Your Title Here

Chaker Benhamed

June 28, 2017

Contents

Chapter 1

Overview

Introduction

This introductory chapter presents the general context of the project and gives an overview of its scope. In the first section, we will specify pedagogical and professional frame of the project, where we will introduce the host company, its business field and the Technical Infrastructure team. Then, in the second section, we will present some general concepts of the project. Finally, we will define the project problematic and present the project goals.

1.1 General context

In the part we will put the project in its frames, both scholar and professional. First we will set the pedagogical. Then we will define the professional frame, i.e Predictix and more specifically the Technical Infrastructure team.

1.1.1 Pedagogical Frame

This project was elaborated in the context of a graduation internship, a final step of obtaining the National Diploma of computer engineering from the Faculty of Science of Tunis The internship was undertook on period of four months, from February 1st to May 31th, 2017, during which, we integrated the Technical Infrastructure team at Predictix Tunisia.

1.1.2 Professional frame

Predictix is an American company headquartered in Atlanta, GA, USA, with a center of operations in Tunisia.

1.1.2.1 Predictix: Foundation and activity domain

Predictix is a fast growing consulting and software company founded in 2005, headquartered in Atlanta, Georgia USA with affiliations in Tunis, Tunisia, London, UK and Amsterdam, Netherlands. It merged with LogicBlox in January 2014 which private company that provide the next generation of smart database. This new database permits to present finished products to the clients in the simplest ways, after performing a lot of complicated cloud and multicore computing.

Predictix is specialized in providing software solutions implementing predictive analytic technologies to solve retail problems and improves retailers' profitability. In fact, it helps Tier 1 TODO retailers and brands make better merchandising decisions by eliminating the traditional silos between varrious buying and selling decision, delivering solutions that evolve easily to adapt to their ever-changing business and simplifying the process and lowering the stakes of investing in technology.

As recognition, Gartner, the world's leading information technology research and advisory company, has classified as a leader, in the Magic Quadrant for Merchandise Assortment

Management Applications in July 2015.

As a matter fact, because of the leading role of Predictix in Cloud based solutions for retailers. It get acquired by Infor, in 2016.

1.1.2.2 Infor

Infor is the third-largest provider of enterprise applications and services behind Oracle and SAP. Think: over 70,000 customers in 194 countries for its industry-specific applications and suites designed for the cloud, on-premises, or both, that brings in about $2.5 billion in revenue. To handle this massive operation, Infor employs 13,000 people across the globe.

Predictix joined Infor as the base of Infor-retail the Infor devision fore retailers.

Infor on Predictix: "Atlanta-based Predictix experienced more than 40 percent growth in SaaS subscriptions in 2015 and counts 5 of the top 15 global retailers as customers, managing more than $60bn in weekly forecasts. The company has an engineering- and science-driven culture with deep expertise in retail and a drive to revolutionize the industry. LogicBlox, the company's technology platform underlying all Predictix applications, revolutionizes the development of next-generation predictive and prescriptive applications, and has attracted funding from DARPA, the Defense Advanced Research Projects Agency."

Predictix is providing SaaS applications to big retailers, among which the largest drug retail chain in the United States in sales and profits with annual sales of over US $76 billion in 2015, employs over 240,000 people, with more than 8,000 stores across all 50 U.S States, the District of Columbia, Puerto Rico and the US Virgin Islands. As a matter of act, for forecasting purposes, this client benefited from $125M from inventory reduction alone, with 20% of ongoing. Some of Predictix clients, whose solutions are already on production, are

- Kiabi
- The Home Depot
- Crate and Barrel
- M.Video
- Target

Predictix offers its clients a variety of solutions. The figure gives an overview of Predictix business units:

TODO

Pricing and Promotions: Retailers, wholesalers and some manufacturers use Predictix services to help them manage their promotions, pricing and markdowns for higher profit bottom lines,by offering a very high flexible and configurable solution that facilitates the user interaction with the results he gets.

- **Forecasting and Replenishment**: The firm provides forecasts and replenishment strategies that enable a retailer to implement the promotional, pricing and assortment strategies established. Additional profit can be gained by minimizing inventory levels and reducing lost sales.
- **Assortment and Space Optimization**: Predictix helps retailers to define the optimal localized assortment for every store and every space. Retailers can expect category margins to improve 100 basis points or more, and to drive higher sales, from optimizing their assortment with Predictix ASO.
- **Planning and Allocation**: Business Consulting at Predictix helps retailers review their current planning and allocation process and design new ones for the future.

In order to provide these solutions , Predictix has forged a number of strategic partnerships with other companies. Some of which are the following companies:

- **LogicBlox**: The platform provider for Predictix' solutions. Predictix use LogicBlox' scalable and declarative database system, programming language and different environment tools to design and build its systems. After a partnership lasting for years, the two companies have merged together.
- **Amazon Web Services**: AWS is a world leader in cloud computing solutions. Predictix is using AWS' IaaS and PaaS solutions to host its different systems. The need for cloud services will be detailed later in the next chapter.

During its life cycle, every client's product goes through a number of steps, from development to testing to deployment. Most of tools and steps of the development pipeline are the responsibility of the Technical Infrastructure team within which we undertook this internship.

1.1.2.3 Technical Infrastructure team

The technical infrastructure team is needed in every step of the client's product life cycle. Inf fact, they gather and collect the data, store it, do batch processing or real-time processing on it, and serve it via an API to a data scientist who can easily query it. They have extensive knowledge on databases and best engineering practices. These include handling and logging errors, monitoring the system, building human-fault-tolerant pipelines, understanding what is necessary to scale up, addressing continuous integration, knowledge of database administration, maintaining data cleaning and ensuring a deterministic pipeline.

The main activities of Predictix' architecture and integration team can be summarized to:

• They intervene, in the beginning of the project when the requirements are defined. They play the role of consultant as they are consulted to discuss the data specification, in collaboration with the data science team, with the client. In fact, their role, at this step is to take all necessary measures in order to make sure that the data is normalized and consistent.

• They intervene, also, in the development of the applications, as they are in charge of the applications architecture design. In fact, they translate complex functional and technical requirements into detailed architecture, design and high performing software with the ability to architect highly scalable distributed systems.

• They are, also, in charge of batch implementation thorough data modelling, rules definitions, protocol buffers services, etc.

• Finally they are responsible of the application being automatically build and deployed on the cloud, with extremely high attention to the application performance, as it world affect the application cost.

• One of the main duties of the Technical Infrastructure team. Is to create internal tools that facilitate the development of project, deploying and monitoring performance and cost.

lb-job-cost: It's a tool for monitoring lb jobs cost of each client's account. It will give the current actual cost for the month along with the projected cost for the whole month.

Ops-console: It's a web application that helps engineers at Predictix to monitor the batch process in real-time.

Command-history: A web applications that monitor the commands used in all

Predictix EC2 machines. It give the user the ability to search for command usage, for certain user.

Nixops-dashboard: SaaS solution that allows users with no technical knowledge, to easily provision Predictix applications on the cloud. Business consultants, testers, and others are now able to deploy complex infrastructures with a click of button.

lamias-servicetester: A DSL for testing web services. It can be used for unit tests and for regression tests.

1.2 Project presentation

Since Predictix works with 1 tier retailers in the US, application data tend to be large enough to face issue in performance. In fact some of Predictix client deliver weekly data of the size of 2 TB of sales records. Developer need to monitor the performance of the application during development phase.

