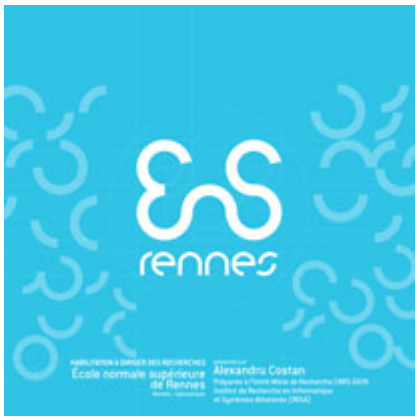
EFFICIENT STREAM DATA MANAGEMENT : FROM BIG DATA TO FAST DATA

le 14 mars 2019 14h00

ENS Rennes, Salle du conseil

Soutenance d'Habilitation à Diriger des Recherches de Alexandru Costan
Spécialité Informatique



This manuscript provides a synthetic overview of my research journey since my PhD defense. The document does not claim to present my work in its entirety, but focuses on the contributions to data management in support of stream processing. These results address all stages of the stream processing pipeline: data collection and in-transit processing at the edge, transfer towards the cloud processing sites, ingestion and persistent storage.

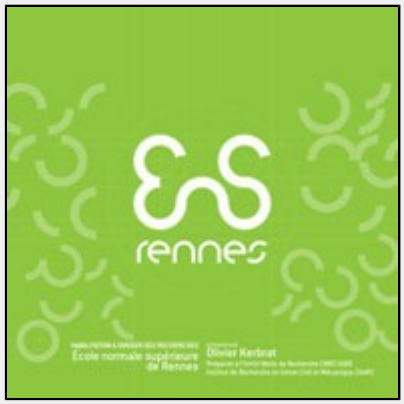
I start by presenting the general context of stream data management in light of the recent transition from Big to Fast Data. After highlighting the challenges at the data level associated with batch and real-time analytics, I introduce a subjective overview of my proposals to address them. They bring solutions to the problems of in-transit stream storage and processing, fast data transfers, distributed metadata management, dynamic ingestion and transactional storage. The integration of these solutions into functional prototypes and the results of the large-scale experimental evaluations on clusters, clouds and supercomputers demonstrate their effectiveness for several real-life applications ranging from neuro-science to LHC nuclear physics. Finally, these contributions are put into the perspective of the High Performance Computing - Big Data convergence.

THÉMATIQUE(S)

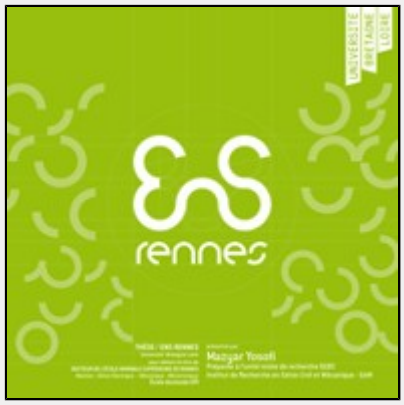
Recherche - Valorisation, Vie des personnels, Vie de l'École

Mise à jour le 12 mars 2019

À LIRE AUSSI



Performance des procédés pour une fabrication innovante et durable



Méthodologie de caractérisation prédictive des procédés de fabrication additive avec une approche technique, économique et environnementale

Modèles mathématiques de type «Hamiltonian Mean-Field» : stabilité et méthodes numériques autour d'états stationnaires