AboutDataGovernance

Demo Notes iGovernance Suite

AboutDataGovernance Srls – Piva 13348991004
Sede Legale: Via Flaminia 670 – 00191 Roma
Sede Operativa: Piazza Gandhi 3 – 00141 Roma
www.aboutdatagovernance.com

iGovernance Suite: 10Clicks2knowledge.

Intro

IGovernance Suite is not a Data Governance Infrastructure and cannot even be classified simply as a Metadata Repository: Is that kind of Application Customers should develop on their own starting from a Technical Metadata Collection (Repository) in order to obtain a valid instrument to satisfy different Business Units needs (CDO, Operation, Software Development, etc.) and gain enormous advantages in terms of Efficiency, Time to Market, Money saved, Data Quality, Risks Reduction, and to be ready to start any kind of future project like Digital Transformation, Infrastructure Changes, Migrations, etc.

AboutDataGovernance during the implementation of several Data Governance Projects solved any kind of criticalities Customers should face with, using different approaches than simply a Technical Metadata Collection in a Repository and Data Lineage Generation, and transformed them in an Application: iGovernance Suite, a ready to go Application.

Main Differentiators:

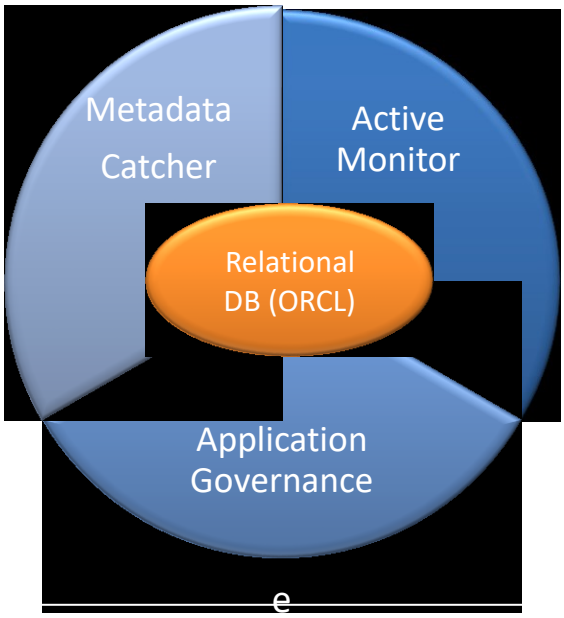
iGovernance Suite first of all is not based on a proper Techical Metadata Repository, the flexibility needed to solve many criticalities pushed us to use a simple relational DB, in order to extend easily the model and satisfy any kind of Customer's need having also the possibility to historicize easily any runned process in the DB (useful for audit pourposes). Usual Metadata Repositories has not that kind of flexibility and functionality.

We also immediately find out that Technical Metadata were not sufficient to solve some criticalities (like understand that Transformation Logic expressed in Parametric Jobs or Dynamic queries, is possible to collect those values only during IT process Run Time) so for any connector (Metadata Catchers) able to understand the differents technologies transformation logic using Technical Metadata, we did implement also Agents able to collect for that Techology those useful Operational Metadata. These has been used to validate Technical Metadata and follow the physical Operational End To End and Multitechnology Data Lineage while the IT process is running and finally historicizing it in the relational DB.

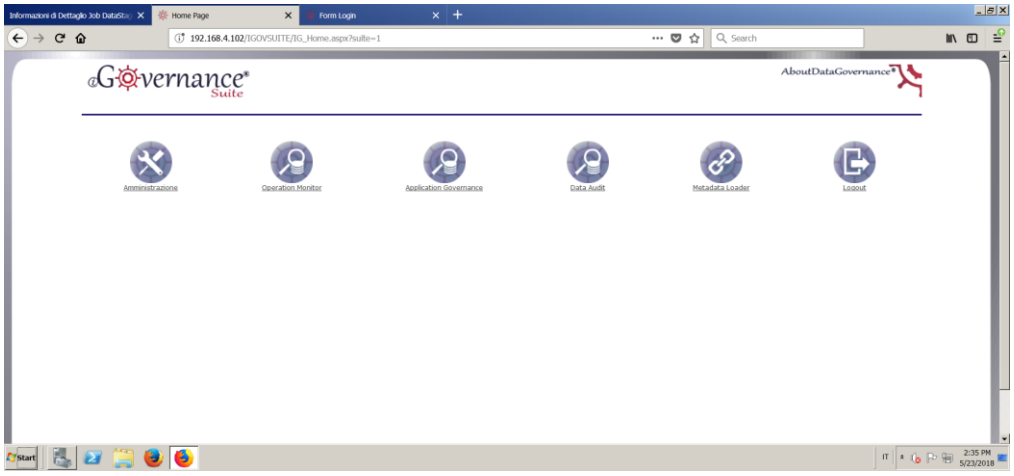
So iGovernance Suite today is composed of:

- A Set of **Metadata Catchers** for differents techologies (DataStage, Powercenter Informatica, Teradata, Netezza, Exadata, Cognos, Microstrategy, Microsoft, PL/SQL, etc.) able to collect both Techical and Operational Metadata and understand any kind of transformation logic (SQL, Stored Procedures, BTEQ, scripts, etc.) Operational Metadata can be collected also at near realtime.
  Metadata Catchers will provide the above certified Transformation Logic Loading it in the Customer's preferred Metadata Repository.