1.2.1 Problematic

Measuring the performance of an application is an operation that has to be done on daily basis or every time changes are introduced in the application codebase or one its dependencies. To automate the benchmarking process, we need first to understand how developer are doing benchmarks currently.

1.2.1.1 Timing services

Compare to traditional 3-tired development stack, Business logic is installed in the LB database itself, Thus simplifying application development by alleviating the need for:

- fetching data to and from DB
- integration between OLAP and OLTP
- simpler language for business logic

So all LB applications expose a set of sevices to the client (mostly the UI).

Conclusion

TODO

Chapter 2

State of the Art

Introduction

We will introduce in this chapter some useful definitions and concepts that she light on the project. We will also assess the current situation be examining the literature as well as the current solutions adopted by Predictix. Next, we will specify the different proposed solutions and choose the one that fits most our needs.

2.1 General concepts

In this section, we will be defining the general concepts that the Technical Infrastructure team work around, and more specifically those that this project evolves around. We will also clarify the use cases for these technologies and concepts in Predictix.

2.2 Declarative programming

In computer science, declarative programming is a programming paradigma style of building the structure and elements of computer programsthat expresses the logic of a computation without describing its control flow.[1]

Many languages that apply this style attempt to minimize or eliminate side effects by describing what the program must accomplish in terms of the problem domain, rather than describe how to accomplish it as a sequence of the programming language primitives[2] (the how being left up to the language's implementation). This is in contrast with imperative programming, which implements algorithms in explicit steps.

Declarative programming often considers programs as theories of a formal logic, and computations as deductions in that logic space. Declarative programming may greatly simplify writing parallel programs.[3]

2.2.1 Datalog

Datalog is a declarative logic programming language that syntactically is a subset of Prolog. It is often used as a query language for deductive databases. In recent years, Datalog has found new application in data integration, information extraction, networking, program analysis, security, and cloud computing.

Its origins date back to the beginning of logic programming, but it became prominent as a separate area around 1977 when Herv Gallaire and Jack Minker organized a workshop on logic and databases.[2] David Maier is credited with coining the term Datalog.

The applications that we will measure the perforamnce of, are all written in commercial implementation of Datalog, LogiQL.

2.3 REST architecture

Rest stands fore REpresntational State Transfer. It is an architecture style for designing networked applications. It permits creating, modifying resources easily. Indeed REST is a lightweight alternative to complex mechanism like RPC, CORBA and SOAP.

Rest is not a 'standard'. In fact, it is a guideline to build an efficient framework for communication between two machines using HTTP protocol. The World Wide Web itself, based on HTTP, can be viewed as a REST-based architecture. REST relies on a stateless, client-server, cacheable communication protocol. It is simple to implement and maintain. In addition, it allows applications to be scalable by supporting multiple backend services at the same time.

Much like Web Services, a REST service is platform-independent, language-independent standard-based asit runs on top of HTTP, and is easily used in the presence of firewalls.

However, there are a few major concepts which make REST unnique from other web services. In fact , its main key principales are the following:

• **Unique URL-Resource mapping**: Every resouce is mapped to a unique URL. That refers to a some logical way to access information.

• **Statelessness**: All information required to precess the request by server is contained along with the request. This means that no informatio of the previes request is maintained by the server. This is inherited from the fact that REST is based on HTTP.

• **Action Verbs**: REST architecture use HTTP verbs to identify the apropriate action. The main HTTP verbs used in a REST architecture are GET, POST, PUT and DELETE. In fact, GET is used by the client to access the resource on the server, PUT to update a resource, POST to create a new one and DELETE to remove resource.

• **Data Exchange formats**: REST architecture does not require any particular encoding for the resource body. JSON and XML are the most used format, but it can be PROTOBUF, YAML etc.

2.4 Cloud

Cloud computing is a type of Internet-based computing that provides shared computer processing resources and data to computers and other devices on demand. It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources

(e.g., computer networks, servers, storage, applications and services), which can be rapidly provisioned and released with minimal management effort. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in either privately owned, or third-party data centers that may be located far from the userranging in distance from across a city to across the world. Cloud computing relies on sharing of resources to achieve coherence and economy of scale, similar to a utility (like the electricity grid) over an electricity network.

2.5 Retail

TODO

2.6 SaaS

Software as a service (SaaS) is a software licensing and delivery model in which software is licensed on a subscription basis and is centrally hosted. It is sometimes referred to as "on-demand software", and was formerly referred to as "software plus services" by Microsoft. SaaS is typically accessed by users using a thin client via a web browser. SaaS has become a common delivery model for many business applications, including office and messaging software, payroll processing software, DBMS software, management software, CAD software, development software, gamification, virtualization,[4] accounting, collaboration, customer relationship management (CRM), Management Information Systems (MIS), enterprise resource planning (ERP), invoicing, human resource management (HRM), talent acquisition, content management (CM), and service desk management. SaaS has been incorporated into the strategy of nearly all leading enterprise software companies.

2.7 Automation

Automation or automatic control, is the use of various control systems for operating equipment such as machinery, processes in factories, boilers and heat treating ovens, switching on telephone networks, steering and stabilization of ships, aircraft and other applications and vehicles with minimal or reduced human intervention. Some processes have been completely automated.

Automation has been achieved by various means including mechanical, hydraulic,

pneumatic, electrical, electronic devices and computers, usually in combination. Complicated systems, such as modern factories, airplanes and ships typically use all these combined techniques. The biggest benefit of automation is that it saves labor; however, it is also used to save energy and materials and to improve quality, accuracy and precision.

The term automation in the software industry has gain a lot of success in the recent years. From the automation of build to the deployment, various steps in the develompent pipeline can be automated. Including benchmarking the applications.

2.8 Benchmark

In computing, a benchmark is the act of running a computer program, a set of programs, or other operations, in order to assess the relative performance of an object, normally by running a number of standard tests and trials against it.[1] The term 'benchmark' is also mostly utilized for the purposes of elaborately designed benchmarking programs themselves.

Benchmarking is usually associated with assessing performance characteristics of computer hardware, for example, the floating point operation performance of a CPU, but there are circumstances when the technique is also applicable to software. Software benchmarks are, for example, run against compilers or database management systems.

Benchmarks provide a method of comparing the performance of various subsystems across different chip/system architectures.

In our case the we are going to benchmark applications along with the database system that powers them. Since they are strongly coupled.

2.9 NoSQL

A NoSQL (originally referring to "non SQL", "non relational" or "not only SQL") database provides a mechanism for storage and retrieval of data which is modeled in means other than the tabular relations used in relational databases. Such databases have existed since the late 1960s, but did not obtain the "NoSQL" moniker until a surge of popularity in the early twenty-first century, triggered by the needs of Web 2.0 companies such as Facebook, Google, and Amazon.com. NoSQL databases are increasingly used in big data and real-time web applications.[6] NoSQL systems are also sometimes called "Not only SQL" to

emphasize that they may support SQL-like query languages.

Motivations for this approach include: simplicity of design, simpler "horizontal" scaling to clusters of machines (which is a problem for relational databases), and finer control over availability. The data structures used by NoSQL databases (e.g. key-value, wide column, graph, or document) are different from those used by default in relational databases, making some operations faster in NoSQL. The particular suitability of a given NoSQL database depends on the problem it must solve. Sometimes the data structures used by NoSQL databases are also viewed as "more flexible" than relational database tables

Conclusion

TODO

Chapter 3

Requirement definitions

Introduction

After imbibing the general context of the project, we are going to focus in the is chapter on providing a full description of out project, that we will be calling Benchmarks Dashboard. To do so, we will start by analyzing the requirements and extracting the functional and non-functional specifications. Then we will define the Product backlog of the project to achieve. Finally we will present the general use cases of the project.

