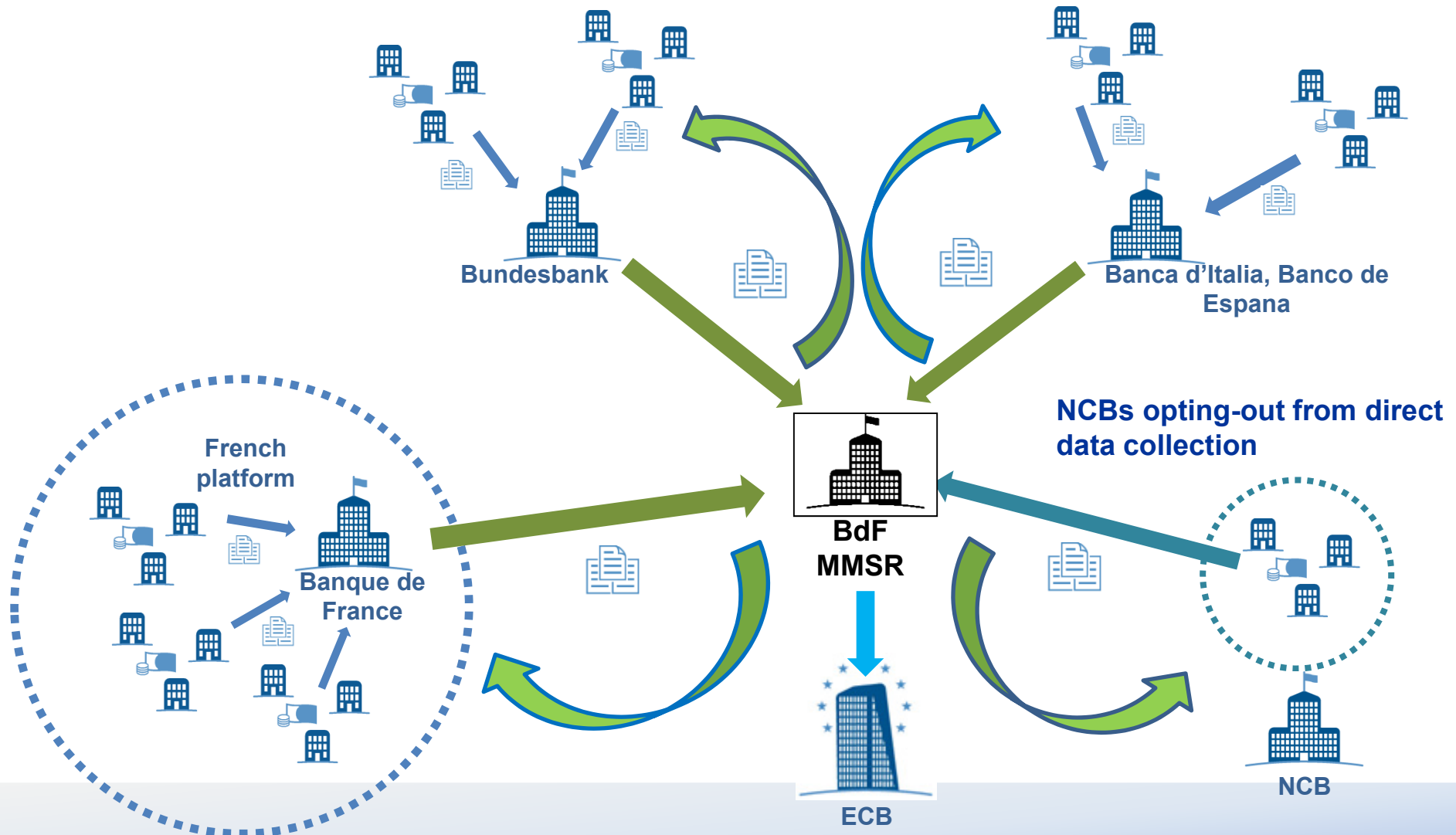
Capacity to collect and process large quantities of data : the Banque de France solution

- **MMSR (Money Market Statistical Reporting)**

On behalf of the ECB, the BDF DGS has developed, and is running, as of April 1, 2016, the transactional platform for the Eurozone data collection for the MMSR

- The 52 largest credit institutions of the Euro Zone
- Reporting of transaction by transaction data, daily frequency
- Banque de France is responsible for the gathering, ‘cleansing’, checking data coming from all countries
- **‘Big Data’ solution for data processing**
- **50 000** transactions recorded / day
- By 19 september : *over 3 500 000 transactions* received and processed by the Banque de France

