

# Multidimensional analysis and prediction of the OfficeObjects® WorkFlow process performance

Anna Staniszczyk, Witold Staniszkis  
Rodan Development sp. z o.o.

## Abstract

*We present a methodology and the associated business process design tools supporting performance-oriented design of workflow processes supporting the administrative procedures executed in the eGovernment environment. First of all we position our work in the context of the process mining technology concentrating on the process conformance checking and the model enhancement aspects. Further we discuss the OfficeObjects® WorkFlow run-time meta model and the associated business process performance model. Subsequently we present the principal steps of the performance-oriented design methodology discussing a real life performance issue. The principal steps of the methodology entail specification of the OLAP view of the process event logs based on the Mondrane engine executing MDX analytical queries and the business process performance prediction with the use of the MVA queueing network model.*

**Keywords:** Business process performance, process mining, online-analytical processing, queueing network models, mean value analysis (MVA)

## Introduction

The ubiquitous business process management platforms have determined the architecture of the enterprise information systems opening new opportunities in the worker productivity management area. As the result competitive advantage of companies, and indeed of entire nations, depends on efficiency of formally-defined business processes controlling the flow of work activities. Management techniques applying business intelligence tools to business process performance analysis are a pre-condition to succeed in the fierce competition to achieve the higher rung in the productivity ladder.

**Process mining** has emerged as the new technology within the business process management realm. Several important initiatives, such as the Process Management Manifesto [3] and the ensuing published research results [2, 13, 15], provide the methodological foundation for our work. We shall apply the terminology proposed in [2] throughout this chapter.

The primary source of information for analysis of business process performance are the workflow process execution logs correlated with the process definition and runtime metadata models. The workflow process models may be implied by the sequence of **events**<sup>1</sup> recorded in execution logs of interoperating information systems, in the case of absence of a business process management platform. In such case the process mining focus is on the **process discovery**<sup>2</sup> aiming at generation of formal process models expressed usually in such formal notations as Petri nets or BPMN.

The transformation of information systems enterprise architecture into the service oriented architecture (SOA) has been the prevailing trend over the past decade. Hence, we usually deal with situations where there exists a BPM platform providing explicit support for execution of formally specified workflow processes. In such case, we deal with two remaining types of process mining, namely the **process conformance checking**<sup>3</sup> and the **model enhancement**<sup>4</sup>.

Our work presented in this chapter pertains to the two latter process mining types. The event logs generated from OfficeObjects® WorkFlow are constructed in accordance with the process run-time meta model representing an implementation of the WfMC [17] specification. We discuss the conformance checking and the model enhancement in more detail in the ensuing sections. The quality of logs, as discussed in [2], is very high, since they follow a well-defined process meta model and they are stored in the corresponding tables of a relational data base.

---

<sup>1</sup> An action recorded in the log, e.g. the start, completion, or cancellation of an activity for a particular process instance.

<sup>2</sup> One of the three basic types of process mining. A process model to be expressed in a chosen formal notation must be learned based on corresponding event logs. Note, that usually several event logs of different information systems must be correlated and analysed.

<sup>3</sup> Analysing whether reality, as recorded in the corresponding event log, conforms to the model and vice versa.

<sup>4</sup> A process model is extended or improved using information extracted from the corresponding event log.

OfficeObjects® is a proprietary JEE (Java Enterprise Edition) BPM platform comprising several specialized components supporting such functionality as the full text search, business intelligence and reporting, business process management, as well as the portal environment. The OfficeObjects® architecture and application development methodology are presented in [14] and the detailed technical descriptions are published in [8,9].

The focus on human-centric resource utilization analysis stems from our extensive experience in the area of the administrative process management [11, 12] as well as the research and development work pertaining to management of knowledge work processes [21]. Significance of the human-centric business process management has been thoroughly presented by Michael zur Muehlen in [16]. In particular, discussion of the human workflow participant role models and scheduling disciplines are relevant both to our process workload as well as resource performance modelling approaches. Simulation of human-centric workflows has been discussed in [4].

The snowflake data model used in dynamic multidimensional analytical views developed and presented in an OLAP platform [15] with the use of the MDX query language [10] provides the basis for our workflow log data analysis. Any number of multidimensional analytical views may be defined as a result of appropriate transformations of the source event logs to match the objectives of the analysis and the specific performance related queries.

Performance analysis of the business process execution history may be performed from various perspectives depending on its scope and objectives. Typical perspectives may focus on such performance aspects as the **control flow** dealing with ordering and repetitions of process activities, **organizational** pertaining to utilization of human resources and load on elements of the organization structure, **case** representing process instance characteristics such as the path in a process and actors participating in activity roles, as well as the **time** perspective concerning timing and frequency of events.

Note, that the concept of a “**case**” is overloaded with another ubiquitous usage, particularly in the realm of Adaptive Case Management (ACM) [18, 19, 20]. In ACM the concept of a case pertains to a long lived transaction, that may last for many months or years, comprising many business processes and the rich electronic documentation stored in a repository.

The snowflake models developed for analytical processing of the event log data may also provide a basis for calibration of predictive models using queuing network models, such as Mean Value Analysis (MVA) [5], underlying the process performance prediction algorithm provided in the OfficeObjects® WorkFlow Process Designer tool.

### OfficeObjects® WorkFlow run-time meta model

The partial OfficeObjects® WorkFlow run-time meta model shown in Figure 1 provides sufficient data for the **organizational** as well as the **time** analysis presented in this chapter. We use Occam’s Razor<sup>5</sup> to define a generalization of the model sufficient from the vantage point of our discussion.

The subset of the process run-time meta data sufficient to perform analyses discussed in this chapter comprises entity classes representing process instances and the associated manual activities characterised by the timestamp attributes. Each manual activity provides reference to the activity performer, i.e. a person meeting the criteria specified for the participant role, defined with the use of the work participant assignment rules. We limit our perspective to the manual activities of a business process because our focus is on the organisational rather than computing performance. This is consistent with the generally accepted approach discussed extensively in [1, 13, 16].

The automatic process activities are executed by the corresponding application functions or the BPM platform services, thus they are of little interest from the point of view of our analysis. The performance problems that may occur usually stem from the computing bottlenecks, due to hardware configuration congestions or by inefficient software algorithms, and are usually easily alleviated.

The **cost** attribute comprised in the Process Instance and the Manual Activity Instance classes is computed with the use of a cost coefficient to be defined for the corresponding activity participant roles. The cost value computed for a process instance represents a sum of costs of all activity instances comprised in the case, i.e. the path executed within the process graph, where the activity cost attribute is the product of the corresponding cost coefficient and the activity duration.

---

<sup>5</sup> Application of the simplest model that can explain the behaviour of a modelled reality.

The **case** supported by the process instance is modelled by the sequence of the executed manual activities including the repeated and concurrent activities represented by the Transition Instance class appropriately related to the manual activity instances.

Note, that the **duration** attribute of the process instance is not a sum of the corresponding duration attributes accrued by the manual activities comprised in the case, albeit the actual difference with respect to the sum of all activity instances, i.e. the automatic and manual activities, may be sufficiently insignificant to be ignored in the process performance analysis.

