Nat 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, APIs, Cloud Apps, AI, Big Data and Streaming Data.

Managing all data related activities & teams. Reporting directly to CxOs and Management.

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 scalable AI and ML 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.

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

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.

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 Chief Data Scientist and Principal 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

- AI and ML on Cloud
- Open Source Stacks
- Spark, Elastic Search, Kafka
- Dataflows Automation
- Customer Profiling
- Recommender Systems
- Fraud Analytics
- Predictive Data Analysis
- Exploratory Data Analysis
- Anomaly Analysis
- 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

Technologies

A *selection* of technologies that I have used in the past recent years:
I have left out those which are not/less relevant as of today.

AI NLP (Bert, GPT-2, spaCy, gensim, LDA, word2vec), **AI Vision** (Inception, Xception, ResNet), **AI models** (Tensorflow, PyTorch, OpenAI Gym RL), **ML** (Scikit-Learn, Spark MLlib), **Data Processing** (Spark, Beam, Pandas, Koalas, Numpy, Scipy, Dask, Ray), **Data Visualization** (Bokeh, Datashader, Plotly), **Databases** (Clickhouse, MSSQL, Oracle, MySQL, Postgres, SQLlite, Cloud/Proprietary: BigQuery, Redshift, Teradata), **ML and Data Management** (TFX, Atlas, MLflow, Kubeflow, catalog APIs), **Data CI/CD** (Prefect, Airflow, Gitlab-CI), **Data Provisioning** (Ansible, Kubernetes), **Data ETL/ML Cloud Platforms** (Google, AWS, Azure), **API/Apps**(Flask, Play framework), **Frontend/Mobile**(Javascript, Angular, Vue, Flutter), **Languages** (C, C++, Python, Scala, R, Dart)

Am i hands on? By all means. Proudly coding everyday and back to strategy and vision. :)
Some projects and demos I have built either entirely or as a first committer/contributor.
https://github.com/natbusa/deepcredit
https://github.com/natbusa/datafaucet
https://github.com/coral-streaming/coral
https://github.com/natbusa/geo-services-tutorial
https://github.com/natbusa/deepnumbers

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)

Results and Deliverables at Teko/VNPAY:

1) **Product 360 with a custom dashboard**,
delivering KPIs for the board on the product portfolio, reducing inventory days.

2) **A forecast sales engine**:
very sophisticated time series predictive engine to across over 100 product categories.

3) **A RL Recommendation framework**: doubled the CTR of autocomplete searches.
Very advanced recommender, combining weblogs, knowledge graph and a mix of collaborative and content based recommendations. Reinforcement Learning engineering to apply multiple policies and experiments in practice. Available as a fast API for autocomplete as well as search.

4) **Customer Segmentation based on user touchpoints:**
Increased cross selling overall, and promotions yield rate by combining multiple customer events and creating very detailed micro clusters, to better make use of promotions

5) **Anomaly detection engine**:
Dramatically reduced the number of manual handling of frauds and manual interventions. Using a combination of statistics, rule based, and ML algorithms for fraud detection.

6) **Redesigned all Data Pipelines, Data Warehouse, and Data Governance**:
A very flexible, integrated and secured new system based on testable BI, automated pipelines via a combination of Gitlab CI/CD and Prefect/Airflow, fully Dev/Test/Stage/Prod data and role separation. Migrated to S3 compatible storage (min.io) and fully re-deployable engines with ansible and docker containers. Data lineage fully instrumented via Apache Atlas, including applied processes and quality KPI analyzing the lineage graph.

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:

1) **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.

2) **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.

3) **Hadoop and distributed R**: predictive forecasting to product/customer classification, aiming for new and improved services and a better and more personalized user experience.

4) **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.

Senior Researcher at Philips Research
March 1998 - January 2006

Leading an 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.

Education

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

Tias-Nimbas Business School
Master in Business Marketing, 2003 - 2004
Eindhoven, The Netherlands

Activities and Engagements

Personal Projects

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://natbusa.github.io
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