|  |  |  |
| --- | --- | --- |
|  | Scalable Data Analytics,  Scalable Algorithms, Software Frameworks  and Visualization ICT-2013 4.2.a | |
| |  |  | | --- | --- | | Project | **FP6-619435/SPEEDD** | | Deliverable | **D6.1** | | Distribution | **Public** | | |  |
|  | |  |
|  | |  |



http://speedd-project.eu

**The Architecture Design of the SPEEDD Prototype**

Alex Kofman, Fabiana Fournier, …

Status: Draft (Version 0.1)

October 2014

|  |  |
| --- | --- |
| **Project** |  |
| Project Ref. no | FP7-619435 |
| Project acronym | SPEEDD |
| Project full title | Scalable ProactivE Event-Driven Decision Making |
| Project site | http://speedd-project.eu/ |
| Project start | February 2014 |
| Project duration | 3 years |
| EC Project Officer | Aleksandra Wesolowska |
|  |  |
| **Deliverable** |  |
| Deliverable type | Report |
| Distribution level | Public |
| Deliverable Number | D6.1 |
| Deliverable Title | The Architecture Design of the SPEEDD Prototype |
| Contractual date of delivery | M9 (October 2014) |
| Actual date of delivery | October 2014 |
| Relevant Task(s) | WP6/Tasks 6.3 |
| Partner Responsible | IBM |
| Other contributors | NCSR, CNRS, FeedZai, ETH, UoB, Technion |
| Number of pages |  |
| Author(s) |  |
| Internal Reviewers | TBD |
| Status & version | Draft |
| Keywords | TBD |
|  |  |

|  |
| --- |
| **Executive Summary** |

In 1-2 pages give a

- 1-sentence punch line about the contribution of the deliverable

- 1-paragraph presentation of the project goals and the work package goals

- 1-paragraph description of the work presented in the deliverable

- 1-paragraph presentation of how the project (and the state-of-the-art if relevant) benefits from the work.

- 1-paragraph description of the main results/findings of the deliverable

- 1-paragraph presentation of the work to follow, based on the deliverable

|  |
| --- |
| **Contents** |

[1 Introduction 6](#_Toc399700440)

[1.1 History of the document 6](#_Toc399700441)

[1.2 Purpose and Scope of the Document 6](#_Toc399700442)

[1.3 Relationship with Other Documents 6](#_Toc399700443)

[2 Main Sections 7](#_Toc399700444)

[2.1 System Requirements 7](#_Toc399700445)

[2.2 Approach 7](#_Toc399700446)

[2.3 Conceptual Architecture 8](#_Toc399700447)

[2.4 SPEEDD Runtime Architecture 10](#_Toc399700448)

[2.4.1 Event Bus 12](#_Toc399700449)

[2.4.2 Event/Data Providers 12](#_Toc399700450)

[2.4.3 Action Consumption – Actuators/Connectors 12](#_Toc399700451)

[2.4.4 Complex Event Processor 12](#_Toc399700452)

[2.4.5 Decision Management 12](#_Toc399700453)

[2.4.6 Dashboard application 12](#_Toc399700454)

[2.5 Build-Time Architecture 12](#_Toc399700455)

[2.5.1 Event Pattern Mining 13](#_Toc399700456)

[2.5.2 Authoring of CEP Rules 13](#_Toc399700457)

[2.5.3 Decision Management – the Offline Part 13](#_Toc399700458)

[2.6 Integration – APIs and Data Formats 13](#_Toc399700459)

[2.7 Non-Functional Aspects 13](#_Toc399700460)

[2.8 Scalability 13](#_Toc399700461)

[2.9 Fault Tolerance 13](#_Toc399700462)

[2.10 Testability 13](#_Toc399700463)

[3 Conclusions 14](#_Toc399700464)

[4 Appendix – Technology Evaluation 14](#_Toc399700465)

[4.1 Stream Processing – requirements and evaluation criteria 14](#_Toc399700466)

[4.2 Storm 14](#_Toc399700467)

[4.3 Akka 14](#_Toc399700468)

[4.4 Spark Streaming 14](#_Toc399700469)

[4.5 Choice of the Messaging Platform 14](#_Toc399700470)

|  |
| --- |
| **List of Tables** |

# Introduction

## History of the document

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Author** | **Change Description** |
| 0.1 | 1/2/2014 | Alexander Artikis (NCSR) | Set up the document |
| 1 | 1/2/2014 | Alexander Artikis (NCSR) | Content adjusted |

## Purpose and Scope of the Document

Briefly state the purpose and scope of the deliverable, and indicate the target readership

## Relationship with Other Documents

Describe how the deliverable relates to other deliverables and papers

# Main Sections

## System Requirements

This section lays out the main requirements for SPEEDD prototype. The requirements are based on the requirements documents provided by the use cases.

