EPAM Cloud Orchestrator

Billing Solution Overview

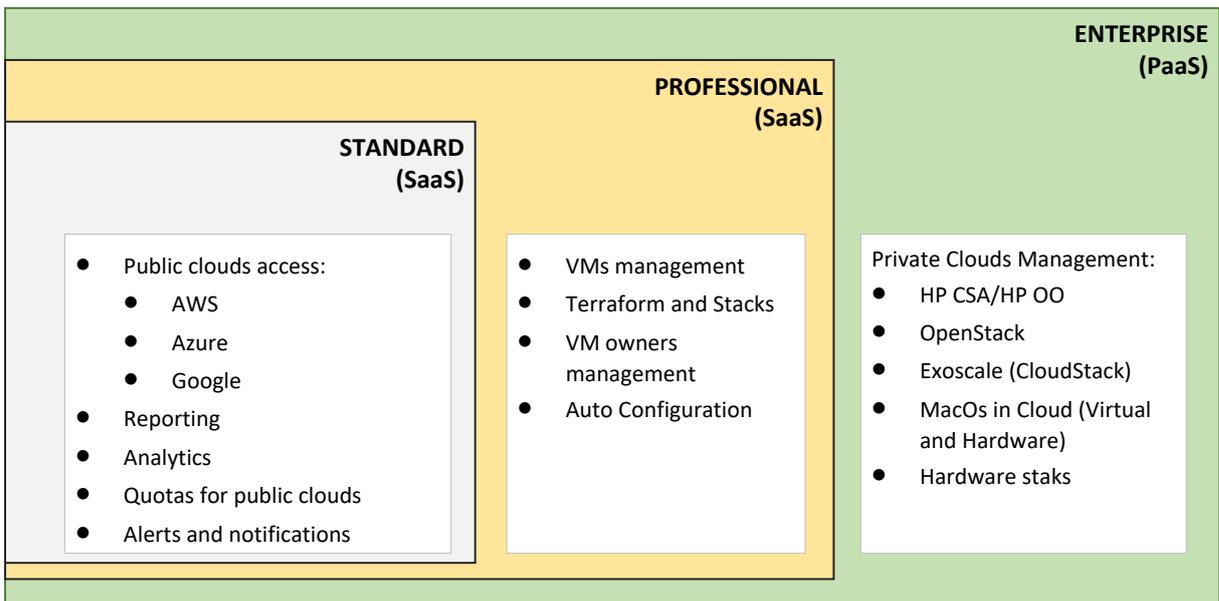**Short Guide**

January 2018

CONTENT

1 INTRODUCING EPAM CLOUD ORCHESTRATOR

EPAM Cloud Orchestrator is a solution that allows to manage, control and monitor virtual infrastructures across public and private clouds.

EPAM Cloud Orchestrator is the result of over six years of EPAM's experience in building hybrid cloud management solutions. It allows users to create, manage, monitor and control virtual infrastructures in terms of self-service, with minimum interaction with the IT operations teams.

It works across multiple cloud platforms, including public cloud providers and virtualization platforms for building private clouds.

EPAM Orchestrator is provided in three options, and the choice depends on the customer's business needs:

ENTERPRISE (PaaS)

PROFESSIONAL (SaaS)

STANDARD (SaaS)

- Public clouds access:
  - AWS
  - Azure
  - Google
- Reporting
- Analytics
- Quotas for public clouds
- Alerts and notifications

- VMs management
- Terraform and Stacks
- VM owners management
- Auto Configuration

Private Clouds Management:
- HP CSA/HP OO
- OpenStack
- Exoscale (CloudStack)
- MacOs in Cloud (Virtual and Hardware)
- Hardware staks

- **The Standard Deployment** model is based on the Software as a Service approach, where the customer is registered under EPAM Orchestrator, and is provided with the following facilities:
  - A single entry point to the unified and simply organized reporting for all customer's resources across all public clouds they use.
  - A set of analytics tools for all virtual resources under the customer's account. This includes analytics on VM and tenant levels.
  - Quotas management tool, that allows to set up the monthly expense limits for virtual infrastructures and the scenarios for different stages of quotas depletion.
  - Alerts and notifications that will inform the customer on the significant events on their resources.
- **The Professional Deployment** model is based on the Software as a Service approach, where the customer is registered under EPAM Orchestrator, and is provided with all the facilities of the Standard model, complemented with the following:
  - Virtual machines management. A VM can be ordered in several clicks, and the process is unified for all the supported clouds.
  - Using Terraform and other stacks solutions to automate infrastructure management.
  - Managing virtual machines owners. This allows to set up the higher level of control on the infrastructure events, and better cooperation with the responsible persons.

