

# Statistiques d'index

Christophe LAPORTE

SQL Server MVP / MCM



SQLSaturday Montreal 2017



**@conseilit**



# Christophe Laporte



Microsoft®  
SQL Server®

~ since 1997

6.5 <= SQL Server <= 2016



christophe\_laporte@hotmail.fr



<http://conseilit.wordpress.com/>



**Microsoft®**  
**CERTIFIED**

*Master*



@conseilit

SQLSaturday Montreal 2017



# Christophe Laporte



Microsoft®  
SQL Server®

~ since 1997

6.5 <= SQL Server <= 2016



**christophe\_laporte@hotmail.fr**



**<http://conseilit.wordpress.com/>**

- Migrations
- Formations
- Remote DBA
- Hébergement BDD



**Microsoft®**  
**CERTIFIED**

*Master*



**@conseilit**

- Conseil
  - Infrastructure / Architecture
  - Virtualisation / Cloud
  - Haute disponibilité / Montée en charge
  - Optimisation / Dépannage
- Audit

# Agenda

---

- 11 démos pour comprendre les statistiques
- Q/R (s'il reste du temps)

# F5 – et la magie opère

---

- Schématiquement, la requête est
  - Compilée (parsing / algébrisation / optimisation)
  - Le plan d'exécution est mis en cache
  - La requête est exécutée
- Démo 1 : plan d'exécution

# Pas de la magie : des statistiques

---

- Informations générales
- Densité
- Histogramme
- Démo 2 : montre-moi une statistique

# Et ça fonctionne

---

- Plutôt bien
- Enfin pas tout le temps !
- Démo 3 : parameter sniffing



# CHECKPOINT - on se pose pour faire un point

---

- Fonctionne bien pour requête ad-hoc
- Malgré le problème de parameter sniffing
- Les statistiques sont utiles pour créer un "bon" plan d'exécution



# Que se passe t'il en l'absence de statistiques

---

- Par défaut, si besoin de statistiques
  - SQL Server va en générer
- Démo 4 :
  - Paramètre AUTO\_CREATE\_STATISTICS
- Pas un bon plan
  - dans tous les sens du terme !

# Les statistiques auto générées

---

- Démo 5 : les statistiques auto générées
- Donc on conserve actif le paramètre AUTO\_CREATE\_STATISTICS

# Le problème des statistiques (1)

---

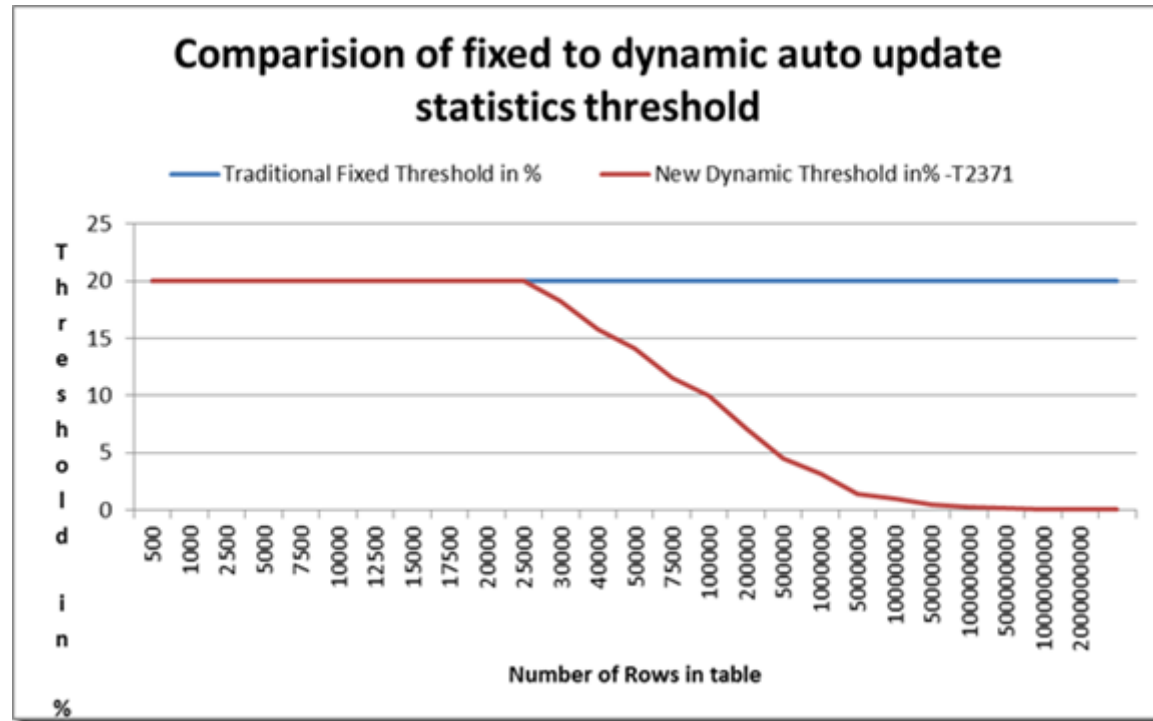
- L'histogramme ...
- Quand ?
  - Nombre important de lignes dans table
  - Nombreuses valeurs distinctes sur une colonne
- 200 valeurs + NULL
  - Problème de précision
  - Problème de distribution
  - Démo 6
- Fonctionne bien
  - si distribution homogène

# Le problème des statistiques (2)

---

- La mise à jour des données
  - Démo 7
- Mise à jour des statistiques
  - Seuil : 20% + 500
  - Rebuild Index
  - Mise à jour manuelle des statistiques
- Autre solution : Trace flag 2371
  - Démo 8

# TF 2371



# Le problème des statistiques (3)

---

- Les nouvelles données
  - Les statistiques se basent sur les données connues
  - Estimer le nombre d'enregistrements hors plage
    - $<$  à la valeur minimale
    - $>$  à la valeur maximale
- Démo 9

# Le problème des statistiques (4)

---

- Présence de statistiques auto générées
  - Peut être un index manquant
  - Je crée l'index ! (merci de ne pas généraliser ...)
  - Mais la statistique auto générée reste
  - Démo 10
- Veiller à ne pas avoir de doublons
  - Index <-> statistique auto générée
  - Consomme des ressources lors de la mise à jour



# Quelques solutions

---

- Update stats + fréquents
  - Attention aux statistiques auto générées
  - Scripts de type [Ola Hallengren](#)
- statistiques filtrées
  - Mais attention !
    - Si données s'étendent sur plusieurs intervalles
    - Démo 11
- Plan guides
  - Attention : en optimisation, tout peut changer
- SQL 2014
  - Nouveau Cost Estimator
  - Statistiques incrémentales (tables partitionnées)

# Récapitulatif

---

- Le plan d'exécution est dépendant des statistiques
- Des statistiques à jour sont donc nécessaires
- Si l'on peut "sniffer" la valeur d'un paramètre
  - l'estimation provient de l'histogramme
  - Estimation relativement précise
  - Si le plan est sauvegardé, cette valeur sera utilisée pour les prochaines exécutions (Procédures stockées)
- Si l'on ne peut pas "sniffer" le paramètre
  - l'estimation provient de la densité
  - Estimation basée sur une "moyenne"
  - Le plan généré convient à la majorité des valeurs distinctes
  - Suivant les cas, ce sera aussi performant, ou moins performant qu'une estimation issue d'un histogramme
  - Si le plan est sauvegardé, aucune valeur spécifique n'est liée au plan
  - Fonctionne plutôt bien lorsque la dispersion des données est homogène

# Q & A

---

- Q & A
- Merci pour votre attention