## Approach

The design of the system architecture for a prototype like SPEEDD is an iterative process that starts with the beginning of the project and continuously evolves, as requirements of the different components are better understood and insights are gained. Therefore, a close iterative and collaborative process was carried out between the architecture team in WP6 “Scalability and System Integration” lead by IBM, and the technical teams of the SPEEDD prototype, specifically the teams of the real-time event recognition and forecasting (WP3), real-time decision making (WP4), real-time visual analytics (WP5), scalability (WP6), and the technical teams from the use cases (WP7 and WP8).

To this end we followed the steps below, as illustrated in Figure ‎2.1:

1. Iterative biweekly virtual meetings that included representatives of all partners involved. A very draft architecture presented at M3 of the project has been frequently updated and refined based on input and feedback to the current architecture (described in sections ‎2.3 - ‎2.5).
2. On a case-by-case basis, bilateral virtual meetings with a specific partner to elaborate on a specific issue (e.g., specific API).
3. Face-to-face meetings during the project meetings in May and September 2014.



Figure ‎2.1- SPEEDD design architecture approach

## Conceptual Architecture

This section provides a high-level overview of SPEEDD prototype. The goal is to introduce the main concepts, high-level components and information flow without getting into implementation and technological details.

Figure ‎2.2 illustrates the conceptual architecture of SPEEDD prototype. We separate between the design time and the run time. The products of the design time activities are event processing definitions and decision management algorithms and configurations that will be deployed and executed at the runtime.



Figure ‎2.2 - Conceptual Architecture of SPEEDD Prototype

Historic data used at design time contains raw events reported during the observed period along with annotations provided by domain experts. These annotations mark important situations that have been observed in past and should be detected automatically in future. Visualization tooling is used to sift through historic data to gain insights and create annotations. Domain experts apply tools and methodologies provided by SPEEDD authoring toolkit to extract complex event pattern definitions from the annotated event history. This is a semi-automatic process involving applying machine learning tools to extract initial set of patterns which is further enhanced and translated with help of the domain experts into deployable CEP artefacts.

The runtime part is composed of the CEP component, the automatic decision management component, and visual decision support tooling. SPEEDD runtime receives raw events emitted by the various event sources (e.g. traffic sensors, transactional systems, etc., - depending on the use case) and emits actions that are consumed by the actuators connected to the operational systems or simulators.

The CEP component is capable of detecting and forecasting complex event patterns under uncertainty. It processes raw as well as derived (detected and forecasted) events to detect and forecast higher-level events, or situations. These serve as triggers for the decision management component, which uses domain-specific algorithms to suggest the next best action to resolve or prevent an undesired situation.

The visualization component (further called the dashboard) facilitates decision making process for business users by providing easily comprehensible visualization of detected or forecasted situations along with output of the automatic decision making component – a list of suggested actions to deal with the situation. The SPEEDD system can be run in either open or closed loop mode. In case of the open loop, the user can approve, reject, or modify the action proposed by the automatic decision maker. The closed loop operation does not require user’s approval, - the action is performed automatically[[1]](#footnote-1). A hybrid mode where some types of actions are taken automatically while other types require human attention is also supported; moreover, we believe that this mode is the most realistic one.

## SPEEDD Runtime Architecture

The architecture of the runtime part of SPEEDD follows the “Event-Driven Architecture” paradigm. Every component functions as an event consumer, or an event producer, or a combination of both. The event bus plays a central role in facilitating inter-component communication which is done via events. Figure ‎2.3 provides a refinement of the conceptual architecture described above where the runtime part is represented as a group of loosely-coupled components interacting through events. The event bus plays a central role as it provides the communication and integration platform for an event-driven system.