- o Auto configuration for complex automated infrastructure setup.
- **The Enterprise Deployment** model is provided as Platform as a Service solution. EPAM Orchestrator is set up on the customer's side, and includes all the features of the Standard and Professional models, as well as the ability to set up private virtual regions. Private regions can use the following technologies:
  - o HP OO/ HP CSA
  - o OpenStack
  - o Special setup for MacOs provisioning (including virtual and hardware Mac instances)
  - o Ability to manage (register, monitor) hardware resources.

The Enterprise deployment model also includes a set of additional features on different layers of the solution. This document describes the main features for PaaS and SaaS models of EPAM Orchestrator provisioning, and allows to compare the feature sets.

2 GENERAL APPROACHES TO BILLING

One of the key parts of Cloud services delivery is effective and transparent monetization by means of flexible billing mechanisms that would allow to quickly respond to market requirements.

EPAM Orchestrator provides a modern, comprehensive, and easy to use billing engine that fully meets the arising challenges. It includes multiple and customizable pricing policies, discounts, costs adjustment and fault management tools, as well as analytics and monitoring mechanisms aimed to simplify resources usage and billing information gathering and processing.

The implemented billing approach is based on the following fundamentals:

- EPAM Cloud Orchestrator is a service-oriented solution, where the pricing for each service (either provided internally or by an external Cloud platform) is calculated according to specific policies or retrieved as is from the third-party service provider.
  The collected costs are aggregated, processed and consolidated into unified bills that are directed to service requestors.

- As a hybrid solution, EPAM Cloud Orchestrator provides **full-scale access to external cloud platforms tools and capabilities** (AWS and Azure), on different levels of hybridization – from creating and managing virtual resources using EPAM Orchestration tools, up to providing direct access to platforms native consoles, with further passing all the resources under Orchestrator's control.
  EPAM Orchestration billing engine allows to aggregate and analyze the chargeback coming from AWS and Azure with negligible delays, process the retrieved information, and put the respective costs to departments' bills in a unified and clear manner.

- The billing engine implemented in EPAM Cloud is based on **chargeback approach**, which means that departments are cross-charged for the Cloud services they use, and then they associate the retrieved costs with respective customers. This approach also includes the pay-as-you-go element.

All the mechanisms and approaches used in EPAM Cloud billing solution, are designed so that it could successfully face the most crucial typical challenges, as well as those caused by EPAM's and its customers' specific requests.

These challenges and their solutions are described further in the document.

3 BILLING CHALLENGES AND SOLUTIONS

Below, you can find the list of the main billing-related challenges that EPAM Orchestrator met, and the main concepts of the solutions that were implemented in order to meet these challenges.

In the next section, you can find the details on each of the described solutions.

Solution	Details	Availability		
		Standard	Professional	Enterprise
Challenge: Chargebacks detailing and allocation. Modern enterprise approach in Cloud provisioning calls for the ability to classify the resources, group them, map to real teams, departments, and customers, and establish respective costs allocation.				
Multi-tenant approach	EPAM Orchestrator is delivered as a multi-tenant solution, where each tenant is bound to a specific unit – a department, a team, a customer, etc. All virtual resources are assigned to specific tenants, each being accessible both as a single unit and as a part of a larger formation. The costs for each resource are assigned to the unit to which the hosting tenant belongs.	+	+	+
Resource tagging	Existing resources can be categorized and logically organized by means of tags. Tags processing is cross-tenant, which allows to push the limits of organizational structure and concentrate on resources purpose, or other conditioning parameters. EPAM Cloud billing engine allows to retrieve and process costs by tags just the same way as it is done by tenants.	-	-	+
Prices detailing	Due to IaaS charging approach, users have access to billing information on different levels – from tenant or cost object total to specific resources and capacities prices.	+	+	+
Common costs and discounts distribution	Additional subscription-based charges (e.g., for support and maintenance) and discounts provided by external Cloud providers to EPAM, are fairy distributed among tenants, according to specific rules, based on analytical and statistic data.	-	-	+
Challenge: Flexibility In order to keep abreast to the constantly changing market conditions and customers' expectations, any Cloud offering should be flexible and customizable, and this is applicable not only to the stack of technological solutions used, but also to the way the provided services are charged.				
Flexible pricing policies	Each virtualization region has its own set of customizable pricing policies that	-	-	+

