

# Big data based fraud risk management at Alibaba

Jidong Chen\*, Ye Tao, Haoran Wang, Tao Chen

*Alipay, Alibaba, China*

Received 15 February 2015; revised 8 March 2015; accepted 20 March 2015

Available online 14 July 2015

---

## Abstract

With development of mobile internet and finance, fraud risk comes in all shapes and sizes. This paper is to introduce the Fraud Risk Management at Alibaba under big data. Alibaba has built a fraud risk monitoring and management system based on real-time big data processing and intelligent risk models. It captures fraud signals directly from huge amount data of user behaviors and network, analyzes them in real-time using machine learning, and accurately predicts the bad users and transactions. To extend the fraud risk prevention ability to external customers, Alibaba also built up a big data based fraud prevention product called AntBuckler. AntBuckler aims to identify and prevent all flavors of malicious behaviors with flexibility and intelligence for online merchants and banks. By combining large amount data of Alibaba and customers', AntBuckler uses the RAIN score engine to quantify risk levels of users or transactions for fraud prevention. It also has a user-friendly visualization UI with risk scores, top reasons and fraud connections.

© 2015, China Science Publishing & Media Ltd. Production and hosting by Elsevier on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Keywords:* Fraud detection and prevention; Risk model; Malicious behavior; Risk score; Big data analysis

---

## 1. Introduction

Big data is an all-encompassing term for any collection of data sets which are large, complex and unstructured, so that it becomes difficult to process using traditional data processing applications. Big data “size” is dynamic and constantly growing, as of 2012 ranging from a few dozen terabytes to many petabytes of data by the time when the article is written. It is also a set of techniques and technologies that to analyze, capture, curate, manage and process data within a tolerable elapsed time (Wikipedia).

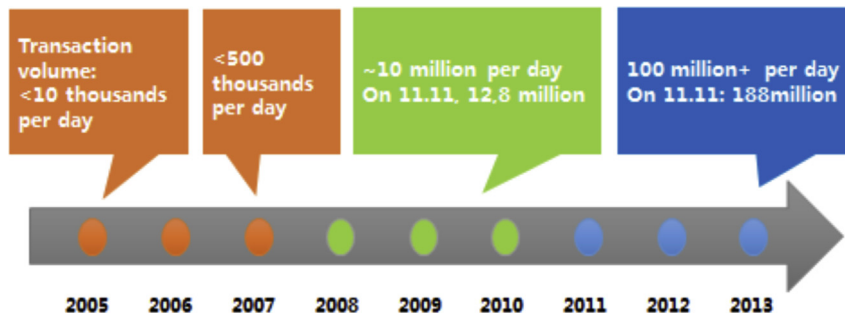
Big data has many different purposes – fraud risk management, web display advertising, call center optimization, social media analysis, intelligent traffic management and among other things. Most of these analytical solutions were not possible previously because data technology were unable to store such huge size of data or processing technologies were not capable of handling large volume of workload or it was too costly to implement the solution in a timely manner.

---

\* Corresponding author.

*E-mail addresses:* [jidong.cjd@alibaba-inc.com](mailto:jidong.cjd@alibaba-inc.com), [chenjd@gmail.com](mailto:chenjd@gmail.com) (J. Chen).

Peer review under responsibility of China Science Publishing & Media Ltd.



Graph 1. Transaction volume increases at Alibaba.

With business needs arose, Alibaba employed optimized system and platform and developed advanced methodology and approach to handle 10 billion level daily volume. It started from data platform of RAC in 2009, via GP (Green Plum, an EMC product, see EMC<sup>2</sup>) and Hadoop (see White<sup>7</sup>), and is using ODPS now. Data processing and analyzing is also improved from T + 1 mode<sup>1</sup> to near real-time mode.

By adapting big data techniques, Alibaba highlights advances made in the area of fraud risk management. It invents a real-time payment fraud prevention monitoring system, called CTU (Counter Terrorist Unit). And CTU becomes one of the most advanced online payment fraud management system in China, which can track and analyze accounts' or users' behavior, identify suspicious activities and can apply different levels of treatments based on intelligent arbitration.

Fraud risk models are one of the supportive layers of CTU<sup>2</sup> (Counter Terrorist Centre). They use statistical and engineering techniques to analyze the aggregated risk of an intermediary (an account, a user or a device etc). Detailed attributes are generated as inputs. Different algorithms are to assess the correlations of these attributes and fraud activities, and to separate good ones from bad ones. Validating and tuning are to make sure models apply to different scenarios. Big data at Alibaba produces thousands and thousands of attributes, and fraud risk models are built to deal with various of fraud activities.

Those big data based fraud models are widely used in almost every procedure within Alibaba to monitor fraud, such as account opening, identity verification, order placement, before and after transaction, withdrawal of money, etc. To build up a safe and clean payment environment, Alibaba decides to expand this ability to external users. A user-friendly product is built, called AntBuckler. AntBuckler is a product to help merchants and banks to identify cyber-crime risks and fraud activities. And a risk score (RAIN Score) is generated based on big data analysis and given to merchants and banks to tell the risk level.

In this paper, we show that Alibaba applies the big data techniques and utilizes those techniques in fraud risk management models and systems. We also present the methodology and application of the big data based fraud prevention product AntBuckler used by Alibaba.

The remainder of the paper is organized as follows. Section 2 introduces big data applications and basic computing process at Alibaba. Section 3 explains fraud risk management at Alibaba in details and fraud risk modeling. Section 4 provides an explanation of AntBuckler. We conclude in Section 5.

## 2. Big data applications at Alibaba

Alibaba grows fast in past 10 years. In 2005, daily transaction volume was less than 10 thousands per day. It reached to 188 million on Nov 11th, 2013 one day. Graph 1 illustrates that the transaction volume at Alibaba changes from 2005 to 2013 on daily basis.

With business growing exponentially, data computing, processing system and data storage are bound to change as well. It started from data computing platform of RAC (Oracle Real Application Clusters (see Oracle white paper<sup>1</sup>)) in 2009, via GP and Hadoop, and is using ODPS now. Data processing and analyzing is also improved from T + 1 mode

<sup>1</sup> T + N mode: T is time, when system runs. N is time interval. T + 1 means the system runs on the second day.

<sup>2</sup> CTU: Alipay's internal risk control system, which is fully developed and designed by Alibaba. This name is inspired by American TV series <<24>>.

Download English Version:

<https://daneshyari.com/en/article/1002011>

Download Persian Version:

<https://daneshyari.com/article/1002011>

[Daneshyari.com](https://daneshyari.com)