Nathan Bijnens

INTRODUCTION

Name	**Nathan Bijnens**
Address	On request 9042 Ghent Belgium
E-mail	nathan@nathan.gs
Website	nathan.gs
GitHub	github.com/nathan-gs
Twitter	@nathan_gs
LinkedIn	linkedin.com/in/nbijnens
Skype	nathan_gs

I am Nathan Bijnens, a developer turned Solution Architect with a passion for great code, Big Data and the Internet of Things. I am interested in programming and system administration, especially where they meet, from scaling platforms to designing the architecture of new and existing products and everything in between.
I am a passionate speaker and evangelist, on Big Data, IoT and Cloud. I am helping my customers with Digital Transformation and becoming Data Driven, as well as all their data in the Cloud needs.

My focus is on data analytics and building Big Data applications. Using Spark, Storm, Hadoop, Cassandra, Azure Data Lake, Azure SQL Data Warehouse, Stream Analytics, EventHubs, Kafka and Azure IoT Hub. I am especially interested in real-time Big Data, actively developing applications on top of Spark Streaming and Storm, designing Lambda and Kappa-like architectures. I advise on Big Data Strategies and evangelize Cloud, Big Data and IoT to clients and at conferences.

I am experienced with Java, Scala and have a general interest in Functional Programming. I am also a passionate Linux system engineer, follower of the DevOps movement, with experience in Puppet, Chef, NixOS and the Cloud.

I enjoy working with clients and partners, from giving advice, talking about the Business and Technological value of Big Data and Internet of Things, to Requirement Analysis.

PRESENTATIONS

Many **presentations to clients and partners**, typically on:

- Digital Transformation
- Cortana Intelligence and Cognitive Services
- The Internet of Things
- Azure and Cloud

The Internet of Things: Things, Connectivity, Data and Insight, bringing it all together at the **Microsoft Services Day March '17.**

Microsoft Advanced Analytics, with a special focus on Cognitive Services and Azure Machine Learning at the **Data Science Belgium Bootcamp October '16.**

Advanced Analytics with Cortana Intelligence at the **Data Science Ghent Meetup April '16.**

Virdata: lessons learned from the Internet of Things and M2M Cloud Services at the **IBM Big Data Developers (San Jose, CA) Meetup of June '14.**

A real-time (Lambda) architecture using Hadoop and Storm at **NoSQL Matters Cologne '14.**

A real-time architecture using Hadoop and Storm at **Devoxx '13**, the biggest European Java conference.

A real-time architecture using Hadoop and Storm at **JAX London '13**, a *Java Enterprise* and Big Data conference.

Big Data, Hadoop and HDInsight, together with Wesley Backelant of Microsoft Belgium at the **SQLUser Group Belgium** meetup of February 2013.

A Series of 3 **Apache Storm Workshops** in the context of **BigData.be**, during 2013.

A real-time architecture using Hadoop & Storm at the 2013 edition of the **FOSDEM NoSqlRoom**, an open source conference, it has been viewed over 24500+ times.

A Vision on Data at the 2013 **Big Data & Security Conference** of **LSEC.**

Getting More out of Your Big Data at **Microsoft Inspirience the Future 2012.**

Hadoop Pig: MapReduce the easy way at the 2011 edition of the **FOSDEM DataDevRoom**, an open source conference, it was featured on the front page of slideshare.net and has been viewed over 18000 times.

EXPERIENCE

On February 2016 I joined **Microsoft** as a **Data Solution Architect.** As a DSA I drive high priority customer initiatives, leveraging Cortana Intelligence on Azure to solve the biggest and most complex data and IoT challenges faced by Microsoft's enterprise customers. It is a technical, customer facing role, accountable for the end-to-end customer deployment and usage experience for cloud data and IoT services

From October 2013 till January 2016 I worked for **Virdata** a start-up funded by **Technicolor**, as a **sr. Big Data and IoT Engineer**, previously as a **DevOps Engineer.** Virdata is an Internet of Things Cloud platform, scalable to millions of devices, from tiny sensors to TVs to Cars.
I work on our Data Processing stack, on top of Spark and Spark Streaming, mostly in Scala. I also worked as the lead on our Spark as a Service Infrastructure.

Data Engineer
From January 2014.
As a Big Data Engineer, I am working on our Data Processing and Analytics stack, creating a Lambda Architecture. Working with Spark, Spark Streaming, Storm, Cassandra, Kafka and HDFS. We mostly use Scala and bits of Java. We deploy using Chef on Mesos in the cloud.

