Architecture of Enterprise Applications 19 Cloud Computing

Haopeng Chen

***RE**liable, **IN**telligent and **S**calable Systems Group (**REINS**)*
Shanghai Jiao Tong University
Shanghai, China
http://reins.se.sjtu.edu.cn/~chenhp
e-mail: chen-hp@sjtu.edu.cn

Agenda

- Cloud Computing
  - What is cloud computing?
  - Core techniques of cloud computing
  - Obstacles and Opportunities
- SaaS
  - Web app to SaaS
  - Multi-tenant Mode

What is cloud computing?

- There is little consensus on how to define the Cloud
  - A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet.
    - In "*Cloud Computing and Grid Computing 360-Degree Compared*"
  - Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The services themselves have long been referred to as Software as a Service (SaaS), so we use that term. The datacenter hardware and software is what we will call a Cloud.
    - In "*Above the Clouds: A Berkeley View of Cloud Computing*"

What is cloud computing?

- There is little consensus on how to define the Cloud
  - Cloud computing promises to radically change the way computer applications and services are constructed, delivered, and managed. Although the term means different things to different people, and includes a bit of marketing hype and technical redefinition, the potential benefits are clear. Large datacenters permit resource sharing across hosted applications and lead to economies of scale at both the hardware and software level. Software services can obtain seemingly infinite scalability and incremental growth to meet customers' elastic demands. The pay-as-you-go model and rapid provisioning can result in more efficient resource utilization and reduced costs.
    - On "*Introduction to Cloud Computing - ACM Tech pack Committee on CC*"

What is cloud computing?

- Cloud computing is a general term for anything that involves delivering hosted services over the Internet.

- These services are broadly divided into three categories:
  - Infrastructure-as-a-Service (IaaS)
  - Platform-as-a-Service (PaaS)
  - and Software-as-a-Service (SaaS).

- A cloud service has three distinct characteristics that differentiate it from traditional hosting.
  - It is sold on demand, typically by the minute or the hour;
  - it is elastic -- a user can have as much or as little of a service as they want at any given time;
  - and the service is fully managed by the provider.

- Significant innovations in virtualization and distributed computing, as well as improved access to high-speed Internet and a weak economy, have accelerated interest in cloud computing.

What is cloud computing?

- Cloud in the real world
  - Cloud as reality, as told by industry partners

SaaS	Software-as-a-Service	Google Apps, Microsoft "Software+Services"
PaaS	Platform-as-a-Service	IBM IT Factory, Google AppEngine, Force.com
IaaS	Infrastructure-as-a-Service	Amazon EC2, IBM Blue Cloud, Sun Grid
dSaaS	data-Storage-as-a-Service	Nirvanix SDN, Amazon S3, Cleversafe dsNet

What is cloud computing?

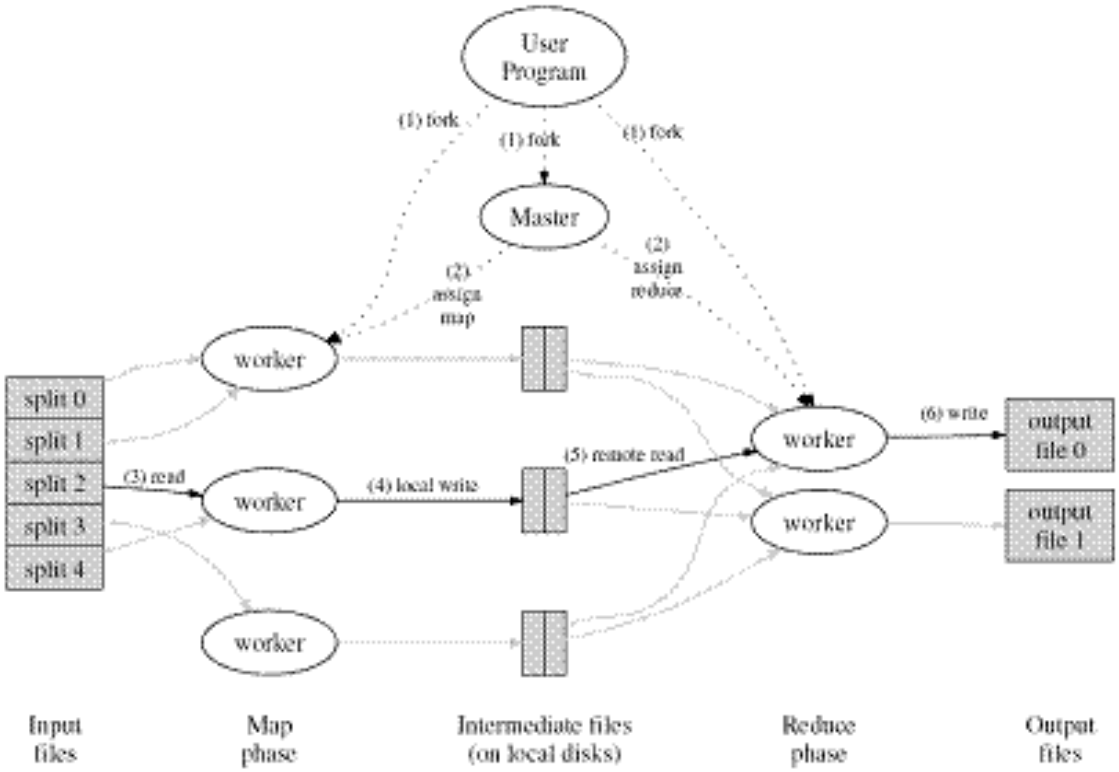
- Cloud in the real world
  - Cloud is different in many dimensions
    - Time to deploy a server
      - Weeks or months -> seconds to minutes
    - Commitment to use service
      - Negotiate & commit year long contract -> select from catalog & Pay as you go
    - Necessary upfront investment
      - $M -> $K
  - Common attributes of clouds
    - Flexible pricing
    - Elastic scaling
    - Rapid provisioning
    - Advanced virtualization

What is cloud computing?

