**Standard for Application on Gold Infrastructure**

MH = MUST HAVE

SH = SHOULD HAVE

# Jobs

## Job documents

* MH : Job has release note
  + A list of embedded features
  + A location for code history with embedded commits
  + A location for release history with features for these releases
  + A list of the impact for any user of the jobs results
  + A list of the impact for launching that job and the parameters changes
* MH : Job has a documentation in Confluence
  + Detail of how the job is working, the full workflow, the dependencies
  + Detail of what the job is producing and why
  + History of modifications (that is not just jira), qualifications, and push in productions
  + Procedures for fixing and resolving incidents (Rollback, Escalation, …)
* MH : Job has code in Stash

## Job building

* MH : Job has been build on a Continuous Integration pipeline
* MH : Job has been released with a Continuous Integration pipeline
* MH : Job artefacts have been stored in a repository (eg nexus)

## Job qualification

* MH : Job had a dry run on a test environment with recent data and has been qualified properly
* MH : Job has a history of stability on a test environment for xx days
* SH : Job has a history of all results since first release

## Job configuration

* MH : Critical Job parmeters are externalized into documented configuration files (file location, dependencies, parallelism, memory settings, servers settings, etc.) with at least default values that would reflect the optimal production environment, stage
* MH : Job follows the recommended settings defined by the « SRE »

## Job reporting

* SH : A job must report its execution results to the central « job result repository » with the following informations :
  + Execution start date
  + Execution stop date
  + Execution status
  + Execution input
  + Execution output

# Datasets generated by a Job

## Naming Normalization

* MH : Datasets name are normalized (and respect the Data Management Confluence page about it)
* MH : Datasets with partitions have the partition naming normalized

## Fields Normalization

* SH : Datasets fields name and data type are normalized (and respect the Data Management Confluence page about it)

## Models

* SH : Datasets should be documented and have a data model explaining
  + lineage,
  + keys of interest (primary, secondary, foreigns),
  + constraints,
  + metadata for each field if needed,
  + specific property of a field (master data, autogenerated, calculated, etc.)

## Data Integrity

* MH : Datasets have an additionnal file or additionnal fields that indicates its integrity status (checksum, timestamp of generation date, etc.)

## Reporting Results

* MH : Jobs have an output that indicate each generated datasets,
* SH : Jobs have information / metrics about each generated datasets

## Historization of Results

* MH : Jobs have a retention policy defined when released
* SH : Jobs have a effective retention policy check system so that there is no data loss

# Scheduling

## Launching Scripts documents / source

* MH : Launching Script has code in Stash or an equivalent with :
  + A location for code history with embedded commits
  + A documentation that details the parameters and the specifcs of the job (idempotency, dependency, mandatory parameters, etc. )
  + As a self documented unit, source code (if applicable) should contains the information detailed above

## Launching Scripts reporting

* MH : A launching script report its execution results to a historized location :
  + Execution start date
  + Execution stop date
  + Jobs launching details
  + Jobs execution status
  + Final status

## Frequency, Chaining, dependency and Data/Jobs Lineage

* SH : Launching scripts are documented in a centralized place with details about their frequency, their starting and landing time (or estimated time to completion) and their dependency to other jobs (be it in a form of a graph, a gantt chart or any other practical information format that would indicate those informations).

# Rights Management

## Application Code / Scheduling Code changes Rights

* MH : Application/Jobs Scheduling Scripts have a list of who is authorized to do changes and this list is enforced and reviewed every X months
* MH : Application/Jobs Scheduling Scripts have a list of who made changes, why and when changes and this list is enforced and reviewed every X months
* MH : There is a process that describe who can gives/remove/change Rights, this process is documented and there is an auditable log of these acces grants / changes / removal.

## Data accessibility Rights

* MH : Datasets (tables or files) have a list of who is authorized to do changes and this list is enforced and reviewed every X months
* MH : : Datasets (tables or files) have a list of who made changes, why and when changes and this list is enforced and reviewed every X months
* MH : There is a process that describe who can gives/remove/change Rights, this process is documented and there is a auditable log of these acces grants / changes / removal.

