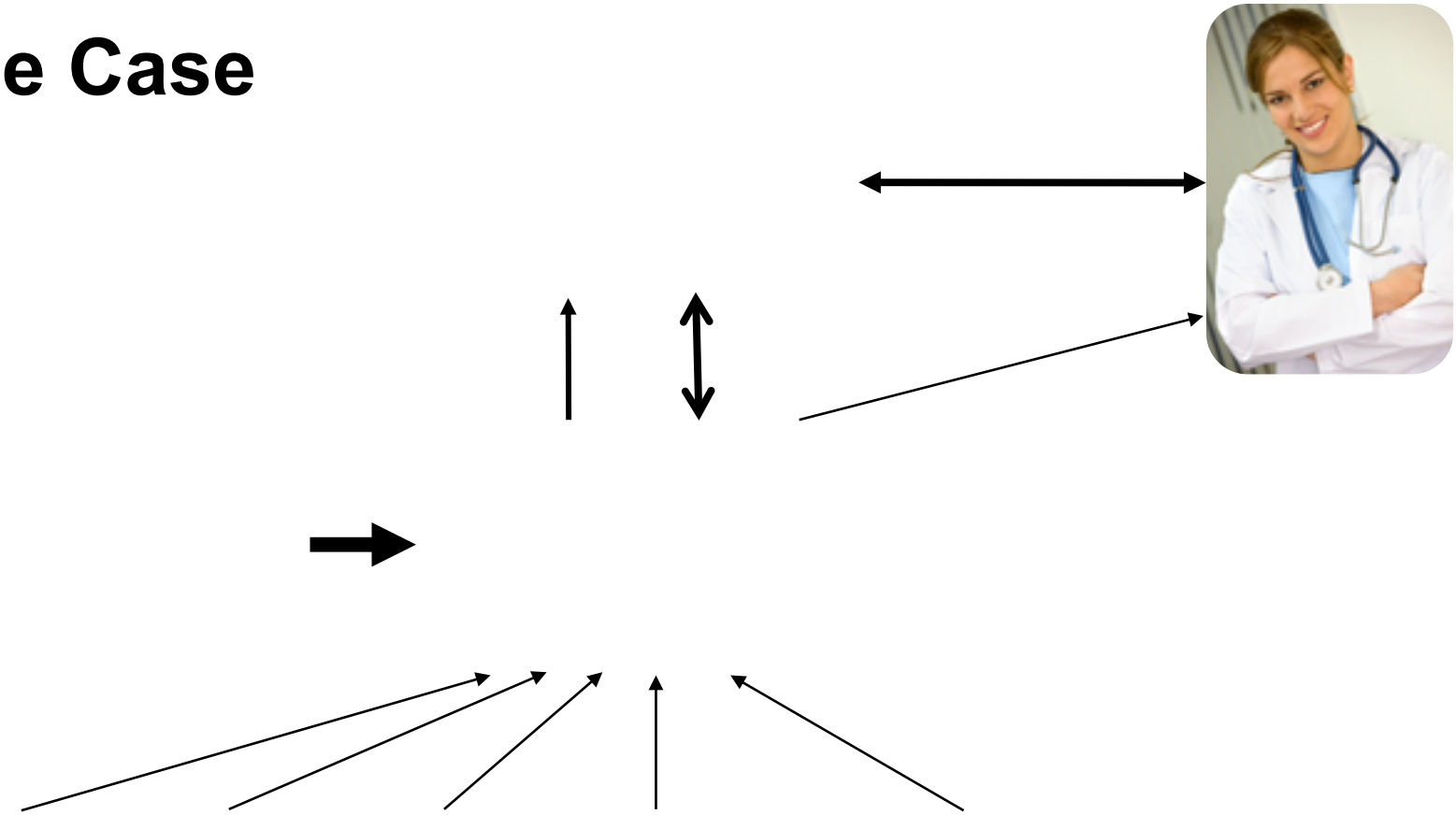
ORACLE®

Fine Grain Provenance Using Temporal Databases

Outline of the talk

- Use case: Classic management of patient data
  - Data types, queries
  - History
  - Security and context information
- Fine grain provenance – I
- Smart management of patient data
  - Facts, knowledge, and information
  - The model
  - Classification and customization
- Fine grain provenance - II
- Implementation details
- Conclusion

