

رشد سریع و نمایی حجم داده‌های دیجیتالی، افزایش تنوع در انواع مختلف داده‌های ذخیره شده و افزایش بسیار سریع نرخ تولید داده‌های دیجیتال منجر به تغییرات مهم در فناوری دیجیتال شده است. این تغییرات باعث باز شدن درهای جدید در علم کامپیوتر شده است که به آن علم داده‌های حجیم یا کلان داده‌ها گفته می‌شود.

ابشرکت‌های اینترنتی با داده‌های حجیم مواجه هستند و توانایی پردازش این داده‌ها از منظر سرعت و کارایی و همچنین انرژی و توان مصرفی چالش بزرگی است که شرکت‌های مربوطه با آن مواجه هستند. برای این منظور سامانه‌ها و بسترهای پردازشی نرم‌افزاری جدیدی به منظور پردازش داده‌های بزرگ طراحی شده است. از طرف دیگر علاوه بر طراحی بسترهای نرم‌افزاری جدید نیاز به طراحی پردازنده‌های مناسب برای پردازش این داده‌ها نیز وجود دارد. اولین قدم برای این منظور شناسایی مشکلات کارایی در سطح پردازنده‌های نوین فعلی است تا بتوان بر اساس نتایج به طراحی پردازنده‌های مناسب پرداخت.

در این پژوهش دو دسته مهم از بسترهای موجود برای پردازش کلان داده‌ها به نام‌های آپاچی اسپارک و آپاچی استورم انتخاب شده‌اند. با استفاده از برنامه‌های محک استاندارد کارایی این سیستم‌ها را بر روی ریزمعماری‌های نوین شرکت اینتل بررسی کردیم. در نهایت گلوگاه‌های کارایی در سطح زیرمعماری برای این دسته از بسترها شناسایی شد و با ذکر دلیل برای حل آنها پیشنهاداتی ارائه کردیم. همچنین انرژی و توان مصرفی این پردازنده‌ها برای پردازش داده‌ها مورد بررسی قرار گرفت و بر اساس مشاهدات و یافته‌ها پیشنهاداتی برای بهبود انرژی مصرفی ارائه شد.