- A **Relational DB** (Oracle) where both Technical and Operational Metadata are integrated stored and ready to be analyzed easily to satisfy any kind of business need.
  In the DB is stored and historicized also any runned IT process (one of our customers has a 2 years historicization in it) for any kind of audit pourposes.
- **Active Monitor**, an application builded on iGovernance Suite DB to solve all Operation Team problems and helping them to mantain and Enhance their Service Level Agreement with a more efficient approach allowing an End To End Multistep and Multitechnology IT processes Active Monitoring and allowing the team to define and set in a centralized way any kind of controls and alarms cross Technologies,
- **Application Governance** a navigation application builded to allow Software Development Team to contextualize with few clicks ANY object in his IT phisical process in a complex environment.



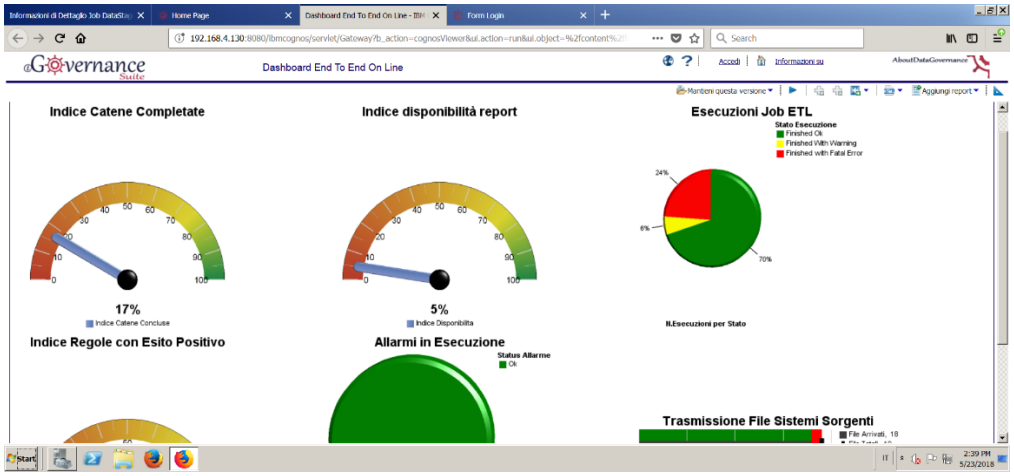
Following demo is runned using real Operational and Technical Metadata used in different projects. The environment used is a DW loading composed by 5 different steps (Staging Area, Primary DW, Secondary DW, Olap Cubes and Reporting Services Loading and Starting) and different Techologies (DataStage, Teradata, PL/SQL, Cognos, Microstrategy and Microsoft Reporting Services).



(Pic. 1)

This Demo will show probably less than 20% of functionalities and analysis availables actually on the system, once ALL Metadata Informations are stored properly in the iGovernance Suite DB, to solve a certain issue or obtain a certain result is just a matter of presentation considering the DB model can be extended to meet any Customer's business issue.

Now let's go into the Operation Active Monitor functionality to start the analysis.



(pic 2)

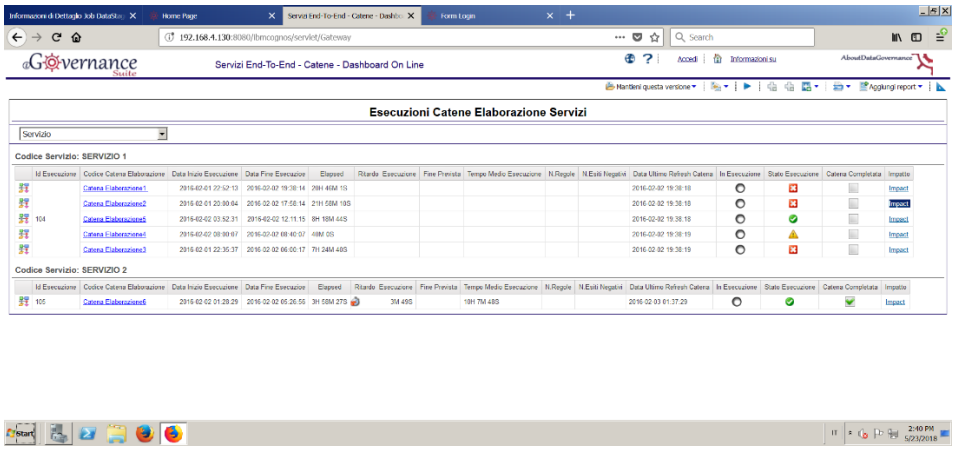
Let's start the navigation from VP Operation's Dashboard. At this point the system has collected all Tech Metadata from all different technologies along the application chain and every "N" minutes (2 minutes in this case but can be setted from seconds to hours) does collect Operational Metadata and integrates them in the DB.

If the need was only to have a Data Lineage for CDO and Regulatory & Compliance pourposes the complete information, decoding in the meanwhile any transformation logic, was sended to the choiced Metadata Repository.

In this case the VP has the view of his important meters: Percentage of completed chains respect those expected at the moment, percentage of report availability respect the report

availability expected at the moment, result of all executions (Jobs, BTEQ, Stored Procedures, SQL procedures and PL/SQL procedures), number succesful rules applied and number of alarms in execution (those 2 are defined directly by customers in the application), finally becouse the entire load is starting when are received files from Main Frame we made the integration of their FTP (Sopra AXWAY) and we can see we were expecting at the moment 19 files but we received only 18.

