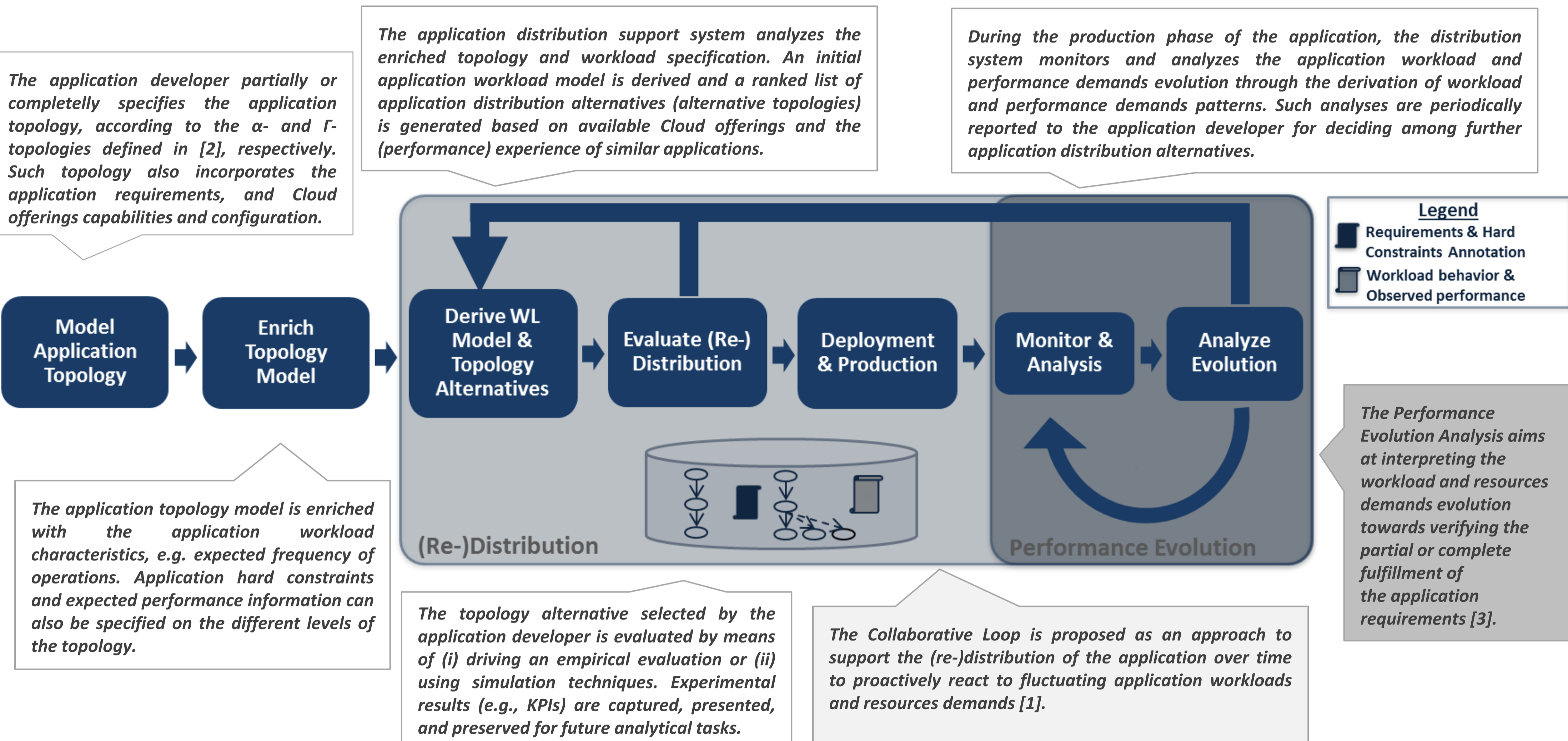# Cloud Application Design Support for Performance Optimization and Cloud Service Selection