# Job CheckList

## Code or Source

* 3D box Code is in Stash, all history of modifications is available, code submition is a process

## Documentation

* 3D box Confluence Documentation Page / A Readme.MD Page / A embedded documentation in the code
* 3D box Release notes and release history
* 3D box Configuration Details

## Building

* 3D box CDS or Equivalent
* 3D box Artefacts (bundle or jar) are stored on a dedicated environment that prevent modifications or is user restricted

## Qualification

* 3D box Dry run on a test environment with recent data
* 3D box Stability on a test/preprod environment for xx days

## Reporting

* 3D box Execution Reports output
* 3D box Reports are stored

# Datasets CheckList

## Naming Normalization and Models

* 3D box Datasets file name / table name are normalized
* 3D box Datasets Fields are normalized
* 3D box Data Model is described

## Data Integrity

* 3D box Datasets have integrity metadata (named sets and metrics) reported

## Historization of Results

* 3D box Datasets have a retention policy defined
* 3D box Datasets have a retention check

# Scheduling CheckList

## Launching Scripts documents / source

* 3D box Scripts Code is in Stash, all history of modifications is available, code submition is a process
* 3D box Scripts have documentation

## Launching Scripts reporting

* 3D box Launching script has historized reports
  + 3D box Execution start date
  + 3D box Execution stop date
  + 3D box Jobs launching details
  + 3D box Jobs execution status
  + 3D box Final status

# Rights Management CheckList

## Application Code / Scheduling Code changes Rights

* 3D box List of who can change the code and scripts is defined and limited
* 3D box Adding / Changing / Removing from this is a process, list are frequently reviewed
* 3D box There is a log of who made changes

## Data accessibility Rights

* 3D box List of who can acces / alter data and datasets is defined and limited
* 3D box Adding / Changing / Removing from this is a process, list are frequently reviewed

# Example on Revenue

|  |  |  |
| --- | --- | --- |
| **Job CheckList** | Code or Source | ☑ Code is in Stash, all history of modifications is available, code submition is a process |
|  | Documentation | ☑ Confluence Documentation Page / A Readme.MD Page / A embedded documentation in the code  ☑ Release notes and release history  ☑ Configuration Details |
|  | Building | ☑ CDS or Equivalent  ☑ Artefacts (bundle or jar) are stored on a dedicated environment that prevent modifications or is user restricted |
|  | Qualification | ☑ Dry run on a test environment with recent data  ☒ Stability on a test/preprod environment for xx days |
|  | Reporting | ☑ Execution Reports output  ☑ Reports are stored |
| **Datasets CheckList** | Naming Normalization and Models | ☒ Datasets file name / table name are normalized  ☒ Datasets Fields are normalized  ☒ Data Model is described |
|  | Data Integrity | ☑ Datasets have metadata (named sets and metrics) reported ☒ Datasets have integrity information reported |
|  | Historization of Results | ☑ Datasets have a retention policy defined  ☑ Datasets have a retention check |
| **Scheduling CheckList** | Launching Scripts documents / source | ☑ Scripts Code is in Stash, all history of modifications is available, code submition is a process  ☒ Scripts have documentation |
|  | Launching Scripts reporting | ☒ Launching script has historized reports  3D box Execution start date  3D box Execution stop date  3D box Jobs launching details  3D box Jobs execution status  3D box Final status |
| **Rights Management CheckList** | Application Code / Scheduling Code changes Rights | ☒ List of who can change the code and scripts is defined and limited  ☒ Adding / Changing / Removing from this is a process, list are frequently reviewed  ☒ There is a log of who made changes |
|  | Data accessibility Rights | ☒ List of who can acces / alter data and datasets is defined and limited  ☒ Adding / Changing / Removing from this is a process, list are frequently reviewed |