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
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

- Repair Application
- Repair Assignment
- Post-repair Visit
- Doc. Archive

Workflow Model

- Only activities are present
- Software development

Biz Proc Execution

- Documents, people, devices, resources, collaboration...

Workflow Execution

- Flow mgmt, services, db’s, software, resources...

BTC'17
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?
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

Flow management, services, databases, software, resources, ... only activities are present

Workflow Model

- Repair Application
- Repair Assignment
- Application Review
- Doc. Archive

Software development

Biz Proc Model

- On-site Repair
- Post-repair Visit

Biz Proc Execution

- On-site Repair
- Post-repair Visit

Workflow Execution

documents, people, devices, resources, collaboration, ...

data & documents

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

CBD'17  2017/8/13
Workflow Management System

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

business processes

data management

(workflow)

(workflow management)
Where Are Data in WfM Systems

- **Typical architecture:**

  - WfMS
  - Local data store
  - Execution Engine
  - Task wrapper
  - Task wrapper
  - Task wrapper
  - Enterprise database

  "Includes all data required for control flow decisions, correlations, ..."

- **Logging often an option:**
  
  - Workflow instance, activity (task), variables, ..., etc
  - all logged separately, sometimes in different databases

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

Biz Processes are missing!
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...
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

Lack flexibility
ETL processes need to change, when data needed changes or the process model changes

[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

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

<table>
<thead>
<tr>
<th>lsn</th>
<th>wid</th>
<th>is-lsn</th>
<th>task name</th>
<th>input map</th>
<th>output map</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>CheckIn</td>
<td>referId=034d1,referState=start,balance=1000</td>
<td>referState=active</td>
</tr>
</tbody>
</table>

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

Equivalent json:
```json
{
    "referState":"active",
}
```
## Log Records within One Instance

<table>
<thead>
<tr>
<th>lsn</th>
<th>wid</th>
<th>is-lsn</th>
<th>task name</th>
<th>input map</th>
<th>output map</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>start</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>GetRefer</td>
<td>-</td>
<td>hospital=&quot;People Hospital&quot;, referId=022f3, referState=start, balance=2000</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>CheckIn</td>
<td>referId=022f3, referState=start, balance=2000</td>
<td>referState=active</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>4</td>
<td>SeeDoctor</td>
<td>referId=022f3, referState=active</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>5</td>
<td>UpdateRefer</td>
<td>referId=022f3, referState=active, balance=2000</td>
<td>balance=5000</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>6</td>
<td>SeeDoctor</td>
<td>referId=022f3, referState=active</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>7</td>
<td>PayFees</td>
<td>referId=022f3, referState=active</td>
<td>receipt1=4560, receipt1State=active</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>8</td>
<td>TakeTreatment</td>
<td>referId=022f3, receipt1=4560</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>9</td>
<td>GetReimburse</td>
<td>referState=active, balance=5000, receipt1=6560, receipt1State=active</td>
<td>amount=6560, balance=0, reimburse=5000, receipt1State=complete</td>
</tr>
</tbody>
</table>
# Log Example: College Hospital Referring Application

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

Log: Example: College Hospital Referring Application
Outline

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

- **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

- 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
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