

BTS Services Informatiques aux Organisations

Barème de notation du rapport de stage (à insérer en première page du rapport de

stage) Nom de l'étudiant : Deniz Sarigul

	Très insuffisant	Insuffisant	Assez bien	Très bien
Respect des consignes	0	0.5	1	2
Mise en page et Orthographe				
Mise en page	0	0	0.5	1
Orthographe	0	0	0	1
Présentation Structuration...	0	0	0.5	1
Evaluation du contenu				
Sommaire et remerciements	0	0	0.5	1
Présentation de l'entreprise	0	0.5	1	2
Présentation du thème du stage	0	0.5	1	2
Analyse de l'existant et des besoins	0	1	2	3
Recherche et mise en place des solutions	0	1	2	3
Conclusion	0	0.5	1	2
Rédaction	0	0.5	1	2

Rapport remis le		Pénalités de retard	
------------------	--	---------------------	--

Appréciation globale	Note / 20
----------------------	-------------------------

Professeur ayant suivi le stage :

Rapport de Stage

Deniz Sarigul
BTS SIO SLAM 1ère année

Du 13 mai au 28 juin 2024 chez Contentsquare



Contentsquare

Sommaire

Remerciements.....	5
Présentation de l'entreprise.....	6
Thème du stage et travail à réaliser.....	7
Analyse de l'existant et des besoins.....	8
Réalisation.....	9
Le déroulé du stage.....	12
Conclusion.....	13

Remerciements

Je tiens à exprimer ma gratitude à l'équipe de Contentsquare pour m'avoir offert l'opportunité de réaliser ce stage au sein de leur entreprise. Je remercie particulièrement mon tuteur de stage, Ali-Firat Kilic, pour son soutien, ses conseils précieux et son encadrement tout au long de ces sept semaines. Je remercie également mon équipe pour leur accueil chaleureux et leur aide précieuse, qui ont grandement contribué à mon apprentissage.

Présentation de l'entreprise

Contentsquare est une entreprise technologique spécialisée dans l'analyse de l'expérience utilisateur sur les sites web et les applications mobiles. Fondée en 2012 par Jonathan Cherki, Contentsquare est devenue le leader mondial dans son domaine, aidant les entreprises à comprendre comment les utilisateurs interagissent avec leurs plateformes numériques.

Leur mission est simple : aider les entreprises de toutes tailles et de tous secteurs à créer un monde digital plus humain.

L'entreprise est présente internationalement notamment, New-York, Londres, Sydney, Tokyo, Barcelone etc.

Les secteurs d'activités des clients de Contentsquare sont majoritairement:

- Automobile
- Luxe
- Banque & Assurance
- Cosmétiques
- Energie
- Tourisme et Hôtellerie

Thème du stage et travail à réaliser

Mon stage au sein de Contentsquare s'est déroulé sur une période de sept semaines, durant laquelle j'ai travaillé en tant que data engineer. Mon principal objectif était d'implémenter une requête SQL permettant de connaître le nombre de requêtes effectuées sur chaque table le jour précédent. Cela va permettre d'avoir des statistiques d'utilisation pour mon équipe.

Travail à réaliser

- Développer et implémenter une requête SQL efficace pour recueillir les données nécessaires.
- Intégrer cette solution dans l'outillage existant.
- Documenter le travail réalisé et former l'équipe à l'utilisation de cette nouvelle requête.

Analyse de l'existant et des besoins

Existant

Contentsquare utilise plusieurs clusters ClickHouse(base de données analytics format colonnes) pour stocker et analyser de grandes quantités de données.Cependant, l'équipe manquait d'une vue claire et consolidée sur l'utilisation quotidienne des tables de la base de données, ce qui compliquerait les choses en cas d'incidents.

Besoins

L'équipe avait besoin d'un outil permettant de :

- Suivre le nombre de requêtes effectuées sur chaque table quotidiennement.
- Identifier rapidement les tables les plus sollicitées pour prioriser les interventions en cas d'incidents.
- Recevoir des rapports quotidiens sous forme de messages Slack pour une meilleure réactivité.

Réalisation

J'ai développé une requête SQL permettant de connaître le nombre de requêtes effectuées sur chaque table le jour précédent, les tables qui n'ont eu aucune requête sont aussi affichées.

```
SELECT
    _table AS use_table_name,
    cnt,
    st.name AS table_name
FROM
(
    SELECT
        splitByChar('.', `table`)[2] AS _table,
        count(1) AS cnt
    FROM system.query_log
    ARRAY JOIN tables AS `table`
    WHERE (event_date = yesterday())
    AND (NOT hasAny(databases, ['system']))
    AND (query_kind = 'Select')
    AND (NOT startsWith(_table, `_tmp`)) AND (type IN (2, 3, 4))
    GROUP BY _table
) AS used_tables
FULL OUTER JOIN system.tables AS st ON st.name = _table
WHERE database = 'default'
ORDER BY cnt DESC ;
```

Un exemple de ce qu'affiche cette requête :

	ABC use_table_name ▼	123 cnt ▼	ABC table_name ▼
1	views_local	2 004	views_local
2	views	1 981	views
3	sessions	105	sessions
4	sessions_local	105	sessions_local
5	projects	2	projects
6		0	devices_names
7		0	dynamic_vars_string
8		0	sessions_products
9		0	sessions_merge
10		0	tokenized_string_mapping_local
11		0	snapshots_local
12		0	goals
13		0	views_contents
14		0	iab_review_local_kafka_mv
15		0	alias_predicates

Régénérer Save Cancel

Les use_table_name représentent les tables utilisées avec leurs nombres de requêtes effectuées et les table_name sont toutes les autres tables qui n'ont pas été utilisées.