Solution	Details	Availability		
		Standard	**Professional**	**Enterprise**
	specify the rules, according to which the resources in these regions are billed.			
Pay-as-you-go	Pay-as-you-go system with hourly costs updates scheme allows to control the Cloud-related expenses dynamically, omitting typical issues and difficulties accompanying prepayment approaches.	-	-	+
Management of inaccurate costs	In case a final billing report includes points at issue, or any erroneous numbers, it is possible to correct the bill or balance the difference by adjusting upcoming costs.	-	-	+
Challenge: Costs unification Effective billing means not only precise calculations and costs distribution among cost objects, but also the ability for users to get their billing information in a unique and clear format, irrespective of the costs origin.				
External clouds costs processing	EPAM Cloud billing engine regularly collects the costs for infrastructures hosted on external platforms, processes them and makes available together with other Cloud costs, within the same tools and in the unified format. The delay in external costs processing is minimum – no more than two hours after the update on the provider's side.	+	+	+
Single entry point for costs review	The costs for all resources, irrespectively of their hosting provider, can be reached by user with the one set of tools and are given in a unified format, which simplifies further chargeback processing and management.	+	+	+
Billing for automated infrastructures and native services	Billing engine includes functionality that adds the resources created by infrastructure automation solutions under costs monitoring systems. In addition, EPAM Orchestration's in-built tools keep track of the resources creation amounts and prohibits excessive automated requests for resources.	+	+	+
Challenge: Expenses control for users Cloud users should be empowered not only to watch the infrastructure usage, performance and costs dynamics, but also to set boundaries that would ensure that the unit stays within some pre-defined budget.				
Infrastructure costs Quotas	The quotas engine allow to perform budget-based control over tenants'	+	+	+

Solution	Details	Availability		
		Standard	**Professional**	**Enterprise**
	resources, by setting limits on infrastructure usage costs.			
Personal permissions management	Besides direct expenses, Managers can implicitly influence costs by changing the set of permissions for users dealing with Cloud-based infrastructures.	-	+	+
Scheduled resources manipulation	As Cloud costs depend not only on the resources number and capacity, but also on their activity rate, infrastructure scheduling solutions allows to significantly cut Cloud costs by establishing automatic start/stop schedules for instances.	-	+	+
Challenge: Visibility and transparency Users, unit managers and customers need to keep abreast of their infrastructures costs, and the information should be at hand at any time.				
Access to pricing details	Specific tools are implemented to provide access to the details of actual pricing policies applied in EPAM Cloud and to estimate the costs of resources to be created, with ability to compare prices for different usage schemes.	-	-	+
Real-time access to billing information	All the information on Cloud costs is available for users in real-time mode. It can be returned to the screen or sent via email for further processing and analysis.	+	+	+
Automatic analytics	EPAM Orchestration analytic engines automatically gather and process virtual infrastructures performance, manipulation and billing information, returning the resulting data in a comprehensive and structured way.	+	+	+
Monthly summary reports	At the end of each month, managers get detailed billing, resources, and services usage reports that allow to keep track of the financial changes and provide statistic data that can be a powerful background for infrastructure optimization.	+	+	+

4 IMPLEMENTED SOLUTIOS DETAILS

4.1 CHARGEBACK DETAILING AND ALLOCATION

Modern enterprise approach needs not only fast and effective provisioning of virtual infrastructures, but also calls for the ability to classify the resources, group them, map to real teams, departments, and customers.

This results into the crucial challenge of enabling precise and transparent costs allocation to correct cost objects.

Within EPAM Cloud, the challenge is faced by four solutions:

- Multi-tenant approach.
- Resources categorization with tags
- Prices detailing
- Common costs and discounts distribution

Each solution is described below in this section.

