

GLOSSARY

Adaptation is a modification of an executed software application by a human operator or an automated routine. Adaptation is typically conducted to react to changes during operations, like increased usage intensity, and is often focused on modifications in component deployment and allocation of computation resources.

Application Manifest is a file that specifies application's name, version, modules, services, and cloud resource requirements.

Application Programming Interface (API), in the context of cloud services, is a collection of code features (e.g., methods, properties, events, and URLs) that a developer can use in their applications to interact with components of a cloud infrastructure.

Architectural Decision (AD) describes a concrete, architecturally significant design issue (a.k.a. design problem, decision required), for which several potential solutions (a.k.a. options, alternatives) exist.

Architectural pattern expresses a fundamental structural organization schema for software systems. It provides a set of predefined subsystems, specifies their responsibilities, and includes rules and guidelines for organizing the relationships between them.

Architectural Refactoring (AR) is a coordinated set of deliberate architectural activities that removes a particular *architectural smell* and improves at least one quality attribute without changing the scope and functionality of the system. An AR can possibly have a negative influence on other quality attributes, due to conflicting requirements and trade-offs.

Architectural Runtime Model is a runtime model that reflects the component-based architecture of an executed software application. It describes the single software components of the application, their composition and deployment. In the context of iObserve, it also describes the resource environment and usage of the application. The architectural runtime model comprises quality-relevant annotations that are required for analyzing the model using simulation or analysis techniques.

Architectural Smell is a suspicion (or indicator) that something in the architecture is no longer adequate under the current requirements and constraints, which may differ from the originally specified ones.

Architecture Description Language (ADL) describes the architecture of a software system with hierarchical components and their communication points.

Artificial Delay is an enforced delay on disruptive tenants usually with the goal to support performance isolation.

Auditable Version Control System (AVCS) is a *version control system* designed to function properly under an adversarial setting, in which the storage server is untrusted and may misbehave.

Auto-scaling means automatically adjusting the amount of resources used by an application in order to fulfill the application's quality of service requirements.

Biodiversity is the variety of species and their individuals whose concrete characteristics vary (e.g., their genetic code and habits); a high level of biodiversity often implies a strong ability of a species to survive and reproduce.

Black List is a list of disruptive tenants in cloud-based systems which will be handled separately to ensure performance isolation.

Certificate Authority (CA) is a trusted entity that issues digital certificates. A certificate is cryptographically signed data asserting the ownership of a public key by a subject. CAs play an important role in the public key infrastructure.

Cloud Auto-scaling is an automatic and elastic process, typically running on a Physical Machine (PM) that adapts software configurations and hardware resources provisioning on-demand according to changing environmental conditions (e.g., the workload).

Cloud Computing is a subscription-based service that delivers computation as a utility.

Cloud Ecosystem is a multitenant cloud environment where the cloud-based services, their software configuration, and hardware provisioning exhibit dynamic and uncertain interactions with each other and the environment, emerging as a system.

Cloud Migration is the process of moving a software application, or part thereof, to the cloud.

Cloud Stability is the extent, to which the cloud is able to prevent violations of Service Level Agreements (SLA) and budget requirements.

Cloud Sustainability is the ability of a cloud to endure the disturbance/stress caused by dynamic and uncertain events, e.g., workload and QoS interference, with the aim to continually optimize QoS attributes of all cloud-based services while minimizing their costs and energy consumptions.

Code Generation is an essential part of a Code Generator that addresses the systematic synthesis of a concrete implementation from an abstract input model.

Complex Event Processing is processing of multiple data streams from different sources to infer meaningful events.

Data Intensive Frameworks is a class of computing applications which focus on distributing the data across different processing elements for a parallel approach to processing large volumes of data.

Data Lake is an analytics system that supports the storing and processing of all types of data.

Data Locality (a.k.a. Disk Locality, Memory Locality) is the strategy of colocating data with relevant processing elements to avoid global network transfers during processing phases. Two substrategies exist, namely Disk Locality (nonvolatile storage medium, e.g., SSD or spinning disks) and Memory Locality (volatile memory, e.g., RAM).

Data Stream Management Systems are systems similar to database management systems that manage data streams and allow continuous queries.

Data/Information Governance ensures that people roles, processes and technology which provides information is protected and managed appropriately to guarantee that an organization gets the maximum value from it.

Deadline is the time limit for the execution of an application.

Delta Encoding is a coding technique that is utilized by modern version control systems to store the data of a version control repository. In delta encoding, only the first version of a file is stored in its entirety, and each subsequent version of the file is stored as the difference from the immediate previous version. These differences are recorded in discrete files called “deltas”.

Design Rules define the reusable design heuristics for designing an application architecture based on selection of features of the family feature model and the reference architecture.

Dataflow Engine (DFE) is a compute device that contains a *Field Programmable Gate Array (FPGA)* as the computation fabric and RAM for bulk storage, and is integrated with a host system or shared by multiple host systems.

Disruptive Tenants are customers which have higher request rate to cloud resources than the permitted request rate.

Domain Specific Language (DSL) is a language designed to be useful for a limited set of tasks, in contrast to general-purpose languages that are supposed to be useful for much more generic tasks, crossing multiple application domains.

Domain-driven design is a process for software development based on understanding the problem domain. Typically, it involves the creation of an abstract model about the problem domain which can then be implemented with a particular set of technologies.

Elasticity is the degree to which a cloud-based system is able to adapt to workload changes by provisioning and deprovisioning resources in an autonomic manner such that the available resources match the current demand as closely as possible.

Enterprise Applications are software applications that are designed to fulfill the business needs of organizations. Examples of such applications are Enterprise Resources Planning (ERP), Customer Relationship Management (CRM), and Supply Chain Management (SCM) applications.

