**Migration of Database and Optimizing Web Queries**

**Executive Summary**

OneTravel is a well-known travel booking website. They cater to millions of travelers allowing them to book their flights, hotels, rental cars. In their nascent stage they created a MySQL database which has grown very big with huge amount of data. The purpose of this project is to migrate the database to SAP S4 HANA.

Since the existing database is inefficient and the queries are slow, they would like to redesign the database. Data clean up would be the first step of the project.

The website queries to the backend database are slow since they are pointing to old database. They would like to rewrite the queries which would now go to a new database and of course be more efficient and faster.

**Business Objectives**

The client would like to achieve following objectives:

1. Data Analysis / Usage Reporting on existing database.
2. Design proposal on new database.
3. Implementation of proof of concept on sandbox system.
4. Implementation and deployment in production environment.

**Background**

In a recent user survey, most of the users who rated their experience negatively complained about slowness of the website. They were not able to locate their bookings fast enough and the flight comparison tool was buggy and slow. In the investigation client found out that the backend database inefficiency was cause of most of the issues. S4 HANA database is a faster database with column-based storage and is expected to be a good fit for the client because they have a huge amount of data to store.

**Scope**

The scope of the project includes:

1. Data Usage Analysis: Run analysis reports on existing database to understand the bottlenecks in data base.
2. Data Clean up: The old database needs to be looked at from data redundancy point of view. There might be duplicate data which can be deleted reducing our data foot print.
3. Database redesign: Since we are moving to a new database we would like to take this opportunity to redesign the database so we don’t carry over same data inefficiencies over to new system.
4. Retrofit queries from website: The queries from website/application need to be re-written so fit the new data base schema correctly.

The project does not include redesigning of the UX.

**Functional requirements**

The current data is being stored in at an SQL Server located in system sysdb01. The database schema can be located at the path /usr/bin/data/database/travelsystem/schema.db . The data is stored in 6 main tables along with some secondary tables.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Description** | **Size (in Millions)** | **Comments** |
| Flights | Contains flight bookings | 3.2 M |  |
| Hotels | Contains Hotel Bookings | 2.3 M |  |
| CarRentals | Contains rental car bookings | 2.03 M |  |
| UserProfile | Contains User Details | 4 M |  |
| FlightCache | Contains active/saved searches by users | 8-10M | Possible duplicates and redundant data. It probably contains fields which are not needed. |
| PriceCache | Contains current prices of flights and discounts | 500K | This data refreshes every 1 hour and is source of prices users see on applications. |

The real time data (about potential bookings, *flights/hotel etc being looked at) is* being stored in tables such as FlightCache, *PriceCache,* HotelsCache. Any time a user queries a particular route of flight or a hotel in a city, these tables get updated . We suspect these tables have duplicate entries and that create an issue for the end user as either the website shows them duplicate records or it’s very slow to load saved searches.

Since we have decided to upgrade to a fast database system like S4 HANA we would like to do an analysis of the data to see if we can reduce the database size. The tables would have to be redesigned and the data redundancy would have to be taken care off.

The project would comprise of following phases:

1. Data Usage Analysis: Using database administration tools we would like to get a report on which data is being accessed and at what frequency. The report should be based on database analysis from last 3 months plus holiday season of Oct-Dec 2021. The report should include which tables are being accessed and what percent of data is not being accessed at all. That would help us make a decision what data can be archived. We would like to get a proposal on how much data can safely be archived.
2. Data Clean up: Apart from archiving the unused data the old database needs to be looked at from data redundancy point of view. There might be duplicate data which can be deleted reducing our data foot print. Smaller tables can be efficiently indexed by S4 HANA database and it will lead to a faster user experience.
3. Database redesign: Moving to S4 HANA , we will have to redesign the database. That would give us opportunity to decouple fields which are rarely accessed from frequently accessed data points. We would like see a proposal of new database schema and relationships between tables before its implemented.
4. Retrofit queries from website: The queries from website/application need to be re-written so fit the new data base schema correctly.

Once all reports and database redesign has been agreed on, a sandbox implementation should be done and loaded with 30% of data from current database. A proof of concept must be implemented in this sandbox system which should show us the expected improvement on data query, data foot print parameters.

Once the proof of concept has been tested, the implantation on development environment can be started.

**Personnel requirements**

The following skillset is required in the team. The team size and exact composition will be determined by technical manager.

1. Database Administration – To provide reports of current data usage
2. Data Analyst – To analyze the current data situation and to report on data redundancy
3. Database developer(s) – S4 HANA
4. Web Development/Application Development – to re-write the queries from website.
5. System Administration – To create the sandbox system and migration of new solution

**Delivery schedule**

The project would be broken into following phases:

1. Data Analysis Report: The report on current database is due in 2 months. The report should include

* what data can be archived,
* what data can be deleted.
* The overall impact of the proposed data clean up should also be quantified.
* The management would want to know how much can the data size be reduced and if such data clean up exercise would be enough to fix issues faced by end users.

1. New Database schema proposal: Based on discussions around the data analysis report, the new proposed database schema is due in 1 month after go-ahead from management.
2. Sandbox system build: Once the database schema is agreed on the system administration then will build the sandbox system and hand over the credentials to database developers in 2 weeks.
3. Database built and prototyping: The database developers will implement the new schema and load the system with 30% data from production environment. The web developers will then re-write the queries and make the prototype ready. This phase of project would be completed over 3 months since system built date.
4. Testing/QA : The testing and QA team will test the prototype and submit bugs to development team. The development team will fix the bugs and retest the system. All major bugs are expected to resolved in 2 months.
5. Development system redesign: Once the sandbox system is deemed satisfactory and the go-ahead is given the development may begin in development system. Estimated time : 4 months (includes test and QA) .
6. Deployment: Once testing dev system is complete, the database schema and data reload may happen in production environment. The total system downtime must be reported prior to this step. And also any risk mitigation should be considered.

**Other requirements**

Apart from redesigning the current database, as extension, the client would like to explore possibilities of migrating the database to cloud. S4 HANA does offer to host the database in the cloud. There is added cost to using their cloud offerings. The client wants to know the cost comparison of maintaining database on their own servers vs using the cloud hosting of database.

**Assumptions**

As the new provider of database services, we have hyper support available from SAP for duration of 1 year. Their experts will be available to us to resolve any technical issues

**Limitations**

The total budget for this project is estimated between 2-2.5 Million dollars. The procurement of S4 HANA services is included in this estimate. The final decision will be made depending on Return on investment (ROI) analysis. Initial analysis shows that the project would break even in 5 years after implementation.

**Risks**

The following risks have been identified. The project management should plan for mitigation of these risks and prepare a contingency report.

1. The migration of database fails: The project management should ensure there is a backup database available and the code has maintained versions so that the system can fall back to original system.
2. The downtime of website should be planned so it does not fall prior to a major holiday/travel dates where the company generates most of the traffic and revenue.