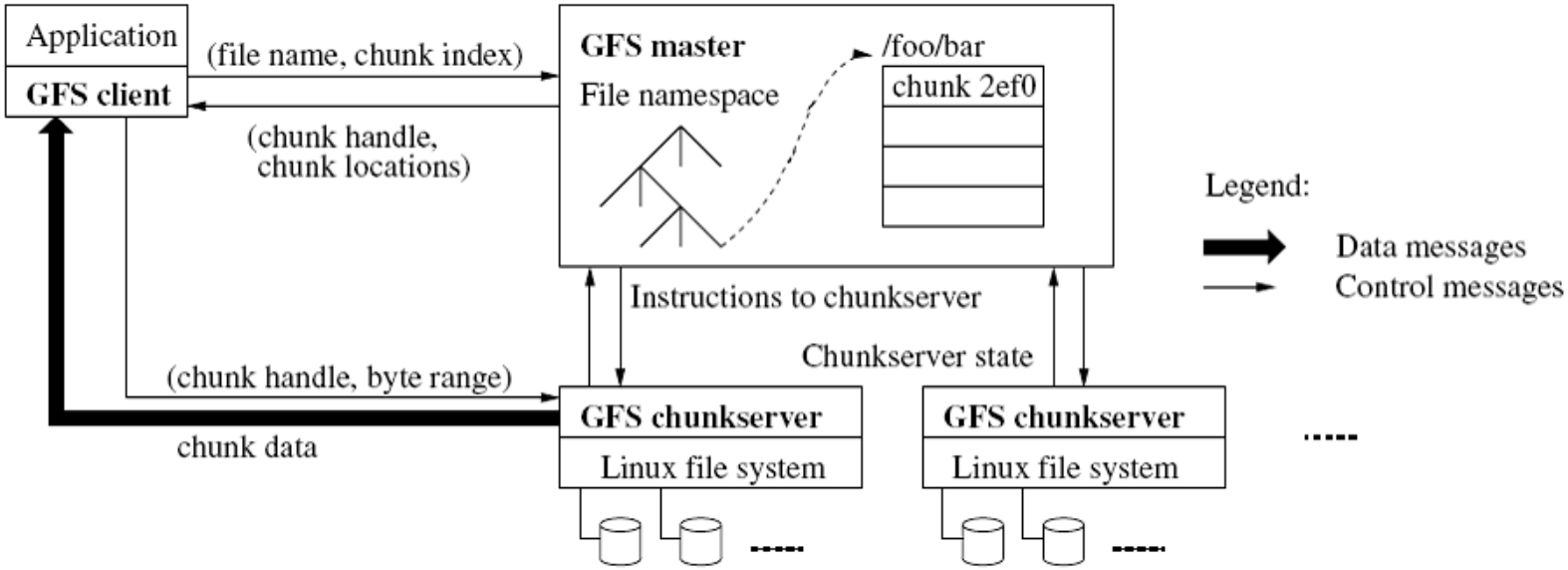
- Cloud in the real world
  - Necessary supprort
    - Workload optimization
    - Integrated service management
      - Service delivery, service request, service monitoring
      - Lowers operational costs, drives efficiency, enhances security
    - Deployment choices
      - Public clouds or private cloud

Core techniques of cloud computing

- Google
- MapReduce
  - parallelizes the computation, distributes the data, and handles failures conspire to obscure the original simple computation with large amounts of complex code to deal with these issues.

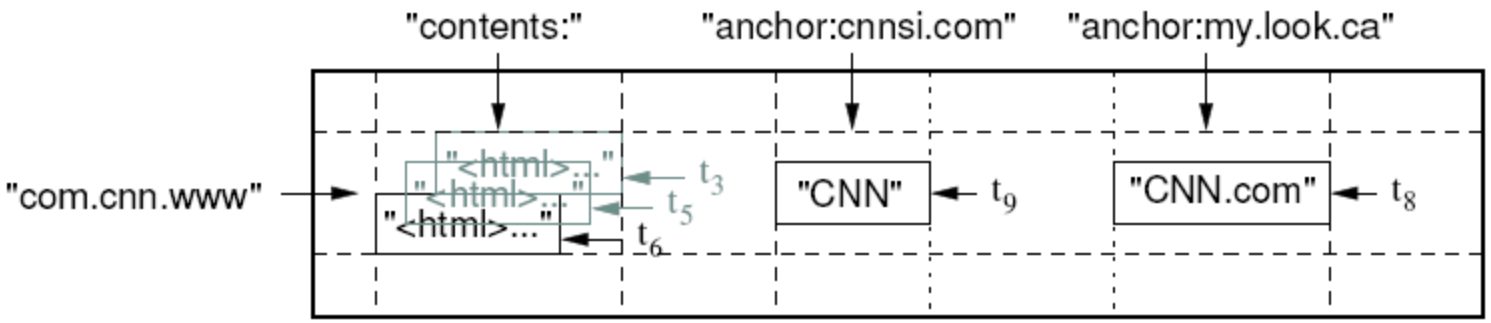


- Google
- Distributed Google File System
  - Google File System(GFS) to meet the rapidly growing demands of Google's data processing needs.



Core techniques of cloud computing

- Google
- Bigtable
  - Bigtable is a distributed storage system for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers.