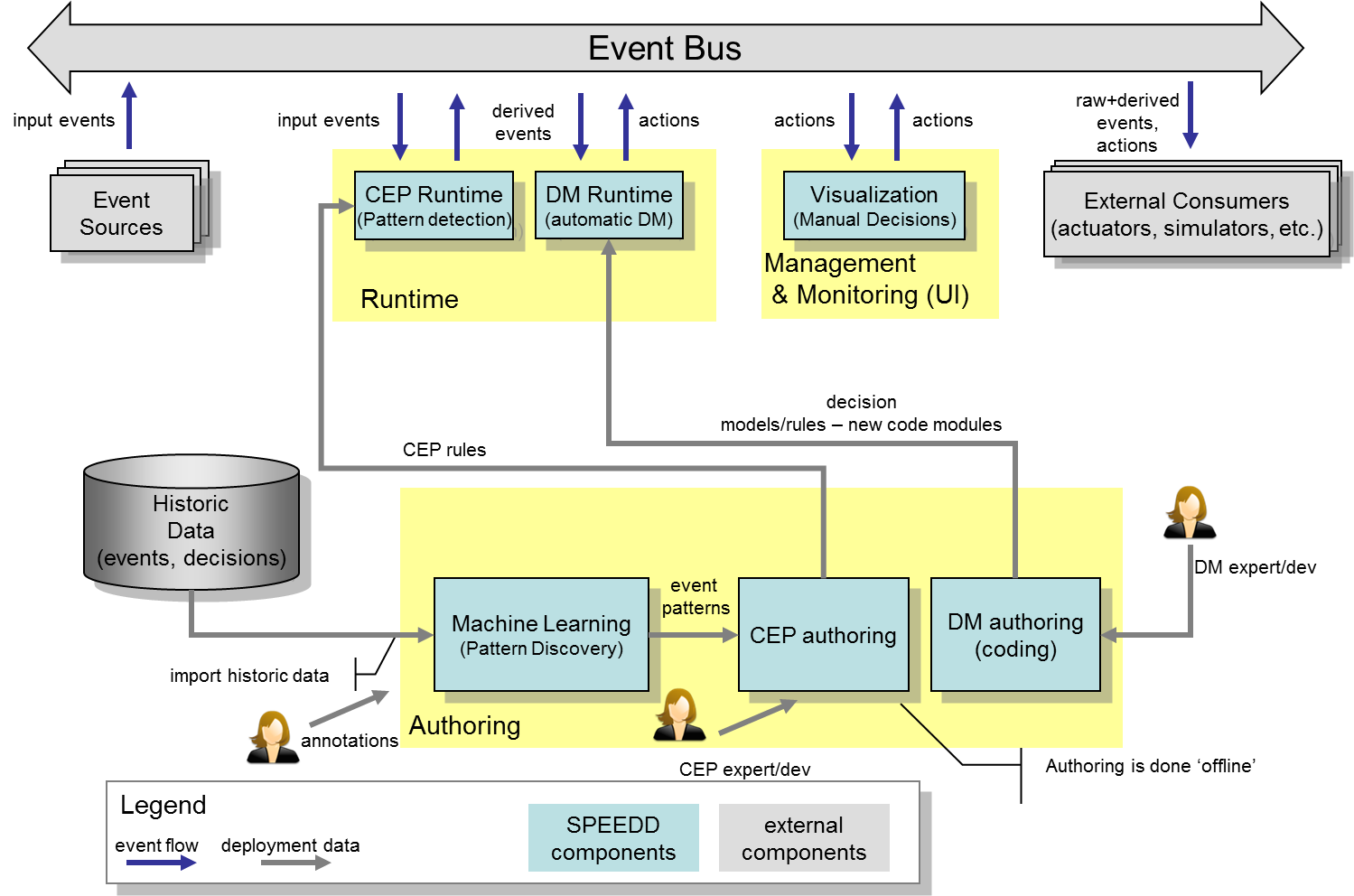


Figure ‎2.3 - SPEEDD - Event-Driven Architecture

In the following subsections we are describe the individual architecture of all the components as well as discuss the technology used to implement them.

Figure ‎2.4 and Figure ‎2.5 illustrate the architecture for the traffic and credit card fraud use cases respectively.