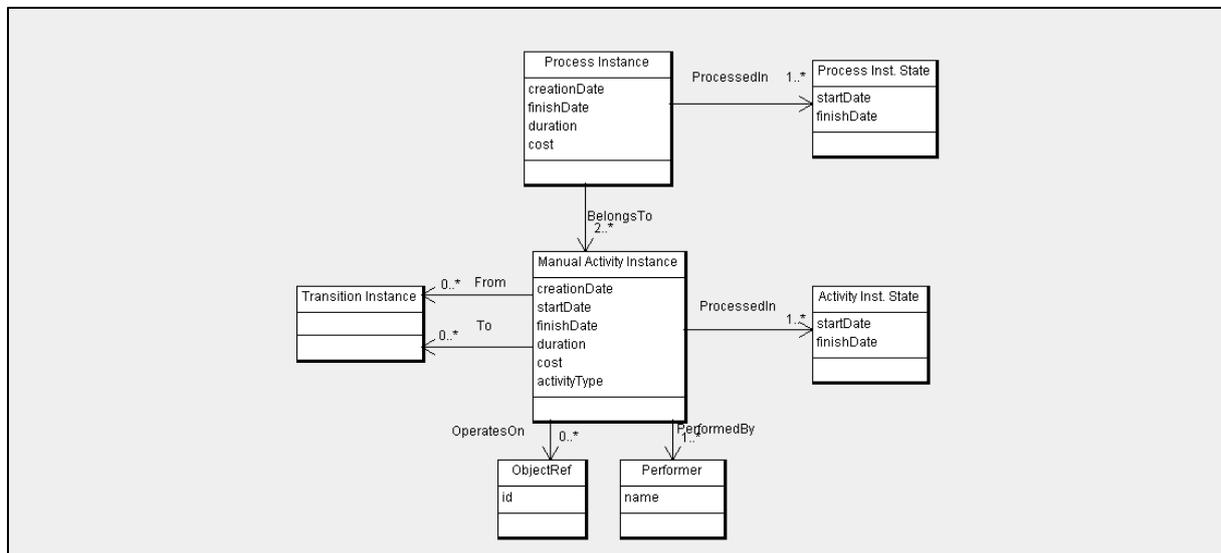


Figure 1. Partial OfficeObjects® WorkFlow run-time meta model

### OfficeObjects® WorkFlow performance evaluation model

The process performance evaluation model defined in [13] closely follows the performance-oriented functions of the OfficeObjects® WorkFlow platform. The performance model presented in Figure 2 is specified as a hierarchy of interacting models representing the resource allocation decisions, the actual execution of the case, modelled respectively by the utilisation of the process resources, and the process enactment model. Note, that consistently with our analysis, the process resources comprise only human resources invoked within the manual activities of a case represented by the corresponding process instance.

The **business process resource allocation model** represents design and management decisions pertaining to the configuration of the human resources within the organisation’s role model used by the work participant assignment rules controlling the business process instance executions. Such decisions are usually determined by the process performance data derived from the analysis of the event logs. They can also be automated with the use of a resource allocation optimisation algorithm interacting with the predictive process performance models.

The **business process resource performance model** represents the human resource configuration within the role model underlying the evaluated business processes and it provides tools to calculate the required performance metrics under the given workload characteristics. The model is presented in more detail in the ensuing section. The formal specification of the resource performance model has been presented in [13].

Roles comprised in the role model may be abstractly viewed as sets of potential work participants selected by the work participant assignment (WPA) rules specified for each manual activity of the process. The actual work assignment is performed by the activity enactment functions of the workflow engine, usually with the use of a semi-random algorithm of a manual activity instance selection by a participant eligible and free to perform the task.

Note, that the participant sets produced by the WPA rules specified within the process manual activity definitions may service several different process activities and any number of concurrent activity instances. The participant sets are not disjoint, hence they may contain the same individuals as the potential participants designated for any number of activities comprised in any process types considered in an performance analysis model.

The WPA characteristics determine the bi-directional workload specification mappings between the business process enactment and the business process resource performance models. The OLAP analytical views provide

means for correct interpretation of the actual process performance data in the context of the process enactment models defined by the BPMN graphs and the associated activity specifications, thus facilitating calibration of the predictive process performance models based on queuing network models (QNM) presented in [13]. The process resource utilisation metrics computed from the vantage point of a QNM service model, shown in Figure 3, discussed in the ensuing section, provide the basis for the resource allocation decisions determining the participant set cardinalities.

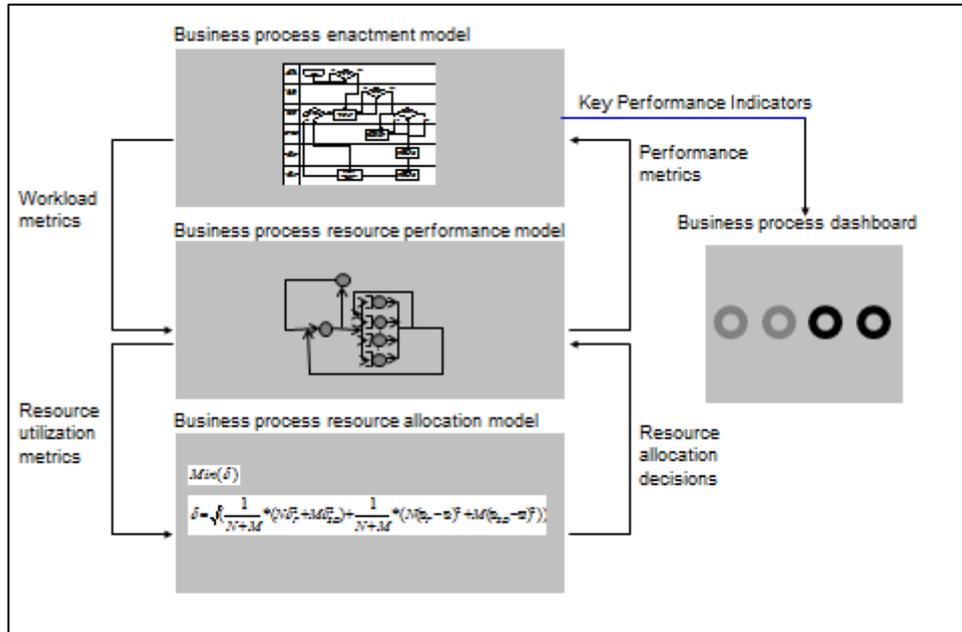


Figure 2. Business process performance-oriented model

The hierarchy of business process performance models facilitates separation of the performance evaluation domains and their respective analysis methods and algorithms. The top level represented by the **business process enactment model** provides means to establish the workload metrics and key performance indicators (KPI's) for each business process class. The KPI's provide input to the **business process dashboard** designed to facilitate monitoring of organization's performance from the business process perspective.

The key performance indicators are metrics representing the application semantics of business processes as defined by the BPMN model and the associated process execution rules.. Such metrics are computed as statistical values providing behavioural characteristics of a population of process instances belonging to a process class over an observation period. Analogically to the balanced scorecard indicators [22], the business KPI's are of interest to the organisation's management and they are usually reported as interactive analytic views presented by the business process dashboard managed by the workflow management system platform.