J'ai intégré cette requête dans un programme écrit en Golang, en voici un extrait:

```

// The runLakeHouseUsageCmd function is executed when the lakehouse-usage command is run.
func runLakeHouseUsageCmd(_ *cobra.Command, _ []string) error {
    // Get the ClickHouse Analytics URL based on the provided region
    clickhouseAnalyticsURL, ok := providers.ClickHouseClusters[providers.LakeHouse][lhRegion]
    if !ok {
        return fmt.Errorf("%s is an invalid ClickHouse analytics region", lhRegion)
    }
    clickhouseAnalyticsURL = fmt.Sprintf(clickhouseAnalyticsURL, lhEnv)

    // Define credentials for connection to Clickhouse Cluster with User and Password.
    sourceCreds := providers.ClickHouseCredentials{User: lhUser, Password: lhPassword}

    // The number of queries successful per table for yesterday.
    Query := "SELECT _table AS use_table_name, " +
        "cnt, " +
        "st.name AS table_name " +
        "FROM " +
        "(" +
        "SELECT splitByChar('.', `_table`)[2] AS _table, " +
        "count(1) AS cnt " +
        "FROM system.query_log " +
        "ARRAY JOIN tables AS `_table` " +
        "WHERE event_date = yesterday() " +
        "AND NOT hasAny(databases, ['system']) " +
        "AND query_kind = 'Select' " +
        "AND NOT startsWith(_table, '_tmp') " +
        "AND type IN (2, 3, 4) " +
        "GROUP BY _table " +
        ") AS used_tables " +
        "FULL OUTER JOIN system.tables AS st ON st.name = _table " +
        "WHERE st.database = 'default' " +
        "ORDER BY used_tables.cnt DESC" +
        BuildSettings(&Settings{true, true})

    // Execute the query with the ExecuteClickHouseQuery function.
    sourceResponse, err := helpers.ExecuteClickHouseQuery[Response](
        Query,
        clickhouseAnalyticsURL,
        sourceCreds)
    if err != nil {
        return err
    }
}

```

Le déroulé du stage

Du 13 mai au 28 juin: Découverte de l'entreprise, de l'équipe, compréhension des outils et technologies utilisés. Nous avons comme exemple ClickHouse qui est un système de gestion de bases de données (SGBD) spécialisé conçu principalement pour l'analyse de données à grande échelle. Contrairement à des SGBD relationnels traditionnels comme MySQL, PostgreSQL ou SQL Server, ClickHouse (base de données orienté colonne) est optimisé pour des performances élevées sur des volumes de données massifs, néanmoins ClickHouse se rapproche des autres SGBD est l'utilisation de SQL pour l'interaction avec la base de données et le modèle de données relationnel.

17 Mai: Début de ma formation en GO, je me suis formé via un cours en ligne suivi d'exercices. Cela me permettra de pouvoir ajouter mes requêtes SQL dans un programme en GO dans l'outillage existant.

30 mai : Exploration et formation sur le format des tables ClickHouse. Je commence réellement mon travail de Data Engineer. Ma première tâche est de réaliser une requête SQL qui compte le nombre total de requêtes effectuées sur chaque table le jour précédent sur leur base de données 'default'.

4 juin : J'ai commencé à réaliser un code en GO pour implémenter cette requête. Cela m'a pris 2 semaines, mais cela était très instructif, j'ai beaucoup appris que ça soit du GO mais aussi travailler avec le terminal.

Mercredi 18 juin : On a fait une réunion dans l'après-midi avec l'équipe pour que je présente le projet que j'ai effectué mais aussi le bilan de mon stage, des attentes que j'avais, de ce que j'ai appris.

Jusqu'à la fin du stage : J'ai commencé à rédiger mon rapport de stage et en même temps je faisais de petites tâches en rapport avec SQL.

Conclusion

Travail réalisé

Au cours de ce stage, j'ai pu analyser les besoins de l'équipe, développer une requête SQL pour suivre l'utilisation des tables, et intégrer cette solution dans l'infrastructure de Contentsquare.

Bénéfices personnels

Ce stage m'a permis de développer mes compétences en SQL et Go, ainsi que ma capacité à analyser et répondre aux besoins techniques d'une équipe. J'ai également appris à travailler au sein d'une équipe agile, à gérer des projets et à communiquer efficacement mes résultats.

Je remercie encore une fois toute l'équipe de Contentsquare pour cette expérience enrichissante, qui a été un tremplin important pour ma carrière future dans le domaine de l'informatique et des données.