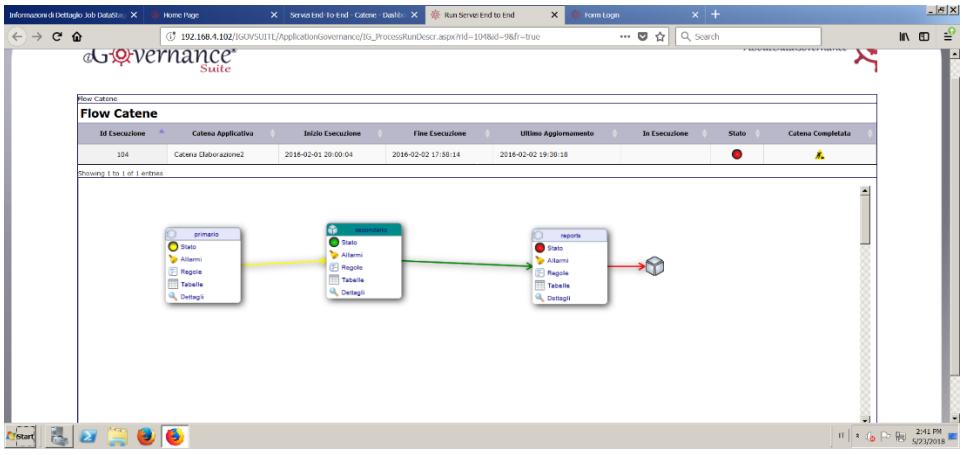
Worried about the bad result start the navigation clicking on completed chains Meter.



(pic. 3)

And finding out in this moment he has 2 different services (DW Loading and CRM loading by example) running with 6 application chains: 2 with positive result, 1 in warning State and 3 with errors.

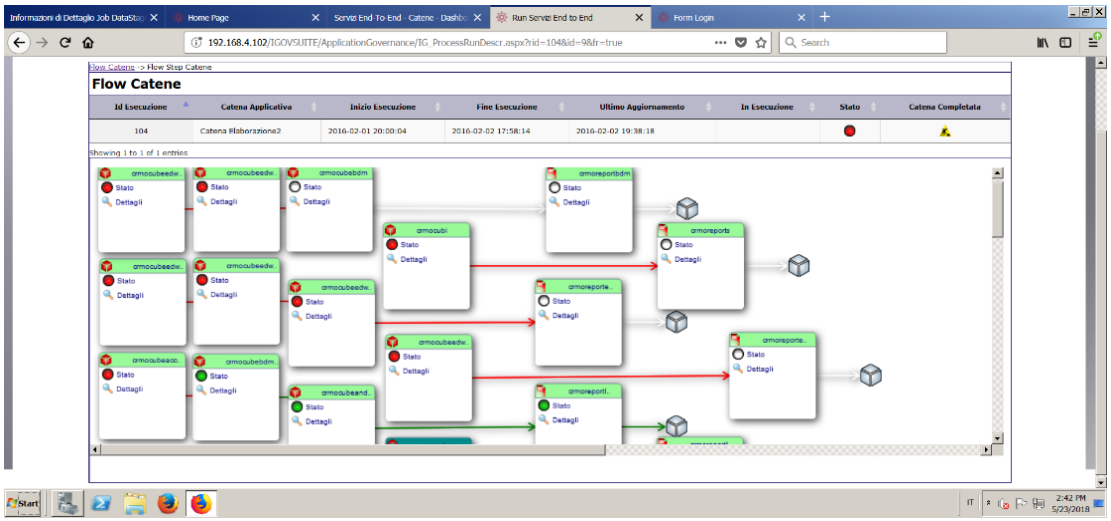
Clicking on the second he wants find out how the process is composed and where is supposed to be the problem:



(pic. 4)

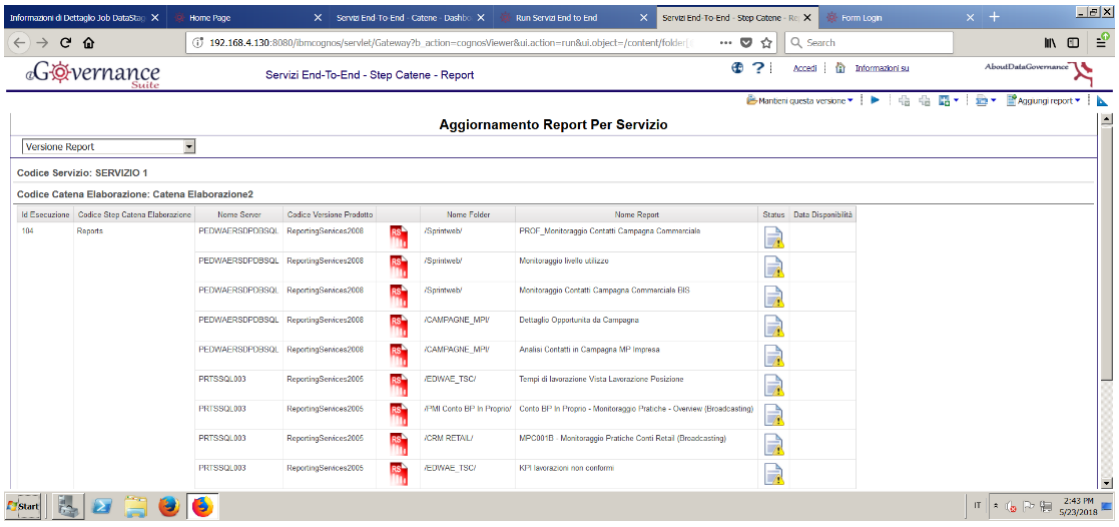
And the system shows steps the process is composed of and the state for any step highlighting there are warnings while loading from primary to secondary DW house (Yellow arrow DataStage on Teradata with BTEQ usage) and errors on the Microsoft Reporting services activity (Red Arrow).

