Exploring Workflow Enactments through Querying Execution Logs

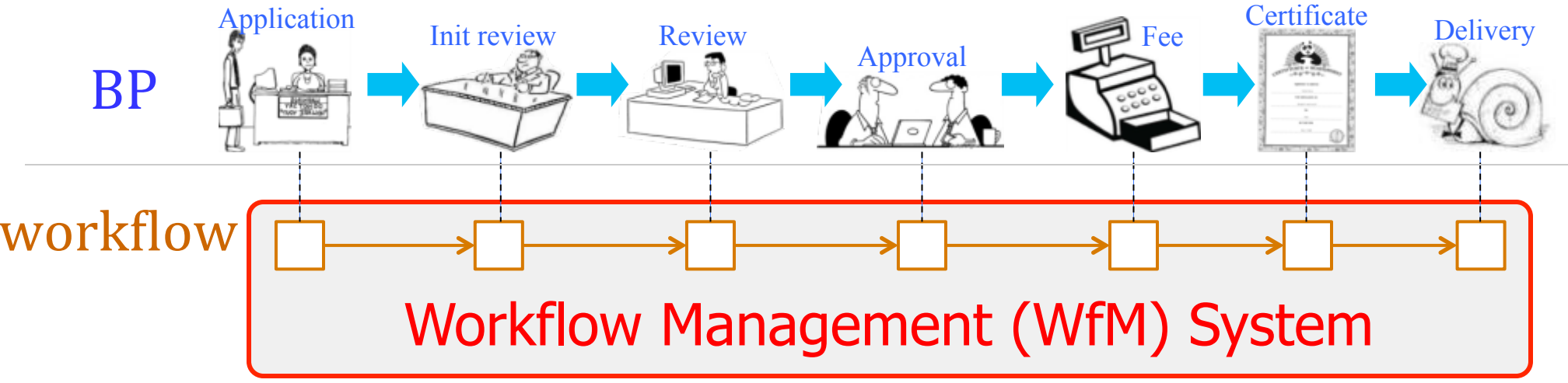
Jianwen Su
Dept. of Computer Science
UC Santa Barbara
(Joint work with Yan Tang)

Outline

- Business Workflow and Analytics (=BI)
- Classical Approach to BI
- Workflow Logs
- Exploration as An Example
- Research Challenges

Business Processes & Workflow Management

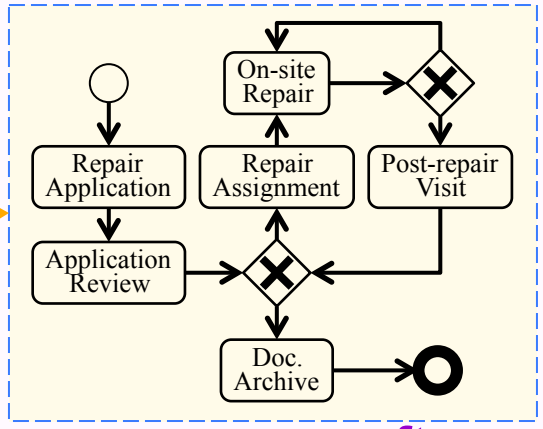
- A BP is an assembly of tasks to accomplish an objective
  - Eg: Obtaining a Permit



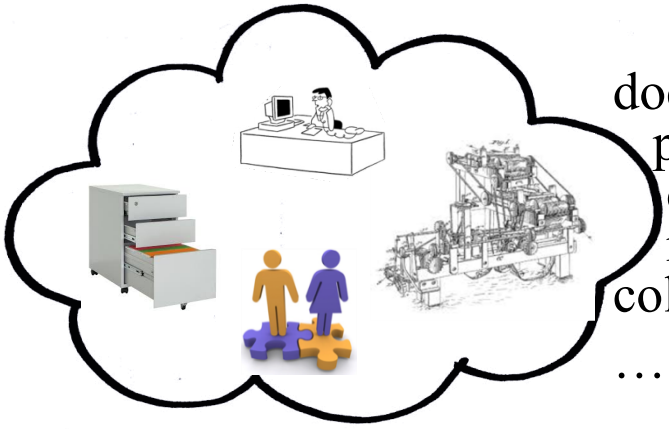
- Each workflow model matches a BP model
- Each workflow activity (□) is a software program that interfaces one task in the BP
- A WfM system manages executions, resources, documents, etc.