IoT | Big Data | Analytics | Lamdba Architecture | Software Architecture | Scala | Java | Spark | Spark Streaming | Storm | HDFS | Cassandra | Mongo | Mesos | Docker

Spark as a Service Lead
From May 2014 till December 2014.

We created a Spark as a Service, where Data Scientists can on demand request a Spark and Mesos cluster in the Cloud. We leveraged our existing Chef & Mesos architecture. We also added tools for ad hoc analytics, like IPython PySpark & Scala Spark notebooks. It is also possible to launch Docker containers.

IoT | DevOps | Big Data | Analytics | Scala | Python | Spark | HDFS | Chef | Mesos | Amazon EC2 | Softlayer | Linux | Docker | Project Management

Cloud & DevOps Engineer

From October 2013 till December 2013.

As a Big Data and DevOps Engineer, I was responsible for the scalability, quality and Operational Intelligence of a new platform. I introduced the Continuous Integration platform, using Jenkins, added unit tests. The platform is build using Chef, tries to be cloud independent (runs on Amazon AWS, Softlayer, ...).

DevOps | IoT | Big Data | Scala | maven | Jenkins | Unit Testing | Storm | Spark | HDFS | ElasticSearch | Ruby | Chef | Mesos | Scalability | Amazon AWS | Amazon IAM | Amazon EC2 | Softlayer | Vagrant | Linux | Docker

At the end of 2013 I co-founded lambda-architecture.net, a website about the **Lambda Architecture**. The Lambda Architecture is a data-processing architecture designed to handle Big Data by combining both batch- and stream-processing, while attempting to balance latency, throughput, and fault-tolerance. Batch Processing is used to provide comprehensive and accurate (although slightly out of date) views, while simultaneously using real-time stream processing to bridge the gap.

From April 2012 till March 2014 I worked for **DataCrunchers**, as a **Hadoop and Big Data Consultant**. DataCrunchers is the leading Belgian consultancy firm around everything Big Data. I co-developed on our internal Semantic Analysis Engine on top of Storm, in Java, I co-presented our Hadoop Ecosystem course, I created a new website using Drupal and I was responsible for our Microsoft Big Data partnership.
I have been working for the following clients: Technicolor, Flemish Government, AWV (Macq), a social media startup, HSHMRK & Ewals Intermodal.

Big Data & DevOps Engineer, Technicolor

From October 2013.

As a Big Data and DevOps Engineer, I was responsible for the scalability, quality and Operational Intelligence of a new platform. I introduced the Continuous Integration platform, using Jenkins, added unit tests. The platform is build using Chef, tries to be cloud independent (runs on Amazon AWS, Softlayer, ...).

DevOps | IoT | Big Data | Java | Scala | maven | Jenkins | Unit Testing | Storm | Spark | HDFS | Ruby | Chef | Scalability | Amazon AWS | Amazon EC2 | Softlayer | Vagrant | Linux

Project Manager ABBAMelda, Flemish Government (Macq)

From June 2013 till October 2013.

Project Manager and Lead Developer of ABBAMelda, a ticket and maintenance management system, originally developed by Siemens. ABBAMelda consists of a Java EE backend, an Informix database, with a PHP/jQuery frontend. Under my lead, we created an enhanced Tablet intranet site, improved the bulk upload possibilities, additional REST services, introduced unit testing and switched to git. I was responsible for coordinating with different teams within the Flemish Government, as well as the contractor (Macq).

Software Architecture | Project Management | Agile | Scrum | Java | JPA3 | EclipseLink | maven | Informix | REST | SOAP | PHP | jQuery | Refactoring | Unit Testing | Puppet | Linux | Mobile Development | ETL

A social media startup

From February 2013 till June 2013.

Defining and implementing the architecture for a social media analytics startup. I designed and implemented a Lambda Architecture (in Java), on top of Storm and Hadoop, using Redis, Voldemort as well as Thrift.

Java | Spring DI | maven | Storm | Hadoop | Thrift | Kafka | Zookeeper | Voldemort | Redis | JSoup | Guava | jUnit | Puppet | Nagios | Logstash | Ganglia | Linux | Software Architecture | Lambda Architecture

HSHMRK

From December 2012 till January 2013.

Defining and implementing the architecture for hshmrk, a data visualization startup. The application backend is written as a Jersey REST (Java), service, using ElasticSearch as storage. The frontend is an AngularJS and D3 web application. This approach allowed us to easily scale.

