Data Recovery for Web Applications

Jorge Costa

Bruno Santos

Resumo

- Registo e rastreio de requests e bugs
- Ajuda os administradores;
- Ajuda a identificar responsáveis pelos danos provocados;
- Reutilização de logs fornecidos pela BD para uma recuperação selectiva;
- Relacionar os requests entre as diferentes camadas (3-tier), para identificar a melhor recuperação;
- Wordpress, Drupal e Gallery2.

Principais objectivos

1. Permitir administradores da aplicação Web diagnosticar falhas;
2. Disponibilizar uma recuperação selectiva aos dados, sem afectar o resto da aplicação.

Dois novos métodos:

1. Combina uma reprodução da aplicação com um método de rastreio usado em modo offline para encontrar as dependências ao nível aplicacional.
2. Explora os benefícios de usar dependências atómicas, para posteriormente abordá-las a um nível superior.

A abordagem:

- O objectivo é ajudar o administrador a identificar dados corrompidos por um bug ou uma má configuração, e que possa recuperar os dados de forma selectiva sem colocar em causa o resto da aplicação.

O modelo da aplicação

- Arquitectura Three Tier (apresentação, aplicação e base de dados);

- Aproveita as seguintes funcionalidades de aplicações Web:
  1. Os dados são guardados numa base de dados para controlo de concorrência;
  2. As aplicações são escritas em alto nível (PHP ou Java), permitindo a fácil monitorização;
  3. Cada servidor Web trata os requests de forma independente, criando muitas vezes um processo separado para assegurar isolamento.

Visão Geral

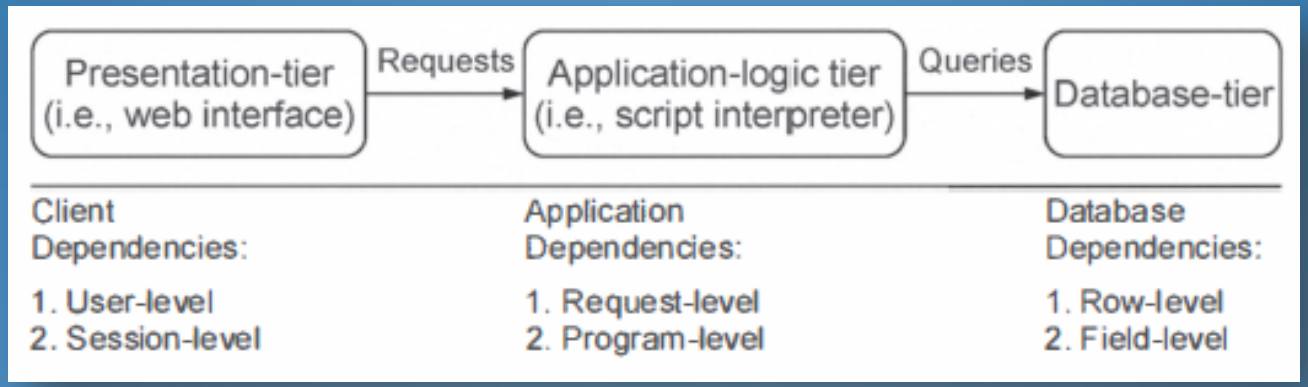
- O sistema é composto por:
  1. Um componente de monitorização em tempo de execução;
  2. Dois componentes, sendo um de analise e outro de recuperação de dados após ser detectado alguma corrupção.

Visão Geral

- Os componentes de analise e recuperação são usados depois do administrador verificar que a sua página não foi apresentada como o esperado. Esses componentes usam dados que foram recolhidos durante a monitorização, para guiar o administrador através do processo de recuperação.

Monitorização

- Os monitores seguem e correlacionam os requests em todas as camadas da aplicação.



- Os monitores criam um *log* com informação suficiente para permitir criar um mapa por cada request a cada transacção da base de dados,e em cada transacção é expecificado as tabelas e as linhas modificadas
- Estes requests e transacções formam em conjunto com a base de dados o undo log.

Análise

- O componente de análise ajuda a determinar a corrupção e/ou perda de dados que são cruciais para uma recuperação eficiente.

- O componente de análise usa os dados colhidos durante a fase monitorização para derivar três tipos de dependências:
    1. Base de dados
    2. Aplicação
    3. Cliente

Recuperação

- O componente de recuperação fornece ferramentas que simplificam este processo.
- Além de fornecer informações como tempo, ou tabelas especificas, ajuda o utilizador a saber a raiz do problema.

Component	Existing Software	Changed Lines
DB Monitor	MySQL	287
Application-logic Monitor	PHP interpreter	219
Application-logic Analysis	PHP interpreter with taint support	519
Query Rewriter	JSQLParser	1850
Recovery Component	-	4757

Implementação

- Foi implementado um protótipo do sistema de recuperação, para PHP, e MYSQL. Em MYSQL foi implementado tainting com o rewrite das querys, modificando um pouco o JSQLParser, em vez de alterar a própria base de dados, que consiste numa tarefa muito mais complexa.

Avaliação

- Politicas de dependência:
  1. Request-level dependency with row-level tainting (request-row)
  2. Program-level dependency with row-level tainting (program-row)
  3. Database-level dependency with row-level tainting (database-row)
  4. Program-level dependency with field-level tainting (program-field)
  5. Database-level dependency with field-level tainting (database-field)

Resultados

Table 2. Recovery accuracy for request-level and program-level dependency policies. The false positives column shows numbers without and with table whitelisting, respectively.

Case	Total Number of Requests	Requests to Undo	Dep. Policy	False Positives	False Negatives
Wordpress - link category rename	109	1	none	0	0
			request-row	60	0
			program-row	8	0
			program-field	6	0
Drupal - lost voting information	118	7	none	0	6
			request-row	111/100	0
			program-row	95/89	0
			program-field	89/0	0
Drupal - lost comments	117	1	none	0	0
			request-row	116/102	0
			program-row	100/93	0
			program-field	95/0	0
Gallery2 - removing permissions	91	1	none	0	0
			request-row	90/13	0
			program-row	88/11	0
			program-field	82/10	0
Gallery2 - resizing images	151	1	none	0	0
			request-row	148/0	0
			program-row	139/0	0
			program-field	119/0	0

Resultados

Table 3. Recovery accuracy of database-level dependency policies. All numbers indicate queries.

Case	Queries to Undo	Dep. Policy	False Positives	False Negatives	Inconsistencies after Undo
Wordpress - link category rename	23	database-row	0	15	The count value does not match the actual number of links.
		database-field	0	21	
Drupal - lost voting information	38	database-row	86	16	The poll_votes table has duplicate entries.
		database-field	0	18	
Drupal - lost comments	24	database-row	116	0	none
		database-field	0	0	
Gallery2 - removing permissions	9	database-row	97	0	The global sequence id has an old value breaking future inserts requiring a new id.
		database-field	9	0	
Gallery2 - resizing images	17	database-row	110	0	
		database-field	20	0	

- Wordpress – é uma aplicação de blogging popular que permite os utilizadores criarem conteúdos e associar categorias para uma apresentação mais organizada.
- Drupal – permite o administrador criar uma votação através de escolha multipla.
- Gallery2 – tem uma mecanismo de controlo de acesso aprofundado. Permite atribuir vários recursos específicos para fotos ou álbuns.

Conclusão

- Através deste sistema foi possível verificar que um esquema baseado em marcas (tainting) mostra ao administrador como recuperar e diagnosticar vários cenários de corrupção e erros.

“FIM