BP and Workflow: Another View

Biz Proc Model



Biz Proc Execution

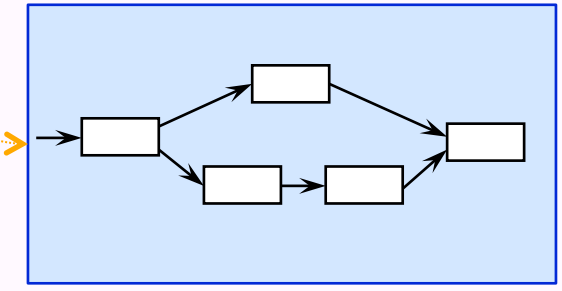


documents,
people,
devices,
resources,
collaboration
…

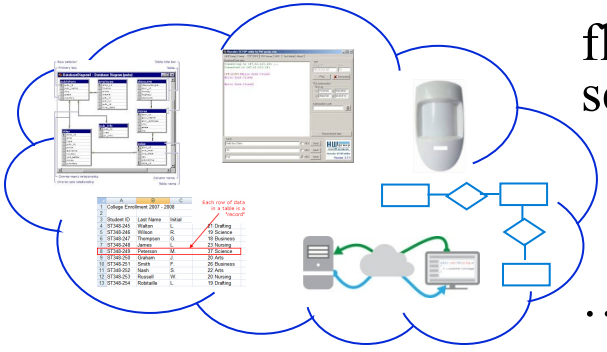
Only activities are present

software development

Workflow Model



Workflow Execution



flow mgmt,
services,
db's,
software,
resources,
…

BP/Workflow: Change is Essential

- Causes of changes: policy/regulation change, environment change, market, improvements, . . .
- To incorporate changes, need to modify:
  - BP/workflow models
  - Databases
  - WfM systems
  - …

  Very hard problem but not the focus of here
- To find opportunity for change from past executions
  - Business intelligence (BI)

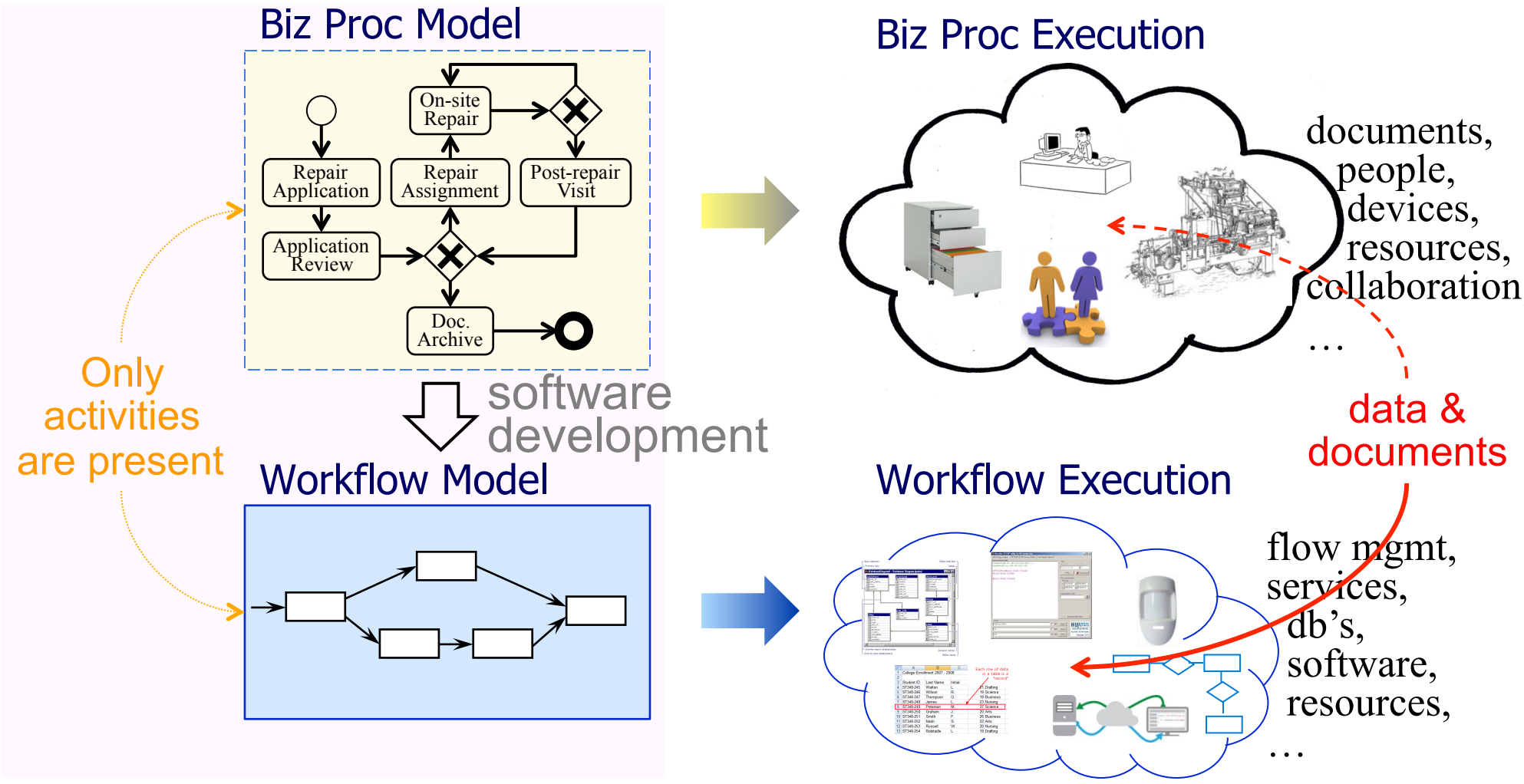
    Needs data, where are data?