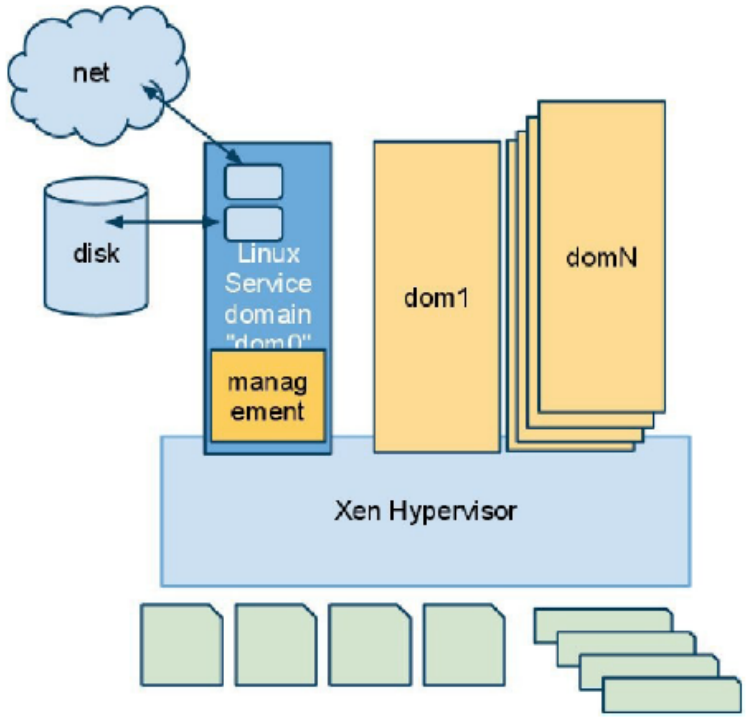
Core techniques of cloud computing

- 

- The Apache Hadoop project develops open-source software for reliable, scalable, distributed computing. Hadoop includes these subprojects:
  - **Hadoop Common**: The common utilities that support the other Hadoop subprojects.
  - **HDFS**: A distributed file system that provides high throughput access to application data.
  - **MapReduce**: A software framework for distributed processing of large data sets on compute clusters.
- IBM, Amazon, Yahoo
  - Base stone

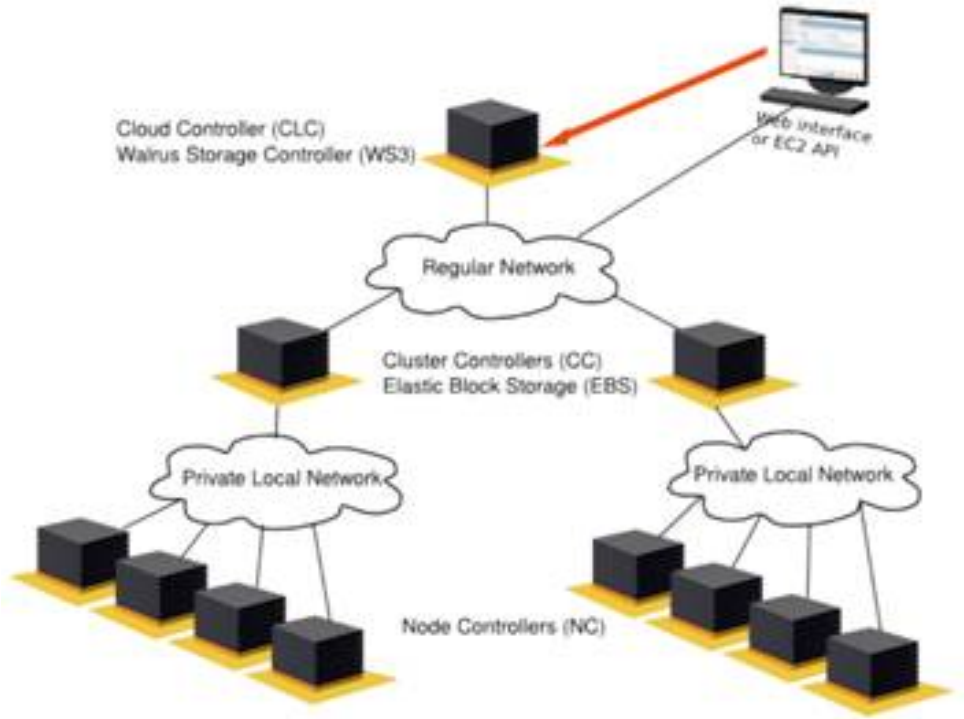
Core techniques of cloud computing

- Xen
- Xen Hypervisor
  - Server Virtualization with the Xen Hypervisor
    - Enterprises looking to increase server utilization, consolidate server farms, reduce complexity, and decrease total cost of ownership are embracing server virtualization.
    - The Xen® hypervisor is the fastest and most secure infrastructure virtualization solution available today, supporting a wide range of guest operating systems including Windows®, Linux®, Solaris®, and various versions of the BSD operating systems.



Core techniques of cloud computing

- Eucalyptus Systems
- **Ubuntu Enterprise Cloud** brings Amazon EC2-like infrastructure capabilities inside the firewall. The Ubuntu Enterprise Cloud is powered by Eucalyptus, an open source implementation for the emerging standard of EC2.



Obstacles and Opportunities

- From "Above the Clouds: A Berkeley View of Cloud Computing"

Table 6: Top 10 Obstacles to and Opportunities for Adoption and Growth of Cloud Computing.

	Obstacle	Opportunity
1	Availability of Service	Use Multiple Cloud Providers to provide Business Continuity; Use Elasticity to Defend Against DDOS attacks
2	Data Lock-In	Standardize APIs; Make compatible software available to enable Surge Computing
3	Data Confidentiality and Auditability	Deploy Encryption, VLANs, and Firewalls; Accommodate National Laws via Geographical Data Storage
4	Data Transfer Bottlenecks	FedExing Disks; Data Backup/Archival; Lower WAN Router Costs; Higher Bandwidth LAN Switches
5	Performance Unpredictability	Improved Virtual Machine Support; Flash Memory; Gang Scheduling VMs for HPC apps
6	Scalable Storage	Invent Scalable Store
7	Bugs in Large-Scale Distributed Systems	Invent Debugger that relies on Distributed VMs
8	Scaling Quickly	Invent Auto-Scaler that relies on Machine Learning; Snapshots to encourage Cloud Computing Conservationism
9	Reputation Fate Sharing	Offer reputation-guarding services like those for email
10	Software Licensing	Pay-for-use licenses; Bulk use sales

SaaS 介绍

- **什么是SaaS**

“软件即服务（SaaS）是一种通过Internet提供软件的模式，厂商将应用软件统一部署在自己的服务器上，客户可以根据自己实际需求，通过互联网向厂商定购所需的应用软件服务，按定购的服务多少和时间长短向厂商支付费用，并通过互联网获得厂商提供的服务。”