Evolution is the modification of a software application by a developer, typically in order to implement new features, migrate the application to a new platform or correct errors.

Failure is defined as any deviation of a component of the system from its intended functionality.

Fault-Tolerance is used to provide correct and continuous operation of fault-prone components.

Feature Constraints prescribe constraints of feature configurations in a feature diagram, with the aim to limit the configuration space, so that invalid configurations are prevented throughout the derivation process.

Feature Model is a domain model that defines the common and variant features of a domain.

Field Programmable Gate Array (FPGA) is an integrated circuit that can be (re)programmed to the required functionality, allowing specific types of workload to be executed orders of magnitude faster than CPUs.

High Performance Computing (HPC) refers to the practice of aggregating computing power from desktops, workstations or servers in a way that delivers much higher performance than one could get out of a single machine.

HTTPS is a protocol for secure communication over a computer network which is widely used on the Internet. It uses *TLS/SSL* to establish an encrypted communication channel, and uses Hypertext Transfer Protocol (HTTP) for communication.

Legacy Software comprises enterprise software applications that are planned to be migrated (in part or in full) to a cloud-based environment.

Machine Learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence.

Megamodel is a model that describes the relationships of models, meta-models and transformations.

Meta-Data is data that provides information about other data.

Migration factor is a quantity or indicator that influences the process of migrating *legacy software* to the cloud. Several example migration factors are: the general measure, benefits, and risks.

Model-Driven Engineering (MDE) is a software development process where models serve as the primary development artifacts. Those models are not only used for documentation purposes but also for creating a system (semi)automatically.

Montage is a scientific workflow application from the astronomy domain designed to compute mosaics of the sky based on a set of input images.

Multitenancy refers to a software system's ability to transparently fulfill the individual requirements of multiple groups of users (i.e., tenants), while sharing the system's resources such as the database or instances of application components.

Nonisolated Cloud System is a system which does not take into account the possibility of disruptive tenants.

Performance Isolation is a condition in which workload or performance of a single tenant cannot affect the performance of the other tenants.

Processing Elements (PEs) refer to processing end points in large/distributed computational systems. Depending on a system configuration, PEs can be a core, a processor, GPU, or FPGA, etc.

Public Key Infrastructure (PKI) is a system for storing, distributing, and authenticating public keys, so that an entity can ascertain that a given public key belongs to a given entity.

Real time processing systems are systems which can guarantee the processing to be done within a tight deadline.

Reengineering is the process of restructuring and adapting the structure and functionality of a software application in order to improve its qualities or to fulfill new requirements.

Reference Architecture is a software architecture that presents a blueprint for the software system which incorporates best practices by various means such as vocabularies, standards, design patterns, and can be employed by software architects throughout the software lifecycle.

Remote Data Integrity Checking (RDC) is a technique that allows a data owner to check the integrity of data outsourced to an untrusted server, and thus to audit whether the server fulfills its contractual obligations.

Request Rate is the amount of a request sent from a tenant in a particular unit of time.

Resilient Distributed Dataset (RDD) is the primary data abstraction in Apache Spark. RDDs are immutable distributed memory abstractions, each being a partitioned collection of elements that can be operated on in memory through parallel processing on large clusters in a fault-tolerant manner.

Runtime model is a model that reflects the current state of an executed software application. It is constructed or updated based on monitoring data gathered while observing the application.

Scientific Workflow Management System is a system responsible for transparently orchestrating the execution of the workflow tasks in a set of distributed compute resources while ensuring the dependencies are preserved.

Scientific Workflow is an application model used to describe scientific applications by defining a series of computations or tasks and the dependencies between them.

Service Composition is a paradigm where basic services are combined to form a higher level service with increased value for the customer.

Service Level Agreement is Contract between a service provider and the customer that defines the level of service expected from the service provider.

Skip-delta Encoding is a type of delta encoding that is further optimized towards reducing the cost of retrieval. In skip-delta encoding, a new file version is still stored as the difference from a previous file version; however, this difference is not necessarily relative to the immediate previous version, but rather relative to a certain previous version.

Service-Level Objective (SLO) is a file that specifies the objectives (e.g., job completion time or throughput) that must be achieved when running an application.

Software-as-a-Service (SaaS) is a software delivery and licensing model which allows users to access cloud hosted software applications on-demand and pay for the usage on a subscription basis.

Stream Computing is a data-centric high-performance computer system that ingests, manages, analyzes, and passes on (e.g., storage or display) multiple data streams from many sources in real or near real time.

Stream processing is a programming paradigm where a sequence of small operations are applied on each element of a stream.

Stream is a sequence of data tuples that are being generated continuously from a source in real time.

Tenants are the customers that are sharing the same resources of the cloud application.

Thick/Fat/Rich client is a client computer in client-server architecture that provides rich functionality independent of the central server.

Thin client is a lightweight client computer that is purpose-built for remote access to a server to fulfill its computational roles.

Transport Layer Security (TLS) as well as its predecessor Secure Sockets Layer (SSL), is a cryptographic protocol providing communications security over a computer network. It is widely used in web browsing, email, instant messaging, and voice-over-IP (VoIP).

Trophic Web is an interaction network in ecology to model the consumer–resource relationship, for example, predator–prey or organism–resources.

Version Control System (VCS) is a system that provides the ability to track and control the changes made to the data over time, automating the process of version control. It records all changes to the data into a data store called repository so that any version of the data can be retrieved at any time in the future.

Workflow Makespan is the time to complete the execution of a workflow application. It is equal to the completion time, including computation and output data transfer times, of the last workflow task.

Workflow is a collection of tasks connected by control and/or data dependencies. Scientific Workflows automate computation for managing complex large scale distributed data analysis and scientific computation.