As Operation team the first thing we are worried about is what we've been not able to produce becouse of that error so at lesast we can advise our internal customer about.



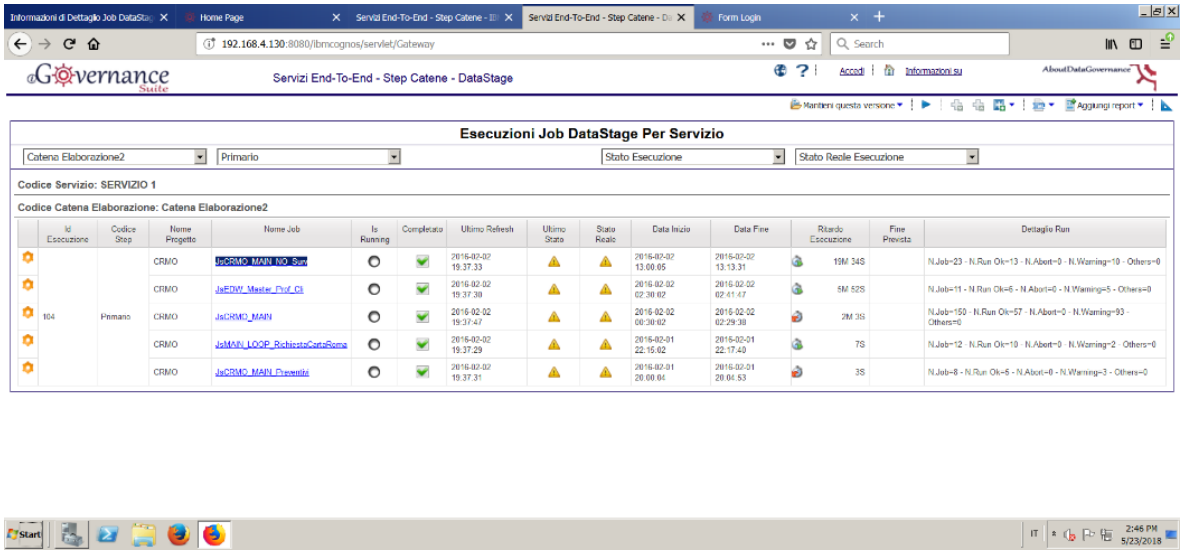
(pic. 5)

By clicking on the last step we can see how many Reporting Processes were supposed to start and their state, realizing the first did not start at all, then 3 aborted and 1 finished well. By clicking on those in error we obtain the complete list of reports aborted and the owners so our internal customer can be at least adviced about the problem.



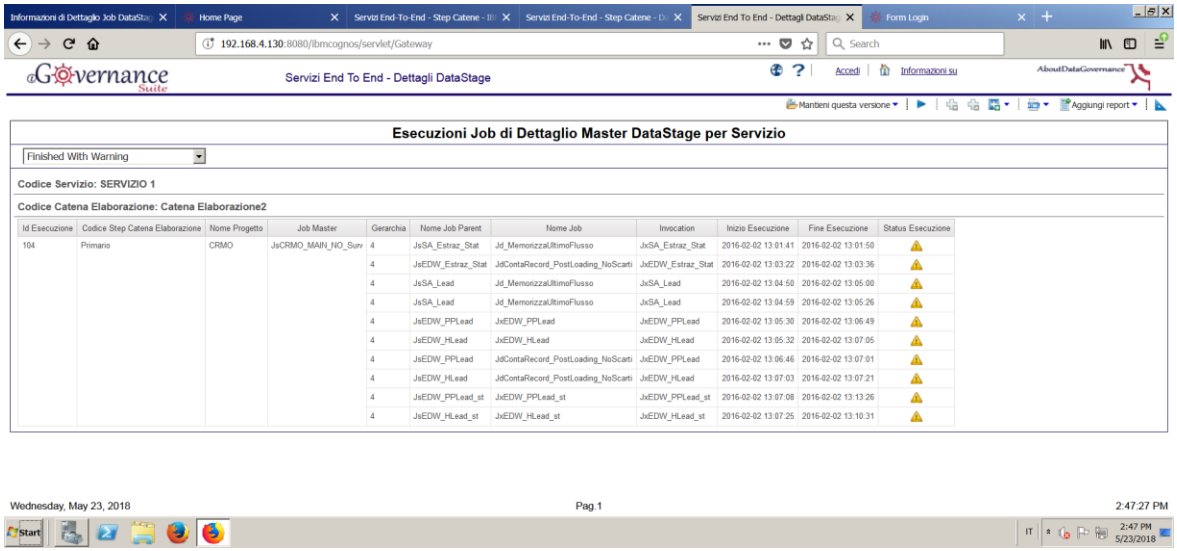
(pic. 6)

Then let's start to investigate about the cause of the problem so clicking on the step with the yellow arrow (on pic. 3) we obtain a list of Steps (containers of jobs in Datastage environment) with some important informations: state, elapsed, if still in execution, advance or late in execution, and number of job contained and their state.