3.1 Requirements analysis

3.1.1 Design session

After the beginning of the internship, we had a on-month training where we were trained about the different technologies that Predictix is using, such as LogiQL, Nix and LB. Also, as interns, we had the opportunity, first, to get familiar with the company's working environment, second, understand better the project's goal and its added value to its users, and third, were able to collect the functional and non functional requirements of the project.

3.1.2 Functional requirements

The purpose of this project is to automate the benchmarking process of the applications developed in Predictix. For that ther are two main requirements that need to be addressed.

3.1.2.1 Automating "benchmark runs"

The application should give the ability to the clients, the developers, to run benchmark with the click of a button. The applicatin must run all the repeated process without user's interactions. For that 3 operations must be automated:

• **Deploy the application**: The applications developed in house are SaaS applicatons, thus they must be deployed to the cloud. The user choose what the application that want to run benchmark for. Then we need to automaticly provison a server in the cloud, install the application on that server along with all dependcies that are needed to run the benchmark.

• **Load test data**: To have a real and consice measure of the perfoarance of the application, a data sample should be loaded into the database. This sample should be the same across all benchmarks runs so it can make sense to compare them. The data should be loaded as soon as the application is installed.

• **Run The benchmark**: After loading test data we will run a series of service test and measure the time each service took to complete the request. Each service test will be run 3 times to unveil any issues with database warmup.

Finally after the benchmarks completed the result of the test should be stored for the developer to check and the instance should be unprovisioned.

3.1.2.2 Visualize the previous benchmarks

They developer should be able to see and investigate the result of the benchmarks. He must be also be able to:

• **Compare benchmarks**: In this case the user can see differnece in the overall perfoaramance of the application, between two different dates.

• **Test history**: The user can see the test history. Thus the state of the test between all benchmarks, or specifly between two bencharmks.

• **Benchmark status**: The user can see the status of the benchmark. Whether they competed successfully or not or they still running. Also they can see who started the benchmark run, and when he started it.

3.2 Product Backlog

From the previous requirements and some other realted to user mangment we can extract the product backlog described in table 3.1

Once the Product backlog was established and validated by the Product Owner, the scrum team has broken down the Product Backlog into four sprints of a duration of 20 days each. The sprint duration was suggested by the Scrum Master as sprint ceremonies were planned every 20 days at the end of each sprint.

The table 3.2 defines the user stories that will be achieved throught each sprint.

3.3 Requirement specifcations

3.3.1 Actors identification

The inteded users of the Benchmarks Dashboard are Predictix employees, The AppDev team (Application development team). They must be able to, easily intereact with the system, run new benchmarks and check the status and result as soon as possible.

- **AppDev team**: The AppDev members are the developer that implement new feature in the application, write the necassry steps to build the application and monitor the performance of the application after every new feature.

Table 3.1: Product Backlog

ID	User Story	Priority	Estimation
1	As a user, I want to login using my username and password and start using the application right away.	2	99
2	As a user, I want to visualize accounts applications and select one to run new benchmark	1	99
3	As a user, I want to visualize the list of accounts.	4	99
4	As a user, I want to visualize the status of accounts benchmarks.	6	99
6	As a user, I want to be able to choose an account and visualize its history	5	99
7	As a user, I want to be able to choose a benchmark and visualize the tests durations.	6	99
8	As a user, I want to be able to choose a test in an account and visualize the history over all benchmarks.	9	99
9	As an account administrator, I want to be able to grant Run benchmark right to users also be able to revoke it.	9	99
10	As an administrator, I want to be able to manage all accounts.	3	99

Table 3.2: User Stories through each sprint

	User stories	Estimation
Sprint 1	As a user, I want to visualize accounts applications and select one to run new benchmark	20
Sprint 2	As a user, I want to be able to choose a test in an account and visualize the history over all benchmarks.	
	As a user, I want to visualize the status of accounts benchmarks.	
	As a user, I want to visualize the list of accounts.	
Sprint 3		
Sprint 4	As an administrator, I want to be able to manage all accounts.	
	As an administrator, I want to be able to manage all accounts.	
	As a user, I want to login using my username and password and start using the application right away.	4

- **Account admin**: The account admin will be responsible for manging account users. Grant run benchmark rights and revoke them.
- **Adminstrator**: the administrator is a person who should be chosen, carefully. in fact, s/he willbe creating, editing and deleting users, accounts. this person should also be able to assing accounts to users.

Conclusion

Chapter 4

Project Design

Introduction

After having specified the project requirements in the previous chapter, we are going to present in the one, the design of the Benchmarks Dashboard. To do so, we will present the project global architecture. Finally, we will zoom into the main layers of the project architecture and explain by providing the accurate diagrams that describe it best.

4.1 UML

4.2 Mockups

In this part we are going to describe and validate the project business requirements in every UI views using mockups. First, we are going to present the Pencil tool. Then we are going to define the project's mockups.

4.2.1 Pencil

Predictix development teams suggested to use Pencil, which is a small graphical tool to sketch out use interfaces, for websites and web / desktop / mobile application. Mockups provide enough interactivity to replace prototypes, and make it easy to collaborate and get feedback on the wireframes.

4.2.1.1 Project mockups

In this section, we are going to describe our project's components mockups.

UI general Description

The Benchmarks dashboard is composed of five main screens which are the account choice, Account's info, Tests choice, Test, New Benchmarks. In the following, we are going to proceed by describing the mockups and the need required in every view.

• **Account choice**: The figure 4.3 represents the landing page of our application, for a administrator user, after the authentication screen which the account's choice. The page should contain a list of the different accounts that the user have access to. In fact, the user should have the possibility to select a certain account, in order to view the all benchmarks run and the list of tests of that project.

• **Account overview**: 4.2 represents the Account overview screen mockup. This page is generated after a user selected an account. This page will contain a chart that represent the overall performance of the account's tests. The chart will contain the total duration it took every benchmark run to finish. It will give a general overview on the state of the account's applications, the general trend for its tests and also any alerts about the the current benchmarks status.

• **New deployment**: The figure 4.3 represents the landing page of our application,