Java ElasticSearch AngularJS D3 Jersey maven Spring DI jUnit Data Visualization Software Architecture Puppet

Ewals Intermodal

From October 2012 till June 2013.
Developing Oracle database views for integration of Greencat and Crystal Reports.

Oracle Greencat Crystal Reports ETL

Microsoft Big Data Partnership

Responsible for the contact with Microsoft. Speaker at the Microsoft Inspirience Day about Big Data.

Client & Partner Relationships Public Speaking Business Strategy Technology Evangelism Big Data Azure Hadoop

Solution Architect

Discussions, presentations and conversations with prospects, clients and partners, discussing potential Business Ideas and architectures. Evangelizing the Belgium market for Big Data.

Client & Partner Relationships Public Speaking Business Strategy Technology Evangelism Big Data IoT

Internal Projects

1. I co-developed on our internal Semantic Analysis Engine on top of Storm and ElasticSearch. The web interface is build around a Java, Jersey backend and a frontend in AngularJS.
2. Setting up an internal Hadoop, ElasticSearch & Storm test cluster using Puppet and Cloudera Manager. I created some Puppet modules to manage Storm, ElasticSearch and Ganglia.
3. I co-developed on IHarvest project. It is a distributed HTTP *Fetcher & Parser* on top of Storm, written in Java, the results are stored on HDFS for more extensive querying using Hadoop.

Cloudera Manager Storm ElasticSearch Puppet Hadoop Hadoop Pig HDFS Ganglia Java AngularJS Jersey Bootstrap

From March 2011 till April 2012 I worked for **iController**, as **Lead Application and Warehouse Developer**. We build a credit management web application, written in PHP. I managed a small group of developers, took the lead on everything technical and coordinated with the directors, partners and clients.
My job involved lots of PHP, refactoring, performance tuning, system engineering and bits of data analysis.

From September 2010 till February 2011 I was employed at **Netlog** as a **Warehouse and Web Developer**, creating an analytical warehouse based on Hbase. I also created large parts of the processing infrastructure using Hadoop and Hadoop Pig. I also advised on new technologies, git migration, best practices & unit tests. Netlog is a social community network, with over 70 million members, mainly active in Europe and the Arabic world.

From October 2009 to August 2010 I started working at **iController**, a small SME in Ghent, as **web application developer**. I developed a new, greatly improved debtor & credit management web application. We used the symfony framework as a starting point to create a stable and easy to maintain application.

From March 2007 till December 2014 I was the **Consultant, Managing Director and founder** of **Servs** BVBA; mainly active as a (part-time) IT consultant in the field of Big Data & Spark, Scala, web applications (PHP, sql, scalability, ..), DevOps & Linux system administration and hosting applications.

From April till August 2011 I worked as a consultant for **Truvo**, the Belgian yellow pages, doing a **hadoop deduplication** job. I used Hadoop & Hadoop Pig, wrote custom UDFs in Java for Ngram matching and solved performance issues. At first Amazon Elastic MapReduce was used, later I setup a Hadoop cluster.

Hadoop Hadoop Pig Cloud Amazon EC2 Amazon Elastic MR Java

In 2009 I was active as a consultant to a Dutch services organization. The assignment mainly consisted of **creating a link between existing systems and a new website**.

From 2008 till present I perform various **short term consultancy** tasks commissioned by **Sinergio** . For a **semi government organization** I set up a new tomcat server and did a general check up on their Linux servers.

For another company I configured an ASP.Net application and installed an IIS server. For **Sio Hosting** (Formerly Sinergio Hosting) I planned and executed the move from traditional servers to a virtualized xen cluster. I also checked the integrity and security of their hosting platform on a regular basis. I advised them on strategic planning.

During my secondary education I created a **PHP CMS**, with features as native PHP templates, module in module support (recursion) following a very basic MVC pattern. It was used and expanded by two web agencies to create over 30 websites.