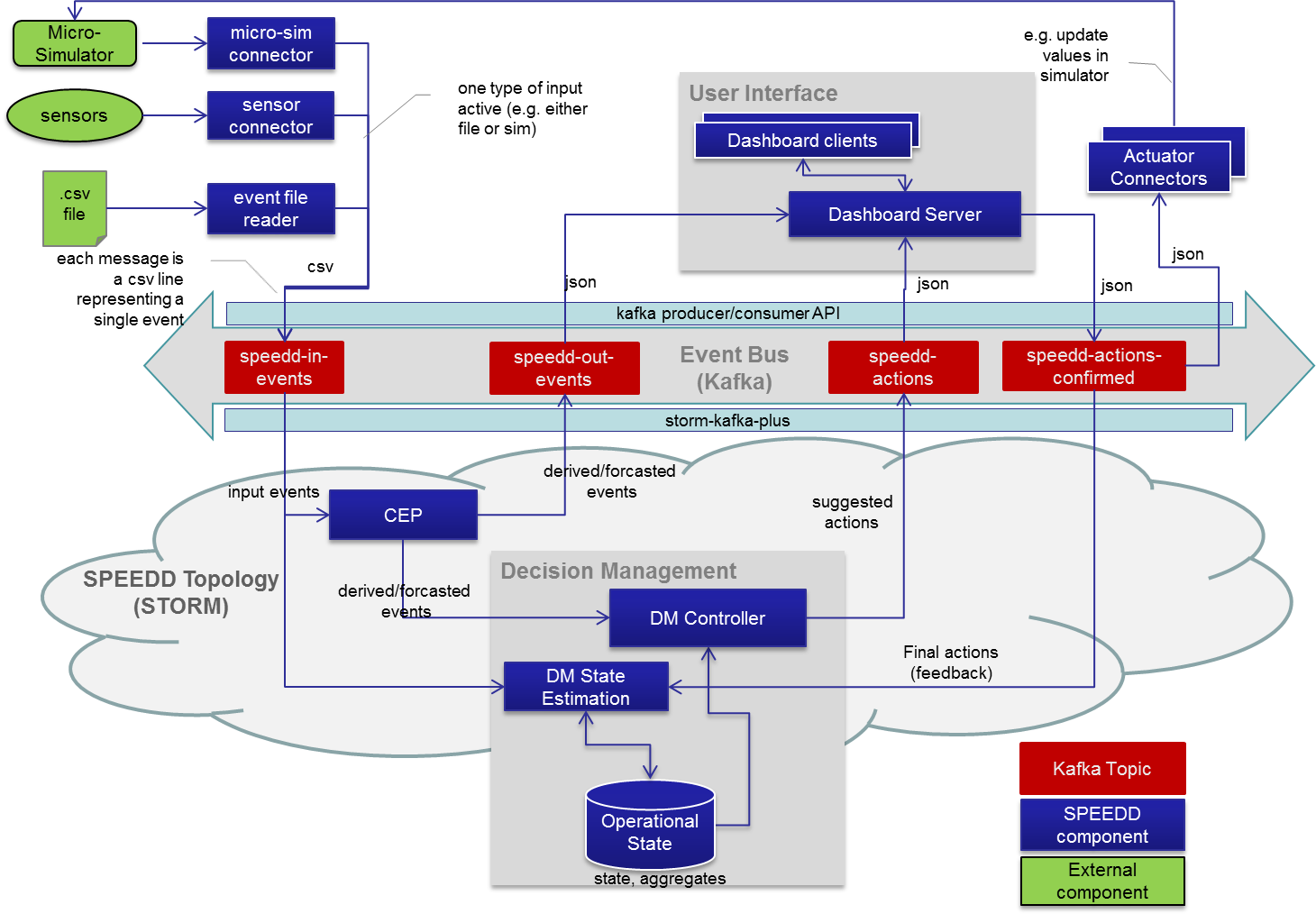


Figure ‎2.4 - SPEEDD Runtime - Event-Driven Architecture (Traffic Use Case)