An Example from Univ of California

- TC Travel Council oversees travel policies, programs, and travel related business services
- Allowing AirBnB: what steps in e.g., reimbursement are affected?
  - State funding
  - Federal funding
  - Gifts and donations
- Defined procedures in 10 campuses, 5 hospitals, & Office of President
- Actual practices vary, need to find from past cases, e.g. travel reimbursement

  ? Where are data

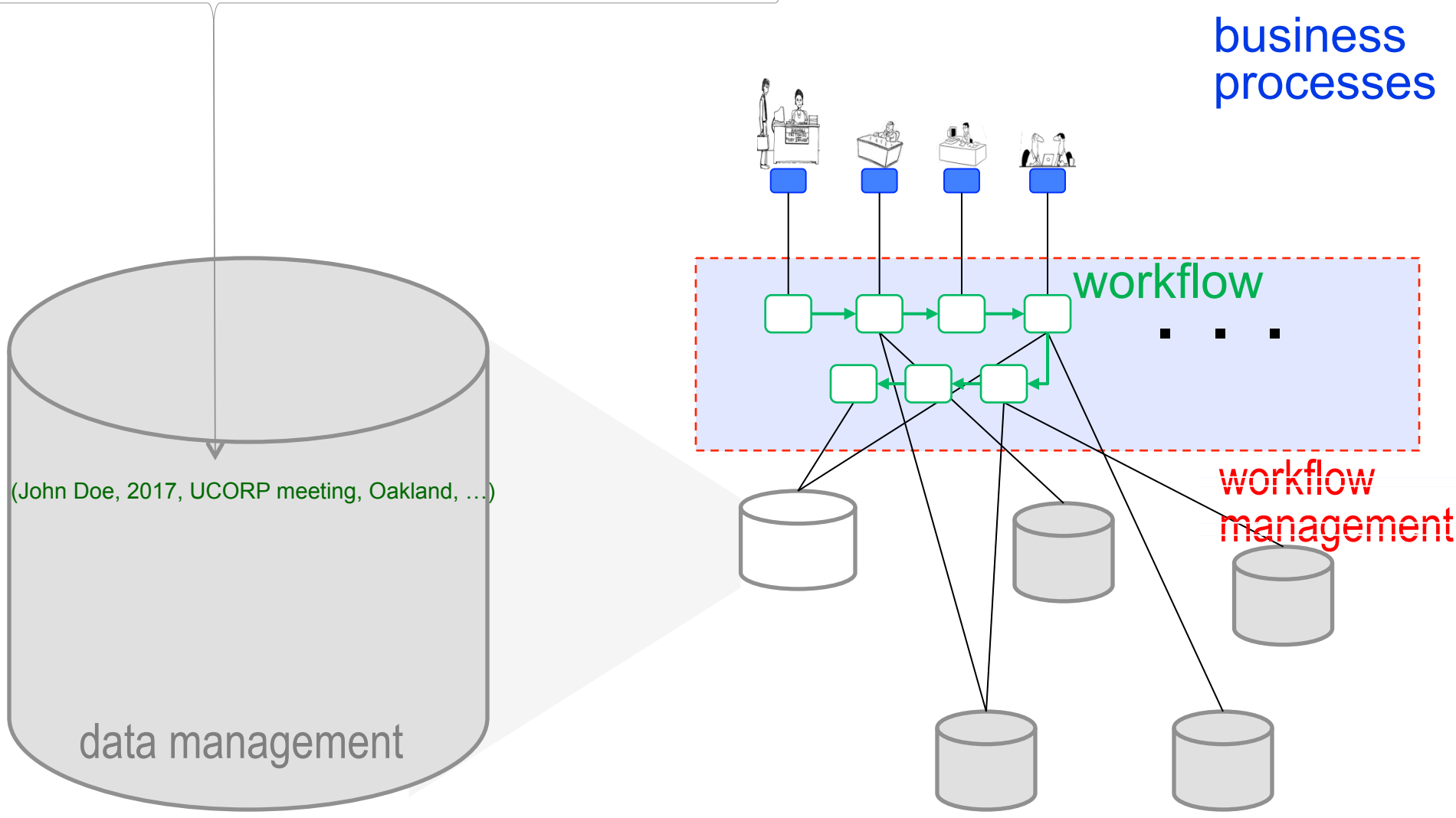
BP and Workflow: Another View



- Workflow execution generates a lot of data: biz data, execution status, resource usages, correlations, …