- **MMSR : global architecture**





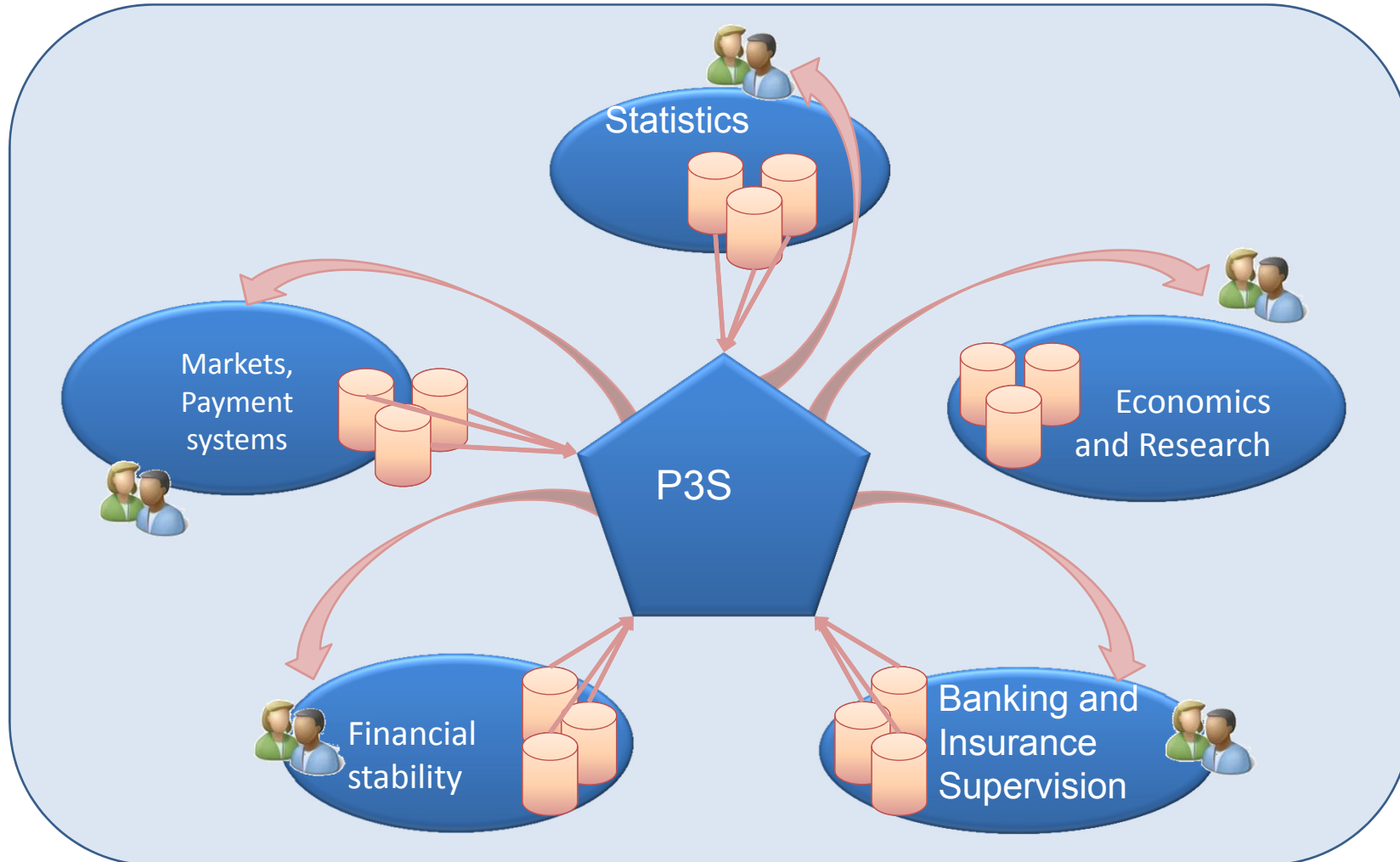
Capacity to collect and process large quantities of data : the Banque de France solution

- **P3S (Pooling and Sharing Statistical Series) – 1/3**
 - **Pooling data within the BDF....**
 - To gather (granular) data on financial institutions and non-financial corporations ...
 - ... collected by the Banque de France and the NSA (ACPR)
 - ... while respecting confidentiality rules
 - **...to allow enhanced analysis for all involved business lines**
 - Offering access to internal users on a ‘need to know’ basis
 - Drop the culture of “silos” in order to take advantage of the multi-purpose nature of data: prudential, monetary, macro-financial, macro-economic.



Capacity to collect and process large quantities of data : the Banque de France solution

- **P3S (Pooling and Sharing Statistical Series) – 2/3**





Capacity to collect and process large quantities of data : the Banque de France solution

- **P3S (Pooling and Sharing Statistical Series) – 3/3**

- **Technological background**

- A solution based on open source « Big Data » technology
- An up-to-date search engine (ElasticSearch)
- A Not Only SQL (NoSQL) DataBase well suited for the storage of heterogeneous data
- Main differentiator : all formats are accepted (SDMX-ML, XBRL,..)