(pic. 7)

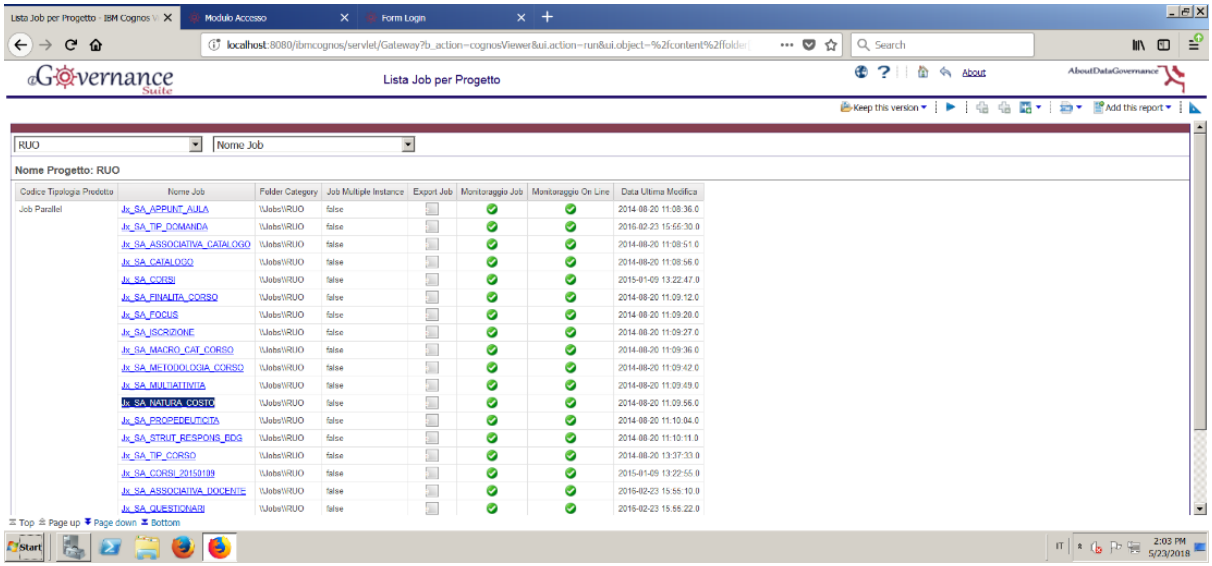
At this point we can find that the first 2 have an important advance in execution (this means probably less data managed and transformed) and an important number of jobs contained are in warning state. So is important to give to Software Development Team the right information about Jobs probably causing these warnings and sequently errors in Reporting activity so clicking on those we obtain the complete list.



(Pic. 8)

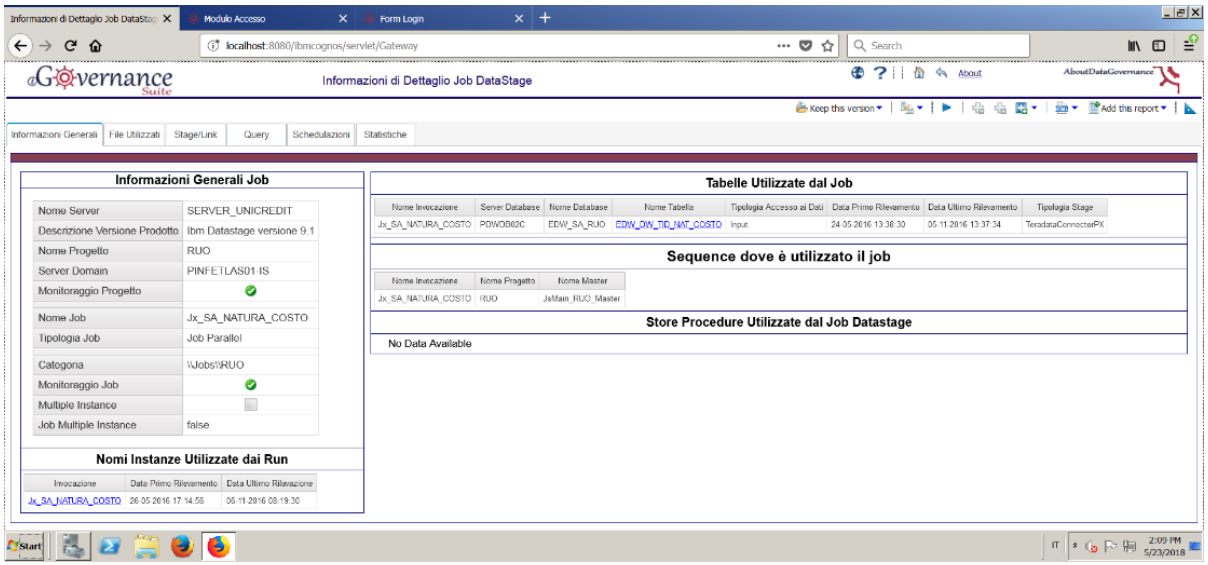
Now let's wear Software Development Team's shoes for a while and let's take by example one of the jobs to be contextualized (becouse an error, a warning or becouse we need to change something for business needs pourposes) is a job called NaturaCosto in the project named RUO and we start the analysis from scratch (knowing nothing more about this job).