Workflow Management System

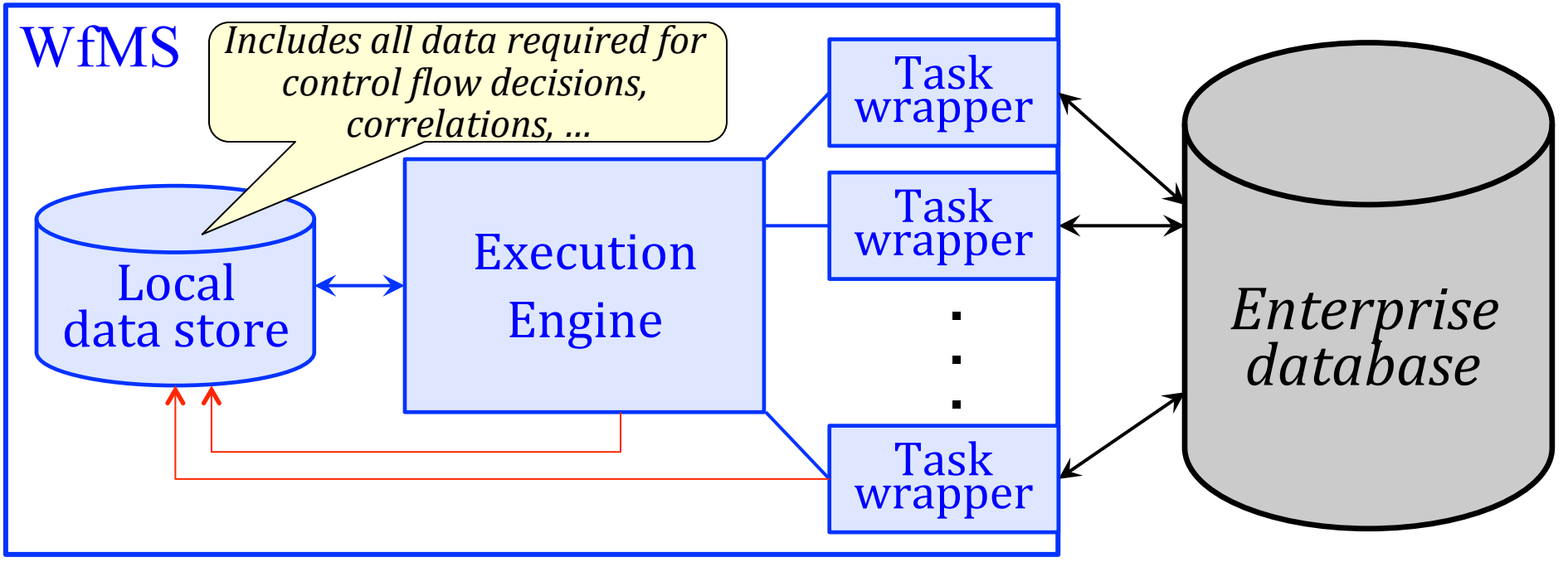
(John Doe, 2017, UCORP meeting, Oakland, …)



Where Are Data in WfM Systems

- Typical architecture:

[van der Aalst-van Hee 2004]
(Pre-architecture [Bussler 1997])



