Item 294

git_comments:

1. we don't project into union as shuffle key.
2. There are totally 5 types of non-primitive. If falling into this branch, it means it is a TIMESTAMP_INSTANT and we will by default treated it as eligible.

git_commits:

1. **summary:** [GOBBLIN-1212] Handle non-Primitive type eligibility-check for shuffle key properly
   message: [GOBBLIN-1212] Handle non-Primitive type eligibility-check for shuffle key properly Closes #3060 from autumnust/embeddedHiveDistcp

github_issues:

github_issues_comments:

github_pulls:

1. **title:** [GOBBLIN-1212] Handle non-Primitive type eligibility-check for shuffle key properly
   body: Dear Gobblin maintainers, Please accept this PR. I understand that it will not be reviewed until I have checked off all the steps below! ### JIRA - [ ] My PR addresses the following [Gobblin JIRA](https://issues.apache.org/jira/browse/GOBBLIN/) issues and references them in the PR title. For example, "[GOBBLIN-XXX] My Gobblin PR" - https://issues.apache.org/jira/browse/GOBBLIN-1212 ### Description - [ ] Here are some details about my PR, including screenshots (if applicable): Before this change, the handling for non-primitive types like map,array and union will hit NPE since they don't have `fieldNames`. This fixed that and making sure the eligibility check happens recursively into those types. ### Tests - [ ] My PR adds the following unit tests __OR__ does not need testing for this extremely good reason: ### Commits - [ ] My commits all reference JIRA issues in their subject lines, and I have squashed multiple commits if they address the same issue. In addition, my commits follow the guidelines from "[How to write a good git commit message](http://chris.beams.io/posts/git-commit/)": 1. Subject is separated from body by a blank line 2. Subject is limited to 50 characters 3. Subject does not end with a period 4. Subject uses the imperative mood ("add", not "adding") 5. Body wraps at 72 characters 6. Body explains "what" and "why", not "how"

github_pulls_comments:

github_pulls_reviews:

1. Should we do a null check here? What happens if originalSchema.getChildren() returns null?
2. Two points: 1. Maybe avoid exposing LinkedIn internal schemas outside, and 2. From the schema definitions, it is hard to see what exactly is being tested. Can you add a comment explaining the difference between struct_a and struct_b and why eligibileForUpConvert should return true?
3. Will add a safety-guard.
4. **body:** minor typo: complextType -> complexType.
   label: documentation
5. Should the error message be: "Illegal format of ORC schema as:" + originalSchema?

jira_issues:

1. **summary:** Handle shuffle-key eligibility check properly
   description:

jira_issues_comments:

1. Issue resolved by pull request #3060 [https://github.com/apache/incubator-gobblin/pull/3060]