So we access to the iGovernance Suite DB asking for a list of jobs by project name.



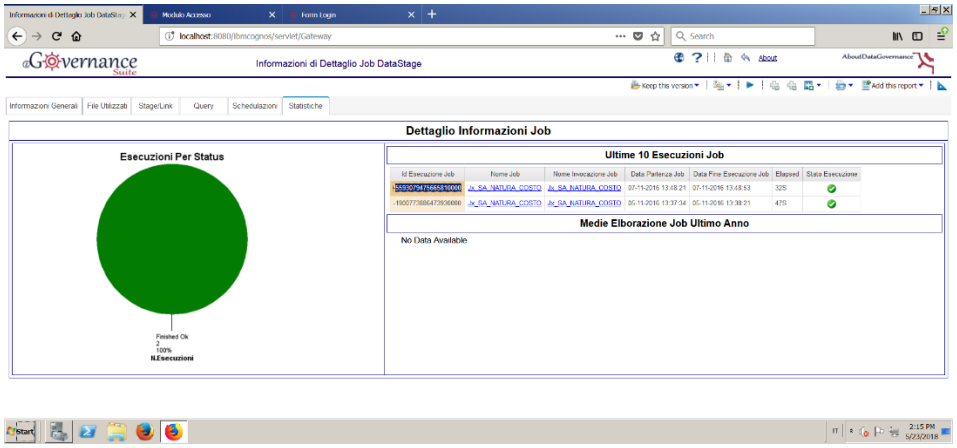
(Pic. 9)

Clicking on the respective NaturaCosto Job we obtain some preliminary useful informations about it.



(Pic. 10)

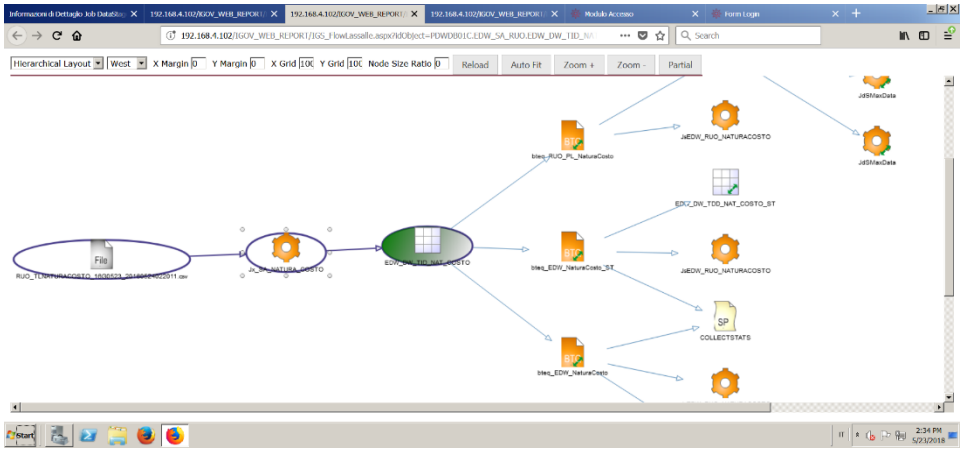
Location (Server Name, Domain, Kind of Job, Release, etc) Instances Names used by Job's Run, Table's Name used by Job, Sequences where the Job is used, Stored Procedures used by Job, Files used by Job (and when, do not forget in DB are stored any IT process execution for a "N" period), Stage/links used by Job, Queries used by Job, When the Job is supposed to be scheduled, and Usage Statistics about the Job.



(Pic. 11)

As we can see this job has been runned twice. The ID Execution number is a unique number we assign to any IT Process execution and to all his components and objects to help historicization and sequently audit activities.

Coming back to previous screen (Pic. 10) at this point we can start contextualization activity and ask to the system to find out the IT process where the named Job NaturaCosto and the Table (a Teradata one) listed has been used and the system answers showing the End-To-End multistep multitechnology Flow executed.



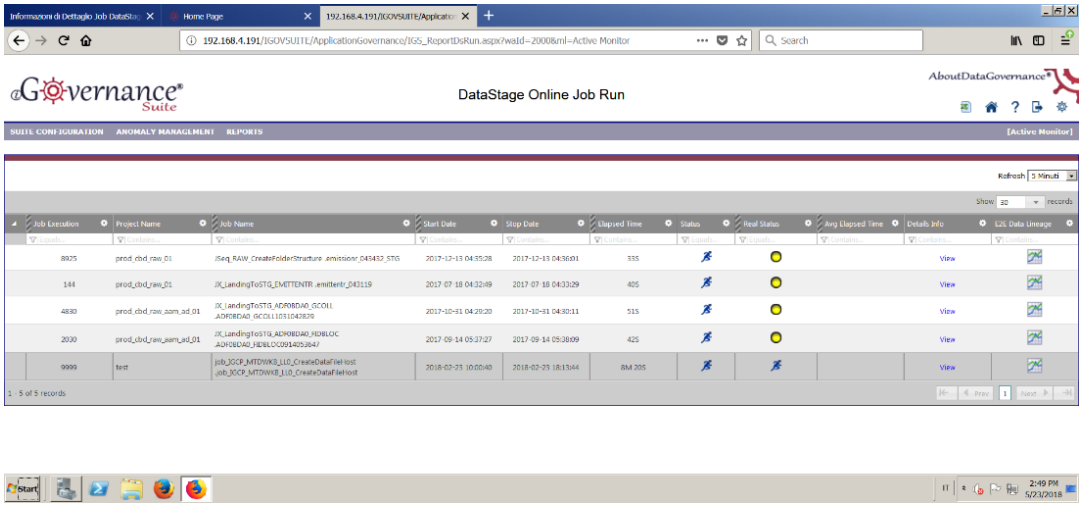
(Pic. 12)