4.1.1 Multi-tenant approach

EPAM Cloud is a multi-tenant solution, where each tenant is bound to a specific unit – a department, a team, a customer, etc.

Each virtual resource belongs to a specific tenant, each being accessible both as a single unit and as a part of a larger formation. The costs for each resource are assigned to the unit to which the hosting tenant is assigned.

This is the basic approach that enables quick and transparent linkage between virtual infrastructures and their requestors.

4.1.2 Resources categorization with tags

Existing resources can be categorized and logically organized by means of tags. Tags processing is cross-tenant and does not depend on the cloud platform on which the tagged resources are hosted. This allows to push the limits of organizational structure and concentrate on resources purpose, or other conditioning parameters.

EPAM Cloud billing engine allows to retrieve and process costs by tags just the same way as it is done by tenants.

4.1.3 Prices detailing

Due to IaaS charging approach, users have access to billing information on different levels – from tenant or cost object total to specific resources and capacities prices.

Thus, it is always possible to get information on each resource's cost, and the elements that comprise it.

Differentiation between prices for active and passive resources is an additional point for infrastructure usage analysis and optimization.

4.1.4 Common costs and discounts distribution

Except costs for resources and Cloud services usage, bills from external cloud providers also include additional expenses for support and maintenance.

These costs are not assigned to any particular tenant, but are taken for the provider's services usage in general. Thus, they are distributed fairy among all tenants that are related to the subject.

The applied formulas can depend on the type of the service or statistics on the service chargeback.

Besides this, AWS and Azure provide discounts for virtual resources usage, and these discounts can influence the resulting costs for all infrastructures hosted on these platforms.

There are two main sources of discounts:

- Based on analysis of infrastructure usage, IT department can make advance order for virtual resources on external Cloud platforms for further usage by all requestors who chose to use EPAM Cloud solution. This allows to get significant discounts that will be applied during the year.
- Specific units (departments, teams, customers) can request resources reservation for their needs, or get a specific discount for one or several months.

In both cases, final costs adjustments are needed. The billing is calculated according to the standard scheme, and after the calculations are done, the discounts are subtracted from the final price for a specific region.

The users get access only to the resulting costs, which allows to omit confusions.

4.2 CHARGEBACK SYSTEM FLEXIBILITY

In order to keep abreast to the constantly changing market conditions and customers' expectations, any Cloud offering should be flexible and customizable, and this is applicable not only to the stack of technological solutions used, but also to the way the provided services are charged.

Meeting this key demand needs a combination of inter-related approaches to be implemented. With EPAM Cloud Orchestration chargeback system, these approaches include:

- Flexible pricing policies, that are independently applied to different virtualization regions
- Pay-as-you-go approach
- Inaccurate costs management system

 All of them are described below in this section.

4.2.1 Flexible pricing policies

EPAM Orchestration billing engine operates with two types of bills – external ones, coming from third-party Cloud providers "as is" and then processed, and internal ones that should be calculated according to resources capacity and usage rate.

Whereas external bills are, basically, only adjusted to EPAM Orchestration reporting format, internal ones are generated by the billing engine according to specific rules, applied to each of the private regions. The set of the billing rules used for costs calculation is called **pricing policy**.

Each region can have one or several pricing policies (but only one is applied at once), which are activated depending on some condition, for example, day of the week.

Pricing policies can be adjusted by request from Cloud change manager, or after their approval, and the changes come into effect within the shortest time.

4.2.2 Pay-as-you-go approach

In many areas, service providers have long-term relations with customers, and these relations are accompanied by prepayments, based on predictable service usage rates.

However, when it comes to using Cloud, such commitments turn out to be ponderous and inconvenient, as the development/production needs can change just over the course of service order processing.

Thus, pay-as-you go approach takes the center stage, as it allows to strictly control the expenses and infrastructure configuration without any obligations, and releases the constraints implied by prepaid models.

The requestors pay only for the resources that were created for them, and only for the time the resources existed.

4.2.3 Management of inaccurate costs

There are situations, when business units get over- or under- billed due to technical or organizational issues.

For such cases, there is a mechanism of applying pre-defined adjustments to business units' monthly reports.

