Natalino Busa
Senior Director Data

Singapore, Amsterdam
natalino.busa@gmail.com
phone/text: +65 8438 2445

Senior Director of Data: Covering the full spectrum of Data Applications as in Engineering, Science, Analytics, Reporting, AI, Big Data and Streaming Data. Managing all data related activities & teams. Reporting directly to CFO/CEO.

Natalino Busa is skilled at defining, designing and implementing custom big/fast data solutions (web, mobile, api, dataflows) for data-driven applications such as predictive analytics, personalized marketing, fraud detection and business event monitoring.

Natalino has successfully led teams across the world with effective, fit-for-purpose and reliable solutions. He provides a unique combination of business vision, applied data science and engineering skills.

Summary

Chief Data Scientist, Data Technologist, Innovator with 15+ years experience in development, management and research of distributed architectures and scalable services and applications.

Implements and executes on CEO strategies and visions delivering state-of-the-art architectures, and products. Proficient at translating marketing and product strategies into customer-centric services and experiences.

Seasoned technology lead, experienced in managing multi-functional local/distributed teams, for faster execution and time-to-delivery. People coach and motivator, stimulating work-and-learn environments, and technology growth in and around the company.

Passionate about data science, data engineering, distributed computing, man-machine interfaces, AI and cloud computing.

Natalino Busa is founder of SELECT COUNT(*) data engineering services. He has worked in the past as chief engineer and scientist for Philips Research and ING Bank in the Netherlands and DBS bank in Singapore.

More recently he has served as CDO and Chief Data Engineer for Teko and VnPay in Vietnam. Author for O'Reilly Media on Spark and Data Science Applications.

Expertise

- Data Engineering
- Data Visualization
- Data Science & AI
- Information Retrieval

Data Specialities

- ETL on Open Source Stacks
- Spark, Elastic Search, Kafka
- Dataflows Automation
- Customer Profiling
- Recommender Systems
- Fraud Analytics
- Predictive Data Analysis
- Exploratory Data Analysis
- (Inventory/Revenue) Forecasting

Worked for

- Teko, VnPay, Phong Vu (Vietnam)
  Director Data Science, Engineering, and Analytics

- DBS (SIngapore)
  SVP of Engineering

- Teradata (Netherlands)
  Head of Data Science Benelux

- ING Retail (Netherlands)
  Data Enterprise Architect

- Civolution (Netherlands)
  Chief Data Engineer

- Philips (Netherlands)
  Senior Researcher

Curriculum Vitae

Senior Director, Chief of Data, Teko, VnPay, Phong Vu
September 2018 - January 2020

Senior Director of Data: Covering the full spectrum of Data Applications as in Engineering, Science, Analytics, Reporting, AI, Big Data and Streaming Data. Managing all data related activities & teams. Reporting directly to CFO/CEO. Setup 5 teams: Data Ops, Data Engineering, Data Science, Data Services, Data Reporting. Provided a coherent vision on how to manage, audit, and access data across the company. Full redesign of the data architecture. Reporting and Data Science data flows running daily on jupyter notebooks powered by a Spark Engine (I have reproduced the Netflix notebook dataflow model)

SVP, Enterprise Data Platform Architect, DBS
September 2017 - May 2018

Responsible for defining and driving the next generation platform for BI, Reporting, Analytics, Data Science and AI. Manage directly a group of architects, data modelers, and lead engineers executing the DBS group level data-first technology program. Define a unified strategy and operation model for data practices at DBS at group level across 18 different country domains. Evolve the current data architecture from on-premise to a multi-cloud infrastructure. Define metadata management, data governance and data access control methodologies. Bring together data science, ai, and etl processes in a single coherent framework by integrating vendor solutions, in-house engineering and open source tools.

Head of Applied Data Science at Teradata Benelux
April 2016 – June 2017

As Senior Solution Architect, I have provided consulting on Strategy, Business scans, Roadmap, Gap Analysis, and Technology Mapping for AI, ML, Fast and Big Data Systems. Always hands-on. Build demonstrators and proof of concept/value on data-driven applications. Engagements with most Belgian and Dutch banks, and Telecom companies.

Promote Open Source Technologies to Teradata customers and within the Teradata organization. Mentor and grow a community of data scientists at Teradata. Speaker and Author on Data Science and Big Data.