The flow willl shows first 300 objects finded, where the green arrow is present it means the flow can be expanded so the user can follow the IT process track interested and realizing that process starts taking infos from a file, executing some transformations and delivering them on a Teradata DW Table, then starting 3 differents BTEQ calls,with 3 different tracks and so on until the infos interested are delivered on their target (file, DB or application).

So starting from scratch with just few clicks the object (a single job in this case) has been contextualized in a complex IT process and environment to take any corrective or adjustment action in a complex environment, what happens usually without such

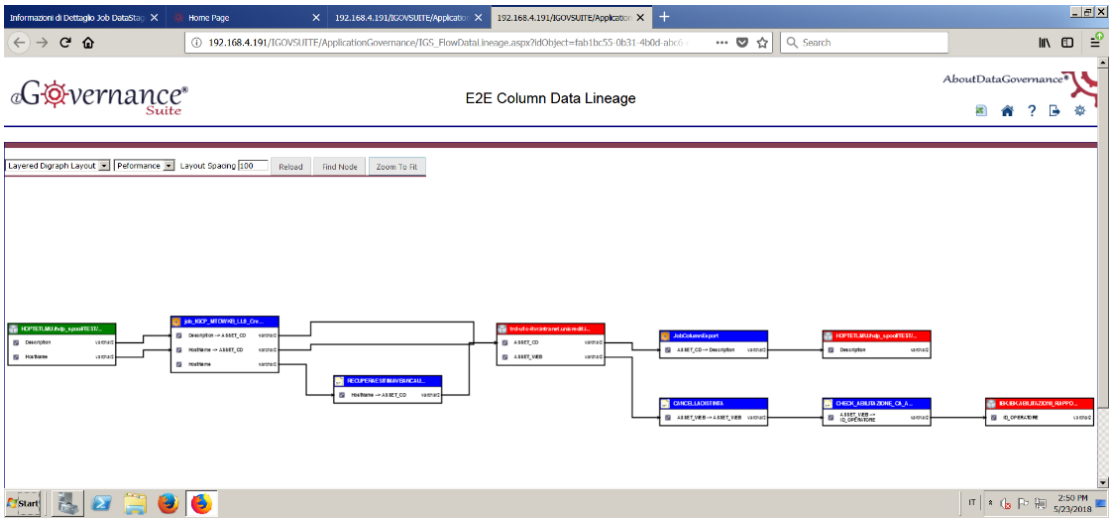
functionality is the classic “war room” with those specialists for each piece of application or technology investing much time and resources.

The above Contualization Phase is made at Table level only, now let’s see an example going down at column level to understand a certain process transformation logic (it was Data Stage calling PL/SQL procedures in this case).



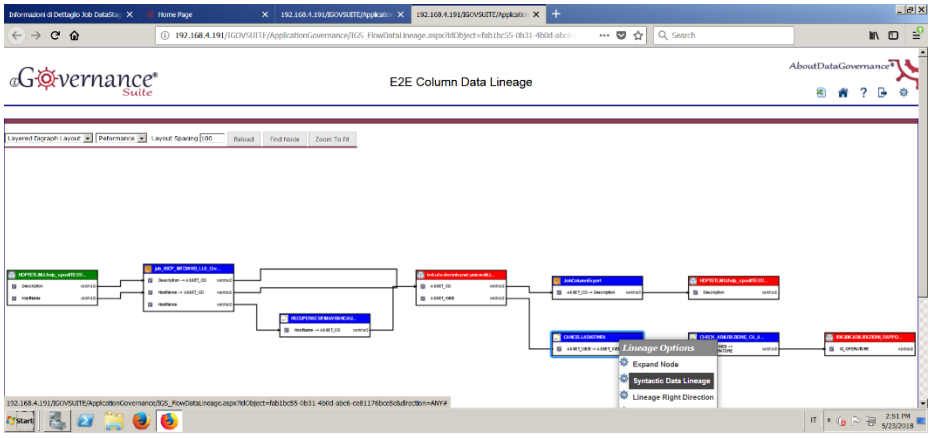
(Pic. 13)

Let’s go back and choice the Application Governance functionality listing all running IT processes (iGovernance Suite Agents are still working in background collecting new informations every “N” minutes) and let’s select the last IT process listed (still in running State btw).