The following business process KPI's may be derived from the multidimensional process performance analysis model:

- Process-oriented KPI's
  - The number of process instances executed within an observation period
  - Mean duration of the business process within an observation period
  - Min/Max durations of the business process within an observation period
  - Mean cost of the business process within an observation period
  - Clustering of the business process instances by the case category (desirable, acceptable, pathological)
    - Process-oriented KPI's for each case category
- Activity-oriented KPI's
  - Mean frequency (number of executions) of the activity instance within a process instance
  - Mean residence (time in queue + in service) time of the activity for an observation period
  - Min/Max residence times for an observation period

## The process resource performance model

The process resource performance model shown in Figure 3 represents a role-focused view of the human resources, i.e. participants of the manual activities, thus taking the organizational perspective of the business process performance analysis. The role model is orthogonal to the hierarchical organisation chart, in the sense that individuals may play roles enabling them to participate in business process instances independently of their affiliations within the organisation's management structure. The role model is amply discussed in [1,23] and its use in the OfficeObjects® WorkFlow environment has been thoroughly discussed in [8].

The Queueing Network Model (QNM) methodology underlying the predictive performance analysis model of the OfficeObjects® WorkFlow platform is based on the Mean Value Analysis (MVA) queueing network analysis algorithm [5,9]. Mean Value Analysis is a recursive technique for computing expected queue length, residence time at queueing nodes and throughput in equilibrium for a closed separable system of queues. The MVA algorithm and methodology have been initially presented in the context of computer system performance analysis in [25].

The predictive analysis is performed by the OfficeObjects® WorkFlow Process Designer tool resource model with the use of the MVA algorithm parameterized with such variables as the activity service time and the role cardinality, i.e. the expected number of the potential work participants. Matching the role paradigm with the QNM methodology in our multidimensional process performance model facilitates the use of the real performance data for calibration of the MVA predictive model.

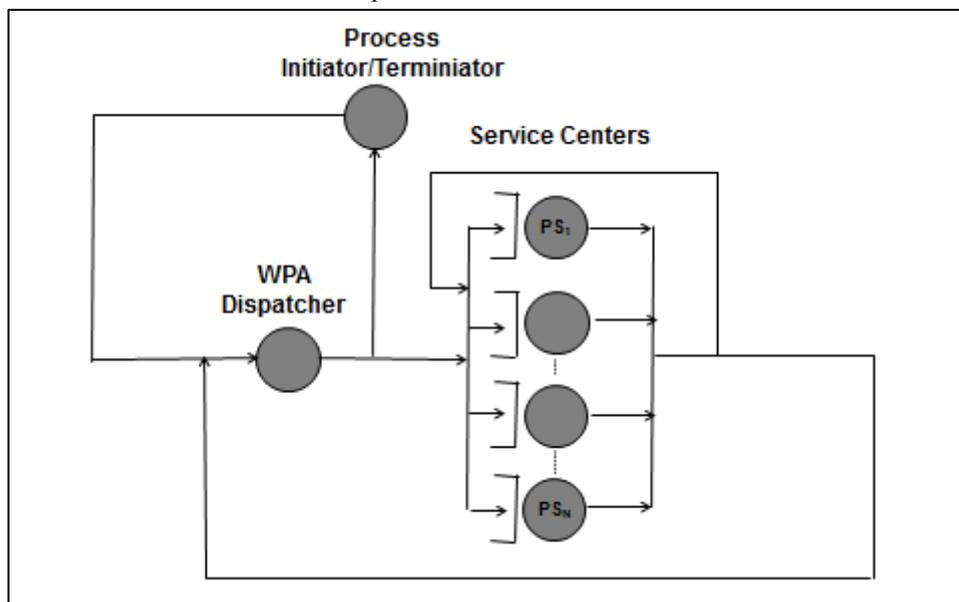


Figure 3. The QNM process resource service model

The human-centric workflow management model is based on an assumption that the work participant assignment is resolved either by a manual decision indicating a participant of the subsequent activity or it is based on an automatic rule selecting the desired role, i.e. the potential participant set, of the activity. At any rate, the queueing network model prove to be a good abstraction for the performance-oriented perspective of the workflow management platform.

The QNM shown in Figure 3 comprises  $N$  service centres representing the roles, for compatibility with our previous discussion called participant sets ( $PS_i$ ), where  $N$  is the number of distinct roles identified within the scope of the performance analysis and  $1 \leq i \leq N$ . As we have mentioned above, any number of activity instances of the same or different processes may call at any of the service centres as many time as required. Also there is a  $N:1$  relationship between the process activity types and the service centres. Hence, it is straightforward to interrelate the performance characteristics of a service centre of the resource service model and the corresponding activities defined in the BPMN model of concurrent process instances evaluated in the performance study.

The service centre, called a WPA Dispatcher, is responsible for selecting the appropriate role service centre for any pending process activity instance. The Process Initiator/Terminator node of the graph initiates or terminated

the process instances. In our predictive MVA model we assume the closed network model parameterized with the number of concurrent process instances.

The performance metrics are computed with the use of the Mean Value Analysis algorithm [5] producing the following estimates for each service centre  $PS_j$ :

- Utilization  $U_j = B_j / T$ , where  $B_j$  is the number of time units the  $j$ -th service centre is busy,  $T$  is the observation period
- Mean residence time  $R_j = (\text{time in service} + \text{time in queue})$  at the  $j$ -th service centre
- Mean queue length  $Q_j = (\text{number of requests in the queue} + \text{the request in service})$  at the  $j$ -th service centre
- Mean process instance cycle time

The MVA model input parameters to be defined in the OfficeObjects® WorkFlow Process Designer tool include the following workload and human resource characteristics:

- Workload parameters:
  - The mean number of concurrent process instances
  - The number of executions (visits) of each manual activity
  - Time required to complete each manual activity
- Process resource parameters:
  - The role model pertaining to evaluated processes
  - The participant set cardinalities for each role
  - The cost coefficient for each role (optional)

The use of queuing network models (QNM), and in particular application of the MVA performance evaluation algorithms, are constrained by formal requirements, i.e. the queuing network model separability constraints, that must be met to obtain mathematically tractable models. The following discussion shows that our process resource performance model meets the QNM separability assumptions defined in [5]:

- *Service centre flow balance assumption* – the number of arrivals at each centre is equal to the number of completions there. This requirement is met by all workflow management systems, since all enacted process activities must be completed.
- *One step behaviour assumption* – no two processes in the system “change state” at exactly the same time. This is clearly a characteristic of all centralised computer systems and also holds for the workflow management systems. In the case of distributed workflow management platforms, we assume such system behaviour, due to the order of magnitude difference of the human service times with respect to state transition functions of the workflow management platform.
- *Routing homogeneity assumption* – the proportion of times that a request completing service at the  $j$ -th centre proceeds directly to the  $k$ -th centre is independent of the current queue lengths at any of the service centres, for all  $j$  and  $k$ . The assumption holds for workflow routing algorithms, which enact the process rules independently of the current business process workload in the system.
- *Service centre homogeneity assumption* - the rate of completion of process activities may vary with the number of tasks at that centre, but otherwise may not be dependent on the number of placements of tasks within the service centre network. The load dependent behaviour usually occurs at service centres below a certain request threshold level, and clearly such behaviour is independent of the workflow situation (i.e. workflow task lists) pertaining to other concrete role participant sets (service centres).
- *Homogeneous external arrival assumption* – the times at which arrivals from outside the network occur may not depend on the number of processes in the network. This is clearly a characteristic of workflow management systems, where the number of process instance enactments is always independent of the current system workload.

## **The multidimensional process performance analysis model**

In this section dealing with the process mining issues, we present the methodology supporting the multidimensional analysis model providing means to produce process performance information useful for calibration and validation of the performance prediction models.

Practical experience, as well as the methodological information that may be found in [2,3,8], indicate that the process event logs may not be sufficient to obtain performance data required. Often a cross-reference with the corresponding workflow platform ontologies is also needed. Due to the limitation of the analytical modelling tools, the workload data must be aggregated and transformed to meet the objectives of a performance study.

