

Une étude de performance et de passage à l'échelle des systèmes de traitement de données dans les smart grids

vendredi 7 octobre 2016 14:45 (45 minutes)

De nos jours, la taille des données générées et collectées au sein des smart grids ne cesse d'augmenter. Ces données proviennent, en grande partie, des compteurs communicants ainsi que les capteurs déployés massivement à travers le réseau électrique. De plus, la génération de plus en plus fréquente de ces données, introduit de nouveaux défis de performance et de passage à l'échelle pour les systèmes traditionnels de gestion de données smart grid. Dans ce travail, nous menons une étude de comparaison afin d'explorer les capacités de différentes approches de gestion et de traitement de données pour le cas de compteurs communicants. A cette fin, nous menons une série d'expériences approfondies afin d'évaluer plusieurs systèmes pour des traitements types sur les données de compteurs. Parmi ces systèmes, nous déployons un système de gestion de bases de données relationnelles réparties (RDBMS), des systèmes basés sur Map Reduce comprenant Hadoop et Spark, et un système NoSQL avec une architecture pair à pair. Les ensembles d'expériences ont été conduits sur la Grid'5000 sur un ensemble de données de 1.4 To et jusqu'à 140 nœuds de calcul (serveurs). Les résultats obtenus démontrent que les RDBMS répartis sont plus adaptés pour les traitements de calculs intensifs d'agrégations mais sont très lents pour charger les données. En revanche, nous démontrons qu'avec une bonne combinaison de modèles et de systèmes, nous atteignons d'excellentes performances à l'échelle et cela pour les traitements les plus coûteux sur les données de compteurs.

Orateur: M. CHIHOUB, Housseem (INPG Entreprise SA)

Classification de Session: Big Data & HPC