按照这个定义，“基于云计算的勘探信息应用PaaS平台”上应该提供大量的SaaS服务，这些服务是通过现有系统的包装而形成的。它们可以为油田各个部门提供基础服务。

- **SaaS对于用户的益处**
  - 部门用户无需自己购买和部署软件，而改为租用SaaS服务来执行业务活动。
  - 部门用户无需对软件进行维护，勘探信息应用PaaS平台会全权管理和维护软件。
  - 部门用户无需大量投资用于硬件、软件和人员，而只需要支出一定的租赁服务费用。
  - 部门用户无需大量时间用于部署系统，通常经过简单配置即可使用。

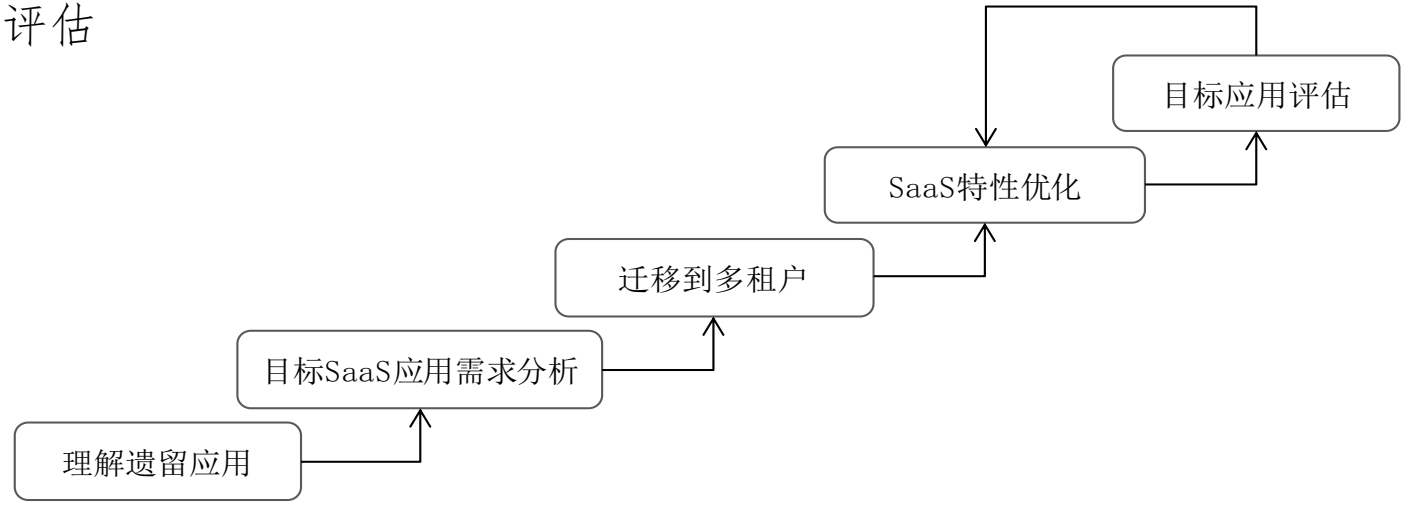
- **SaaS对于服务提供商的益处**
  - 效益高：SaaS 采取一对多的软件开发模式。一套软件，在边际成本几乎为0的情况下，可以租给多个租户，并获得成倍的收益。
  - 交付简单：不用提供物理形式的程序，只在后台给用户开通账号（数量和权限依用户需求而定）
  - 升级简化：不用向四面八方提供升级包，只需对自己的后台进行升级，所有用户都可以用到最新版本。
  - 技术支持远程化：无需到客户公司，进行实地技术支持。
  - 不再有盗版困扰：因为没有提供软件源码给外界，所以不存在盗版问题

现有应用迁移为SaaS

- **W2S（Web to SaaS）软件迁移原则**
  - 忠实于遗留应用
    - 不应改变其设计风格，要使熟悉该遗留软件的用户面对新版本的软件也能很快上手。
  - 迁移后的软件必须能够反映遗留软件的所有特性
    - 即达到内涵上的等同。
  - 迁移后的软件要不拘泥于遗留应用
    - 对不同机型、不同软件平台灵活采用有效的IT技术，从而充分发挥新的商业模式、软硬件平台的优势。
- **W2S迁移的挑战**
  - 遗留系统的理解
  - 多租户架构的迁移
  - 在线定制能力的增强

现有应用迁移为SaaS

- **W2S迁移步骤**
  - 理解遗留Web应用
  - 目标SaaS应用需求分析
  - 迁移到多租户模式
  - SaaS特性优化
  - 目标应用评估



Saas成熟度模型分级

– 根据可配置性，是否高性能，是否可伸缩，将Saas分为四个层级

	可配置	高性能	可伸缩
Level1 可定制	不满足	不满足	不满足
Level2 可配置	满足	不满足	不满足
Level3 高性能	满足	满足	不满足
Level4 可伸缩	满足	满足	满足