Use Case



Classic Management of Patient Data

Data Types

- Structured Data – SQL
- Semi structured data – XML
  - **HL7 -** Health Level-7
- **DICOM** - Digital Imaging and Communications in Medicine
- Text
- **Any mix**

Data Manipulation and (continuous) queries

- SQL 92 and 99
- XQuery
  - HL/7 verbs
- DICOM verbs
- Text processing verbs
- Mixed use of languages

History

- Data management for patient history
  - No extended data model
    - Simplifies programming significantly
  - Standard update, insert, delete
- Queries
  - The current values
  - The values/images at a specific time
  - The values/images as seen at a specific time
  - The evolution of values/images

Security and Context Information

- All queries and temporal queries support (fine grain) security
  - A doctor/nurse can only access data from patients s/he is currently treating
- Additional information recorded by the data base
  - The transactional context of any change or query
  - The transactional context includes host, database/OS user, program

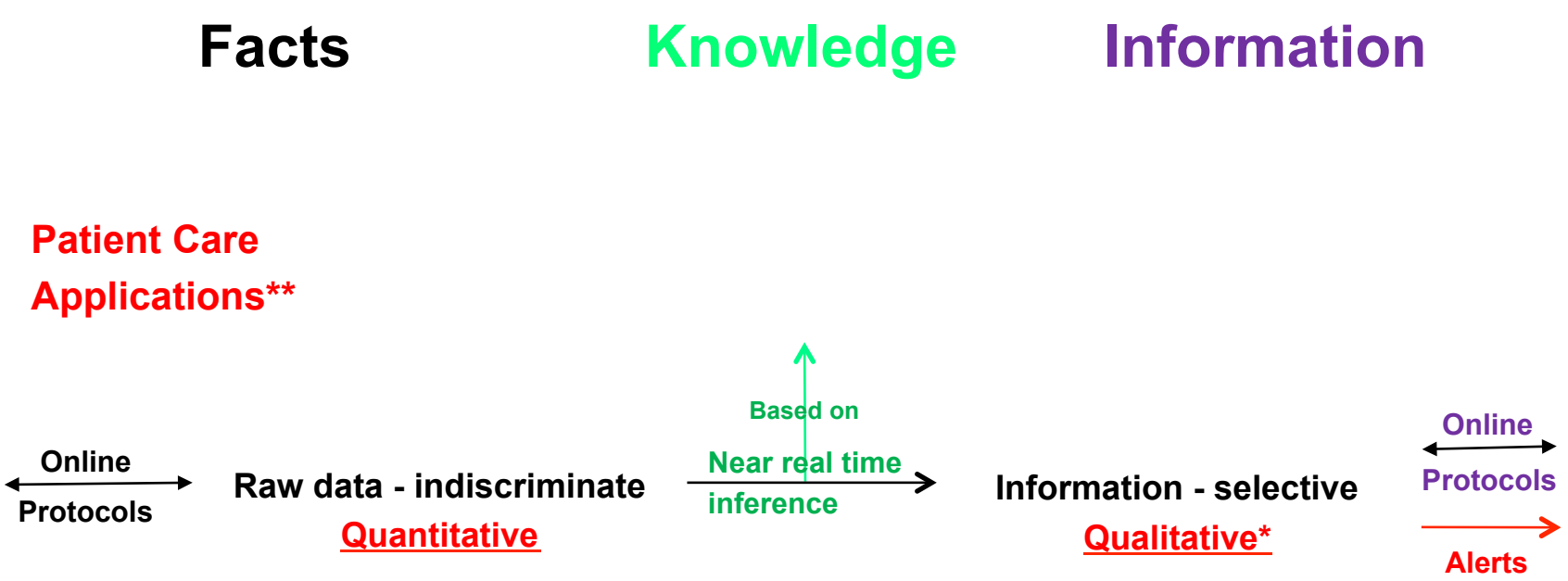
Fine Grain Provenance - I

- The database is able to answer the following questions
  - What was a single or set of values at a given time – from the current perspective?
    - What was a single or set of values at a given time from an earlier perspective – imported to deal with corrections
    - What is the history of a single or a set of values
  - Was a value ever corrected?
    - What is the history of correction?
  - Who was responsible for providing/deleting a value?
  - Which program created the value?
  - Who looked at specific values?

