EDB Postgres
HA構築のベスト・プラクティス

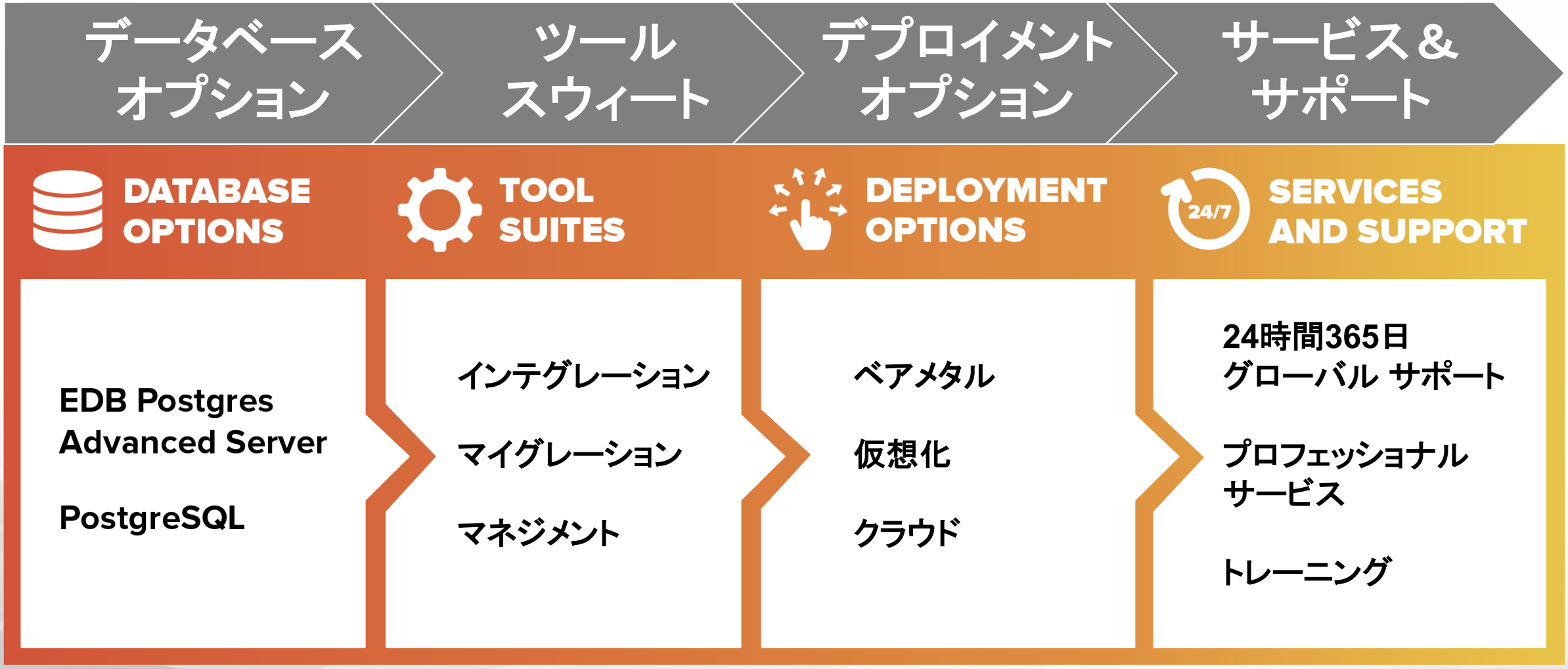
EnterpriseDB
高鶴 勝治

Agenda

1. EDB Postgres Platform
2. EDB Postgres Failover Manager (EFM)
3. EDB Postgres Backup and Recovery (BART)
4. EDB Postgres High Availability(高可用性) アーキテクチャ
5. プラチナ構成(例)