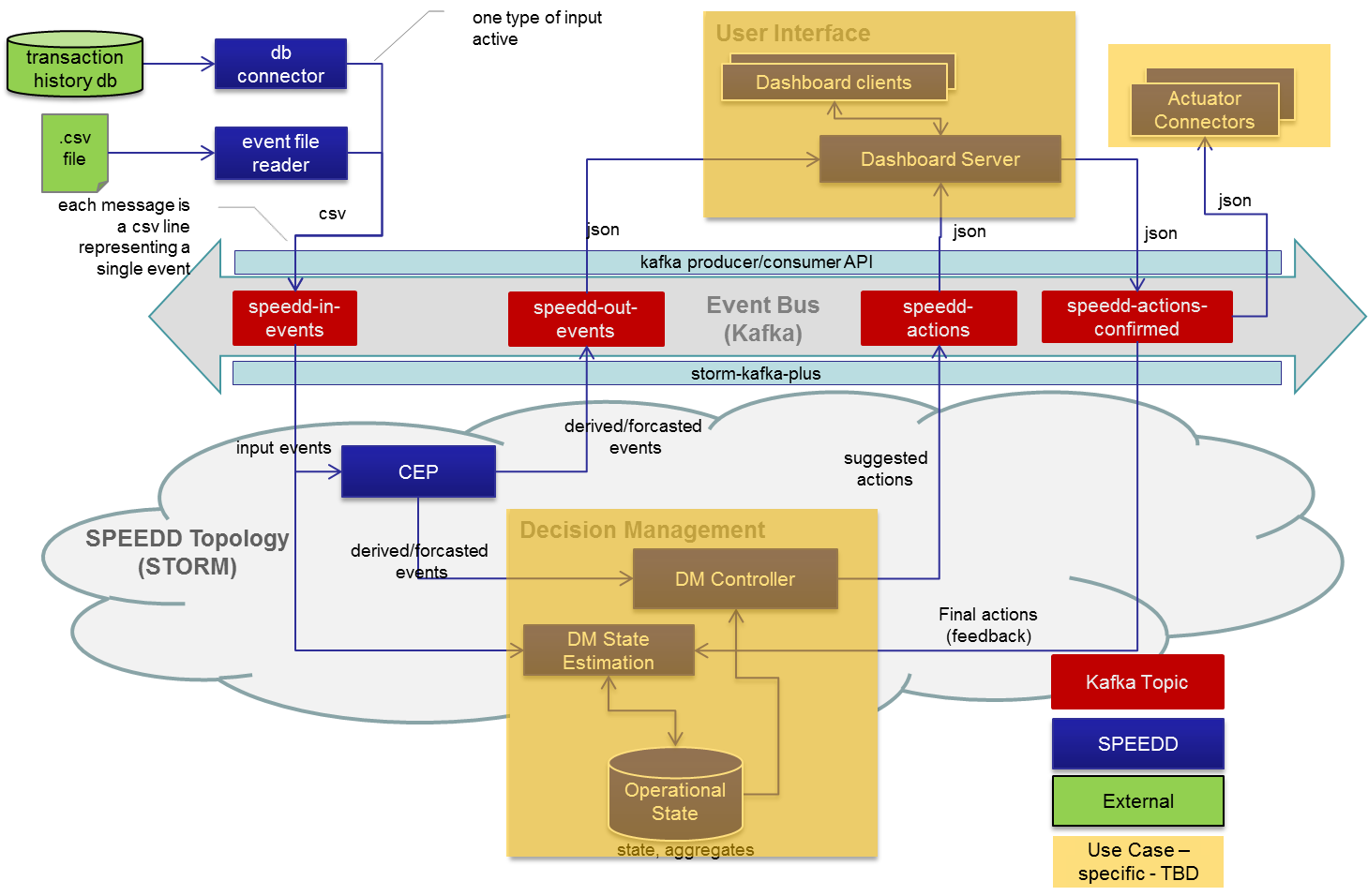


Figure ‎2.5 - SPEEDD Runtime - Event-Driven Architecture (Credit Card Fraud Use Case)

### Event Bus

Describe our implementation of the event bus using KAFKA technology.

### Event/Data Providers

Explain how we’re going to stream the data into the system, for each use case.

### Action Consumption – Actuators/Connectors

Describe who and how would consume action messages issued by the system. Per use case.

### Complex Event Processor

Describe the architecture of PROTON, including implementation on STORM.

### Decision Management

Architecture of the decision management component(s), approaches, issues, etc. Possibly discuss separately the design for every use case.

### Dashboard application

Describe the architecture of the dashboard application.

## Build-Time Architecture

Describe the build-time path and the architecture in details (based on the conceptual view presented earlier).

### Event Pattern Mining

Describe how machine learning approach is used to extract complex event patterns from annotated historic data.

### Authoring of CEP Rules

Describe the process, challenges, and approach to translation of the CEP patterns discovered in using machine learning into Proton EPN definition.

### Decision Management – the Offline Part

Should see if this is relevant. If it is, describe how decision management component is configured or adjusted based on the exploration of the historic data.

## Integration – APIs and Data Formats

Describe the integration mechanisms between different components and between the system and the outside world. List and describe the APIs and data formats in use.

## Non-Functional Aspects

## Scalability

Explain why is the proposed architecture is scalable. Describe how the system will scale up and out to match the load.

## Fault Tolerance

Explain what types of failures the system is designed to stand. Describe the designed behavior of the system in case of such failures.

## Testability

Describe the approach to testing the system. Address the functional testing as well as performance testing approach as designed.

# Conclusions

TBD

# Appendix – Technology Evaluation

This could be one or more appendix parts. Here we’ll explain the approach, criteria, and the final choice of the technology stack that was made.

## Stream Processing – requirements and evaluation criteria

## Storm

## Akka

## Spark Streaming

## Choice of the Messaging Platform

1. Actuators are out of scope of SPEEDD prototype. Under automatic action we mean that the message representing the action type and parameters is emitted by SPEEDD, so that the actual operational system listening to action events is supposed to execute it. [↑](#footnote-ref-1)