

عنوان مقاله:

مروری بر اندازه گیری صحت و کیفیت داده ها

محل انتشار:

سومین کنفرانس ملی نوآوری و تحقیق در مهندسی برق و مهندسی کامپیوتر و مکانیک ایران (سال: 1396)

تعداد صفحات اصل مقاله: 6

نویسنده:

الهه السادات مصطفوی - کارشناسی ارشد مهندسی کامپیوتر گرایش نرم افزار دانشگاه آزاد اسلامی واحد ملارد

خلاصه مقاله:

یکی از کاربردهای عمده نسل آتی سیستم های موازی و توزیع شده، مربوط به تحلیل داده های عظیم است. مخازن داده برای چنین کاربردهایی امروزه بیش از چندین آگزابایت بوده و به سرعت نیز در حال افزایش هستند. علیرغم حجم بسیار زیاد این مخازن، این دیتاست ها و همچنین اپلیکیشن های نظیر آنها، چالش های عمده ای را برای متدها و نرم افزارهای برنامه نویسی مربوطه ایجاد کرده اند. دیتاست ها معمولا توزیع شده بوده و نیز حجم آنها و دسترسی مجاز به آنها توسط تکنیک های توزیع شده تضمین شده است. داده ها معمولا روی یک پلت فورم با قابلیت محاسباتی و شبکه ای بالا، مقیم هستند. توجه به میزان تحمل خطا، امنیت، و کنترل دسترسی موضوع مهمی در بسیاری از کاربردهاست. برای بیشتر کاربردهای درحال ظهور، مدل ها و متدهای مبتنی بر داده، که قادر به عملیات در مقیاس های مختلف هستند، هنوز برایمان ناشناخته است. حتی در صورتی که متدهای شناخته شده مقیاس پذیر باشند، اعتبارسنجی نتایج آنها موضوع مهمی خواهد بود. مشخصات پلت فورم های سخت افزاری و نیز پشته های نرم افزاری، اساسا تحلیل داده ها را تحت تاثیر قرار داده اند. در این پژوهش ابتدا به بررسی مدیریت کیفیت داده ها و تولید آنها می پردازیم. در بخش های بعدی مقیاس و محدوده تحلیل داده و ویژگی های حجم کار را مورد بررسی قرار می دهیم و با ارایه نتیجه گیری بحث خود را پایان می دهیم.

کلمات کلیدی:

اندازه گیری داده، صحت داده ها، کیفیت داده، داده های بزرگ، تجزیه و تحلیل، مراکز داده

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/741108>

