
DATA ARTISANS

Name

FlinkNote

Email

note@data-artisans.com

Website

<https://note.data-artisans.com>

The current problem to solve

In most of the existing distributed data-intensive systems, especially Flink Streaming, only provided rich imperative code makes developers or users less productive to build a big data applications

- *Non-standard*: Each framework has a different set of API so that developers must learn how to use its API properly to take the advantage of the framework. Probably those API is not standardised. Therefore, developer even takes long time to master it.
- *Less Compacted and Optimized*: given that most big data applications are fairly simple application-wise, a block of API codes might be less optimal to use for most the popular queries. In addition, the imperative code is usually more complicated and less compact than other declarative language such as SQL
- *Interaction*: During development cycle, developer usually compiles the source code and work on command line which is less interactive and definitely not friendly
- *Collaboration and Sharing*: The only way to share your work is to pack it as a library or publish a service. Moreover, to collaborate with your colleagues simultaneously, you have no choice but sitting in front of a same machine.

About the company and product:

Data Artisan: a spin-off company from Technische Universitat Berlin, Germany. The founder team has a high level of expertise in distributed data-intensive system. They are aiming to develop the next generation technology of Big Data Analytics and Processing via Apache Flink

Apache Flink: is a top-level project of the Apache Software Foundation with a community of over 100 contributors from industry and academia. Flink is an open source system for expressive, declarative, fast, and efficient Big Data analysis. It originates from EU-granted Stratosphere project since 2000s

Flink Notebook: an interactive , collaborative web-based environment inspired by iPython Notebook and Google Docs. Flink Notebook allows to build an analytics application on Flink easier and more productive in a very friendly way.

The market need, customer segment:

The market needs:

We are going to supply services to German market at the beginning. Even though the German big data market till appears to be at an early stage, it is expected to from EUR 650 million in 2013 to almost EUR 1.7 billion in 2016. Additionally, Software and Services would account for 70% of that value

DATA ARTISANS

Customer Segments:

We target Big Data developers in German market. Those developers are in charge of building and maintaining a data analytics platform for companies in different industrial sectors such as finance, e-commerce, Internet Of Thing. They are looking for a scalable , reliable and high-available framework with excellent UX

Competitors, and how to compete with them:

The idea of Complex Event Processing was declared from early 1900s. Therefore , we are challenged by many services from both all major software vendors (IBM, Microsoft, SAP, Oracle...) and new players (DataBricks, Adatao ...). However, our product owns several crucial winning features that do not exist or still under developing at other products. We outperform them by a true streaming processing with highly optimised core. With Flink Notebook's excellent environment, developing a big data application never be easier than that.

Partners:

Recently , we are making a partnership agreement with Google. We now enable Google Cloud Dataflow users to leverage Apache Flink as a backend. Another agreement with Amazon web service and Microsoft Azure is on process

History, references, prior results:

Meetup group in Berlin (> 300 members), in Stockholm

Explanation:

Big Data: a popular term used to describe the exponential growth and availability of data, both structured and unstructured. Three most important features of Big Data is Volume - Velocity - Variety.

Data Stream: Uninterrupted flow of a long sequence of data, such as in audio and video data files

Stream Processing: analyse one or multiple data streaming to identify meaningful events and respond to them as quickly as possible

Alternative to grow (size of capital, other resources, timing, manpower, etc.):

Manpower: ? committers , 100 contributors, academic supervisors (TU Berlin, Sztaki, KTH), Apache support

Size of capital : VC-funded (I am asking for the name)

Timing: First release (May 2015): SQL syntax. Second release (July 2015): Web-based. Third release(Sept 2015): full collaborative