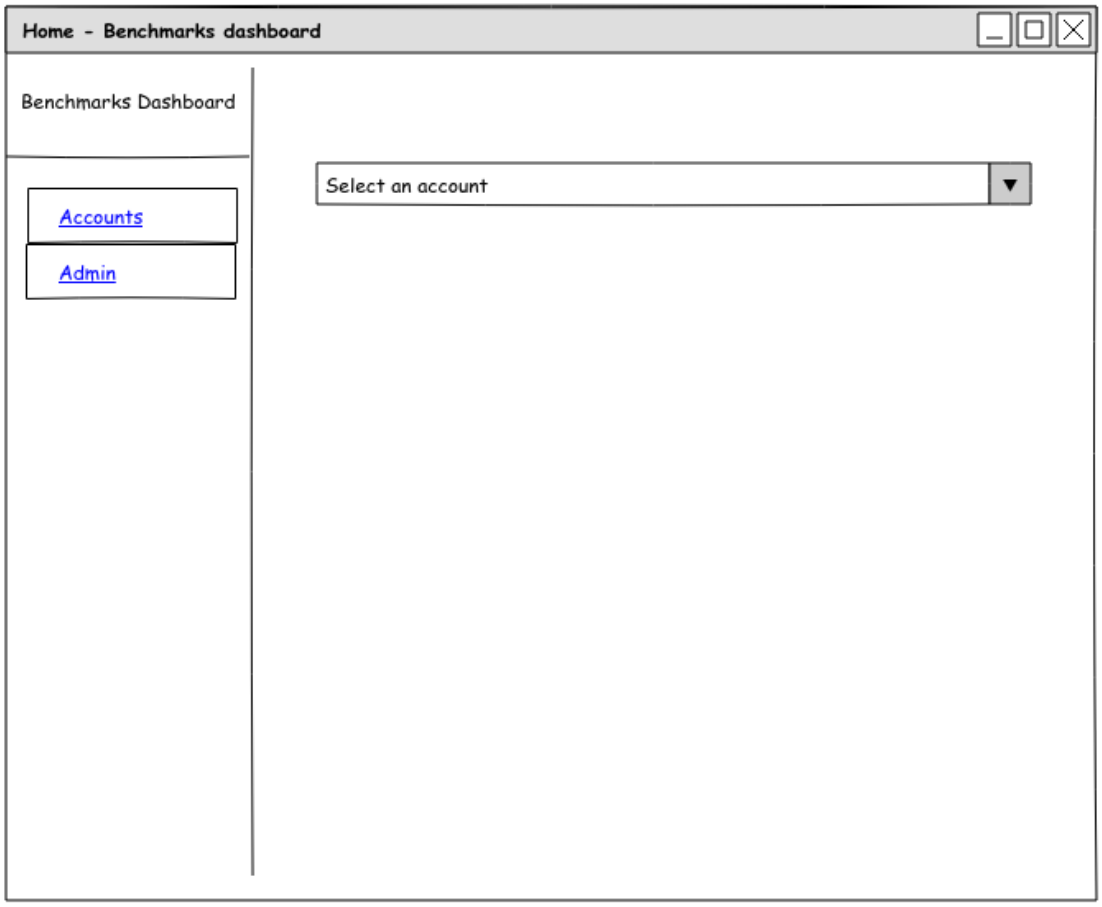

Figure 4.1: Benchmarks dashboard Home

for a administrator user, after the authentication screen which the account's choice. The page should contain a list of the different accounts that the user have access to. In fact, the user should have the possibility to select a certain account, in order to view the all benchmarks run and the list of tests of that project.

• **Account' tests**: 4.4 represents the account's tests mockup. This screen will present the list of all account's tests. This list will show the test name, the percentage of time difference between the latests two runs for that test and the difference in seconds. The list tests that exceed 15 seconds in difference will be highlighted in red and those that exceed 5 seconds will be highlighted in yellow. This will help the user to identify malfunctioning tests. Th

