

Requerimientos de Big Data y tecnologías seleccionadas en 3 sistemas reales

Martín Chaia



Jornadas de Definición Estratégica en Big Data 2013

Agenda

- 1 Los 3 sistemas
- 2 Requerimientos comunes
- 3 Requerimientos diferenciadores
- 4 Selección de tecnologías

PAM

Price & Availability Manager

Sistema que administra y provee precios y disponibilidad para hoteles con *contrato directo*

- Requerimientos funcionales
 - ▶ Búsquedas interactivas de precios y disponibilidad de hoteles
 - ▶ Retener, confirmar y cancelar reservas sobre habitaciones
 - ▶ Guardar precios y disponibilidad de dos años hacia adelante y un año hacia atrás

- Algunos números
 - ▶ 20,000 hoteles ($\approx 15\%$ de nuestro catálogo total)
 - ▶ Cantidad de precios = 14,000,000 (muchas “dimensiones”)

Capa de servicios que sirve a la GUI gestionando las búsquedas de vuelos

- Requerimientos funcionales
 - ▶ Paging, filtering & sorting
 - ▶ Cache de búsquedas

- Algunos números
 - ▶ 1,800,000 búsquedas diarias
 - ▶ 30 búsquedas por segundo en horas pico
 - ▶ 1 resultado de búsqueda \approx 1 Mb

Herramienta de automatización de pruebas que usamos para nuestros procesos de calidad

- Requerimientos funcionales
 - ▶ Almacenar definiciones de tests
 - ▶ Ejecutar tests en forma automática
 - ▶ Almacenar resultados de ejecuciones de tests
 - ▶ Categorización de tests
 - ▶ Interfaz gráfica de administración y consulta
- Algunos números
 - ▶ 50 aplicaciones
 - ▶ 400 “baterías” de tests
 - ▶ 30,000 resultados de tests automáticos diarios

Requerimientos comunes

- Todos necesitan una “base de datos” ...
 - ▶ ... en el sentido de *repositorio*, no necesariamente *relacional*
- Volumen
- Escalabilidad
- Performance (R/W)
- Disponibilidad
- Modelos semi-estructurados (*schema-less*, *key/value*)
- [Commodity | High end virtualized] HW

Requerimientos diferenciadores

PAM

- Stock
- Transacciones

FSM

- Cache
- Estado
- “Sesión”

Henry

- Gran colección de resultados
- Heterogeneidad de resultados
- Resultados de gran tamaño

Productos Big Data



- No SPOF
- Consistencia “ajustable”
- Escrituras super-rápidas
- CQL
- Map/Reduce
- Monolítico



- Consistencia “estricta”
- MVCC, ICV, CAS
- HDFS & Map/Reduce
- SPOF failover
- Modular



- Orientado a “documento”
- JSON/BSON
- SPOF failover
- GridFS
- Modular

Un jueguito

Unir según corresponda

PAM



Henry



FSM



Un juegoito

Unir según corresponda

PAM



A P A C H E
HBASE

Henry



FSM



Un juegoito

Unir según corresponda

PAM



APACHE
HBASE

Henry



FSM



Un juegoito

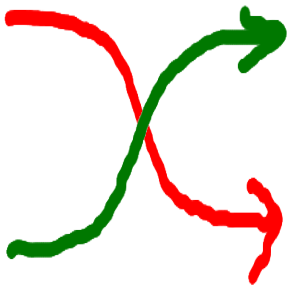
Unir según corresponda

PAM



APACHE
HBASE

Henry



FSM



Disclaimer

- *Mapping* “requerimientos → tecnología” no siempre es tan “lineal”
- Otros *mappings* válidos (PAM → Cassandra ??)
- Otras tecnologías (PAM → RDBMS, FSM → Redis)

¡Gracias!

- Contacto:

- ▶ Martín Chaia - mchaia@despegar.com
- ▶ Research @ Despegar - research2@despegar.com