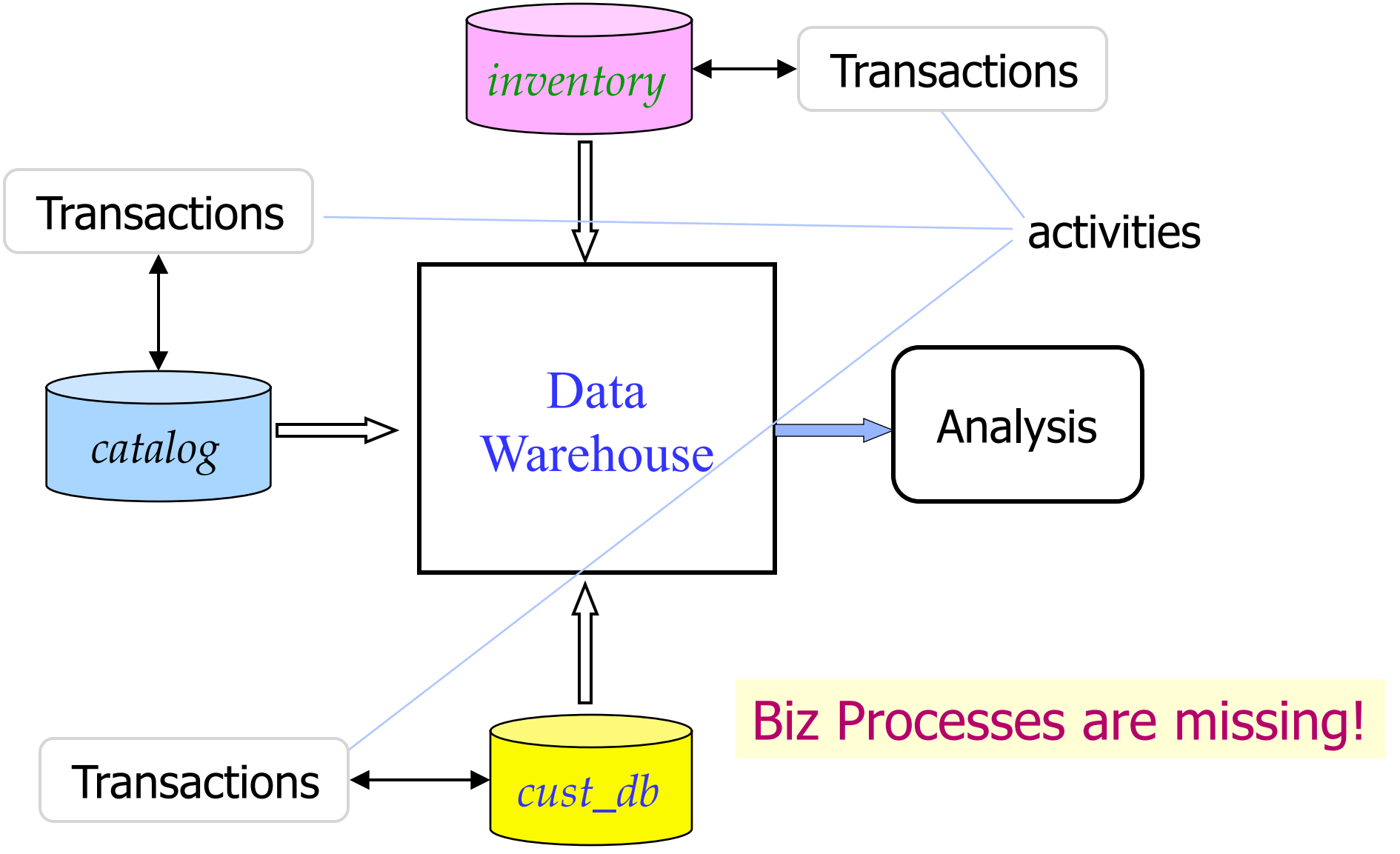
- Logging often an option:
  - Workflow instance, activity (task), variables, …, etc all logged separately, sometimes in difference dbs

Outline

- Business Workflow and Analytics (=BI)
- Classical Approach to BI
- Workflow Logs
- Exploration as An Example
- Research Challenges

Traditional Approach to BI (Big Data)

- Biz analytics (intelligence): Extract-Transform-Load



A Few Details on ETL-OLAP

- Typically, extract via relational queries (SQL, relational algebra, etc.) on other relational databases
  - Could be from other types
- Relational queries may be chained together and combined, results loaded into data warehouse
- The types of queries are predetermined:
  - What data to get
  - How data from different sources are combined
  - Fixed semantics and representations of data