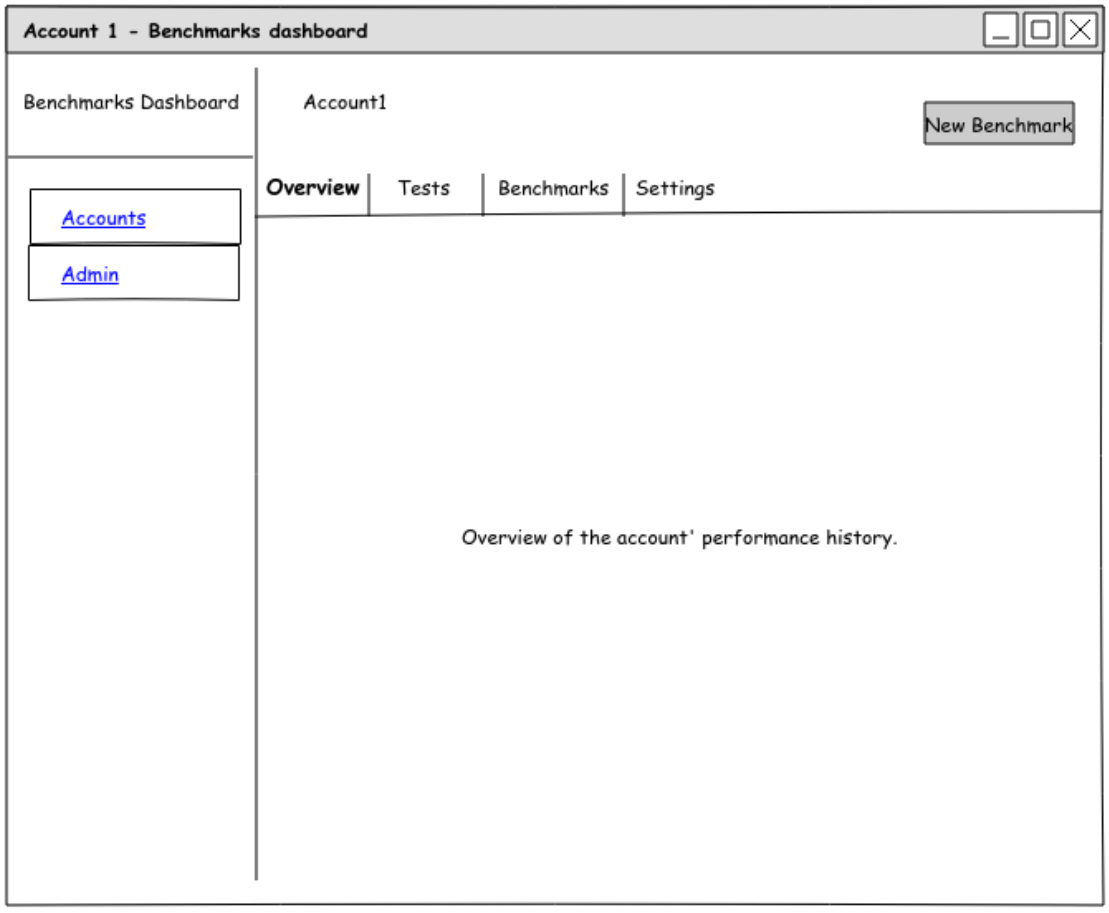

Figure 4.2: Account overview

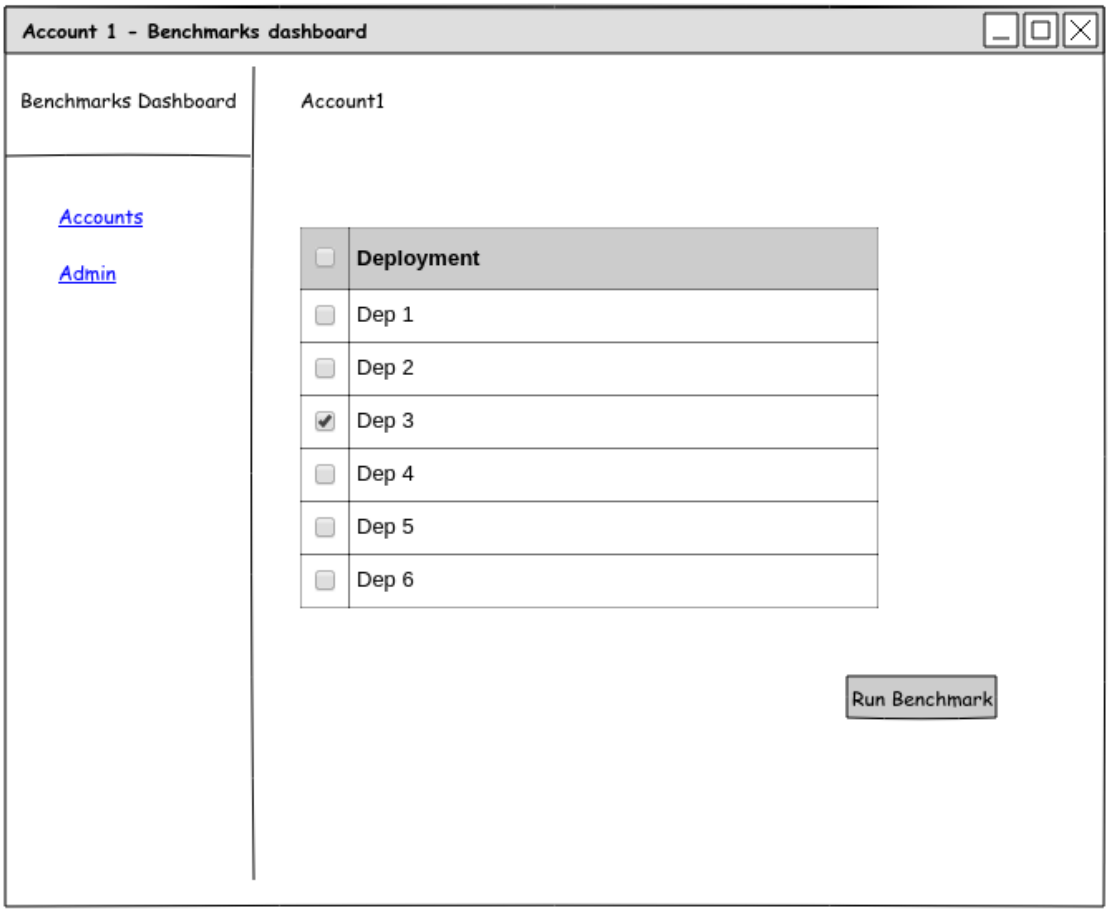

Figure 4.3: New benchmark screen

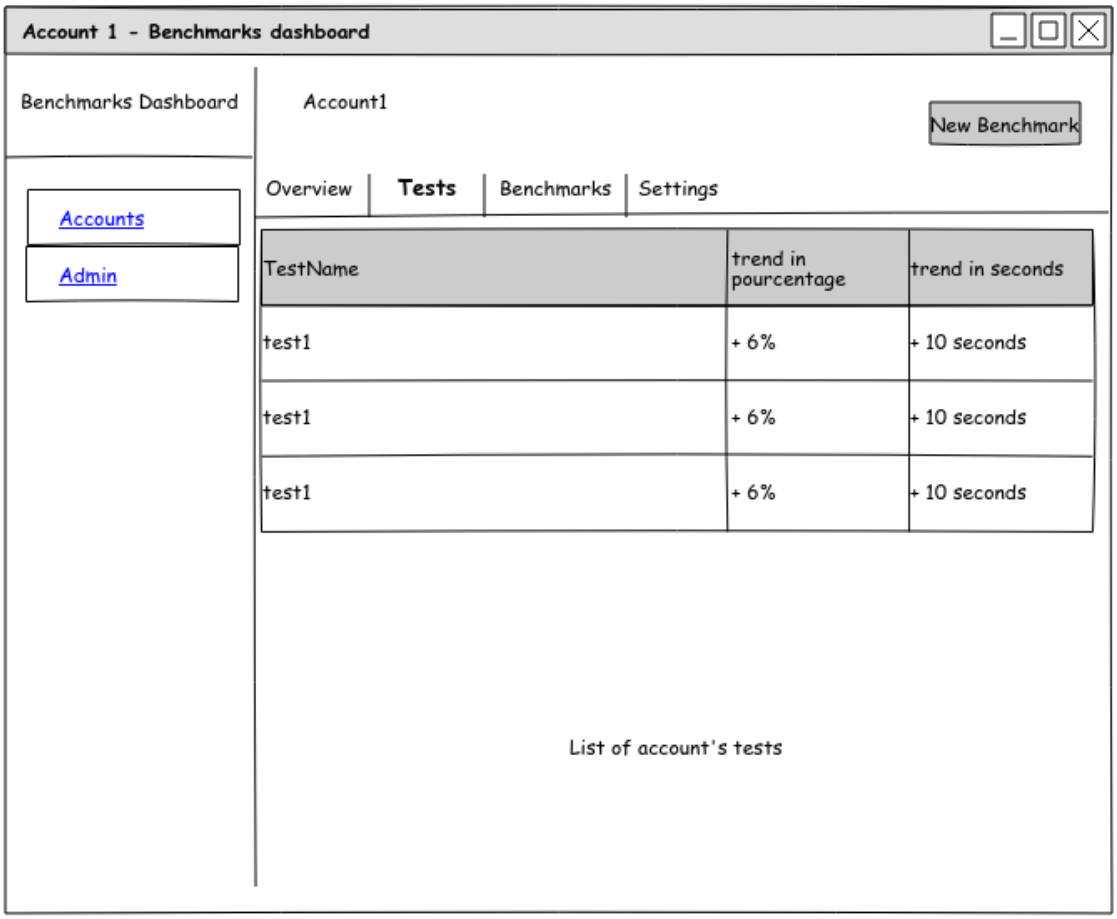

Figure 4.4: Account overview

Chapter 5

Used technologies

Introduction

In this chapter we will be defining the technologies, tools and services that we worked with during the internship.

5.1 Programming languages

5.1.1 Python

Python is an interpreted, interactive, object-oriented programming language. It incorporates modules, exceptions, dynamic typing, very high level dynamic data types, and classes. Python combines remarkable power with very clear syntax. It has interfaces to many system calls and libraries, as well as to various window systems, and is extensible in C or C++. It is also usable as an extension language for applications that need a programmable interface. Finally, Python is portable: it runs on many Unix variants, on the Mac, and on PCs under MS-DOS, Windows, Windows NT, and OS/2.

Python is a high-level general-purpose programming language that can be applied to many different classes of problems. The language comes with a large standard library that covers areas such as string processing (regular expressions, Unicode, calculating differences between files), Internet protocols (HTTP, FTP, SMTP, XML-RPC, POP, IMAP, CGI programming), software engineering (unit testing, logging, profiling, parsing Python code), and operating system interfaces (system calls, filesystems, TCP/IP sockets). Look

at the table of contents for The Python Standard Library to get an idea of whats available. A wide variety of third-party extensions are also available. TODO(Add link Python https://docs.python.org/2/faq/general.html)

5.1.2 Javascript

JavaScript (shortened to JS) is an interpreted, object-oriented, weakly typed, programming language with first-class functions most known as scripting language for Web pages. It was originally implemented so that client-side scripts could interact with the user, control the browser, communicate asynchronously and alter the web page content that was displayed, and now it has evaluated to involve game development and applications creation. JavaScript's main purpose was the creation and embedding of scripts in the HTML client side in order to guarantee more interactivity with the user. It was created to offer dynamic tasks such as the transformation of the page elements, visual effects and immediate reactions to the user actions. But now JS for server-side web is gaining popularity with the appearance of JavaScript frameworks for that purpose. JavaScript was a purely interpreted language. This means that scripts execute without preliminary compilation (without conversion of the script text into machine code). The user's browser interprets the script, that is, analyzes and immediately executes it. In modern implemented, The user's browser interprets the script, that is, analyzes and immediatley TODO(Add link Mozilla developer network, https://developer.mozilla.org/en-us/docs/web/javascript)

5.2 Frameworks

5.2.1 Pyramid

Pyramid is a fast and low-level Python web framework. It is developed as part of the Pylons Projects. Even Pyramid isn't famous like Flask and Django, it minimal, unobtrusive and agnostic design helps in developing more robust and fast web applications. TODO(Complete)

5.2.2 ReactJS

React is an open-source front end library developed by Facebook. It's used for handling view layer for web and mobile apps. ReactJS allows us to create large web applications

that use data which can change over time, without reloading the page, using reusable UI components. Its main goal is to be fast, simple and scalable. It is currently one of the most popular JavaScript libraries and it has strong foundation and large community behind it. React components are typically written in JSX, a JavaScript extension syntax allowing quoting of HTML and using HTML tag syntax to render subcomponents.