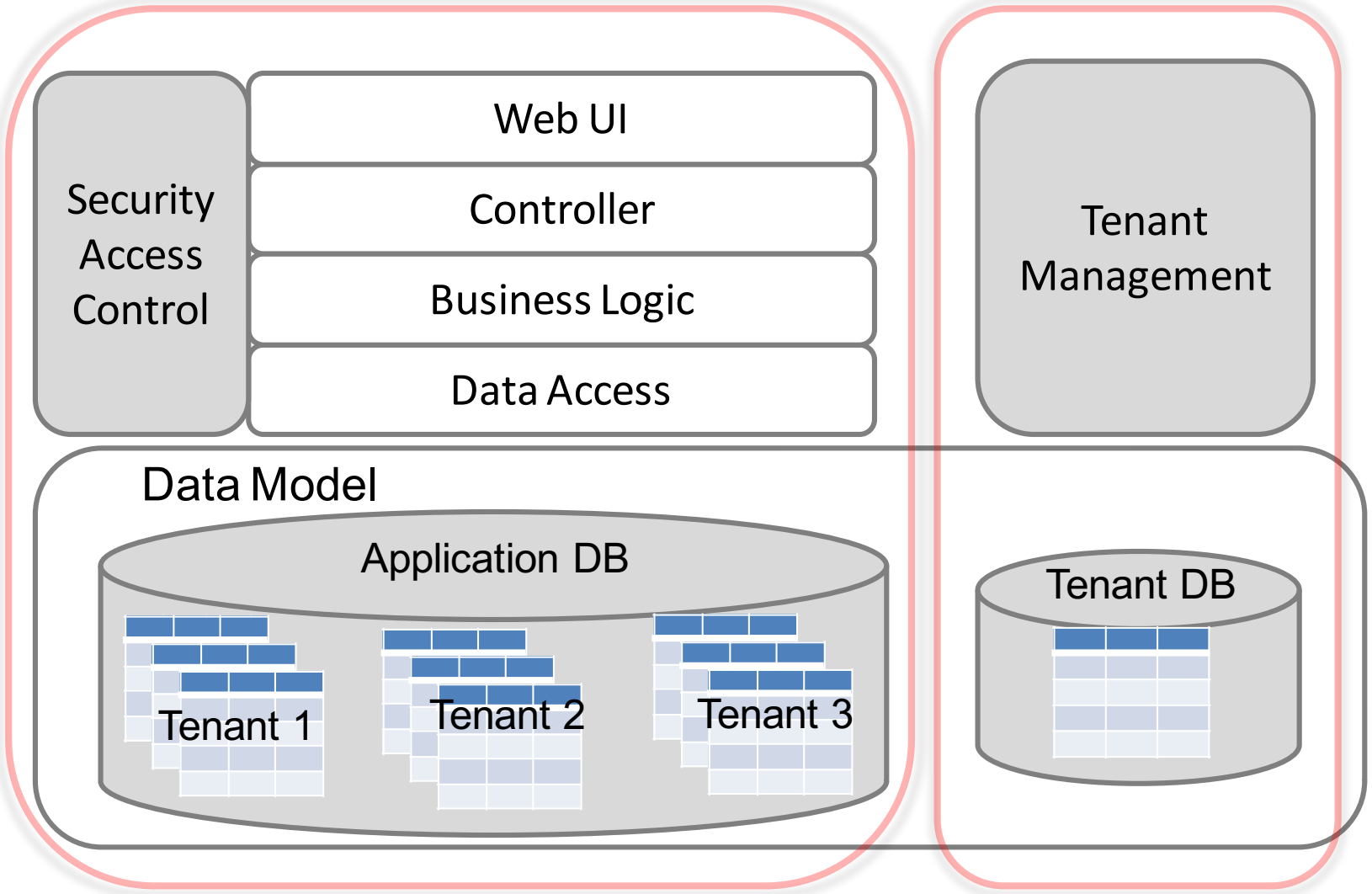
SaaS成熟度模型分级

- **Level1可定制**
  - 有一个客户项目，就根据客户需求定制一个版本，每个客户软件都有一份独立的代码，不同客户软件中可重用的只是少量代码和少量的可重用组件、库及开发人员的经验。
- **Level2 可配置**
  - 客户可以通过简单的配置，让通用的软件可以满足自己的一些个性需求，为每个客户独立部署运行实例，只不过实例云顶的是同一份代码。
- **Level3 高性能**
  - 多租户单实例的架构才是真正意义上的Saas应用架构，也就是通常所说的Multi-telant架构。
- **Level4 可伸缩**
  - 在用户数大量增长的情况下，无需更改架构，只需通过硬件设备的增加，直接适应需求

SaaS特性优化

- **SaaS可定制性**
  - 数据与表单定制
  - 功能定制与用户权限定制
  - 界面定制
  - 业务逻辑定制
- **定制增强可用技术**
  - 重构
  - 包装器和适配器
  - 模块和组件替换
  - 面向方面的编程

多租户架构


Security Access Control
Web UI
Controller
Business Logic
Data Access
Tenant Management
Data Model
Application DB
Tenant 1
Tenant 2
Tenant 3
Tenant DB

SaaS数据库设计的三种方法

- 改造Saas的重点在于数据的管理

- 共享数据（shared data)和独立数据(isolated data)的区别

这是一个连续的过程，其中可能有很多种变化，由于Saas是针对多租户开发的一种模式，所以数据存储架构和策略十分重要。

- 三种方法

根据数据的隔离程度分为三种开发策略，从左至右分别为独立数据库，共享数据库独立表架构，共享数据库共享表架构隔离数据架构三种模式模式，在实际生产应用中主要使用后两种策略。

SaaS数据库设计的三种方法

- **独立数据库模式**

  如下图所示，这种模式为每个租户（Tenant)分配独立的数据库，实现了最简单的数据隔离，防止其它租户对数据的恶意破坏，但是这种方式导致维护开发成本高，维护成本高，对用户的数据备份成本高。