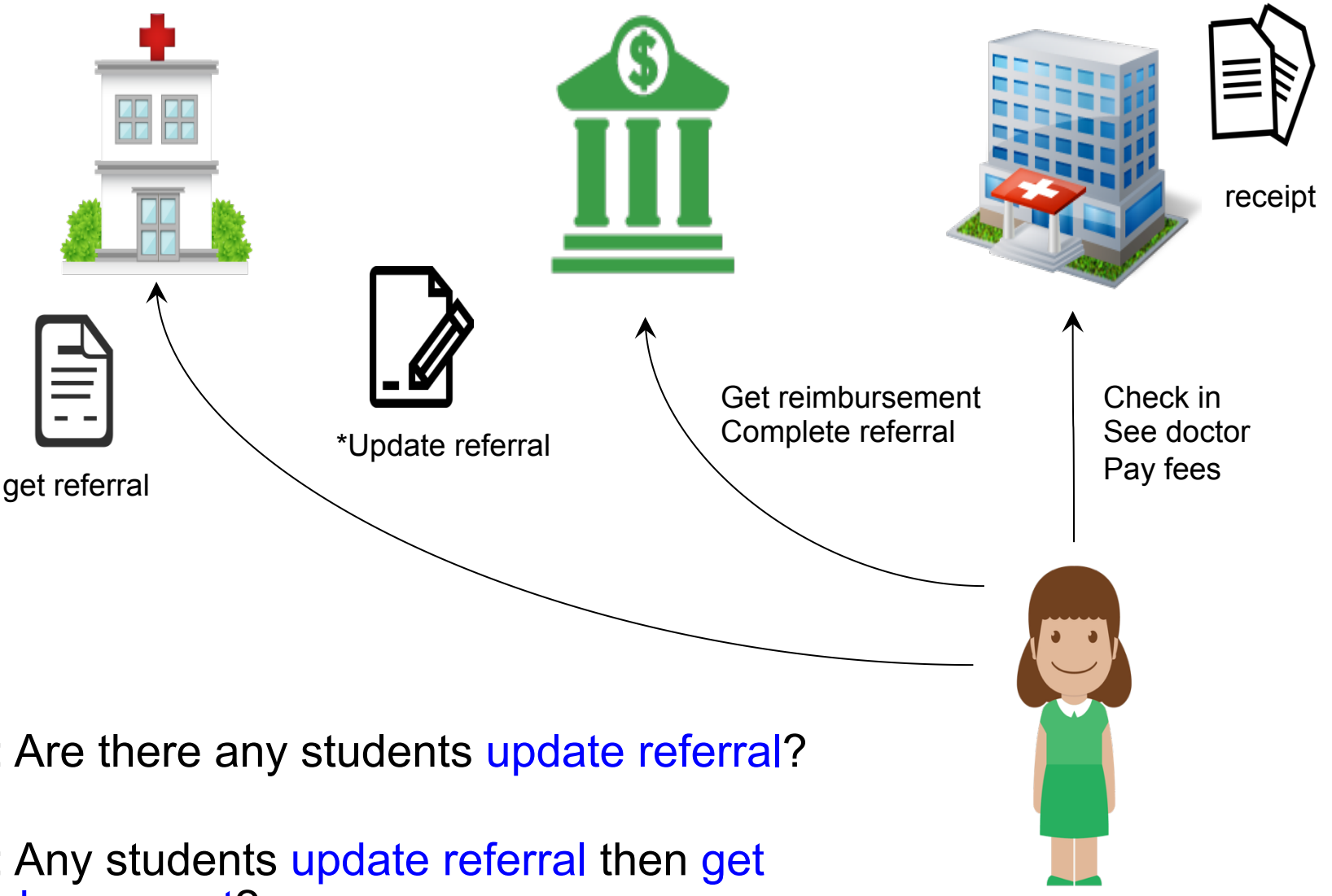
Example: TaoMart

- Whole sales, dominant online sales
- Could like to do market predictions based on past sales
- Fact table: sales transactions
- Dimensions:
  - Shopper demographics
  - Locations
  - Seasons
  - Classifications of goods
  - Possible others
- OLAP: data cube operations

- Does an excellent job for the types of queries, but...

Hospital Referring Example

[Tang-S. PED 17]



Q1: Are there any students update referral?

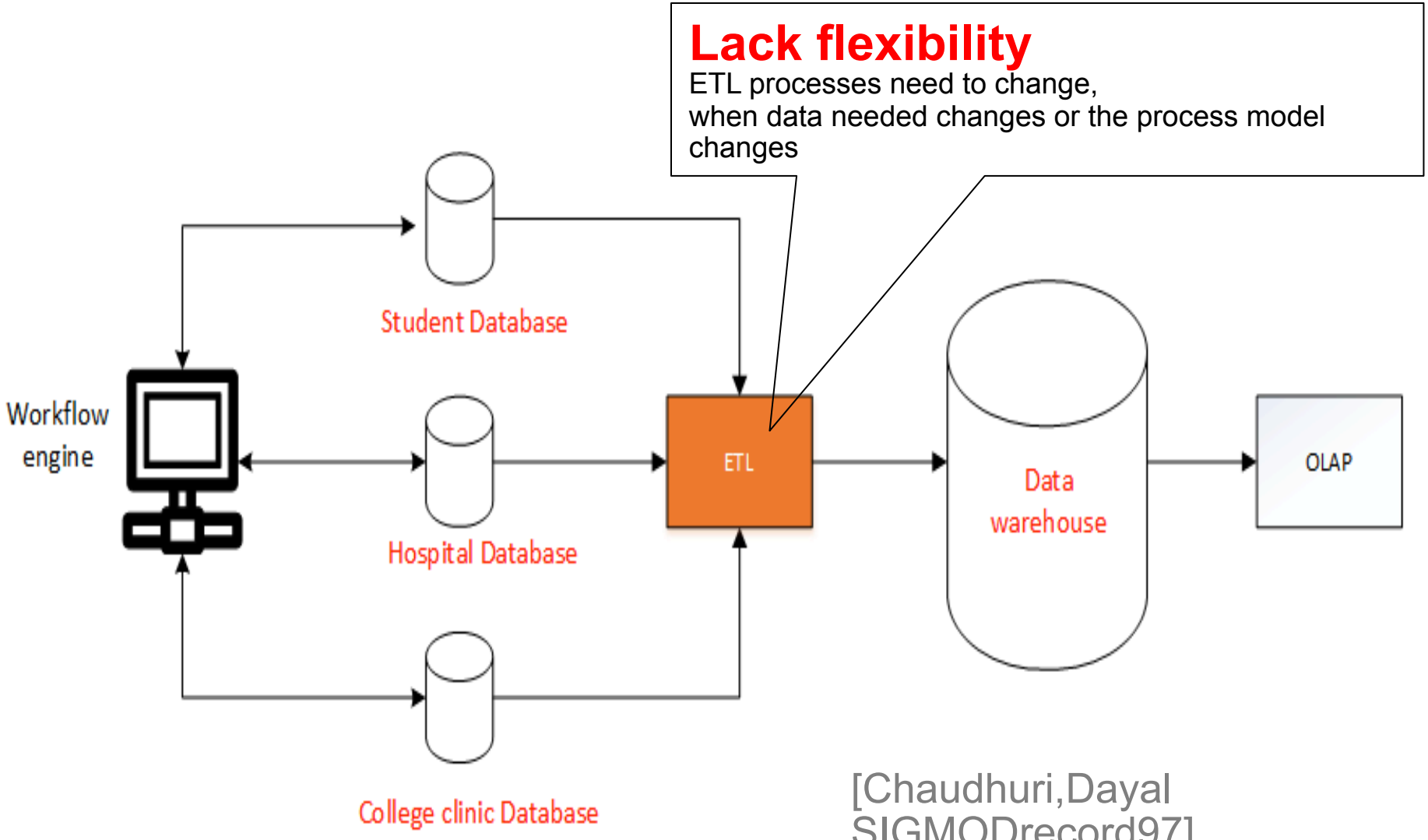
Q2: Any students update referral then get reimbursement?