5.2.2.1 Grommet

Grommet is a UX framework for enterprise applications developed by HP, on top of ReactJS. It was created to align the desing and developer workflow into one seamless experience. The Grommet components library is designed to be perfect complement to the design system. With little setup, developer will be able to create a Grommet application from scratch in minutes and take the pain out of translating desing files into unning application code.

5.3 Tools and services

5.3.1 AWS

Amazon Web Services (AWS) is a subsidiary of Amazon.com that provides on-demand cloud computing platforms. These services operate from many global geographical regions including 6 in North America. They include Amazon Elastic Compute Cloud, also known as "EC2", and Amazon Simple Storage Service, also known as "S3". As of 2016, AWS has more than 70 services, spanning a wide range, including compute, storage, networking, database, analytics, application services, deployment, management, mobile, developer tools and tools for the Internet of Things. Amazon markets AWS as a service to provide large computing capacity quicker and cheaper than a client company building an actual physical server farm.

5.3.2 Nix

Nix is a purely functional package manager. This means that it treats packages like values in purely functional programming languages such as Haskell they are built by functions that dont have side-effects, and they never change after they have been built. Nix stores packages in the Nix store, usually the directory /nix/store, where each package has its own unique subdirectory such as

`nix/store/b6gvzjyb2pg0kjfwrjmg1vfhh54ad73z−firefox−33.1/`

Where `b6gvzjyb2pg0...` is a unique identifier for the package that captures all its dependencies (its a cryptographic hash of the packages build dependency graph). This enables many powerful features.

Nix was created by Eelco Dolstra as he's PHD. Eelco is, along with other main contributors to the Nix ecosystem, a Predictix employment.

5.3.3 Nix language

The Nix expression language is a pure, lazy, functional language. Purity means that operations in the language don't have side-effects (for instance, there is no variable assignment). Laziness means that arguments to functions are evaluated only when they are needed. Functional means that functions are normal values that can be passed around and manipulated in interesting ways. The language is not a full-featured, general purpose language. Its main job is to describe packages, compositions of packages, and the variability within packages.

5.3.3.1 NixOS

Nixos is Linux distribution built on top of the Nix package manager. It uses the Nix language for declarative configuration which allows realiable system upgrades. Although NixOS started as a research project, it is a fully functional and usable operating system.

5.3.3.2 NixOps

NixOps is a tool for deploying sets of NixOS Linux Machines, either to real hardware or to virtual machines. It extends NixOS's declarative approach to system configuration to networks and adds provisioning.

The NixOps Dashboard is a built-in house extension to NixOps and offers the same content and functionality via a Web UI and a rich RESTful API. Core features are:

- User-friendly web-based interface
- Improved security
- Scheduled operations (deployments, backups, ...etc)

- Permanent traceability
- Clear audit trail
- Role based access

We will be using NixOps Dashboard API to provision servers and deploy the applications in the cloud in order to benchmark them in a production-like systems.

Conclusion

TODO

Chapter 6

Project realization

6.1 Introduction

This chapter covers the implementation, testing and the deployment processes of the Benchmarks dashboard project. To deal wit each of these processes, we will introduce the technologies used and describe the way we used them. In addition, we will presente the project planning.

6.2 Developmnet environment

6.2.1 Devlopment machine characteristics

Below are the characteristics of the developmnet machine we used during the project implementation.

- **Processor**: Intel Core i7-3540M CP
- **RAM**: 16 GO DDR3
- **System**: Ubuntu Linux 14.04 LTS

6.3 Project management

This section gives an overview of the project management process of the Benchmarks Dashboard application. At first, we will start by introducing the project team and then we will describe the tracking process of the different implementation tasks.

6.3.1 Project team

The table TODO introduces the project team members and their roles.

Scrum role	Person
Product owner	Oussama Elkaceh
Scrum Master	Wassel Msehli
Team members	Chaker Benhamed

6.3.2 Jira

Predictix development teams use Jira for managing the projects. Jira is a commercial software product developed by Atlassian. It is used for bug tracking, issue tracking and project management. Jira allow prioritizing, assigning, tracking, reporting and auditing issues. Indeed, it improves productivity by cutting down on time wasted on tracking issues and coordination. In fact, it keeps the team on track and allows the project manager to monitor the progress on projects. Besides, it improves quality be ensuring that all tasks are recorded down with all details and followed up until accomplishment. Moreover, Jira is an

extensible platform which means that it offers workflow customization to match more the business process. TODO(Add link Atlassian, jira software provider, www.atlassian.com.)

6.4 Implementation

6.4.1 Implementation process

This section describes the whole implementation process step by step TODO

6.4.2 Achieved work

This section is a description of the four sprints and the results we achieved at the end of each one of them.

6.4.3 Sprint 1

In the first sprint, our goal is to start by automating the benchmark of one of Predictix's applications, thus we can understand how to generalize for all other project and solve issue of consistency. The application we worked on is THD-AP. We started by installing the application locally understand how to build it, deploy it and how to benchmark it.

After we get acquainted to the benchmark process, we started righting defining the steps in the benchmark pipeline.

6.4.3.1 Deploying

To be tested in production-like environment, the application need to be deployed to the cloud. In order to compare between benchmarks runs the server in which the application will be deployed need to be consistent across all the runs The user mustn't have any interaction with the provisioning of the server nor the deployment of the application.

• **Provisioning**: Since Predictix application are memory-intensive, we chose to deploy them in a r3.4xlarge EC2 instances from Amazon AWS. This type of instances offer lower price of GiB of RAM. r3.4xlarge has 122 GB of RAM with 16 cores CPU and 320 GB of SSD Storage.

• **Installing**:After provisioning the server. We need to deploy the application to that server. The closure of the application will be already built in Hydra. We will be using

an existent nix configuration to install the application. It uses a systemd service called *install-app*. This service will start the database, the web-server, install workspaces and it will make sure that the application is up and running.

For starting the deployment we use NixOps Dashboard, it will take some arguments like the type of the instance TODO.