The multidimensional workflow process performance analysis model has been constructed as the result of transformation and enhancement of the OfficeObjects® WorkFlow process execution event logs comprising historical data based on the run-time meta model. The data model of the OLAP cube defined with the use of the Mondrane OLAP engine [15] is presented in Figure 4. The cube dimensions support the process performance analysis in terms of the organizational and the time perspectives. The OLAP model provides the basis for definition and materialization of analysis views specified with the use of the Multidimensional Expressions (MDX) queries [10].

The process performance OLAP cube schema, corresponding to the class diagram shown in Figure 4, is presented in Figure 5 and formal definitions of the cube measures in terms of the OfficeObjects® WorkFlow event log attributes are provided in Table 1. A partial presentation of the event log database class diagram is shown in Figure 6.

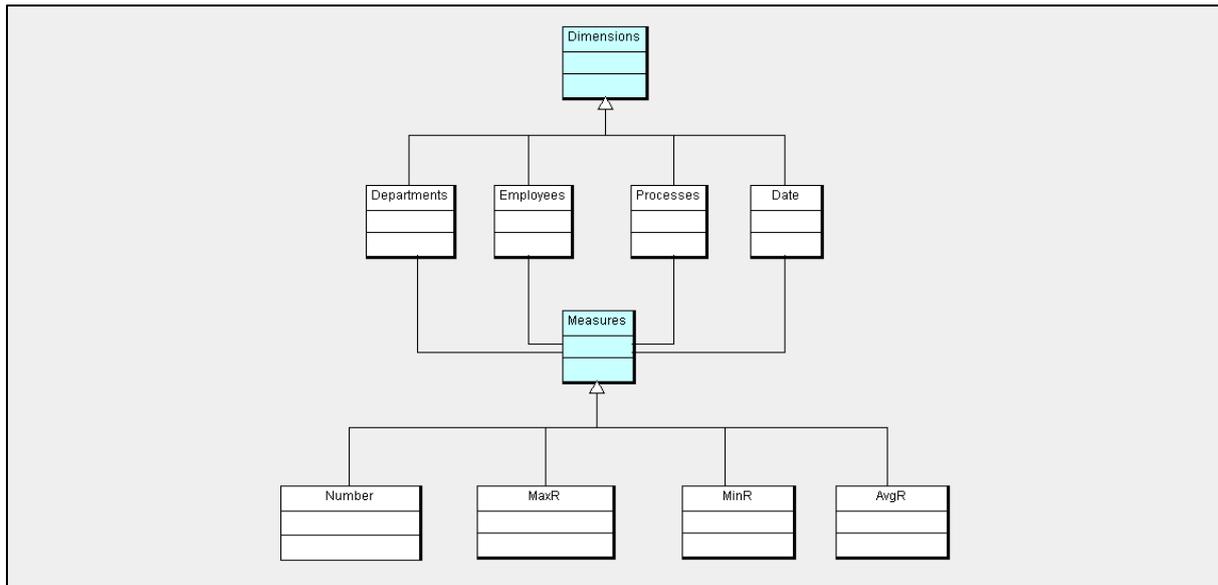


Figure 4. Class diagram of the of the OfficeObjects® WorkFlow event log cube.

The fact table of the *OfficeObjects Workflow Processes* cube has been defined as the view *v\_cz\_finished\_manual* over the event log table of the generic workflow processes. Among others, the analytical view comprises the event log data of the generic workflow process responsible for *distribution and management of incoming documents* (Polish: Proces obsługi korespondencji w komórce) covering the observation period of 5 days between Monday the 5<sup>th</sup> of May 2014 through to Friday the 9<sup>th</sup> of May 2014.

The cube dimensions, as presented in the class diagram presented in Figure 4, comprise the organisational structure (*Departments*), employees (*Employees*), processes (*Processes*), and date (*Date*). Semantics of all syntactic elements of the dimension schema definitions may be found in [15].

The **Departments dimension** is based on the tables *as\_zw\_komorek\_aktywne* and *as\_zw\_komorek\_aktywne\_closure* created in the event log database to represent the organisation structure covered by the OLAP model. The built\_in closure algorithm of the Mondrane platform [15] materialises the dimension hierarchy referencing the cube fact table via the foreign key *icz\_podmiot\_attr2*.

The **Employees dimension** is derived from the fact table by referencing the table column *icz\_podmiot\_nazwa*.

The **Processes dimension** provides a hierarchy *Process->Activity->Instance* based on a join table materialized from *tb\_act\_inst\_finished\_manual* and *tb\_process\_def* base tables by referencing columns *pd\_name*, *icz\_name*, and *icz\_id* respectively.

The **Date dimension** has been defined on the base table *date\_time* as the dimension type *TimeDimension* provided by the Mondrane platform referencing the cube fact table via the foreign keys *icz\_data\_rozp*.

Four cube measures, namely the number of activity executions (*Number*) and the maximal, minimal, and mean residence times, denoted *MaxR*, *MinR*, and *AvgR* respectively, have been defined in the cube schema. The aggregation functions to be used while materializing the analytical view hierarchies are *count*, *max*, *min*, and *avg* respectively. The *icz\_finished* and *icz\_residence\_time* columns of the fact table are used to obtain the measure values.

```

Schema name="docman">
<Cube name="OfficeObjects WorkFlow Processes" cache="true" enabled="true">
<Table name="v_cz_finished_manual" schema="docman"/>
<Dimension name="Departments" foreignKey="icz_podmiot_attr2">
<Hierarchy hasAll="true" allMemberName="All Departments" primaryKey="child_id">
<Table name="as_zw_komorek_aktwyne" schema="docman"/>
<Level name="Department" uniqueMembers="true" column="child_id" nameColumn="name" type="Numeric"
parentColumn="parent_id" nullParentValue="0">
<Closure parentColumn="parent_id" childColumn="child_id">
<Table name="as_zw_komorek_aktwyne_closure"/>
</Closure>
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" name="Employees">
<Hierarchy hasAll="true" allMemberName="All Employees">
<Level name="Employee" column="icz_podmiot_nazwa" type="String" uniqueMembers="true" levelType="Regular"
hideMemberIf="Never"/>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" foreignKey="icz_id" name="Processes">
<Hierarchy hasAll="true" allMemberName="All Processes" primaryKey="icz_id"
primaryKeyTable="tb_act_inst_finished_manual">
<Join leftAlias="tb_act_inst_finished_manual" leftKey="icz_pp_id" rightAlias="tb_process_def" rightKey="pd_id">
<Table name="tb_act_inst_finished_manual" schema="docman"/>
<Table name="tb_process_def" schema="docman"/>
</Join>
<Level name="Process" table="tb_process_def" column="pd_name" type="String" uniqueMembers="true" levelType="Regular"
hideMemberIf="Never"/>
<Level name="Activity" table="tb_act_inst_finished_manual" column="icz_name" type="String" uniqueMembers="true"
levelType="Regular" hideMemberIf="Never"/>
<Level name="Instance" table="tb_act_inst_finished_manual" column="icz_id" type="String" uniqueMembers="true"
levelType="Regular" hideMemberIf="Never"/>
</Hierarchy>
</Dimension>
<Dimension type="TimeDimension" foreignKey="icz_data_rozp" name="Date">
<Hierarchy name="YQMD" hasAll="true" allMemberName="All Dates" primaryKey="time_stp">
<Table name="date_time" schema="docman"/>
<Level name="Year" column="year" type="Numeric" uniqueMembers="true" levelType="TimeYears" hideMemberIf="Never"/>
<Level name="Quarter" column="quoter" type="Numeric" uniqueMembers="false" levelType="TimeQuarters"
hideMemberIf="Never" captionColumn="quoter_label"/>
<Level name="Month" column="month" type="Numeric" uniqueMembers="false" levelType="TimeMonths"
hideMemberIf="Never" captionColumn="month_label"/>
<Level name="Day" column="day" type="Numeric" uniqueMembers="false" levelType="TimeDays" hideMemberIf="Never"/>
</Hierarchy>
</Dimension>
<Measure name="Number" column="icz_finished" datatype="Integer" aggregator="count" visible="true"/>
<Measure name="MaxR" column="icz_residence_time" datatype="Numeric" formatString="#,###0.00" aggregator="max"
visible="true"/>
<Measure name="MinR" column="icz_residence_time" datatype="Numeric" formatString="#,###0.00" aggregator="min"
visible="true"/>
<Measure name="AvgR" column="icz_residence_time" datatype="Numeric" formatString="#,###0.00" aggregator="avg"
visible="true"/>
</Cube>
</Schema>

