

Top 60 Talend Interview Questions with Answers 2024

Top Answers to Talend Interview Questions

1. Talend Characteristics

Criteria	Result
Distinguishing feature	First Data integration software as a service
Deployment	Business modeling, graphical development
ETL functionality	Makes ETL mapping faster and simpler for diverse data sources

2. What Talend stands for?

Talend stands for Talend Open Studio.

3. What do you mean by Talend?

Talend open studio is the open source data integration product produced by Talend and it is designed to convert, combine and update data in various areas across a business.

4. When was Talend open studio launched?

Talend launched in October 2006

5. Talend is written in which language?

It is written in Java language.

6. Tell the latest version of Talend open studio.

The latest version is 5.6.0

7. Differentiate between ETL and ELT.

ETL stands for Extract, Transform and Load which is a process that involves gaining data from exterior source, converting it to get fit into operational requirement, then load it into the end target database.

ELT stands for Extract, Load and Transform which is the process in which data is get, then loaded into the staging table in the database and then data is converted according to the need.

8. What is the significance of tLoqateAddressRow component in Talend?

It is a component for mailing correct address belongs to the respective customer data to make sure a single customer view and good delivery for their respective customer mailing.

9. Could we change the background color of the job designer ?

Yes, we can change the background color of job designer.

10. How we can change the background color of job designer?

We can change the background of job designer by clicking on the preferences of the window menus, after the talend, appearance, designer then click on color menu.

11. Can we define a variable which can be accessed by many jobs?

Yes, we can declare a static variable in the routine and add the setter method for the respective variable in the routine. Then this variable can be accessed from various jobs.

12. Can we save our personal settings in the DQ Portal?

No, we can not save our personal setting in the DQ Portal.

13. Can we change the generated code directly?

No, this is not possible we cannot generate code directly for Talend.

14. Which method we should use to include our own Java code in a Job?

We can use tJava, tJavaFlex component, tJavaRow, etc to include our own Java code in a Job.

15. Can we use Binary Transfer mode in SFTP ?

No, in SFTP we cannot use the binary transfer mode because SFTP is not like the FTP. Hence, we cannot apply the concept like 'current mode directory' and 'transfer mode'.

16. For sorting data which component we generally use?

We can use tExternalSortRow and tSortRow.synthesizing sorted input

17. In talend what is the fixed pattern of date?

By default the date pattern is dd-MM-yyyy.

18. What do you mean by component?

Component is simply a functional piece which is used for a single operation. It is a bundle of file kept within a folder named followed by component name.

19. Differentiate between 'insert or update' and 'update or insert'.

Insert or update means first we insert a record, but if a record is matching with the primary key then the record is updated.

Update or insert means first we update the record with same primary and if the record doesn't exists then we insert the record.

20. Differentiate between Repository and Built-In ?

In Built-In we can manually edit the data as data is kept locally in the job whereas in repository all the data is stored there only. We can extract only Read-only-information into the job from repository.

21. Which option is better Built-in or Repository?

It simply depends on the way we use it. We should use Built-in for the data which we use rarely and use the Repository for the data which we use repeatedly.

22. Differentiate between OnComponentOk and OnSubjectOk ?

They both are trigger links which can connect to another subject job. The major difference between both of them is that they both lies in the execution order of the connected subjobs.

23. In talend how the delimited data be normalized?

We can normalized the delimited data by clicking on the tNormalized component.

24. Define tMap ?

tMap is the latest component which simply converts and routes data from one or many sources to one or many destinations.

25. Tmap component support which types of joins?

TMap supports inner,unique,outer, and all joins.

26. Define tDenormalizeSortedRow?

tDenormalizeSortedRow is bundled in a group of all input sorted rows. It helps in saving the memory by synthesizing sorted input flow.

27. For data transform using built in .Net classes which Talend component is used?

For transforming the data by utilizing custom we can use tDotNETRow component.

28. What do you mean by tJoin?

By exact matching the several columns of tables the tJoin joins the two tables.

29. Define MDM in talend ?

It is a management by which an organization makes and manage a single, consistent and correct view of key enterprise data.

30. In 5.6 version of Talend what is new?

The new feature in talend 5.6 is that it have more techical notes. It also have enterprise and open studio solution.

31. Write the advantages of talend ?

It is highly versatile, cost effective, user friendly and readily adaptable.

32. Define project ?

It is the bundle of technical resources and their respective metadata. All the jobs and business items which we design is known as project only.

33. What do you understand by the term workspace?

It is kind of repository where we can store our folders. It is mandatory to have one workspace repository per connection.

34. Define an item?

An item is a fundamental technical part in a project. They are bundled according to their types as code, metadata, context, etc.

35. What do you understand by the term Migration Task in Talend?

It is done to ensure the worth fullness of a project which we have developed with the previous version of Talend.

36. What is the use of Palette setting in talend?

It allow us to launch the studio more fastly because by using this only the current component is loaded in the project.

37. Define Talend data generator routine?

It is a function which allow us to create group of set data. They are based on the entry of first name, address, town, etc.

38. What are the steps to replace an element in a string?

We can replace one element with another in a string by using Change routine along with tJava components.