6.4.3.2 Load data

Loading data is project specific, each application need to provide the steps to download data from S3, extract them and run the necessary workflows to load the data into the database. Those steps will be written in a script called *load-perf-data.sh* by the application developers, the script will be ran as soon as the application in installed. It will be expected to be at the script directory within the application root directory. The figure 6.1 shows the definition of the systemd service *load-perf-data.sh*. And the figure **??** Shows the log of systemd service and how the service is invoked automatically.

```
systemd.services.load-perf-data = {
  description = "Load Perf data";
  after = [ "install-app.service" ];
  script = "source /etc/profile ; sh /data/lb_deployment/installed-app/scripts/load-perf-data.sh";
  serviceConfig = {
    Type = "oneshot";
    RemainAfterExit = true;
    User = "logicblox";
  };
};
```

Figure 6.1: Load perf data nix expression

```
load-perf-data-start[5513]: [load-perf-data] Import workspace - BEGIN
load-perf-data-start[5513]: imported workspace '/thd-ap'
load-perf-data-start[5513]: [load-perf-data] Import workspace - END
load-perf-data-start[5513]: [load-perf-data] run workflow - BEGIN
load-perf-data-start[5513]: 2017-06-20 09:12:13,53800+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > lb.JsonService (maste
load-perf-data-start[5513]: 2017-06-20 09:12:14,00400+00:00  INFO Driver            - Executing task 'dev.deploy_initial_data > master.deploy_initial_data > lb.JsonService (master.wf, line
load-perf-data-start[5513]: 2017-06-20 09:12:15,25400+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > lb.JsonService (maste
load-perf-data-start[5513]: 2017-06-20 09:12:16,00900+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > lb.StartServices (mas
load-perf-data-start[5513]: 2017-06-20 09:12:16,22100+00:00  INFO Driver            - Executing task 'dev.deploy_initial_data > master.deploy_initial_data > lb.StartServices (master.wf, li
load-perf-data-start[5513]: 2017-06-20 09:12:21,89800+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > lb.StartServices (mas
load-perf-data-start[5513]: 2017-06-20 09:12:22,51600+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_hierarc
load-perf-data-start[5513]: 2017-06-20 09:12:22,74900+00:00  INFO Driver            - Executing task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_hierarchies > mas
load-perf-data-start[5513]: 2017-06-20 09:12:22,75800+00:00  INFO lb.tdx.v2.Import  - 'dev.deploy_initial_data > master.deploy_initial_data > master.import_hierarchies > master.TdxImport >
load-perf-data-start[5513]: 2017-06-20 09:12:22,97400+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_users >
load-perf-data-start[5513]: 2017-06-20 09:12:23,27800+00:00  INFO Driver            - Executing task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_users > master.Td
load-perf-data-start[5513]: 2017-06-20 09:12:23,27900+00:00  INFO lb.tdx.v2.Import  - 'dev.deploy_initial_data > master.deploy_initial_data > master.import_users > master.TdxImport > lb.td
load-perf-data-start[5513]: 2017-06-20 09:12:33,73400+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_hierarc
load-perf-data-start[5513]: 2017-06-20 09:12:34,18800+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_users >
load-perf-data-start[5513]: 2017-06-20 09:12:34,60200+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_users >
load-perf-data-start[5513]: 2017-06-20 09:12:34,79200+00:00  INFO Driver            - Executing task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_users > master.Td
load-perf-data-start[5513]: 2017-06-20 09:12:34,79300+00:00  INFO lb.tdx.v2.Import  - 'dev.deploy_initial_data > master.deploy_initial_data > master.import_users > master.TdxImport > lb.td
load-perf-data-start[5513]: 2017-06-20 09:12:45,47000+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_users >
load-perf-data-start[5513]: 2017-06-20 09:12:45,83800+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_measure
load-perf-data-start[5513]: 2017-06-20 09:12:46,08600+00:00  INFO Driver            - Executing task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_measures > master
load-perf-data-start[5513]: 2017-06-20 09:12:46,10300+00:00  INFO lb.tdx.v2.Import  - 'dev.deploy_initial_data > master.deploy_initial_data > master.import_measures > master.TdxImport > lb
load-perf-data-start[5513]: v, /tmp/dev_data/initial/measures/COGS_LY_C_AP.csv, /tmp/dev_data/initial/measures/Inv_TY_U_AP.csv, /tmp/dev_data/initial/measures/MktdStrPos_LY_R_AP.csv, /tmp/de
load-perf-data-start[5513]: 2017-06-20 09:12:57,10000+00:00  INFO Driver            - Processing request: task 'dev.deploy_initial_data > master.deploy_initial_data > master.import_measure
load-perf-data-start[5513]: 2017-06-20 09:12:57,24100+00:00  INFO Driver            - Terminating driver due to 1 queries without work to do.
load-perf-data-start[5513]: 2017-06-20 09:12:57,25400+00:00  INFO Driver            - Scheduler terminated.
load-perf-data-start[5513]: [load-perf-data] run workflow - END
```

Figure 6.2: load-perf-data log

6.4.3.3 Running tests

For running tests we use a generic script called *run-perf*. This script will define the instructions for running the tests, collect the logs, generate the report and upload it to S3. The figure 6.3 shows the definition of the systemd service *perf-runner* responsible for running the script, it will be launched as soon as the data is loaded with the *load-perf-data* service.

```
# service that will execute the performance test suite and collect the logs.
systemd.services.perf-runner = {
  description = "execute the performance test suites";
  after = [ "load-perf-data.service" ];
  requires = [ "load-perf-data.service" ];
  wantedBy = [ "multi-user.target" ];
  script = "source /etc/profile && sh ${<perf_src>}/scripts/run-perf";
  serviceConfig = {
    Type = "oneshot";
    RemainAfterExit = true;
    User = "logicblox";
  };
};
```

Figure 6.3: perf-runner nix expression

```
systemd[1]: Starting execute the performance test suites...
Starting Run 1
Done Run 1
compress the logs
[perf runner] - checking that all inputs are available ...
[perf runner] - file: 'lb-server.log' found.
[perf runner] - file: 'lb-web-server.log' found.
[perf runner] - file: 'results.csv' found.
[perf runner] - Generating the Benchmark website ...
splitting /tmp/test_results/lb-server.log in sections ...
splitting /tmp/test_results/lb-web-server.log in sections ...
creating section tests_attribute_mappings_SHOULD_assign_an_attributeId_for_each_Product ...
creating section tests_attribute_mappings_SHOULD_assign_an_attributeLabel_for_each_Product ...
creating section tests_attribute_mappings_SHOULD_assign_an_attributeLabel_for_each_SKU ...
..................
..................
..................
creating section tests_Start_New_Plan_Start_New_Plan_Setup_and_Service_call_SNP_Second_Transaction ...
[perf runner] - Benchmark website generated.
[perf runner] - Publishing the Benchmark website ...
[perf runner] - Benchmark website generated.
[perf runner] - Generating the benchmark report ...
[perf runner] - Checking the existence of the history of benchmark runs in AWS S3 ...
[perf runner] - Checking predictix/perf-history.csv ...
[perf runner] - Updating the full history of runs in S3 ...
[perf runner] - Full history is updated.
[perf runner] - Compressing the results tarball ...
[perf runner] - Uploading the results tarball to AWS S3 ...
[perf runner] - email is sent to 'chaker.benhamed@infor.com'
DONE.
systemd[1]: Started execute the performance test suites.
```

Figure 6.4: perf-runner log

The figure 6.3

6.5 Deployment

Hydra is a Nix-based continuous build system that constantly checks out code sources of software projects from version management systems such as Mercurial, to build, test and release them. The build tasks are described using Nix expressions. This allows a Hydra build task to specify all the dependencies needed to build or test a project.

In fact, the code of our application is, currently and constantly pushed in a repository in BitBucket, which is a web-based hosting service for projects that use either the Mercurial or Git revision control systems, such as GitHub. In order to have our application ready to be deployed, we added a file called default.nix to our project, that actually defines our projects nix-expressions. Afterwards, we created a project under Hydra, that we called *Benchmarks Dashboard*. Thus, the *Benchmarks Dashboard* project on Hydra, will be pulling our code from our BitBucket repository along with the default.nix file in order to perform the automated build and unit-tests.

The following are some screen shots about our Hydra project build. The figure 6.5 represents the configuration for our project. We can notice that the links to the different project's dependencies are defined, such as our source code and the nixpkgs repository which contains the definitions for all packages available through the nix package manager.

