New Frontiers in Business Intelligence: 
*Distribution and Personalization*

Matteo Golfarelli  
University of Bologna - Italy

---

Summary

- The challenges of BI 2.0
- Motivating scenario and envisioned architecture: *Business Intelligence Networks*
- Distribution
  - Research issues
  - A mapping language
  - Query reformulation
- Personalization → Patrick Marcel
From BI 1.0 to BI 2.0

- *Business intelligence* (BI) transformed the role of computer science in companies from a technology for storing data into a discipline for timely detecting key business factors and effectively solving strategic decisional problems.
- In the current changeable and unpredictable market scenarios, the needs of decision makers are rapidly evolving.
- To meet the new, more sophisticated user needs, a new generation of BI systems (*BI 2.0*) has been emerging.

Issues in BI 1.0

- Performance optimization (query plans, materialized views, indexing, etc.)
- Logical design
- Conceptual design methodologies and formalisms
- ETL modeling and automation
- Testing the DW
- ....
Issues in BI 2.0

- BI as a service
- On-demand BI
- Real-time BI
- Situational BI
- Collaborative BI
- Pervasive BI
- ....

Motivating scenario

- Cooperation is seen today by companies as one of the major means for increasing flexibility and innovating so as to survive in today uncertain and changing market
- Companies need strategic information about the outer world, for instance about trading partners and related business areas
- It is estimated that above 80% of waste in inter-company and supply-chain processes is due to a lack of communication between the companies involved
Motivating scenario

- In such a distributed business scenario, where multiple partner companies/organizations cooperate towards a common goal, traditional BI systems are no longer sufficient to maximize the effectiveness of decision making processes.
- Two new significant requirements arise:
  - Cross-organization monitoring and decision making
    Accessing local information is no more enough, users need to transparently and uniformly access information scattered across several heterogeneous BI platforms.
  - Pervasive and personalized access to information
    Users require that information can be easily and timely accessed through devices with different computation and visualization capabilities, and with sophisticated and customizable presentations.

Envisioned architecture

- Business Intelligence Network (BIN): a dynamic, collaborative network of peers, each hosting a local, autonomous BI platform
  1. Each peer relies on a local multidimensional schema that represents the peer's view of the business, and it offers monitoring and decision support functionalities to the other peers.
  2. Users transparently access business information distributed over the network in a pervasive and personalized fashion.
  3. Access is secure, depending on the access control and privacy policies adopted by each peer.
  4. Participants are collaborative, even if with different grades.
  5. Inclination to collaboration does not reduce autonomy of participants, who are not subject to a shared schema.
  6. A BIN is decentralized and scalable because the number of participants, the complexity of business models, and the workload can change.
Envisioned architecture

interacts with the peer’s BI platform to obtain results from the local data
Envisioned architecture

uses the semantic mappings established towards the peer neighbors to reformulate queries accordingly

applies query routing policies to select the most relevant peers to forward a query to
Envisioned architecture

- **Local BI platform** collects and integrates the results coming from the peers.
- **Local query processing**
- **Query result reconciliation**
- **Access policies resolution**
- **Local MD schema**
- **Mappings**
- **Query reformulation**
- **Query forwarding**

**Business Intelligence Network**

*Peer 1*

*Peer N*

Sets policies for data sharing depending on the degree of trust between participants.
A typical user interaction

A user formulates an OLAP query $q$ by accessing the local multidimensional schema of her peer.

She can annotate $q$ by a preference that enables her to rank the returned information according to her specific interests.
A typical user interaction

To enhance the decision making process, \( q \) is forwarded to the network and reformulated on the other peers in terms of their own multidimensional schemata.

A typical user interaction

Each involved peer locally processes the reformulated query and returns its (possibly partial or approximate) results to the querying peer.
A typical user interaction

The results are integrated, ranked according to the preference expressed by the user, and returned to the user based on the lexicon used to formulate q

Research issues

- **Query reformulation** on peers is a challenging task due to the presence of aggregation and to the possibility of having information represented at different granularities in each peer.
- To optimize query answering across the network, **query routing** strategies that forward queries to the most promising peers only are needed.
- The strategic nature of the exchanged information and its multidimensional structure require advanced approaches for **security**.
- Mechanisms for controlling **data provenance and quality** in order to provide users with information they can rely on should be devised.
- A unified, integrated vision of the heterogeneous information collected must be returned to users through **object fusion** techniques.
Query reformulation

- **Mapping language:**
  - Handling the asymmetry between dimensions and measures
  - Specifying the relationship between two attributes of different multidimensional schemata in terms of their granularity
  - Considering aggregation operators to avoid the risk of inconsistent query reformulations
  - Expressing also mappings at the instance level to transcode data

(Golfarelli et al., 2010)
Query reformulation

Mapping language:

- HOSPITALIZATION
- cost
durationOfStay
- patient
- region
birthDate
- unit
- patientGender
- patientCity
- patientNation
- month
- year
- disease
- organ
- LHD
- ward
- gender
- segment
- role-up
- same
diagnosis
- category
- roll-up
- patientGender
- patientCity
- patientNation
- patientBirthYear
- equivalent
- year
- month
- day
- equivalent
- category
- diagnosis
- same
- roll-up
- equivalent
- patientBirthYear

Framework:

- To translate semantic mappings we use a logical formalism called source-to-target tuple generating dependencies (ten Cate & Kolaitis, 2010)
Example: Schema translation

Example: Query translation

- Total hospitalization costs for region and year

\[ \pi_{\text{region}, \text{year}, \text{SUM(cost)}} (\text{HospFT} \Join \text{DateDT} \Join \text{PatientDT}) \]

\[ q(R, Y, \text{SUM}(C)) \leftarrow \text{HospFT}(\_\_, D\_\_, P\_\_, C\_\_), \]
\[ \text{DateDT}(D\_\_, Y), \]
\[ \text{PatientDT}(P\_\_, R\_\_, \_\_\_) \]
Example: Mapping translation

∀ S, E, C (AdmFT(..., S, E, ...), C = S + E → HospFT(..., C, ...))

Example: The rewriting

- The group-by is reformulated using the roll-up mapping from region to patientCity, while measure cost is derived using the same mapping

π_{year, patientCity, SUM(totStayCost + totExamCost)} (AdmFT × DateDT × PatientCityDT)
Personalization

- The goal of personalization is to deliver information that is relevant to an individual or a group of individuals in the most appropriate format and layout
  - **Recommendation:** the system suggests new queries to support users in navigating the cube (Giacometti et al., 2009)
  - **Personalized visualization:** the user specifies constraints that are used to determine a preferred visualization according to a user profile (Bellatreche et al., 2005)
  - **Ranking:** query results are organized in a total or partial order so that the user visualizes only the “most relevant” tuples (Golfarelli et al., 2011).
  - **Contextualization:** the query is enhanced by adding predicates that depend on the context (Jerbi et al. 2008)

Thank you for your attention

Questions?
Related readings

- Cui, Y., & Widom, J. Lineage Tracing for General Data Warehouse Transformations. JVLDB, 12(1), 2003
- Torlone, R. Two approaches to the integration of heterogeneous data warehouses. Int. Journ. on Distributed and Parallel Databases, 23(1), 2008