39. How we can store a string in an alphabetical order?

We can store a string in an alphabetical order by using ALPHA routine with tJava component.

40. What is the use of String Handling Routines?

It allow us to take out many operations and test on alphanumeric expressions relay on Java methods.

41. What is the use of Numeric Routine?

It allow us to revisit whole or decimal numbers in order to use them as setting in one or more job mechanism.

42. What is the use of Job view?

It shows many information belongs to the open job on the design workspace.

43. Define scheduler ?

This view is used to arrange a task in a sequence that will launch one by one the job which we select through the crontab program.

44. Define configuration tabs ?

It is situated in the bottom half of the design workspace. Every tab open a view which shows the properties of the selected elements in the design workspace.

45. What do you understand by the term Routines?

They are the somewhat complicated Java functions, mostly used to factorize code. It recover Job capacities and optimized data procedure.

46. What is the use of tXML map operation?

With using this we are able to add various input and output flow as needed into the visual map editor to execute.

47. How we can access global and contex variable?

By clicking Cntrl+ Space key we can access global and contex variable.

48. How we can use inner join?

This join is a specific type of join which differentiate itself by the way refusal is performed.

49. What are the operations which tMap allows?

data transformation on any type of fields

data multiplexing and demultiplexing,

fields concatenation and interchange,

data rejecting

field filtering using constraints

50. What is Talend Open Studio?

Talend Open Studio for Data Integration is an open source data integration product developed by Talend and designed to combine, convert and update data in various locations across a business.

51. What is the difference between the ETL and ELT?

ETL:

Extract, Transform, and load(ETL) is a process that involves extracting data from outside source, transforming it to fit operational needs (sometimes using staging tables), then loading it into the end target database or data warehouse. This approach is reasonable as long as many different databases are involved in your data warehouse landscape. In this scenario you have to transport data from one place to another anyway, so it's a legitimate way to do the transformation work in a separate specialized engine.

ELT:

Extract, Load, Transform(ELT) is a process where data is extracted, then loaded into staging table in the database, transforming it. Where it sits in the database and then loading it into the target database or data warehouse.

52. What is the use of tLoqateAddressRow component in Talend?

This Component is used to correct mailing addresses associated with customer data to ensure a single customer view and better delivery for their customer

mailings.

53. What do you understand by MDM in Talend?

Master Data Management, through which an organization builds and manages a single, consistent, accurate view of key enterprise data, has demonstrated substantial business value including improvements to operational efficiency, marketing effectiveness, strategic planning and regulatory compliance. To data, however, MDM has been the privilege of a relatively small number of large, resource-rich organizations. Thwarted by the prohibitive costs of proprietary MDM software and the great difficulty of building and maintaining an in-house MDM solution, most organization have had to forego MDM despite its clear value.

54. What's new in v5.6?

This technical note highlights the important new features and capabilities of version 5.6 of Talend's comprehensive suite of Platform, Enterprise and Open Studio solutions.

With version 5.6 Talend:

- Extends its big data leadership position enabling firms to move beyond batch processing and into real-time big data by providing technical previews of the **Apache Spark**, **Apache Spark Streaming** and Apache Storm frameworks.
- Enhances its support for the Internet of Things (IoT) by introducing support for key IoT protocols (MQTT, AMQP) to gather and collect information from machines, sensors, or other devices.
- Improves Big Data performance: Map Reduce executes on average 24% faster in v5.6 and 53% faster than in v5.4, while Big Data profiling performance is typically 20 times faster in v5.6 compared to v5.5.
- Enables faster updates to MDM data models and provides deeper control of data lineage, more visibility and control.
- Offers further enterprise application connectivity and support by continuing to add to its extensive list of over 800 connectors and components with enhanced support for enterprise applications such as SAP BAPI and Tables, Oracle 12 GoldenGate CDC, Microsoft HDInsight, Marketo and Salesforce.com

55. What is the advantage of Talend?

Talend is cost-effective, easy to use, readily adaptable and extremely versatile. With the help of the graphical user interface we can easily and quickly link up a large number of source systems using the standard connectors.

56. Describe the ETL process?

Extraction, Transformation and Loading (ETL) processes are critical components for feeding a data warehouse, a business intelligence system, or a big data platform. While mostly invisible to users of a business intelligence platform, an ETL process retrieves data from operational systems and pre-processes it for further analysis by reporting and analytics tools. The accuracy and timeliness of the entire business intelligence platform rely on ETL processes, specifically:

- Extraction of the data from production applications and databases (ERP, CRM, RDBMS, files, etc.)
- Transformation of this data to reconcile it across source systems, perform calculations or string parsing, enrich it with external lookup information, and also match the format required by the target system (third normal form, star schema, slowly changing dimensions, etc.)
- Loading of the resulting data into The business intelligence (BI) applications: Data Warehouse or Enterprise Data Warehouse, Data Marts, Online Analytical Processing (OLAP) applications or "cubes", etc.

57. What is tJoin?

tJoin joins two tables by doing an exact match on several columns. It compares columns from the main flow with reference columns from the lookup flows and outputs the main flow data and/or the rejected data.