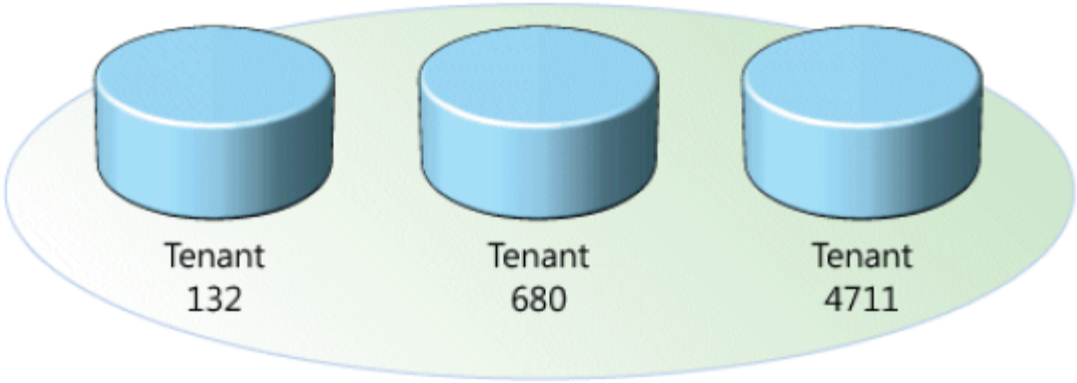


Figure 1. This approach uses a different database for each tenant

多租户 SaaS 的实现

- **共享数据库，独立表架构模式**

这种模式令多个租户共享同一个数据库，开发者为每个租户按照其需求开发不同的表，与其他租户隔离，并建立自己的表架构。这种模型易于扩展，但是难以发生故障时恢复。

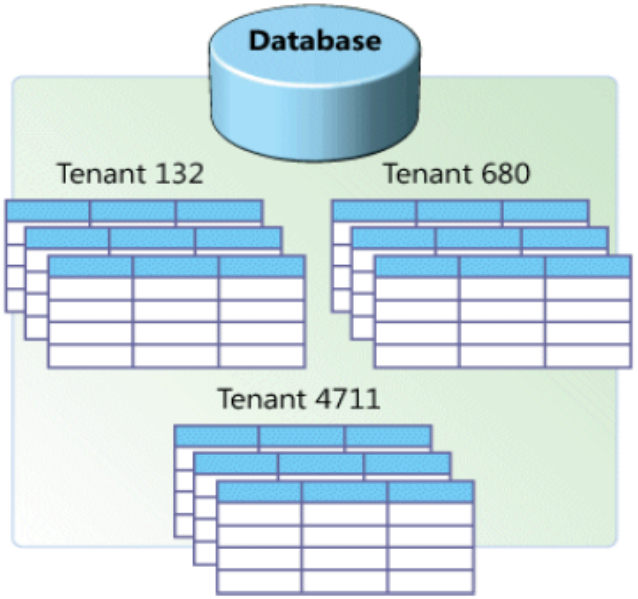


Figure 2. In this approach each tenant has its own separate set of tables in a common database

多租户 SaaS 的实现

- **共享数据库，共享表架构，独立表数据模式**

第三种模式是使用最多也最符合规范话的Saas的模式，这种模式使用相同的数据库和相同的一组表存储多个租户的数据。每个租户的ID与其相关的一行数据连接。这种方法能用最低的硬件成本和备份成本为最多的租户服务。

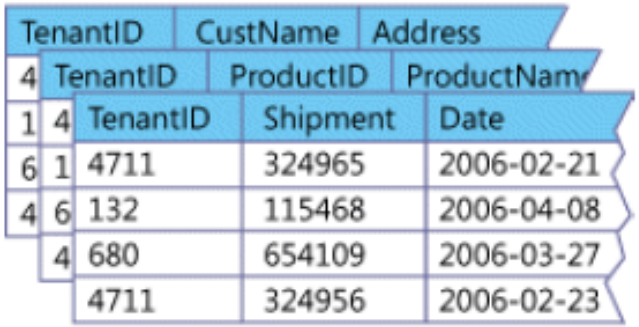


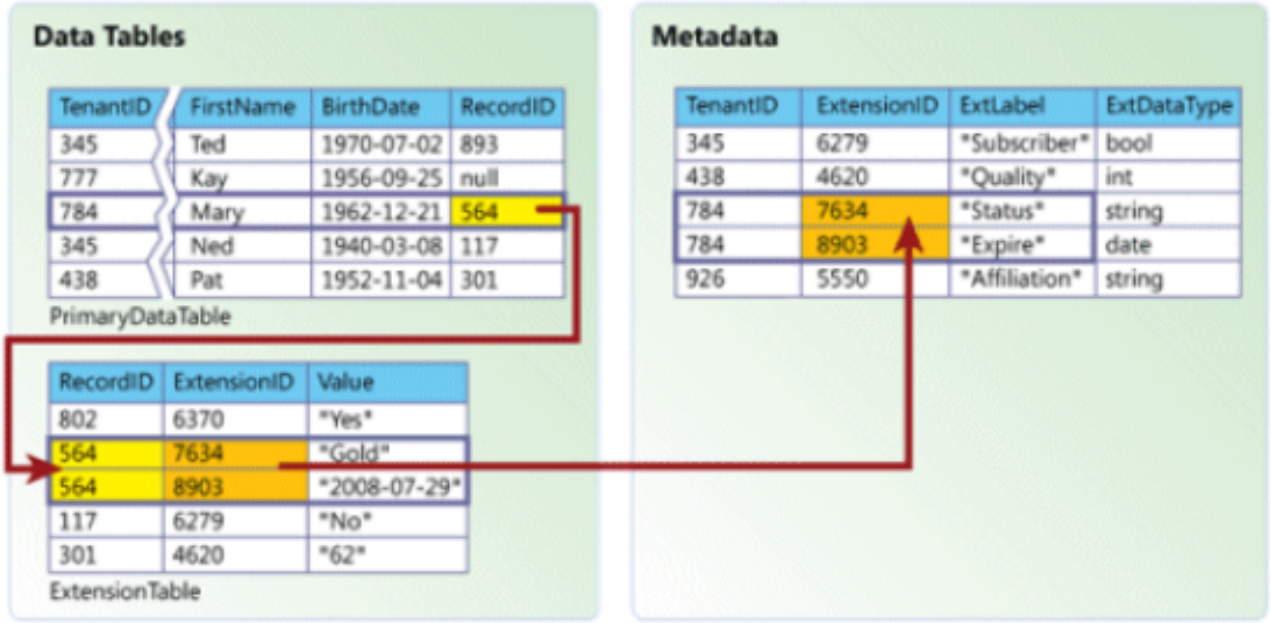
Figure 3. In this approach, all tenants share the same set of tables, and a Tenant ID associates each tenant with the rows that it owns