Smart Management of Patient Data

- The issue:
  - Rapidly increasing amount and complexity of data
  - Rapidly increasing amount and complexity domain knowledge
  - Data and knowledge have grown way beyond the capacity of a human cognitive system
- A solution
  - Capture knowledge and personal preferences
    - Vocabularies, rules/models, classifications, customizations
  - Capture facts – as done in classic support
  - Transform data (facts) into information using captured knowledge
    - Alert medical personnel about time critical adverse conditions

The Model

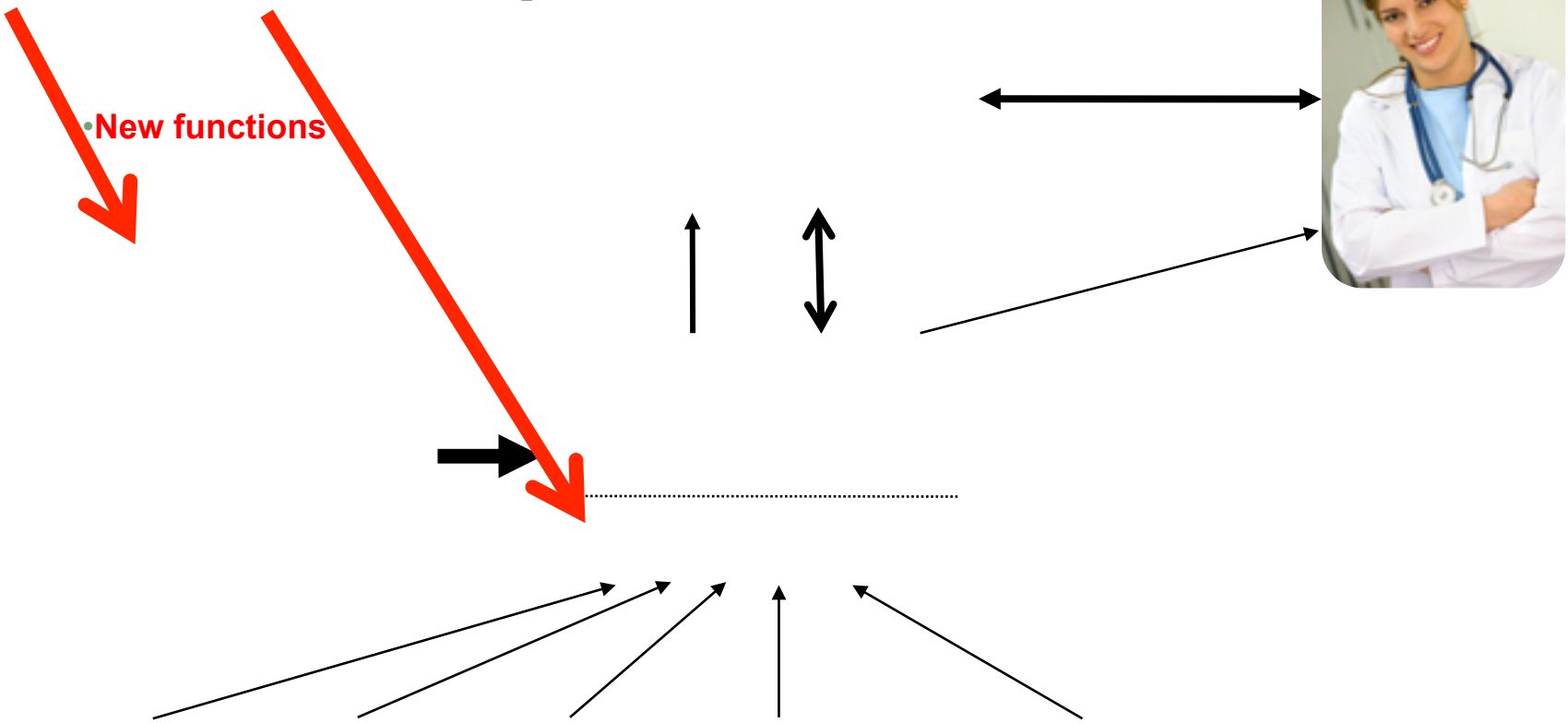


* Qualitative information is preferred by the human cognitive system

** The application is as declarative as possible

Use Case - Updated

- New functions



Information and Incidents

- Information is created as soon as new data/facts or new knowledge become available
  - The information is a compact and qualitative representation of important facts
    - The temperature is critical
    - The blood chemistry indicates a high probability of a cardiac arrest
    - The information has a high uncertainty, additional tests are recommended
  - Information is bundled as incidents
    - Alert is issued for time critical information
  - Doctors can review the status of the patient on a qualitative level
    - What is important; i.e., show incidents with certain characteristics
    - Show the history of selected incidents
      - Is the patient improving as expected?
    - If needed the doctor can also look at the quantitative data

Fine Grain Provenance - II

- Full auditing and tracking of facts
  - Implies full auditing and tracking of information
- Full Description and versioning of
  - Knowledge – rules, queries, model, programs, ..
    - Who developed/tested/deployed/changed the knowledge elements and when
  - Classifications
    - Who developed/tested/deployed/changed the classification and when
  - Customizations
    - Who deployed/changed the customization
- The evolution of the information is now visible
  - What are the facts and knowledge behind information and incidents?
  - Do I accept the information?
  - Why did a colleague come to a (different) conclusion?
  - Why was the information (diagnosis) changed?

Conclusions

- Databases support management of and access to a wide variety of data
- Temporal databases provide full support for auditing and tracking – no user programming required
- Adding knowledge management to data management provides full support for provenance - no user programming is required

Questions
Comments
Suggestions

Read Consistency - Oracle's Flashback

- One of the main features of Oracle is consistent read
  - No read locks are taken