Santiago Gómez Sáez, Vasilios Andrikopoulos, Frank Leymann

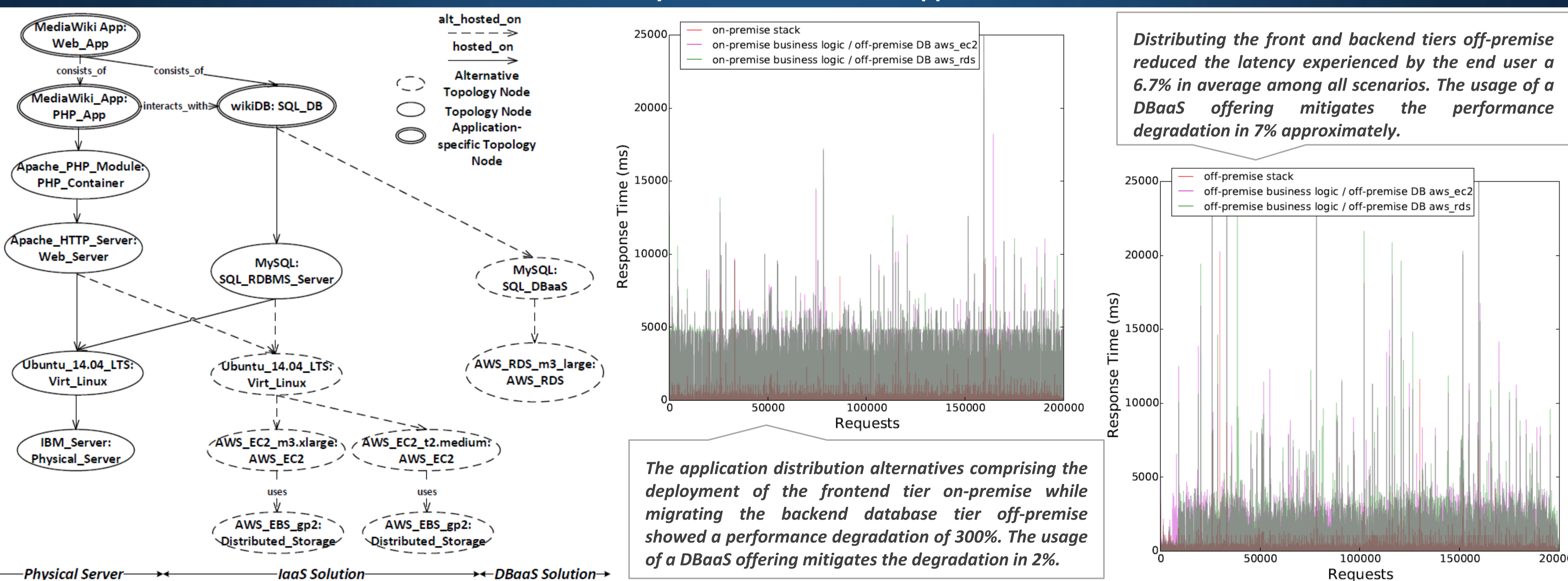Institute of Architecture of Application Systems, University of Stuttgart

Motivation & Problem Statement

The number and types of available Cloud offerings as a service has exponentially increased in the last years, allowing application developers to partially or completely migrate their applications to a highly scalable and pay-per-use infrastructure. The existence of such a technological landscape, however, arises several decision challenges related to efficiently selecting the Cloud offerings to host the application components. Several standards enable the modeling and management of application topology models in a portable manner, potentially facilitating its deployment in a multi-cloud environment. However, there is a lack of design support to application developers towards efficiently selecting and configuring the required by the application underlying Cloud resources to cope in a proactive manner with fluctuating and evolving workloads [1, 3].

Performance Aware Cloud Application (Re-)Distribution Process

The application developer partially or completely specifies the application topology, according to the α- and Γ-topologies defined in [2], respectively. Such topology also incorporates the application requirements, and Cloud offerings capabilities and configuration.

The application distribution support system analyzes the enriched topology and workload specification. An initial application workload model is derived and a ranked list of application distribution alternatives (alternative topologies) is generated based on available Cloud offerings and the (performance) experience of similar applications.

During the production phase of the application, the distribution system monitors and analyzes the application workload and performance demands evolution through the derivation of workload and performance demands patterns. Such analyses are periodically reported to the application developer for deciding among further application distribution alternatives.

Legend
- Requirements & Hard Constraints Annotation
- Workload behavior & Observed performance

Model Application Topology → Enrich Topology Model → Derive WL Model & Topology Alternatives → Evaluate (Re-)Distribution → Deployment & Production → Monitor & Analysis → Analyze Evolution

(Re-)Distribution / Performance Evolution

The application topology model is enriched with the application workload characteristics, e.g. expected frequency of operations. Application hard constraints and expected performance information can also be specified on the different levels of the topology.

The topology alternative selected by the application developer is evaluated by means of (i) driving an empirical evaluation or (ii) using simulation techniques. Experimental results (e.g., KPIs) are captured, presented, and preserved for future analytical tasks.

The Collaborative Loop is proposed as an approach to support the (re-)distribution of the application over time to proactively react to fluctuating application workloads and resources demands [1].

The Performance Evolution Analysis aims at interpreting the workload and resources demands evolution towards verifying the partial or complete fulfillment of the application requirements [3].

Preliminary Results – MediaWiki Application

Distributing the front and backend tiers off-premise reduced the latency experienced by the end user a 6.7% in average among all scenarios. The usage of a DBaaS offering mitigates the performance degradation in 7% approximately.

The application distribution alternatives comprising the deployment of the frontend tier on-premise while migrating the backend database tier off-premise showed a performance degradation of 300%. The usage of a DBaaS offering mitigates the degradation in 2%.

Selected Publications

[1] S. Gómez Sáez; V. Andrikopoulos; F. Leymann; S. Strauch: *Towards Dynamic Application Distribution Support for Performance Optimization in the Cloud*. Proceedings of CLOUD'14.

[2] V. Andrikopoulos; S. Gómez Sáez; F. Leymann; J. Wettinger: *Optimal Distribution of Applications in the Cloud*. Proceedings of CAiSE'14.

[3] S. Gómez Sáez: *Design Support for Performance-aware Cloud Application (Re-)Distribution*. Proceedings of the PhD Symposium at ESOCC'14.

Santiago Gómez Sáez
gomez-saez@iaas.uni-stuttgart.de

University of Stuttgart Germany

IAAS Research

ALLOW Ensembles