

# Definizione ed Implementazione di un Sistema di Raccomandazione Distribuito per film e Modellazione di Eventi Complessi

*Prof. Ing.* Tommaso di Noia

*Prof.ssa* Marina Mongiello

Mauro Losciale

Pietro Tedeschi



Logica e Intelligenza Artificiale  
Ingegneria del Software Avanzata  
Laurea Magistrale in Ingegneria Informatica  
Politecnico di Bari  
A.A 2015 - 2016

# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Stato dell'arte</b>	<b>1</b>
2.1	Introduzione ai sistemi CEP . . . . .	1
2.2	Sistemi di Raccomandazione . . . . .	1
2.2.1	Filtro collaborativo . . . . .	1
2.3	Introduzione allo Stream Processing . . . . .	1
2.3.1	Il paradigma Publish-Subscribe . . . . .	1
2.4	Il pattern Facade . . . . .	1
2.5	Il pattern Singleton . . . . .	1
2.6	Il pattern Model-View-Controller (MVC) . . . . .	1
2.7	La tecnologia WebSocket . . . . .	1
<b>3</b>	<b>Analisi del progetto</b>	<b>1</b>
<b>4</b>	<b>Soluzione proposta</b>	<b>1</b>
4.1	La libreria Spark . . . . .	1
4.1.1	Spark Streaming . . . . .	1
4.1.2	Spark mllib . . . . .	1
4.1.3	Spark SQL . . . . .	1
4.2	Apache Kafka . . . . .	1
4.2.1	Panoramica . . . . .	1
4.2.2	Integrazione con Spark Streaming . . . . .	1
4.3	Il framework Node.js . . . . .	1
4.3.1	Panoramica . . . . .	1
4.3.2	Kafka Client per Node.js . . . . .	1
4.3.3	Il framework Angular.js . . . . .	1
4.4	La libreria socket.IO . . . . .	1
<b>5</b>	<b>Conclusioni e sviluppi futuri</b>	<b>1</b>
	<b>Bibliografia</b>	<b>2</b>

<b>1</b>	<b>Introduzione</b>
<b>2</b>	<b>Stato dell'arte</b>
2.1	Introduzione ai sistemi CEP
2.2	Sistemi di Raccomandazione
2.2.1	Filtro collaborativo
2.3	Introduzione allo Stream Processing
2.3.1	Il paradigma Publish-Subscribe
2.4	Il pattern Facade
2.5	Il pattern Singleton
2.6	Il pattern Model-View-Controller (MVC)
2.7	La tecnologia WebSocket
<b>3</b>	<b>Analisi del progetto</b>
<b>4</b>	<b>Soluzione proposta</b>
4.1	La libreria Spark
4.1.1	Spark Streaming
4.1.2	Spark mllib
4.1.3	Spark SQL
4.2	Apache Kafka
4.2.1	Panoramica
4.2.2	Integrazione con Spark Streaming
4.3	Il framework Node.js
4.3.1	Panoramica
4.3.2	Kafka Client per Node.js
4.3.3	Il framework Angular.js
4.4	La libreria socket.IO
<b>5</b>	<b>Conclusioni e sviluppi futuri</b>

## Riferimenti bibliografici

- [1] Apache<sup>TM</sup>. Hadoop Official Website. <https://hadoop.apache.org/>, 2015. [Online; Ultimo accesso 10 Ottobre 2015].
- [2] Fernando Kakugawa, Liria Sato, Mathias Brito. Architecture to integrating heterogeneous databases using grid computing. In *21st International Conference on Systems Engineering*, 2011.
- [3] Leonardo Candela, Donatella Castelli, and Pasquale Pagano. gCube: a service-oriented application framework on the grid. *ERCIM News*, 72:48–49, 2008.
- [4] Leonardo Candela, Donatella Castelli, and Pasquale Pagano. Managing big data through hybrid data infrastructures. *ERCIM News*, 89:37–38, 2012.
- [5] Carmela Comito and Domenico Talia. GDIS: A service-based architecture for data integration on grids. In *On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops*, pages 88–98. Springer, 2004.
- [6] J.C.S. Dos Anjos, G. Fedak, and C.F.R. Geyer. Bighybrid – a toolkit for simulating mapreduce in hybrid infrastructures. In *Computer Architecture and High Performance Computing Workshop (SBAC-PADW), 2014 International Symposium on*, pages 132–137, Oct 2014.
- [7] D. Garlasu, V. Sandulescu, I. Halcu, G. Neculoiu, O. Grigoriu, M. Marinescu, and V. Marinescu. A big data implementation based on grid computing. In *Roedunet International Conference (RoEduNet), 2013 11th*, pages 1–4, Jan 2013.
- [8] L. Hluchy, O. Habala, V. Tran, P. Krammer, and B. Simo. Using ADMIRE framework and language for data mining and integration in environmental application scenarios. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on*, volume 4, pages 2437–2441, July 2011.
- [9] University of Chicago. Globus Toolkit Home Page. <http://toolkit.globus.org/toolkit/>, 2015. [Online; Ultimo accesso 10 Ottobre 2015].
- [10] The University of Edinburgh. OGSA-DAI Official Website. <http://www.ogsadai.org.uk/>, 2015. [Online; Ultimo accesso 10 Ottobre 2015].
- [11] Tapio Niemi, Marko Niinimäki, Vesa Sivunen. Integrating distributed heterogeneous databases and distributed grid computing. In *ICEIS 5th International Conference on Enterprise Information Systems*, 2003.
- [12] T. M. Sloan, A. Carter, P. J. Graham, D. Unwin, and I. Gregory. First data investigation on the grid: Firstdig.
- [13] Sourceforge. FUSE : File system in userspace. <http://fuse.sourceforge.net/>, 2015. [Online; Ultimo accesso 10 Ottobre 2015].
- [14] Yukako Tohsato, Takahiro Kosaka, Susumu Date, Shinji Shimojo, and Hideo Matsuda. H: Heterogeneous database federation using grid technology for drug discovery process. In *In Grid Computing in Life Science (LSGRID2004) Edited by: Konagaya A, Satou K*. Springer.
- [15] Nicolas Viennot, Mathias Lécuyer, Jonathan Bell, Roxana Geambasu, and Jason Nieh. Synapse: A microservices architecture for heterogeneous-database web applications. In *Proceedings of the Tenth European Conference on Computer Systems*, EuroSys '15, pages 21:1–21:16, New York, NY, USA, 2015. ACM.

- [16] Ćorgi Kakaševski, Anastas Mišev, Boro Jakimovski. Heterogeneous distributed databases and distributed query processing in grid computing. *ICT Innovations Web Proceedings ISSN 1857-7288*, 2011.