```

Figure 5. The multidimensional schema of the OfficeObjects® WorkFlow event log cube.

The analysis view is derived from the production event logs of the document management system based on the OfficeObjects® platform installed in an organization responsible for management of roads and parks of the city of Gdańsk. The document flow within the organization is controlled by workflow processes implemented in the system. In order to illustrate the performance analysis methodology, we present an example of the actual performance issue resolved with the use of the performance analysis tools. The process performance data available in the analytical view provided sufficient insight into the performance issues suffered by one of the organization's departments, namely the Roadside Area Lease Department (Polish: Dział Ewidencji Zajęć Pasa Drogowego).

The top screen view of the *OfficeObjects® WorkFlow Processes* OLAP model is presented in Figure 7. All cells marked with the  $\oplus$  sign may be expanded to provide more detailed view supporting the drill down feature of the analytical view.

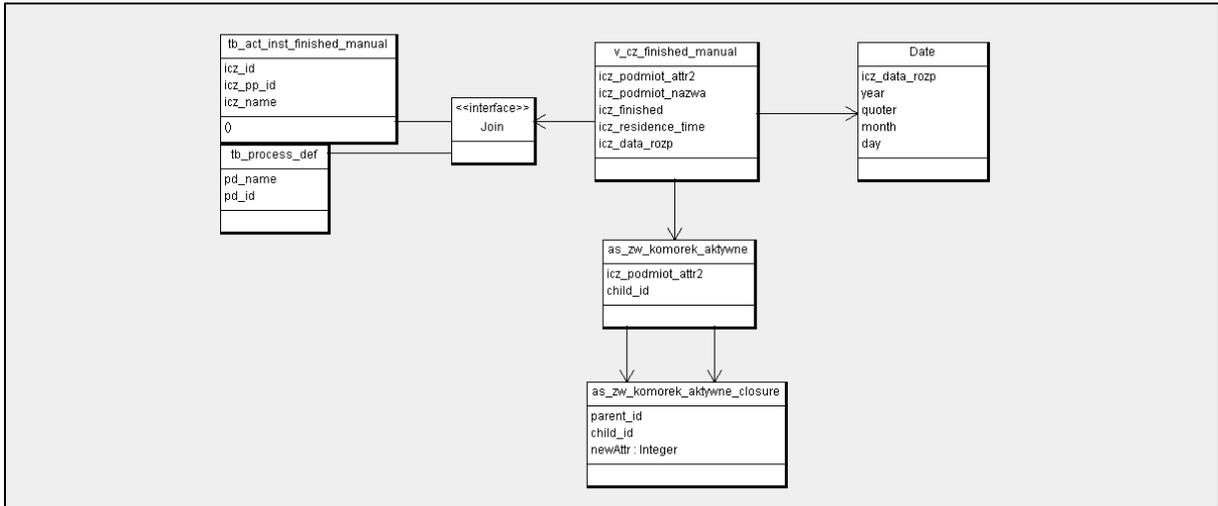


Figure 6. The event log database logical structure

The analytical view displays the top level screen of the organisational structure, as well as the workflow processes and the employees, corresponding to the dimensions of the cube schema, with the appropriate drill down markings and the measure values aggregated accordingly.

The indicated top row of measure values comprises the top level aggregation, for all processes and employees, corresponding to the number of workflow tasks executed during the observation period as well as the aggregations of the residence times. Expansions of the view may either be effected by clicking at the selected + expansion marks or by defining the MDX query. The query editor icon is indicated in Figure 7 and the MDX query generating the analytical view screen shown in Figure 9 is presented in Figure 8.

The query expression selects the required measures to be presented at various levels of aggregation in the analytical view columns. The rows comprise the levels of dimensions defined in the cube schema invoked by the query. Information helpful for MDX code reading may be found in [10]. The result of the query representing the required aggregation of the analytical view measures is shown in Figure 9.

The analytical view aggregated at the department, process, and activity levels for all employees provides sufficient data, pertaining to executions and the corresponding activity residence times, required to parameterize the predictive performance model used to resolve the target performance issue.

Dimensions				Measures			
Departments	Department	Processes	Employees	Number	MaxR	MinR	AvgR
All Departments	DN - Dyrektor	All Processes	All Employees	2 722	34,83	0,01	0,17
	DN - Dyrektor	All Processes	All Employees	586	11,74	0,01	0,15
	DU - Zastępca Dyrektora	All Processes	All Employees	367	2,91	0,01	0,02
	NA - Dział Administracyjno - Osobowy	All Processes	All Employees	84	10,56	0,01	0,16
	NI - Dział Inwestycji	All Processes	All Employees	68	0,05	0,01	0,01
	NZ - Dział Zamówień Publicznych	All Processes	All Employees	465	20,76	0,01	0,23
	NF - Główny Księgowy	All Processes	All Employees	1 102	34,83	0,01	0,21
	DZ - Z.ca Dyrektora ds. zarządzania	All Processes	All Employees	22	1,60	0,01	0,09
	NG - Dział Gospodarki Odpadami Komunalnymi	All Processes	All Employees				

Figure 7. The top screen of the multidimensional analytical view

The measure values representing the mean residence times of the activities executed in the target workflow process, indicated in Figure 9, have been used within the predictive model discussed in the ensuing section.

Information characterizing the actual cases serviced by the target process provided the basis for selection of the partial workflow models. The actual mean residence times, equal to 36, 252, 2484 seconds respectively, have been mapped to the service centres A1, A2, A3, shown in Figure 11, derived from the OLAP model.

Measure	Definition	Formula
Number	The number of the i-th activity instances executed by the j-th employee	$K_{ij}$
MaxR	Maximum residence time of an activity enacted by the employee	$MAX(RT_{ij} = finishDate_{ijk} - startDate_{ijk})$
MinR	Minimum residence time of an activity enacted by the employee	$MIN(RT_{ij} = finishDate_{ijk} - startDate_{ijk})$
AvgR	Mean residence time of an activity enacted by the employee	$AVG((RT_{ij} = finishDate_{ijk} - startDate_{ijk})/K_{ij})$
	Where: M is the number of activities executed in the observation period N is the number of employees executing process activities $K_{ij}$ is the number of i-th activity instances executed by the j-th employee i, such that $1 \leq i \leq M$ , denotes the i-th activity j, such that $1 \leq j \leq N$ , denotes the j-th employee k, such that $1 \leq k \leq K_{ij}$ , denotes the k-th instance of the i-th activity executed by the j-th employee finishDate, startDate and creationDate are attributes of the Manual Activity Instance class The residence time is the time a service request spends in queue plus the time of execution by the service centre	

Table 1. Definitions of the OfficeObjects® WorkFlow event log cube measures.