Weaknesses of ETL-OLAP

- Analysis only limited to data extracted
- Hard to support exploration/ad hoc queries
- Loss of enactment information
  - Only some actions are represented in the loaded data
  - Temporal information is lost
  - Usually no correlations
- When processes change, ETL must be reconfigured

- Also, process mining techniques: no data

Traditional BI Framework is NOT Flexible

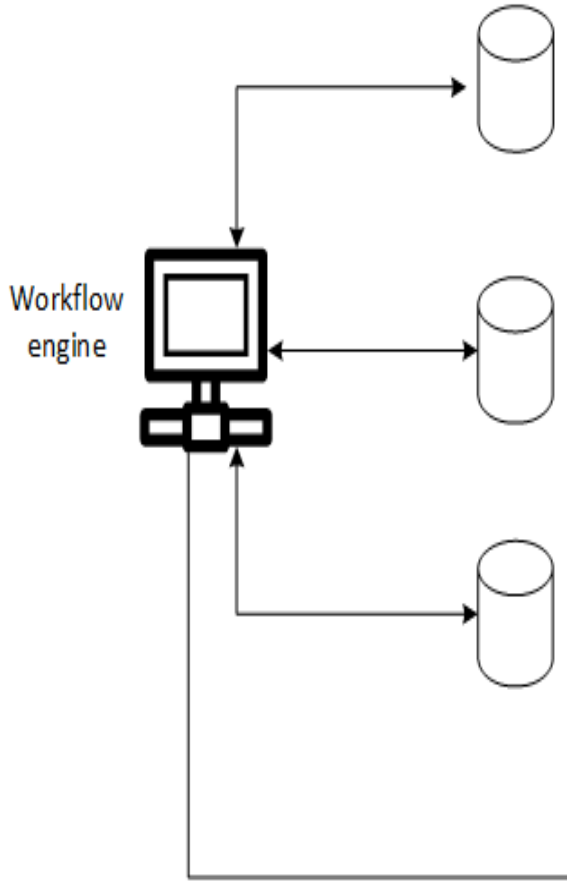


[Chaudhuri,Dayal SIGMODrecord97]
[Vassiliadis IJDWM09]

Outline

- Business Workflow and Analytics (=BI)
- Classical Approach to BI
- Workflow Logs
- Exploration as An Example
- Research Challenges

A Framework for Flexible Process Analytics



Workflow log

General-purpose log structure

Q1: How many students every year get referred with balance larger than 5000?

Q2: Any students update referral then get reimbursement?

Q3: Any students check in with balance larger than 1000, then update the referral, and finally pay additional fees to her/his treatments?

...

No data selection

Easy-to-use query language

Workflow Logs

- Faithfully captures workflow executions
- Includes the following information
  - workflow id
  - workflow instance id
  - activity id
  - activity instance id
  - timestamps
  - data (read/written)
  - correlation information
  - ...
- Granularity of logging

Example Log: A Concrete Log Record

- One log record per activity execution
- Logical timestamp based log sequence numbers

lsn	wid	is-lsn	task name	input map	output map
4	1	3	CheckIn	referId=034d1,referState=start,balance=1000	referState=active

Equivalent json:

```
{
   "referId":"034d1",
   "referState":"start",
   "balance":1000
}
```