  - Instead data is read as of a point in time in the past before all uncommitted changes (using undo)
- Flashback extends CR to be able to read data as of a point in time in the recent past (using undo)
- Total Recall extends flashback to go back far in the past
- Using flashback, it is possible to see data/information/ knowledge as it was at any point in time, providing the main building block for provenance

Temporal Database Support – Oracle's Total Recall

- Total recall provides a way to enable transaction time history on a table for a specified retention
- Using total recall it is possible to do flashback queries
- "As of" queries enable the user to read a row/table as of a point in time
- "Versions" enable the user to get all committed versions of a row/table between a range of time
  - Provides the transaction start/end time of version, transaction context of creator of version
- Audit used for tracking queries
- Valid time support can also be added in future

A Classification Model

Uniform classification of data	• Value: Normal, guarded, serious, critical • Urgency: Stat, ASAP, none
Uniform classification of change	• Type: deteriorating, improving • Rate: rapid, slow
Statistical temporal change model	• Patient is not improving as expected by *model* M_1
Uniform classification simplifies queries	• Find all patients with critical condition lasting more than 2 hrs in the last 5 years • Identify important incidences/adverse conditions

Classification - Design Principles

Uniform classification

- Simplifies aggregating elementary quantitative information into highly compact representation
- Reduces the number of queries, rules, and models significantly
  - **A vital is deteriorating fast**
  - **The patient does not improve as expected**

Personalized classification rules

- Adjust to the preferences of a group, a doctor, or specific condition of a patient
- Adjusts to the specific situation of a patient

Classification Methods

- Decision tables, rules, models, manual

Classification With a Decision Table

	Lower Range				Upper Range			
	Critical	Serious	Guarded	Normal	Normal	Guarded	Serious	Critical
...								
TEMPERATURE	34.5	36	37	37.0	38.4	38.4	40	42
HEART_RATE	40	50	60	60	100	100	125	150
SYSTOLIC_BP	70	80	90	90	140	140	160	190
DIASTOLIC_BP	40	50	60	60	90	90	100	110
MEAN_ARTERIAL_PRESSURE	60	65	70	70	105	105	110	115
RESPIRATORY_RATE	8	10	14	14	26	26	30	35
OXYGEN_SATURATION	80	85	90	90	100			
WEIGHT								
EKG								
CO	3		4	4.0	6.0	6		8
CI	2.2		2.6	2.6	4.2	4.2		6
SVR	600	700	800	800	1200	1200	1400	1600
CWP				4	12			
INTRA_ABD_PRESSURE				5	15	15	20	30
...								

Note: Columns **Guarded** and **Normal** contain intentionally the same information