**OfficeObjects® WorkFlow Processes**

**MDX Query Editor**

```

select {[Measures].[Number], [Measures].[MaxR], [Measures].[MinR], [Measures].[AvgR]} ON
COLUMNS,
NON EMPTY Hierarchize(Union(Crossjoin([Departments].[DN - Dyrektor].[DU - Zastepca
Dyrektora].Children, {[Processes].[All Processes], [Employees].[All Employees]})),
Union(Crossjoin([Departments].[DN - Dyrektor].[DU - Zastepca Dyrektora].[ZA - Dzial Ewidencji i
Zajec Pasa Drogowego], Crossjoin([Processes].[All Processes].Children, {[Employees].[All
Employees]})), Crossjoin([Departments].[DN - Dyrektor].[DU - Zastepca Dyrektora].[ZA - Dzial
Ewidencji i Zajec Pasa Drogowego], Crossjoin([Processes].[Proces obslugi korespondencji w
komorce].Children, {[Employees].[All Employees]})))))) ON ROWS
from [OfficeObjects WorkFlow Processes]
  
```

Figure 8. The MDX query materializing the analytical view shown in Figure 9

**OfficeObjects® WorkFlow Processes**

Dimensions							Measures				
Departments	Department			Processes	Process	Activity	Employees	Number	MaxR	MinR	AvgR
All Departments	DN - Dyrektor	DU - Zastepca Dyrektora	ZA - Dzial Ewidencji i Zajec Pasa Drogowego	All Processes			All Employees	587	14.00	0.01	0.09
					Proces obslugi korespondencji w komorce		All Employees	297	14.00	0.01	0.16
						Dekretacja w komorce	All Employees	124	0.02	0.01	0.01
						Obsluga korespondencji (prac. wiodacy)	All Employees	57	14.00	0.01	0.69
						Rejestracja w sprawie	All Employees	116	2.51	0.01	0.06
						Proces obslugi korespondencji wychodzacej	All Employees	290	0.01	0.01	0.01
			UC - Dzial Oczyszczania	All Processes			All Employees	38	7.12	0.01	0.20
			ZD - Dzial Uzgodnień	All Processes			All Employees	220	11.51	0.01	0.14
			UE - Dzial Energetyczny	All Processes			All Employees	93	4.50	0.01	0.07
			ZI - Dzial Inzynierii Ruchu	All Processes			All Employees	220	34.83	0.01	0.62

Figure 9. The multidimensional workflow process performance analytical view.

## Process performance prediction and optimization model

The process performance prediction and optimization analysis methodology entails the following principal steps performed iteratively:

- Identification of significant cases (process execution paths)
- Parametrization of the abstract process models
- Mapping the abstract process models onto the QNM representations
- Performing the MVA algorithm
- Selection of the “best” target process design

The process targeted by the performance analysis has been based on a generic workflow process responsible for distribution and management of incoming documents (Polish: Proces obsługi korespondencji w komórce) pertaining to the roadside sites managed by the department. The birds eye view of the complete process model and the derived analysis and optimization models are presented in Figure 10.

The BPMN models have been developed with the use of the OfficeObjects® Process Designer tool and parametrized by the performance measures obtained from the OLAP analytical view provided the basis for the ensuing process optimization design.

**Identification of significant cases (process execution paths)** is based on the analytical view data obtained from the event log OLAP cube. As defined above, a case instance is represented by a collection of manual activities performed by system users, i.e. the target department employees, participating in various roles in process instances executed during the observation period. In our analysis the target case comprises three process activities, namely the “Analysis and responsible officer assignment (Polish: Dekretacja w komórce)” activity, the “Case categorization and registering (Polish: Rejestracja w sprawie)”, and the „Decision taking (Polish: Obsługa korespondencji)” activity.

Note, that we have obtained two abstract process models (b) and (c), representing the actual process model, represented in the OLAP analytical view, and the optimized process model respectively. Both models have been derived from the complete process model (a) comprising many more manual and automatic activities on the basis of the OLAP analysis results indicating the manual activities actually performed in the analysed cases.

**Parameters of the abstract process models** are derived from the OLAP analytical view data, augmented with additional log data such as the number of concurrent process instances occurring on average during the observation period. Two log table views stored in the relational database containing the activity execution data and the process execution data pertaining to the target department work performed during the observation period have been analysed. The first provided input to the OLAP cube as shown in Figure 5 comprising the cube schema, and the latter has been analysed with the use of SQL queries to establish the mean number of concurrent processes.

The process log table view has been used to build a histogram representing the average numbers of concurrently executed processes for each working day of the observation period from Monday the 5<sup>th</sup> of May 2014 through to Friday the 9<sup>th</sup> of May 2014. The rounded average number of concurrent processes was equal to 43 for the target observation period.

The numbers of activity instances occurring during the observation period for each activity identified within the case relative to the number of the completed process instances served for calculation of routing probabilities for the abstract process models as shown in Figure 10.

The routing probabilities defined for an abstract process model determine the number of executions (visits) of each activity within one instance of the process required as the workload characterisation parameters by the MVA model.

The following algorithm, first published in the context of the network database data manipulation algorithm [24], has been used to establish the number of visits at each service centre representing the model process activities for one process instance calculated on the basis of the routing probabilities.

Let us have a flow graph comprising  $k$  nodes, where the  $k$ -th node represents the STOP/START activity terminating the process instance. The control flow in the graph is represented by the branching probability matrix  $P$ . The branching probability  $p_{ik}$  represents the fraction of requests to terminate the process directly from the  $i$ -th node and the branching probability  $p_{kj}$  indicates the fraction of input requests to be first routed to the  $j$ -th node.

The matrix P provides sufficient information to calculate the number of visits at each of the flow graph nodes, thus establishing the execution frequencies, ie. the number of visits, of each of the abstract process model activities.