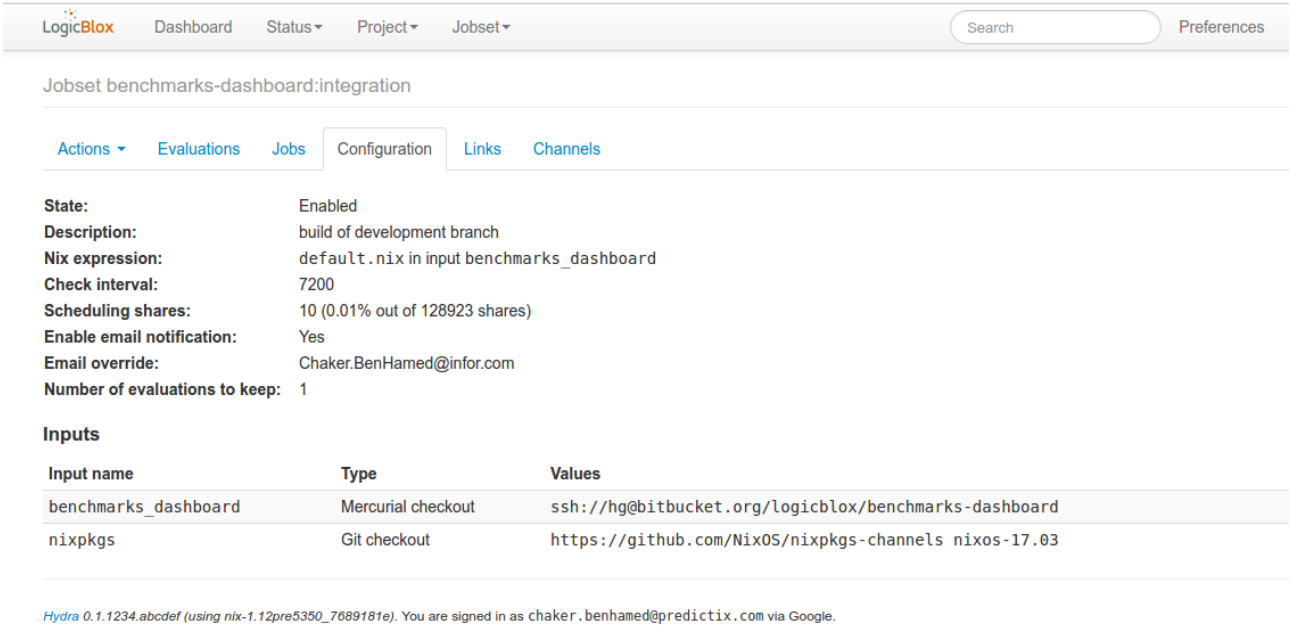

Figure 6.5: Benchmarks Dashboard Hydra configuration screen

The figure 6.6 represents the Hydra evaluation screen, where we can see the history if the build and whether there are erros in the build or not.

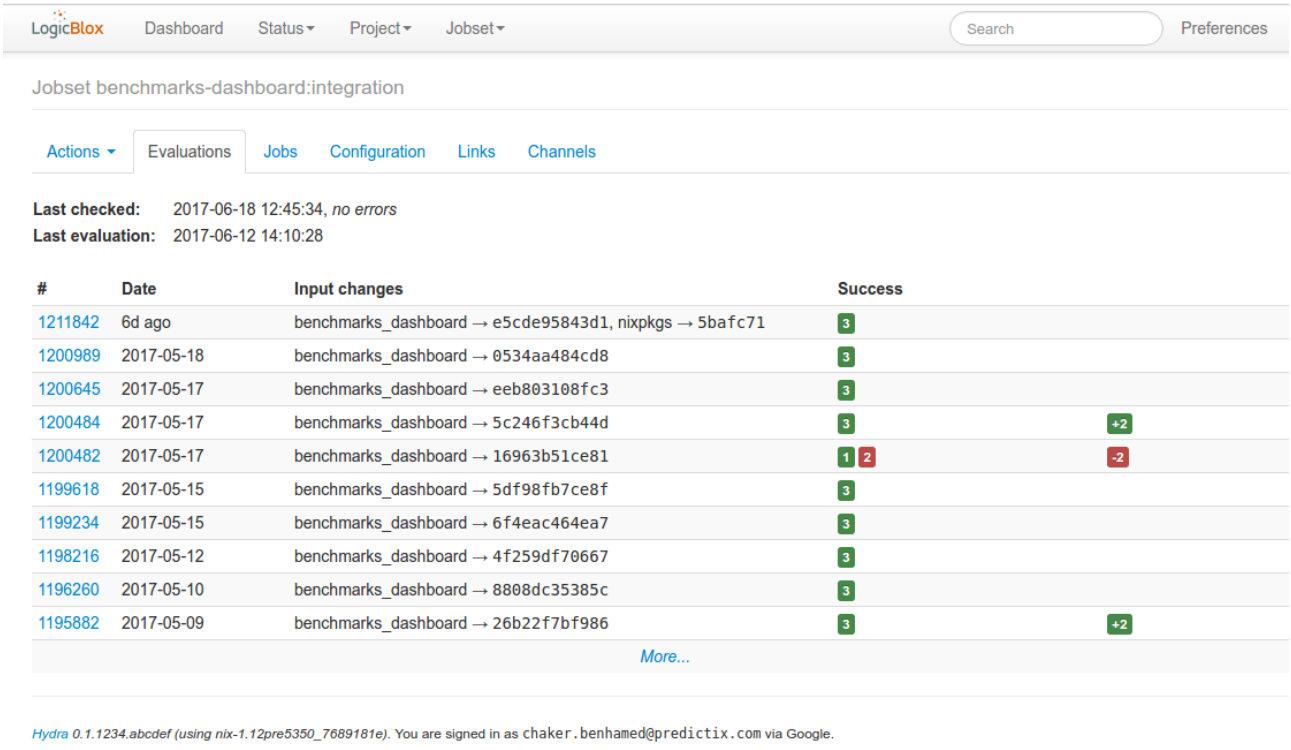

Figure 6.6: Benchmarks Dashboard Hydra configuration screen

In Hydra we can also run unit test and report various indicator about the quality of the application. The figure 6.7 shows the covarge report of the test of our application.

Coverage report: 88%

Module	*statements*	*missing*	*excluded*	*coverage*
BD_backend/__init__.py	46	3	2	93%
BD_backend/api/__init__.py	0	0	0	100%
BD_backend/api/account.py	88	7	0	92%
BD_backend/api/admin.py	9	0	0	100%
BD_backend/api/benchmarks.py	26	0	0	100%
BD_backend/api/home.py	21	0	0	100%
BD_backend/api/tests.py	55	36	0	35%
BD_backend/api_doc.py	5	0	4	100%
BD_backend/models/__init__.py	28	0	0	100%
BD_backend/models/account.py	21	0	0	100%
BD_backend/models/benchmark.py	33	8	0	76%
BD_backend/models/meta.py	23	0	0	100%
BD_backend/models/user.py	41	0	0	100%
BD_backend/models/user_role.py	14	0	0	100%
BD_backend/mongodb/__init__.py	12	7	0	42%
BD_backend/nixops_dashboard.py	97	7	1	93%
BD_backend/routes.py	4	0	0	100%
BD_backend/security.py	18	2	0	89%
BD_backend/serializers.py	29	0	0	100%
BD_backend/tasks/__init__.py	12	2	0	83%
BD_backend/vars.py	2	0	0	100%
Total	**584**	**72**	**7**	**88%**

coverage.py v4.0.1, created at 2017-06-12 14:17

Figure 6.7: Coverage report

Conclusion

TODO