多租户 SaaS 的实现

存储策略（第三种）：

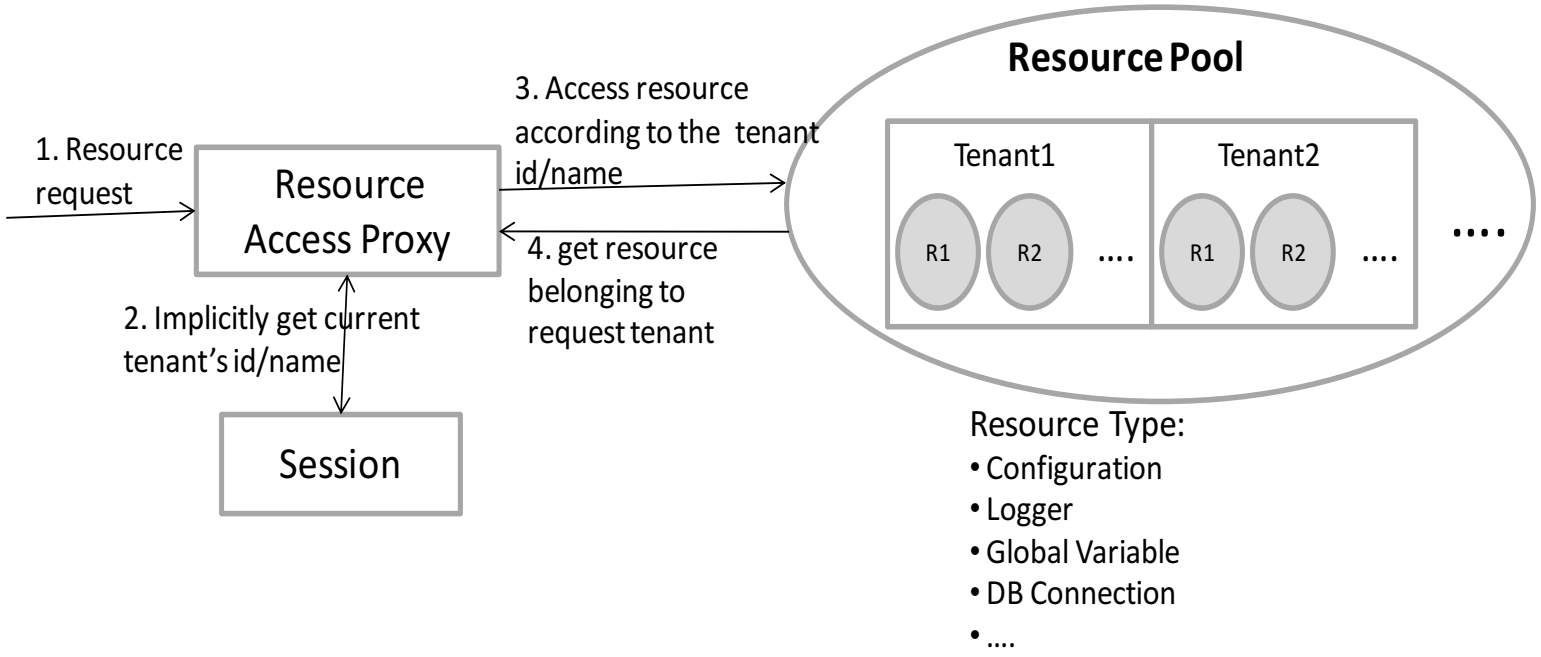
自定义的数据存储在一个单独的表中，并使用元数据来定义标签和每个租户的自定义字段的数据类型，让客户可任意扩展数据模型