In case an incorrect costs issue is detected after the payment was made, the difference is balanced by applying respective adjustments to the bills invoiced in the coming month or several months.

Any chargeback updates and changes are made only after approval from IT Department side.

4.3 COSTS UNIFICATION

Effective billing means not only precise calculations and costs distribution among cost objects, but also the ability for users to get their billing information in a unique and clear format, irrespective of the costs origin.

To establish this unification, EPAM Orchestration supports three solutions:

- A specific flow for external costs processing and formatting
- A single tool for users to access billing information on any part of their infrastructure
- A solution that brings resources, created automatically, under Orchestration's billing

The descriptions of these solutions are given further in this section.

4.3.1 External clouds costs processing

When a unit requests resources in AWS or Azure clouds, EPAM Orchestration automatically starts to collect billing information on these resources directly from providers.

Each provider has its own standards of billing data generation and provisioning. Orchestration billing engine processes this information and passes it further to the database in the unified EPAM Orchestration billing format. EPAM Cloud service does not imply any additional charges on the resources created on third-party platforms.

Any billing-related change on external provider's side is detected within two hours, and passed to Orchestrator's billing engine, which processes it and saves the update to the database.

4.3.2 Single entry point for costs review

The users get access to virtual resources billing information as soon as this information gets to EPAM Orchestration's billing database. Following the idea of unification and transparency, user access to billing information is enabled via the same tools, no matter whether the resources are hosted in private or public platforms.

In case a user's team or department has resources distributed among different Cloud providers, virtualization regions and/or tenants, all the chargeback information is gathered in one place, with scaling facilities enabled to allow to concentrate on general infrastructure costs as well as on specific branches or resources.

4.3.3 Billing for automated infrastructures and native services

Each provider supports additional services that can be used by application, thus increasing the effectiveness of Cloud usage.

It is necessary to provide the users with the ability to use these services, and to organize proper billing for them. EPAM Orchestrator supports two approaches for this:

- **Native Automation Tools via EPAM Orchestrator.** EPAM Orchestrator supports tools that allow to automate configuring infrastructures. These are AWS CloudFormation, EPAM Maestro Stacks

and Azure Resource Groups solutions, having similar approach: all the instructions are stored as templates, and are applied automatically at user's call.

The templates can be stored in specific libraries, so that users can request the template activation, specifying all the necessary parameters for them.

Before a template is run, a specific mechanism verifies that creating new resources during the template processing, will not lead to resource quotas exceeding.

- **Access to native control tools.** If the functionality supported by EPAM Orchestrator is not enough, users (if the respective permission is granted by a responsible person) can request access to native AWS or Azure consoles.

The consoles allow to use any native tools of external Cloud providers.

Virtual resources created either with EPAM Cloud or provider's native tools, immediately get under Orchestrator's control and monitoring, and it starts gathering billing information about them and related services.

4.4 EXPENSES CONTROL FOR USERS

The other important requirement, meeting which is crucial for EPAM Orchestrator to be called a modern, up-to-date and cutting-edge Cloud platform, is providing the users with the power not only to watch the infrastructure usage, performance and costs dynamics, but also to set boundaries that would ensure that a unit stays within some pre-defined budget.

The aim is reached by two main solutions:

- Resource utilization quotas mechanism, designed to control expenses
- Permissions management tool, which allows to control infrastructure manipulation by users, thus, implicitly, influencing the final costs.

The details on these solutions are given below.

4.4.1 Resource utilization quotas

Monthly resource utilization (cost) quotas allow to set the amount of money that can be spent on Cloud each month, and the actions that should be automatically taken when the quota is depleted. Orchestrator allows to check and update quota rules in real time, and one can always receive information on the current quota utilization rate.

Besides this, Orchestrator can perform regular checks of quota utilization, and notify a responsible person on the quota depletion level, so that respective infrastructure optimization actions can be taken timely.

In EPAM Cloud, billing information is gathered each hour, same AWS and Azure billing information is updated each two hours. Orchestrator collects the billing information from all platforms, processes it, and reacts respective to the quota rules, specified for each tenant. This can be forced shutdown of all tenant-related resources, restricting new resources creation, or just sending a warning to users.

