**CHARM: A Cost-efficient Multi-cloud Data**

**Hosting Scheme with High Availability**

**Abstract:**

More and more enterprises and organizations are hosting their data into the cloud, in order to reduce the IT maintenance cost and enhance the data reliability. However, facing the numerous cloud vendors as well as their heterogenous pricing policies, customers may well be perplexed with which cloud(s) are suitable for storing their data and what hosting strategy is cheaper. The general status quo is that customers usually put their data into a single cloud (which is subject to the vendor lock-in risk) and then simply trust to luck. Based on comprehensive analysis of various state-of-the-art cloud vendors, this paper proposes a novel data hosting scheme (named CHARM) which integrates two key functions desired. The first is selecting several suitable clouds and an appropriate redundancy strategy to store data with minimized monetary cost and guaranteed availability. The second is triggering a transition process to re-distribute data according to the variations of data access pattern and pricing of clouds. We evaluate the performance of CHARM using both trace-driven simulations and prototype experiments. The results show that compared with the major existing schemes, CHARM not only saves around 20% of monetary cost but also exhibits sound adaptability to data and price adjustments.

**Existing System**

In existing industrial data hosting systems, data availability (and reliability) are usually guaranteed by replication or erasure coding. In the multi-cloud scenario, we also use them to meet different availability requirements, but the implementation is different. For replication, replicas are put into several clouds, and a read access is only served (unless this cloud is unavailable then) by the “cheapest” cloud that charges minimal for out-going bandwidth and GET operation. For erasure coding, data is encoded into n blocks including m data blocks and n􀀀m coding blocks, and these blocks are put into n different clouds. In this case, though data availability can be guaranteed with lower storage space (compared with replication), a read access has to be served by multiple clouds that store the corresponding data blocks. Consequently, erasure coding cannot make full use of the cheapest cloud as what replication does. Still worse, this shortcoming will be amplified in the multi-cloud scenario where bandwidth is generally (much) more expensive than storage space.

**Proposed System:**

**The proposed CHARM scheme.** In this paper, the system is proposed a novel cost-efficient data hosting scheme with high availability in heterogeneous multi-cloud, named “CHARM”. It intelligently puts data into multiple clouds with minimized monetary cost and guaranteed availability. Specifically, we combine the two widely used redundancy mechanisms, i.e., replication and erasure coding, into a uniform model to meet the required availability in the presence of different data access patterns. Next, we design an efficient heuristic-based algorithm to choose proper data storage modes (involving both clouds and redundancy mechanisms). Moreover, we implement the necessary procedure for storage mode transition (for efficiently re-distributing data) by monitoring the variations of data access patterns and pricing policies. We evaluate the performance of CHARM using both trace driven simulations and prototype experiments. The traces are collected from two online storage systems:, both of which possess hundreds of thousands of users.

**Advantages:**

* Replication mechanism when the file’s size is small. That is why gray level 4 puts its feet into the region of lower read count and smaller file size.
* This storage mode table only depends on prices of the available clouds and required availability. If the prices change, the table will change accordingly, becoming a different one**.**

**Problem Statement**

* Nevertheless, as for multi-cloud people still encounter the two critical problems:
* How to choose appropriate clouds to minimize monetary cost in the presence of heterogeneous pricing policies?
* How to meet the different availability requirements of different services?
* As to monetary cost, it mainly depends on the data-level usage, particularly storage capacity consumption and network bandwidth consumption.
* As to availability requirement, the major concern lies in which redundancy mechanism (i.e., replication or erasure coding) is more economical based on specific data access patterns. In other words, here the fundamental challenge is:
* How to combine the two mechanisms elegantly so as to greatly reduce monetary cost and meanwhile guarantee required availability?
* Data Hosting and SMS are two important modules in CHARM. Data Hosting decides storage mode and the clouds that the data should be stored in.

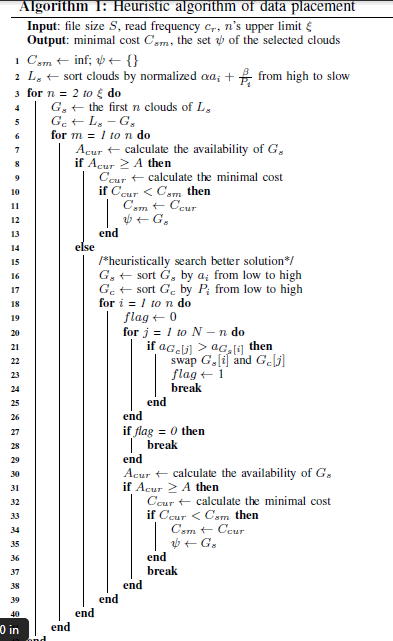
**Scope**

As a holistic storage system, there are several other factors to be considered, such as cache strategies, geographical data consistency, etc. However, we only focus on the data hosting strategy to minimize monetary cost while meeting flexible availability requirements. Though we have considered the complexity and feasibility when designing this strategy, the system design is out of the scope of this paper, and we put the detailed system design of multi-cloud data hosting into future work. the complexity of this algorithm is mainly the first loop, and the worst case complexity is O(Fn), where Fn is the number of files. In order to reduce the complexity further, we can classify files with similar access patterns into groups, and implement transition in the unit of group. This is out of the scope of this paper.

**Algorithm**:

The key idea of this heuristic algorithm can be described as follows:

We first assign each cloud a value which is calculated based on four factors (i.e., availability, storage, bandwidth, and operation prices) to indicate the preference of a cloud. We choose the most preferred n clouds, and then heuristically exchange the cloud in the preferred set with the cloud in the complementary set to search better solution. This is similar to the idea of Kernighan-Lin heuristic algorithm , which is applied to effectively partition graphs to minimize the sum of the costs on all edges cut. The preference of a cloud is impacted by the four factors, and they have different weights. The availability is the higher the better, and the price is the lower the better.



**SYSTEM SPECIFICATION**

**Hardware Requirements:**

* System : Pentium IV 3.5 GHz or Latest Version.
* Hard Disk : 40 GB.
* Monitor : 14’ Colour Monitor.
* Mouse : Optical Mouse.
* Ram : 1 GB.

**Software Requirements:**

* Operating system : Windows XP or Windows 7, Windows 8.
* Coding Language : Java – AWT,Swings,Networking
* Data Base : My Sql / MS Access.
* Documentation : MS Office
* IDE : Eclipse Galileo
* Development Kit : JDK 1.6

**Conclusion:**

Cloud services are experiencing rapid development and the services based on multi-cloud also become prevailing. One of the most concerns, when moving services into clouds, is capital expenditure. So, in this paper, we design a novel storage scheme CHARM, which guides customers to distribute data among clouds cost-effectively. CHARM makes fine-grained decisions about which storage mode to use and which clouds to place data in. The evaluation proves the efficiency of CHARM.