Author at O'Reilly Media
Oct 2015 – present

Authoring education material, such as articles, videos, tutorials on Big Data, Fast Data, Spark, Machine Learning, AI and Data-Driven Applications.

Enterprise Architect at ING
February 2012 – March 2016

Tech lead and scrum master of a team of data and technology senior analysts, programmers, engineers, and statisticians. Exploring and Implementing first-of-a-kind big data solutions aiming to unravel predictive banking and personalized customer intimacy. Transforming services into experiences using data science, statistical languages, distributed computing, and dataflow architectures.

Explore and prototype innovative solutions for data analysis with a particular emphasis on data science, fault tolerance and distributed streaming processing. Provide advice and critical analysis for the selection of strategic and novel technologies and data processing platforms. Public speaker and tech community builder.

Projects at ING:

Streaming analytics: Coral is a real-time analytics and data science platform. It transforms streaming events and extract patterns from data via RESTful APIs. Built on Spray, Akka and Spark.

Real time Predictive Analytics: Define a new vertical architecture which enable the bank IT backend with real time predictive insight, using Spray, Akka, Scala, and R. Built on top of Big Data analytics as a real time extension to the already deployed distributed R framework.

Hadoop and distributed R: Rather than hyping big data, we rather put the framework on the workbench and measure how it performs. This project makes use of various big data technologies such as hadoop, hive, R on map-reduce, etc. This environment is meant as a workbench for testing and rating various applications and solutions ranging from predictive forecasting to product/customer classification, aiming for new and improved services and a better and more personalized user experience.

Redefine Fraud and Cyber Security with new components for streaming analytics for anomaly detection and pattern recognition of web traffic, internal threats and card/account usage.

Lead Software Developer / Web Services Architect at Civolution
October 2011 – January 2013

Design, architect, configure, prototype and test new products and applications related to video/audio fingerprinting, detection and monetization. Architectural lead for a scalable distributed hadoop like web service. Technology scouting for high-throughput, high-load, complex web services.

Engineering Lead at Sagantec
July 2006 - September 2011

Managing a group of seasoned/senior SW designers. Later managing the full operations of the european office, leading the full SW development cycle, from concept to release. Bridging the development team with Sagantec's sales, support and marketing teams.

Periodically interacting with key customers in Japan and US. Senior software engineer and product owner for Sagantec's analog migration and compaction tools. Leading an engineering team for the development of innovative EDA tools and features.

Senior Researcher at Philips Research
March 1998 - January 2006

Leading a 8 man project on advanced software techniques for video compression, aiming at low-power and predictive techniques for a better quality on band and power –limited hand-held devices. Researching new architectural solutions and compilers for DSP systems. High Level synthesis of code to hardware specification and RTL / gate level design.

Education

Università di Catania
Master Degree, Electronic Engineering, 1991 - 1997
Catania, Italy

TiasNimbas Business School
Attended: Master in Business Marketing, 2003 - 2004
Eindhoven, The Netherlands

Activities and Engagements

Founder/Developer SELECT COUNT(*)
August 2018 - Present

I have started this initiative (selectcountstar.com) with the goal of providing tools and practices to simplify data engineering and increase the productivity of data teams. I have noticed that most of the time is spent in packaging and coding glue logic rather than concentrating on the data processing. The first outcomes of this initiative are two packages: databox and datafaucet.

Databox allows a team to jump start a data system in a single laptop and modularly scaling it up to multi node clusters. It's based on docker technology and a kubernetes version is on its way.

Datafaucet is a set of utilities to speed up data preparation, ingestion, etl and reporting which are built on top of a configurable engine (currently supporting spark, dask, and pandas)

Author on O'Reilly Media:
2016-2019

Provide articles, tutorials and live talks/presentations on the topics of Data Engineering, Data Science, AI, Machine Learning, Distributed Computing, Algorithms, API design and Frameworks.

Check the overview here:
https://ssearch.oreilly.com/?i=1;q=busa;q1=Natalino+Busa;x1=author&act=fc_author

Blogging on Data Driven Solutions
2015-Present

Big Data and Data Science: http://www.natalinobusa.com
Streaming Analytics: http://www.odbms.org/2017/01/streaming-analytics-for-chain-monitoring/
Image AI classification: https://conferences.oreilly.com/strata/strata-eu/public/schedule/detail/57661