In addition to costs quotas, there are also resource creation quotas. They specify the amount of resources that can be created by assigned to one tenant within 24 hours, and can be used as additional resource usage control tool, thus influencing the resulting chargeback.

4.4.2 Personal permission management

Self-service approach is an effective way to facilitate and accelerate virtual resources provisioning. However, it increases the risk of unreasonable infrastructure expansion, and, as a result, its costs increase, as Cloud billing directly depends on the number, capacity and activity rate of the resources created for a units' needs.

As a solution to this point, EPAM Orchestration supports a system of permissions, that can be customized for each user individually or for groups of users, so that the manager can define who can request new resources creation, and who can only use them, thus implicitly influencing the final chargebacks.

4.4.3 Scheduled resources manipulation

Virtual infrastructures price depends not only on the number and the capacity of the resources involved, but also on their activity rate, i.e., on the time they spend in running or stopped states.

EPAM Orchestration includes a CRON-based scheduling solution that allows to establish rules, according to which the resources will be stopped and started automatically, so that they stay in the active state only during the hours when teams actually need them.

According to investigation, held by EPAM Cloud team, establishing effective scheduling for a virtual infrastructure can cut its cost in half or even more.

This explains the growing popularity of the scheduling approach among Cloud users.

4.5 VISIBILITY AND TRANSPARENCY

Among essential requirements for billing mechanisms, accompanying enterprise-level Cloud services, are transparency and visibility for end users. Users, unit managers and customers need to keep abreast of their infrastructures costs, and the information should be at hand any time.

Billing and resource usage information transparency is ensured by the following set of solutions:

- Real-time access to pricing details
- Automatic analytics of infrastructure usage and costs
- Automatic monthly summary reports

For the details on these solutions, please see the sections below.

4.5.1 Access to pricing details

When it comes to pay-as-you-go approach, it is important to get not only the total costs, but also the detailed bills for each infrastructure item and service, charged by the cloud provider.

EPAM Orchestration tools allow users to retrieve the costs information with different details – from overall unit's chargeback, to specific prices for an item's elements. Billing information for the current month is updated on hourly basis, and it is possible to reference chargeback history for up to 5 months back.

In addition, EPAM Orchestration provides the Cost Estimator tool which allows to perform preliminary analysis of the infrastructure planned for creation, and retrieve its approximate costs.

4.5.2 Automatic analytics

The value of billing and infrastructure usage information can be higher, if it is properly analyzed, so that the results can be used for making well-founded further decisions.

Thus, EPAM Orchestration is equipped with analytics tools that automatically gather and process resources utilization data, and return statistical and trend information that can be useful in further resources management.

Billing mechanisms gather information on current expenses and alert users on excessive costs grow, analytical tools monitor resources activity, so that the resulting information can be used for infrastructure usage optimization and costs reducing.

All the results obtained by analytical mechanisms, are open and available for users at any time.

4.5.3 Monthly summary reports

At the end of each month, EPAM Orchestration verifies and synchronizes all the billing data from EPAM Orchestration, Amazon Web Services and Microsoft Azure Cloud and makes sure that all the bills made out for clients are correct.

When the verification succeeds, Orchestration sends the Summary report containing billing and resource usage information for the tenants assigned to the receiver (who is typically a manager).

The report covers the following summaries:

- **Monthly billing report by customers** - returns total chargeback by the customers you work with
- **VM Lifetime Report** - provides the information on the lifetime of the VMs existing in your projects. Each column contains the number of VMs of a specified age.
- **Account Optimization Ratio Report** - report provides the information on the time your projects' VMs spend in running state. Each column contains the number of VMs that have running state for the specified time (in % of the overall lifetime during last month).
- **Service usage report** - provides the information on the number of VMs engaged in EPC Services and covered by them. Here, you can find the overall number of VMs in projects and the amounts of service-related VMs in percentage of the overall number.
- **Monthly billing report** - provides the total chargeback for the projects in your account.

VERSION HISTORY

Version	Date	Summary
1.2	January 28, 2018	Added deployment options, Removed Suspend operation mentioning due to its deprecation.
1.1	December 16, 2016	Classification changed from Confidential to Public, approved by Dzmitry Pliushch
1.0	February 26, 2016	Initial version published