



USO DE TECNOLOGIAS E PADRÕES NO DESENVOLVIMENTO DE APLICAÇÕES JAVA/SCALA PARA BIG DATA

Filipe Paulo Coutinho¹, Adalberto Cajueiro de Farias²

RESUMO

O uso de Analytics nos diversos setores e segmentos é um diferencial competitivo que traz diversas melhorias para seus usuários. Entretanto, a atividade de explorar dados requer conhecimentos técnicos especializados e alta necessidade de lidar com grandes volumes de dados (Big Data). Nesse contexto, o desenvolvimento de aplicações requer uso de tecnologias, linguagens e ferramentas que proporcionem transparência no acesso e na exploração dos dados, permitindo o desenvolvedor se concentrar exclusivamente nos algoritmos relacionados ao negócio. Nesta pesquisa exploramos as ferramentas Hadoop e Spark como bibliotecas de acesso a grandes volumes de dados e as linguagens Java e Scala como plataformas de desenvolvimento de aplicações. Realizamos também o uso combinado delas: Java-Hadoop, Java-Spark, Scala-Hadoop e Scala-Spark. Concluiu-se que a última dessas combinações se apresentou como melhor infraestrutura para desenvolvimento de aplicações para Big Data, pois combinou e expressividade e suporte a concorrência de Scala, com a transparência de manipulação dos dados e performance de Spark.

Palavras-chave: Engenharia de Software, Big Data, Scala, Hadoop, Spark.

USE OF TECHNOLOGIES AND PATTERNS IN THE DEVELOPMENT OF JAVA/SCALA APPLICATIONS FOR BIG DATA

ABSTRACT

The use of Analytics in many fields is a competitive feature that brings several improvements to its users. However, the data exploration activity requires a high technical knowledge as well as the capability of dealing with huge volumes of data (Big Data). In this context, application development involves Technologies, languages and tools providing transparency in terms of data access and exploration, allowing, therefore, in such a way that the developer concentrates only on the business algorithms. In this research, we explore the Hadoop and Spark tools as access libraries for huge amounts of data and the languages Java and Scala as application development platforms. We also have combined them: Java-Hadoop, Java-Spark, Scala-Hadoop e Scala-Spark. We conclude that the last combination provided the best infrastructure for the development of applications related to Big Data, because it uses the expressiveness and support for concurrency of Scala with the transparency for data manipulation and performance of Spark.

Keywords: Software Engineering, Big Data, Scala, Hadoop, Spark.

¹Aluno do Curso de Ciência da Computação, Departamento de Sistemas e computação, UFCCG, Campina Grande, PB, e-mail: filipe.coutinho@ccc.ufcg.edu.br

²Professor Doutor, Departamento de Ssistemas e Computação, UFCCG, Campina Grande, PB, e-mail: adalberto@computacao.ufcg.edu.br