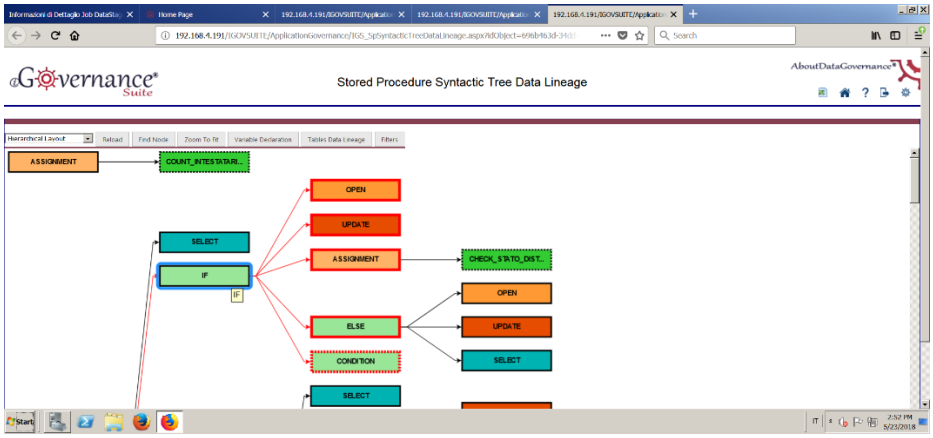
(Pic. 14)

The system it represents his complete Column level flow, in green are expressed sources, in blue transformation objects and in Red targets. A small icon expressed on transformation objects explain the kind (Jobs, Stored Procedures, etc.).



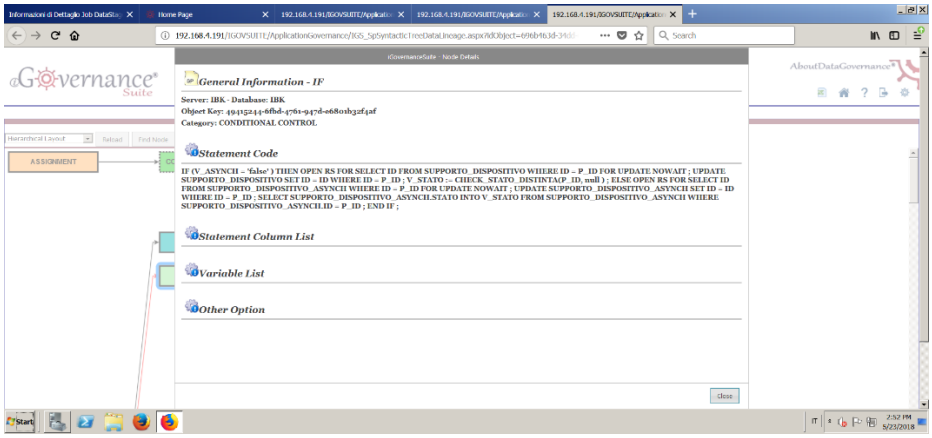
(Pic. 15)

Selecting by example the Stored Procedures called CancellaDistinta we can ask the system to show the complete sintactic tree of the same Stored Procedures



(Pic. 16)

And if necessary by clicking on any object (an “IF” by example) see the complete transformation logic behind it



(Pic. 17)

What above was the result of Data Governance experiences we had and is just an example of results can be obtained with a different approach than based only on technical Metadata, a more flexible approach where the concept of metadata is extended to any kind of

information on data (Metadata) present in the company and where “rigid” metadata models can satisfy usually only a subset of real actual business needs. Where the money invested in Data Governance project can be used to help many different areas in their day by day job, more helpful then a simple “documentation” of a certain process (what Data Lineage represent for Operation and SW development teams) and becouse of that opportunity, with faster ROI.

It’s enough think about some advantages coming from the operational metadata usage by example:

- A Data Lineage cannot be completed if parametric Jobs and Dynamic queries are used (very often).
- Many times we found that Datalineage coming from a technical metadata declaration was not aligned to the real process running. That becouse the usage of SQL and/or scripts and/or Stored Procedures, if a change is made at SQL level inside a Job by example, that change is not automatically reflected in the Technical declaration and the choiced technical metadata repository is generating a data lineage not in line with the respective IT process running. That’s why is necessary to find the real *Operational Data Lineage* to validate Technical declaration and when a process running is not included in those already registered the user needs to be advised and that new IT process needs to be validated by owners.
- Once all DW House IT processes have been registered by the system thanks to operational metadata usage, customers can know exactly which objects and which data in his system are really used and by contrary knowing at this point wich objects and which data are in his systems and are not used at all (obsolete). On our experiences we found our customers having from 20% up to 45% of objects and data obsolete with big savings in maintenance costs and far more savings in case of migrations or new projects. (btw sometimes we found also pieces of IT processes running without generating any result at all).
- Talking about Data Quality, everybody knows the importance of it in a company’s decision taking activity and the need to monitor constantly Data Quality levels. What about once we realize quality level of a certain information is not sufficiently good? We need to understand and verify how the IT process does transform and manage the content of that Data.

So finally we can define Technical Data Lineage is not the landing point of a Data Governance Project (like most people is considering) but only the starting point and what companies needs mainly is to recover that deep knowledge, management and control of IT processes lost during the time becouse of many reasons (different System integrators developing during lasts 10 years, knowledge lost becouse of people left or changed role etc.) and is proven that Technical Metadata are not sufficient and more Metadata types and a more flexible approach to the problemi is needed to satisfy business and digital transformation market requirements.