- 例如下表中所示：



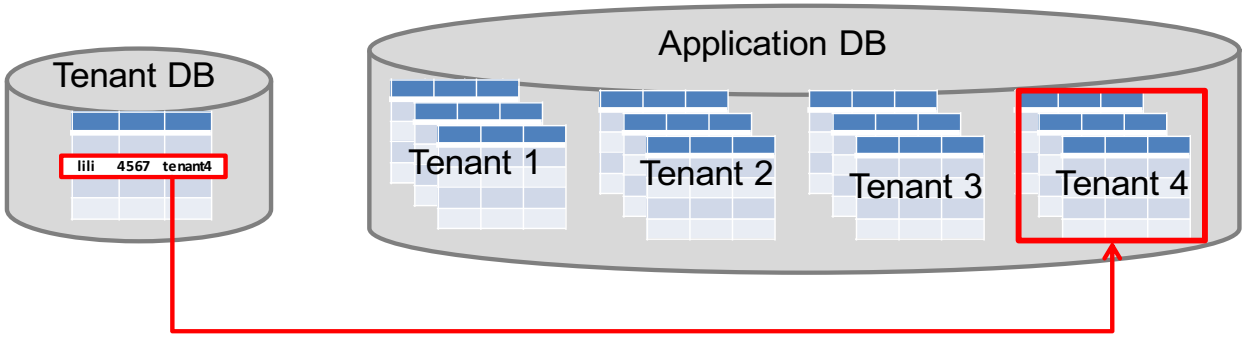
全局访问权限控制

- 多租户应用通常需要增加全局访问权限控制



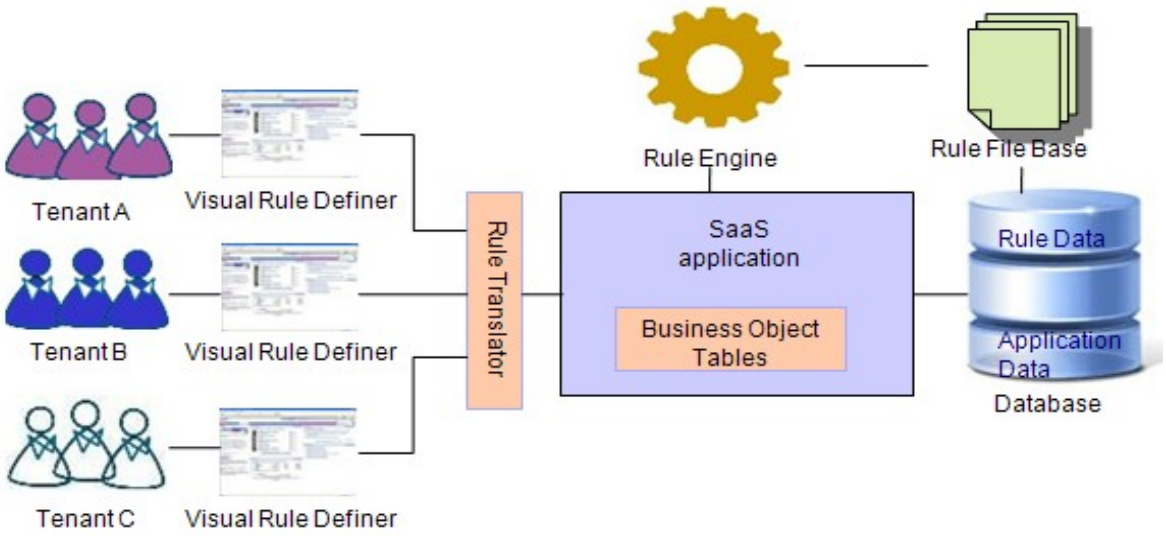
多租户管理与监控

- 多租户应用通常使用两级的管理架构——租户级和用户级



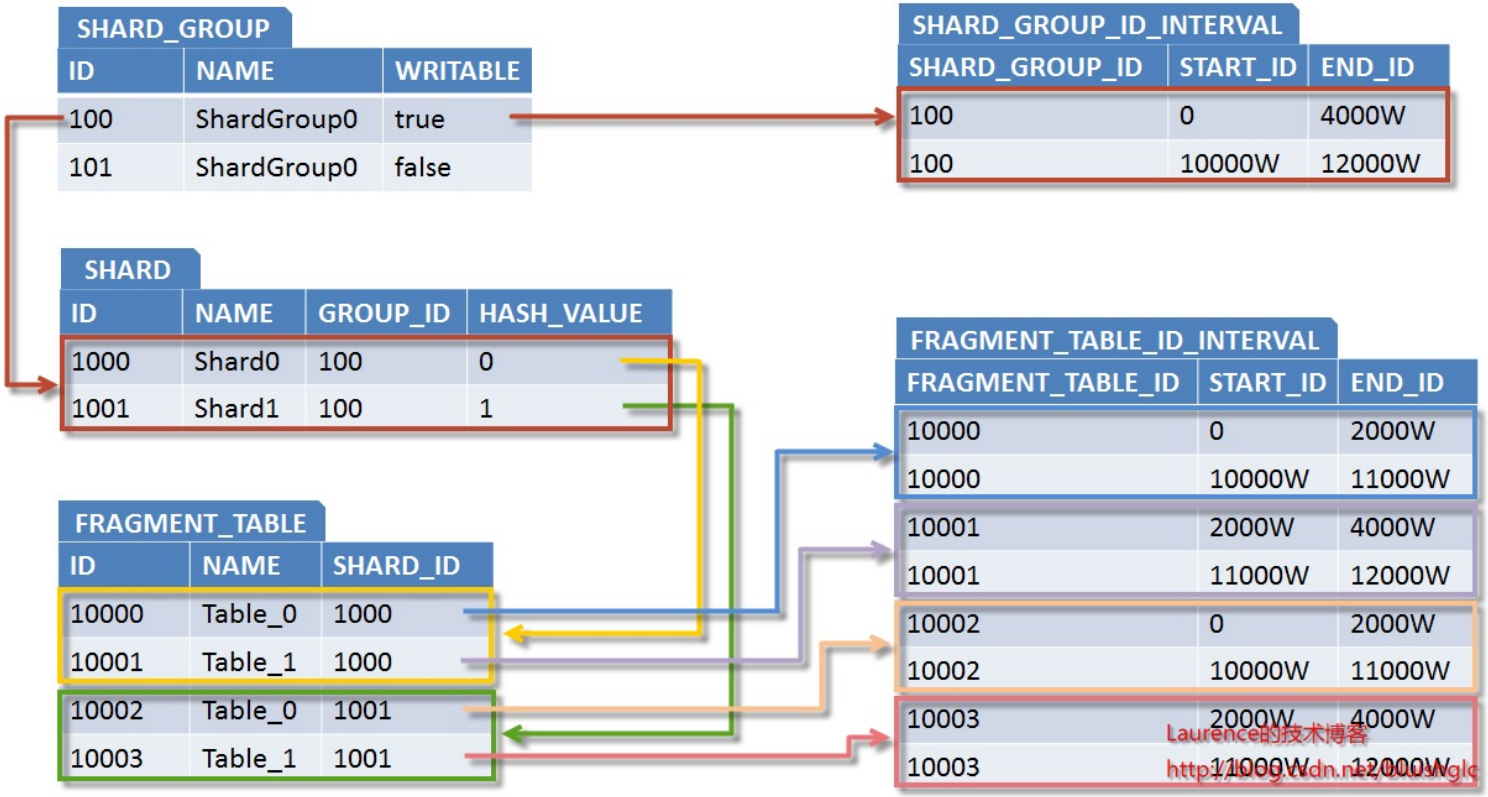
面向SaaS应用的业务逻辑定制框架

- 通过该框架可以定制业务逻辑



海量数据的存储与处理

- 下面来自网上的图片解释了海量数据Sharding存储的机制
  - 该机制正是SaaS应用所必需的数据存储方式



Project

- Requirement
  - Suppose you want to migrate your BookStore website into a cloud and provide SaaS to users. Please give the refactoring design of your BookStore website and tell us how your SaaS supports customization of workflow, DB and UI.

References

- Cloud Computing and Grid Computing 360-Degree Compared
  - Ian Foster, Yong Zhao, Ioan Raicu, Shiyong Lu
  - IEEE Grid Computing Environment 2008
- Above the Clouds: A Berkeley View of Cloud Computing
  - Michael Armbrust, Armando Fox, et.al,
  - http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf, 2009
- Introduction to Cloud Computing
  - Doug Terry: Chairman, ACM Tech Pack Committee on Cloud Computing
  - http://techpack.acm.org/cloud/
- Cloud Computing
  - http://searchcloudcomputing.techtarget.com/definition/cloud-computing

References

- MapReduce: Simplied Data Processing on Large Clusters
  - Jeffrey Dean and Sanjay Ghemawat
  - OSDI 2004
- The Google File System
  - Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung
  - SOSP 2003
- Bigtable: A Distributed Storage System for Structured Data
  - Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Mike Burrows, Tushar Chandra, Andrew Fikes, Robert E. Gruber
  - OSDI 2006
- Apache Hadoop
  - http://hadoop.apache.org/#What+Is+Hadoop%3F
- Xen
  - http://www.xen.org/products/xenhyp.html
- Ubuntu Enterprise Cloud
  - http://www.ubuntu.org.cn/products/whatisubuntu/serveredition/cloud/uec/

Thank You!