SKILLS

I am very interested in and work with **Big Data**, from the processing and storage of large volumes, to real-time stream processing, and machine learning. I read, tweet, and try out as much as I can about new Big Data technologies, like Spark, Samza, Storm and Kafka, as well as more established technologies like Hadoop and Cassandra and learning as much as I can in the process. I use Scala, Java and bits of Python and Hadoop Pig Latin. I setup, administered and monitored Hadoop, Mesos, Storm, Cassandra & Zookeeper clusters.
Big Data related skills:

- Do interactive ad-hoc analytics using Python based Jupyter Notebooks, as well as Scala based notebooks, combined with Spark SQL.
- Query the data with Spark, Spark SQL, Hadoop Pig & Hive.
- Using & administrating Spark, Hadoop and Mesos clusters, including Storm, ElasticSearch, Cassandra, Kafka and Zookeeper.
  - Deploying Hadoop on Azure, Amazon & Softlayer.
  - Using HDInsight, combined with Azure Data Lake
  - Azure Machine Learning, Azure Notebooks and Cognitive Services
  - Chef, Puppet and NixOps to deploy Mesos and Hadoop to the Cloud.
- Developing solutions for real-time Big Data using Spark Streaming, Storm, Azure Stream Analytics, EventHubs, Azure IoT Hub and Kafka.
- Combining batch and real-time technologies to create a Lambda architecture (of Nathan Marz), that is resilient to failure, scalable and fast.

I have extensive exposure to **Azure** and **Cortana Intelligence**, including, but not limited to:

- Azure Data Lake
- HDInsight
- Azure Machine Learning
- Azure Notebooks
- Cognitive Services
- Azure EventHubs
- Azure IoT Hub
- Azure Stream Analytics
- Azure SQL DataWarehouse
- Azure SQL
- Power BI
- ...

I currently mostly develop using **Scala**. Scala is a JVM-based language, with both full support for OO as well as Functional Programming approach. I started with Scala at the end of 2013.

- Using Scala to create Spark and Spark Streaming jobs.
- Using Akka and Akka Persistence to create a CQRS application, in Scala.
- Creating Thrift based services, using Finagle.
- Unit testing using Scalatest and Cucumber.

I am interested in **Functional Programming**, mostly looking at Scala, and a bit at Clojure. Especially functional programming in relation to Big Data has my focus.

I am using **Java**, mostly using Spring, maven and Jersey in combination with the JavaScript MVC framework AngularJS.

- Developing a Java EE application, with Glassfish, using JPA with EclipseLink.
- Creating Threaded servers, using Thrift.
- Autowiring & Dependency Injection using Spring.

As well as using only Stream Processing to build a Kappa Architecture.

Next to programming I have always been passionate about **Linux** and open source. I have used over the last 10+ years several distributions from Debian & Ubuntu, CentOS over Gentoo to Linux From Scratch, lately I am very interested in NixOS. NixOS is a purely functional Linux. I have setup countless servers from Cloud to virtualized to bare metal servers.
I am following the DevOps movement. I am using Puppet, Chef and NixOps to automate and Ganglia to monitor critical infrastructure. I have open sourced and contributed to several NixOS and Puppet modules.

I follow and try out with great interest Cloud related techniques and technologies, in all its forms: IaaS, PaaS, SaaS, .. I have used as test or in production Azure, OpenStack, Amazon AWS, Google BigQuery (private beta tester) and Softlayer.

- Consuming the Twitter & LinkedIn APIs, using OAuth.
- Creating a REST and SOAP based services, using Jersey or JAX-WS.
- Unit testing using JUnit and Mockito.

I have used **PHP** and **JavaScript** for over 10 years, mostly in a combination with the symfony & Symfony2 frameworks as well as a multitude of SQL and NoSQL databases, as MySQL, PostgreSQL, Oracle, Informix and Redis, Membase, Memcached and ...

- html5, CSS & jQuery
- Creating a Drupal site, using Bootstrap and deploying on Azure.
- Creating browser based applications using AngularJS.

Some basic knowledge, but willing to improve:

- Graph databases: Neo4J, HyperGraphDB & Titan.
- Various NoSQL data stores, like MongoDB, Voldemort, riak, ...
- Clojure, Ruby and Python.
- Scrum, Extreme Programming
- Security & penetration testing.

Languages

Dutch Mother tongue

English Very fluent

Other skills

In possession of a driver's license.

EDUCATION

Microsoft Design and Implement Big Data Analytics Solutions (70-475)
Microsoft | February 2016

IBM Big Data Fundamentals Technical Mastery (N32)
IBM | 000-28363001 | January 2013

IBM InfoSphere BigInsights (Hadoop & Big Data) Technical Professional
IBM | 000-28146057 | October 2012

SARA Hadoop workshop
SURF SARA, Amsterdam | December 2010

Institute University of Antwerp

Program Social-Economic Sciences

Period 2005 - 2007

Achieved Not Achieved

Institute Hibernia Steinerschool in Antwerp

Program ASO, Higher Secondary education

Period 1999 - 2005

Achieved 2005