Equivalent json:

```
{
   "referState":"active",
}
```

Log Records within One Instance

lsn	wid	is-lsn	task name	input map	output map
1	2	1	start	-	-
5	2	2	GetRefer	-	hospital=”People Hospital”, referId=022f3, referState=start,balance=2000
8	2	3	CheckIn	referId=022f3,referState=start, balance=2000	referState=active
13	2	4	SeeDoctor	referId=022f3, referState=active	-
14	2	5	UpdateRefer	referId=022f3, referState=active, balance=2000	balance=5000
17	2	6	SeeDoctor	referId=022f3, referState=active	-
18	2	7	PayFees	referId=022f3, referState=active	receipt1=4560, receipt1State=active
19	2	8	TakeTreatment	referId=022f3, receipt1=4560	-
20	2	9	GetReimburse	referState=active, balance=5000, receipt1=6560,receipt1State=active	amount=6560, balance=0, reimburse=5000, receipt1State=complete

Log Example: College Hospital Referring Application

lsn	wid	is-lsn	task name	input map	output map
1	1	1	start	-	-
2	2	1	start	-	-
3	1	2	GetRefer	-	hospital=”Public Hospital”,referId=034d1,referState=start,balance=1000
4	1	3	CheckIn	referId=034d1,referState=start,balance=1000	referState=active
5	2	2	GetRefer	-	hospital=”People Hospital”,referId=022f3,referState=start,balance=2000
6	3	1	start	-	-
7	3	2	GetRefer	-	hospital=”Public Hospital”,referId=048s1,referState=start,balance=500
8	2	3	CheckIn	referId=022f3,referState=start,balance=2000	referState=active
9	1	4	SeeDoctor	referId=034d1,referState=active	-
10	1	5	PayFeest	referId=034d1,referState=active	receipt1=560,receipt1State=active
11	1	6	SeeDoctor	referId=034d1, referState=active	-
12	1	7	PayFees	referId=034d1, referState=active	receipt2=460, receipt2State=active
13	2	4	SeeDoctor	referId=022f3, referState=active	-
14	2	5	UpdateRefer	referId=022f3, referState=active,balance=2000	balance=5000
15	1	8	GetReimburse	referState=active, balance=1000,receipt1=560, receipt1State=active, receipt2=460, receipt2State=active	amount=1020, balance=0, reimburse=1000, receipt1State=complete, receipt2State=complete
16	1	9	CompleteRefer	referState=active, balance=0	referState=complete
…	…	…	…	…	…

Outline

- Business Workflow and Analytics (=BI)
- Classical Approach to BI
- Workflow Logs
- Exploration as An Example
- Research Challenges

Example Queries

[Tang-S. 17]

- Find instance ids where activity *UpdateRefer* occurs before activity *CheckIn*

```
FOR      INSTANCE L IN HospitalLog
SELECT X.wid
FROM    UpdateRefer@L X, CheckIn@L Y
WHERE X << Y
```

- Find the value of attribute balance where the activity *GetRefer* occurs before *CheckIn* with balance greater than 500 right after the *getRefer* activity

```
FOR      INSTANCE L IN HospitalLog
SELECT X.wid, X.out.balance
FROM    GetRefer@L X, CheckIn@L Y
WHERE X[balance>500] << Y
```

Incident Query Language

[Tang-S. 17]

- Basic query language implemented
- Optimization based on costs
- Preliminary evaluation study (on the cost most and optimization
- Still to do:
  - Multi-instance, multi-log queries
  - Aggregates

Outline

- Business Workflow and Analytics (=BI)
- Classical Approach to BI
- Workflow Logs
- Exploration as An Example
- Research Challenges

Research Problems and Challenges

- Workflow logs and (re-)construction
- Query languages for workflow logs
- Process mining with data
- Application techniques and tools

Workflow Logs and (Re-)construction

- No current standards on workflow logs/logging
- Is there a universal model for logging?

- Existing systems all have logging utility but
  - Spread through several types of logs (activity, instance, variables, etc.)
  - Sometimes incomplete
- It seems possible to construct workflow log from these logs, general tools and techniques?

Query Languages for Logs

- Development of query languages
  - Understanding the properties, expressiveness, usefulness, etc.
  - Equivalence of log query languages
  - Indexing and optimization techniques
- Aggregation? Multi-instances? Multi-log queries?

Process Mining with Data

- Existing process mining techniques not compatible with data
- Expressions in log query languages provide abstractions for data
  - Possibly combined with existing process mining algorithms

Application Techniques and Tools

- Many existing tools are developed for ad hoc environment
- Reporting tools, e.g.,
  - Ministry of Housing & Urban-Rural Development needs reports from local Housing Management agencies
  - Could be helped by reporting tools based on logs
- Medical fraud detection
- Staff training
- Many other possibilities

Conclusions

- Growing need for business intelligence beyond the traditional types (retail transactions)
- ETL-OLAP is limited
- Workflow logs lead to a general framework for business analytics
- Many research problems and challenges