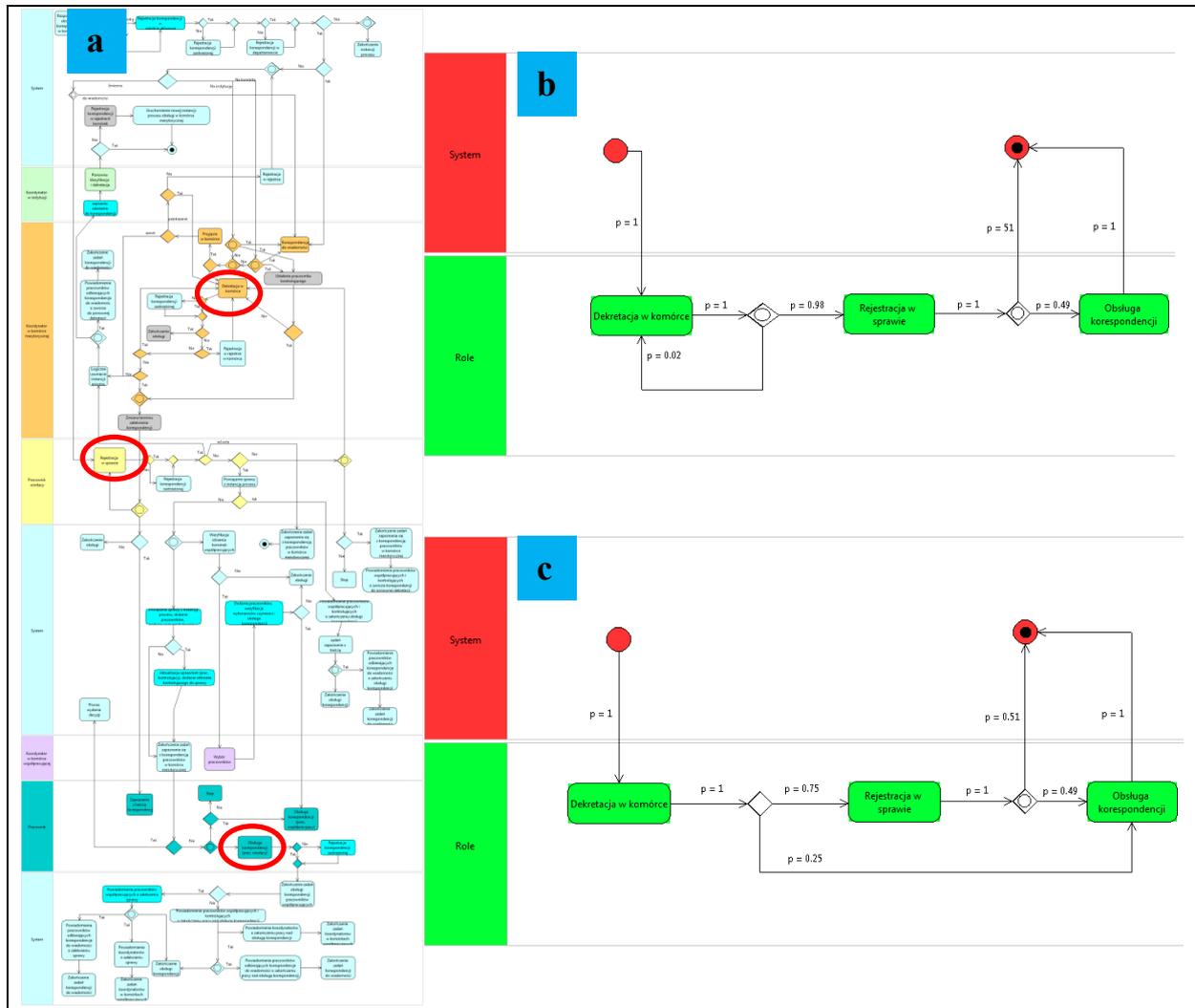
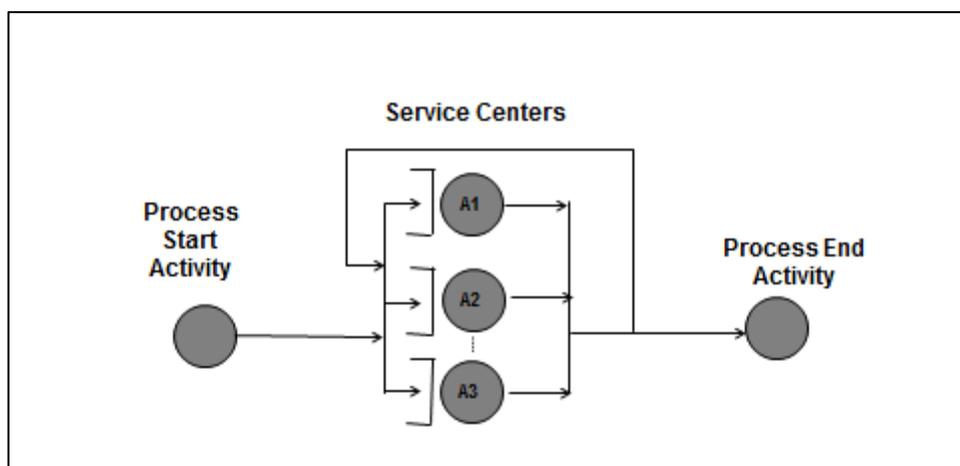


Figure 10. Optimized workflow process BPMN models.



- A1 – Analysis and responsible officer assignment (Polish: Dekretacja w komórce)
- A2 – Case categorization and registering (Polish: Rejestracja w sprawie)
- A3- Decision taking (Polish: Obsługa korespondencji)

Figure 11. The resource service model generated from the abstract BPMN model.

The arcs of the flow graph labelled with the branching probability values, as shown in Figure 10, represent the control flow in the abstract process model representing the analysed case. The branching probability value  $p_{ij}$  represents the fraction of requests proceeding next to the  $j$ -th activity on completion of the  $i$ -th activity.

Under the assumption that the graph is operationally connected, that is each graph node is visited at least once during the process execution, and it complies to the balanced flow principle meaning that no requests are lost, we may calculate the number of visits at each flow graph node by resolving the following system of balanced flow equations:

$$V_j = \sum_{i=1}^k V_i * p_{ij}, \quad j = 1, \dots, k - 1$$

$$V_k = 1$$

Where  
 $V_j$  is the number of visits at the  $j$ -th node of the flow graph  
 $p_{ij}$  is the branching probability from the  $i$ -th to the  $j$ -th node

**Mapping the abstract process models onto the QNM representations** is performed by the following semi-automatic algorithm using the OLAP analytical view presented in

Figure 9:

1. Identification of the potential participant sets, i.e. roles, for each of the activities of the abstract process model
2. Grouping process activities by role-affinity, i.e. by common roles
3. Defining the MVA model service centres for each role-affinity group with the following performance measures:
  - a. The mean service time (seconds)
  - b. The number of service centre participants
  - c. The number of visits
4. All of the measures (a) and (b) are computed for each role-affinity groups, i.e. for each service centre, as the average values weighted with the relative frequencies of the group members. The measure (c) is the sum of the numbers of visits of all activities belonging to the role-affinity group.

In the case of our example all activity roles are disjoint, hence the mapping of the performance measures onto the service centres is straightforward.

The graphic representation of the mappings pertaining to the abstract process models (b) and (c) is shown in Figure 11. The mapping is controlled by the *Role*, i.e. the potential activity participant set, underlying the MVA model Service Centres. The exhaustive specification of the mapping model is presented in [13]. The use of Occam's Razor to derive the abstract process models does not inhibit the performance prediction results. In fact an attempt to parameterize the entire process BPMN model is impractical, due to the lack of the actual measures as well as the complexity of the routing connections.

**Performing the MVA algorithm** is triggered by the "simulation" function of the OfficeObjects® Process Designer tool computing the performance quantities shown in Table 2. Three steps of performance optimisation, corresponding to the base line process (b) with results established in Step 1, and to two optimized processes based of the abstract models (b) and (c), as shown in the results of the Step 2 and Step 3 respectively.

**Selection of the "best" target process design** entails analysis of performance prediction quantities generated by the MVA model. The abstract process model (b) shown in Step 1 corresponds to the actual state of the principal process (a) and the human resource allocations within each Role derived from the OLAP view. The number of visits characterize the workload on the service centres of the closed queuing network processing 43 concurrent processes.