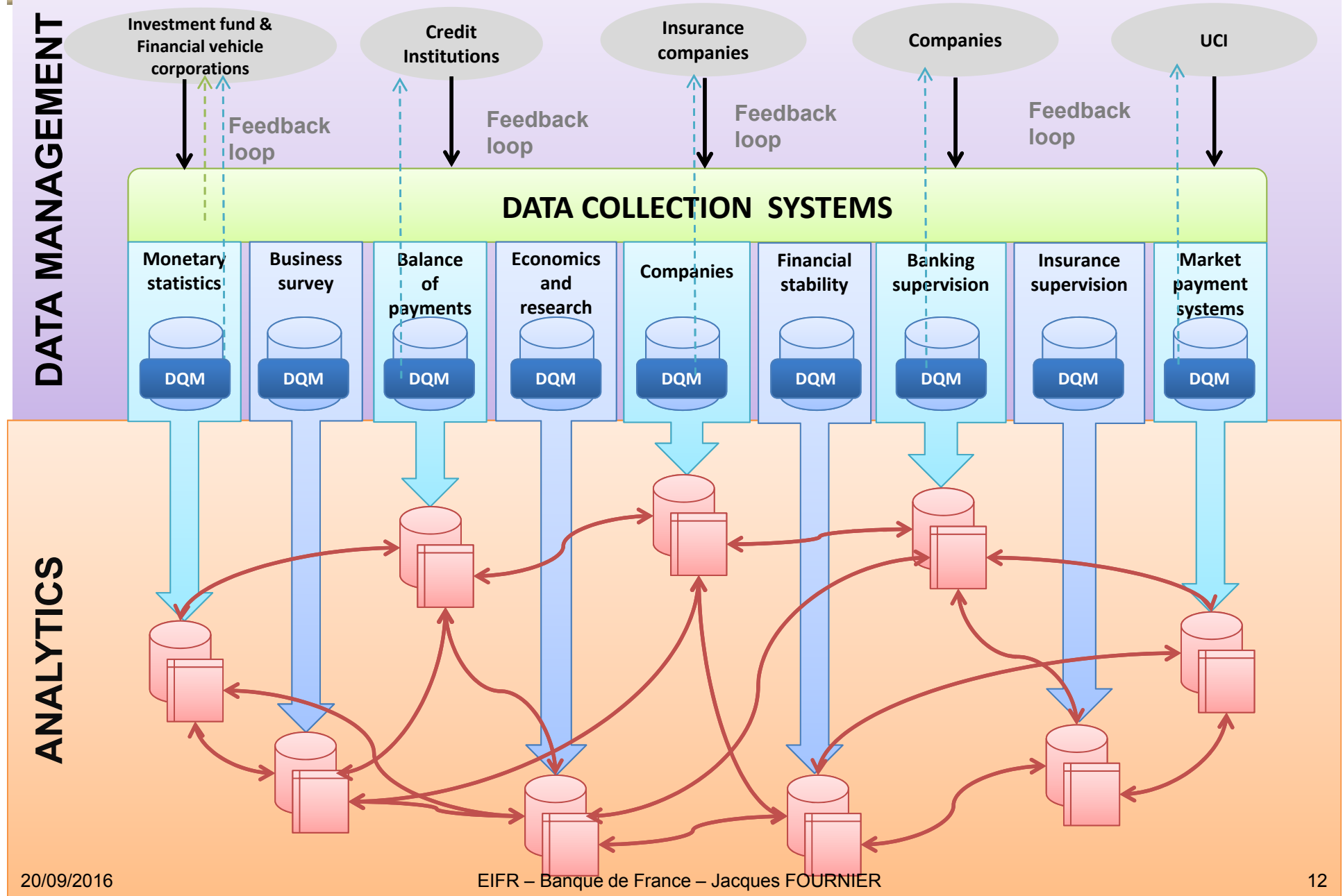
- **400 million series already stored in the database**



Our next step: the DATA LAKE :

- **The industrialization of the data management process** throughout the production cycle in a context of massive data must be broadened. This requirement is at the heart of the Datalake project:
 - **Development of automated systems for immediate data quality management with feed-back loops to reporting agents**
 - **Development of analytical tools** tailored to the large datasets on financial institutions and non-financial corporations collected by the Banque de France and the ACPR (eg, machine learning algorithms, data visualization tools)
- The DG-Statistics is in charge of the project for the BDF lato sensu
- The BDF will dialogue with the industry at all stages of the development

THE INFORMATION SYSTEM BEFORE THE 'BIG DATA REVOLUTION'



DATA LAKE PROJECT : OVERVIEW