The abstract model (b) has been derived from the process model (a) on the basis of the case information pertaining to the target department obtained from the OLAP model. In fact only three manual activities have been visited by all observed processes. The routing probabilities have been derived from the case log data presented by the OLAP model. For the 43 concurrent processes the service centres A2 and start/stop have been visited once for each process execution. The service centre A1 has been visited 43.88 times due to the officer assignment errors that forced repetition of 2% of the assignment tasks. All of the assignment errors occurred

when an account file had already existed. The service centre A3 has been visited 21.07 times due to the fact that 51% of cases did not require any further activity after being registered in the respective account files. The service centre A3 is the process bottleneck, notwithstanding the reduced visit rate, due to the highest mean service time amounting to 2484 seconds and the utilization equal to 100%. As the result the process cycle time has been estimated as 4,85 hours.

Activity	Role Cardinality	Visits	Residence Time (h)	Queue Length	Utilization %
<b>Step 1 - Process (b)</b>					
Process start		43.00	0	0	0,00
A1	1	43.88	0.01	0,10	10
A2	5	43.00	0.02	0.14	14
A3	3	21.07	9.83	42.75	100
Process end		43,00	0	0	0,00
<b>Process cycle time (h): 4,85</b>					
<b>Step 2 - Process: (b)</b>					
Process start		43.00	0	0	0
A1	1	43.88	0.01	0.17	17
A2	1	43.00	2.84	40.60	100
A3	7	21.07	0.32	2.23	100
Process end		43,00	0	0	0
<b>Process cycle time (h): 3.01</b>					
<b>Step 3 - Process: (c)</b>					
Process start		43.00	0	0,00	0
A1	1	43.00	0.01	0.20	20
A2	1	32.25	0.50	6.21	100
A3	7	26.55	3.61	36.59	100
Process end		43.00	0	0,00	0
<b>Process cycle time (h): 2.62</b>					

Table 2. Process optimization results.

Note, that the process cycle time represents the time period, when the process activities are either in service or in queue at any of the service centres. We do not account for the time lag experienced by the process while activities are in the ready state corresponding to time period between an activity being created, i.e. put into all task lists of potential work participants of the corresponding role, but has not been selected for execution by any of them.

Two subsequent optimization steps, based on the abstract cases (b) (c), entail optimization of the abstract model (b) respectively by modifying role cardinalities, and by further redesign of the process graph as shown in the abstract process model (c). In the first case improvement of process cycle time amounted to 38% with respect to Step 1.

The process model optimisation resulting in modification of the process graph took advantage of the fact, that 25% of cases pertained to existing customers and as the result they could be automatically registered in the corresponding account files. The automatic activity registering the cases eliminated errors in case registration experienced by 2% of cases in Step 1. The optimisation step rendered further improvement in the process cycle time amounting to 13%.

## Conclusions

The presented performance evaluation and prediction methodology and tools based on the OLAP analytical views and the QNM prediction models represent a powerful technique of the workflow process performance-oriented design.

The analytical models employed for performance predictions, although less precise with respect to stochastic simulation models, prove sufficient in many design situations. The lack of precision is more than offset by the ease of use in terms of model parametrization and design.

Combination of the multidimensional performance data analysis with the power of MVA performance prediction has, so far, been sufficient in most performance optimization projects in the OfficeObjects® WorkFlow environments.

## References

- [1] Van der Aalst, W., Van Hee, K., "Workflow Management, Models, Methods, and Systems.", The MIT Press, 2002.
- [2] Van der Aalst, W., Process Mining – Discovery, Conformance, and Enhancement of Business Processes, Springer Verlag, Berlin, 2011.
- [3] Van der Aalst, W., et al., Process mining manifesto, in BPM 2011 Workshop Proceedings, Springer-Verlag, 2011.
- [4] T. Jaeger, A. Prakash, Management and Utilization of Knowledge for the Automatic Improvement of Workflow Performance, Proc. of the ACM COOCS '95, California, USA, ACM 1995.
- [5] Lazowska E.D., Zahorian, J., Graham, G.S., Sevcik, K.C., Quantitative System Performance, Computer System Analysis Using Queuing Network Models, Prentice-Hall, Inc., USA, 1984.
- [6] J.A. Miller, A.P. Sheth, K.J. Kochut, X. Wang, A. Murugan, Simulation Modeling within Workflow Technology, Proc. of the 1995 Winter Simulation Conference, ACM, 1995.
- [7] Object Management Group, Business Process Modeling Notation (BPMN), version 1.0, February 2006.
- [8] OfficeObjects® WorkFlow BPM Handbook, Rodan Systems S.A., Warsaw, 2005
- [9] OfficeObjects® Technical Documentation Set, Rodan Systems S.A., Warsaw, 2013.
- [10] Spofford, G., MDX Solutions, John Wiley & Sons, Inc. USA, 2001.
- [11] Staniszki W., Supporting Administrative Knowledge Processes, EGOV 2002, Springer Verlag LNCS 2456.
- [12] Staniszki W., Staniszki, E., Administrative Process Management – Setting the Scene, in Knowledge Transfer in eGovernment, R. Traunmueller (Ed.), Trauner Verlag, Austria, 2006.
- [13] Staniszki W., Modelling Business Process Performance, in Proc. of the Int. Conference on Business Process and Services Computing BPSC '07, Leipzig, Germany, September 2007.
- [14] Staniszki W., Empowering the Knowledge Worker – End-User Software Engineering in Knowledge Management, Proc. of the 17<sup>th</sup> International Conference on Enterprise Information Systems, Barcelona, Spain, May, 2015.
- [15] Mondrian 3.0.4 Technical Guide, Pentaho Org., USA, 2009
- [16] M. zur Muehlen, Organizational Management in Workflow Applications – Issues and Perspectives, in Information Technology and Management, (5) 2004, Kluwer Academic Publishers, The Netherlands, 2004.
- [17] Workflow Management Coalition; Workflow Reference Model; WfMC-TC-1003, version 1.1, Jan 1995.
- [18] Palmer, N., BPM and ACM, in Taming the Unpredictable – Real World Adaptive Case Management: Case Studies and Practical Guidance, L.Fischer (Ed.), Future Strategies Inc., USA, 2011
- [19] Palmer, N., Case Management Megatrends, in How Knowledge Workers Get Things Done – Real World Adaptive Case Management, L.Fischer (Ed.), Future Strategies Inc., USA, 2012
- [20] Palmer, N., Where is ACM Today? Realities and Opportunities, in Empowering Knowledge Workers – New Ways to Leverage Case Management, L. Fischer (Ed.), Future Strategies Inc., USA, 2014.
- [21] ICONS Project IST 2001-32429, The ICONS Project Results Report, 2004, <http://www.rodan.pl>
- [22] Kaplan, Robert S; Norton, D. P. (1996). The Balanced Scorecard: Translating Strategy into Action. Boston, MA.: Harvard Business School Press. [ISBN 978-0875846514](https://doi.org/10.1007/978-0-87584-651-4).
- [23] Harrison-Broninski, K.D., Human Interactions – The Heart and Soul of Business Process Management, Meghan-Kiffer Press, USA, 2005
- [24] Orlando, S., Perri, V., Scrivano, S., Staniszki, W., Database Analyser And Predictor – An Overview, Proceedings of the Fifth International Conference on Data Engineering, February 6-10, 1989, Los Angeles, California, USA.
- [25] Reiser, M.; Lavenberg, S. S. (1980). "Mean-Value Analysis of Closed Multichain Queuing Networks". Journal of the ACM 27 (2): 313.