EDB Postgres Platform

EDB POSTGRES PLATFORM



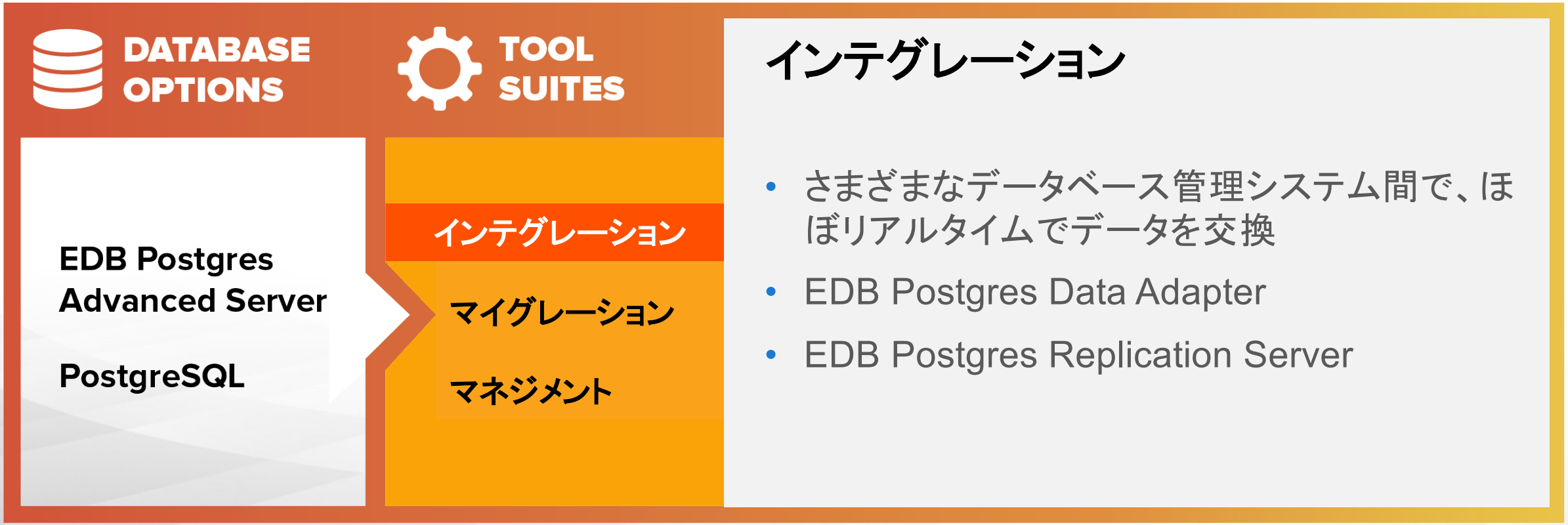
EDB POSTGRES PLATOFORM
POSTGRESQL サーバー

EDB POSTGRES PLATFORM
EDB POSTGRES ADVANCED SERVER

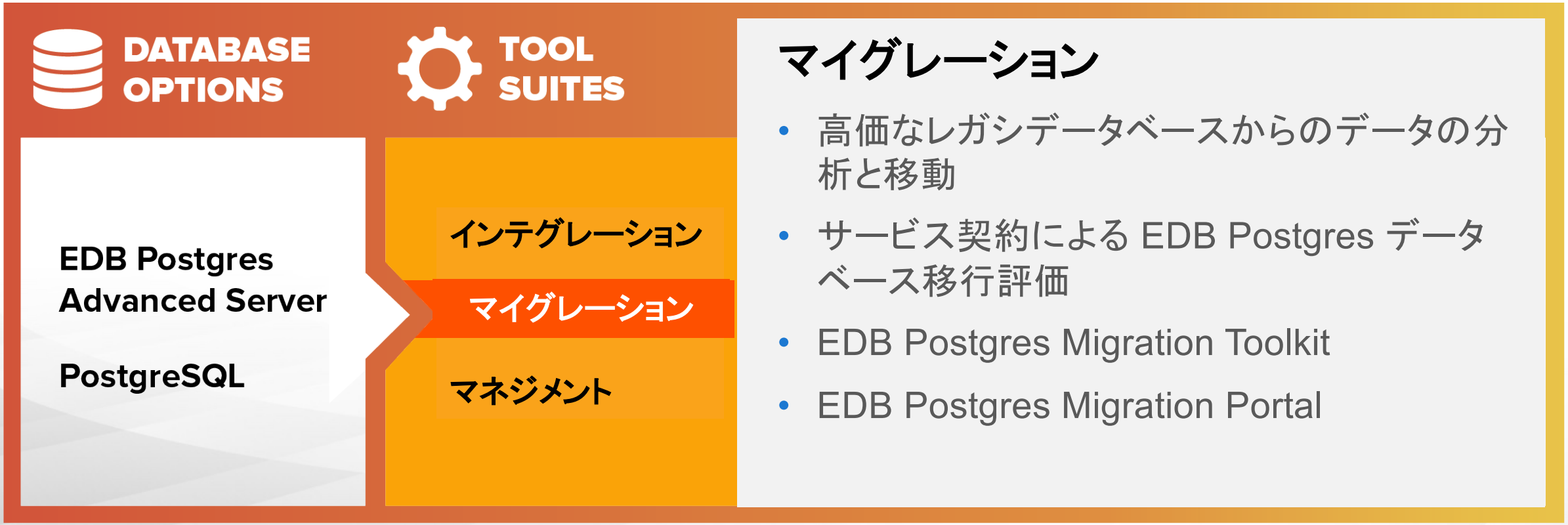
EDB Postgres Advanced Server (EPAS)

- オープンソース PostgreSQL のすべての利点
- パフォーマンス、セキュリティ、および Oracle とのデータベース互換
- 開発者および DBA 向け追加機能

EDB POSTGRES PLATFORM INTEGRATION SUITES



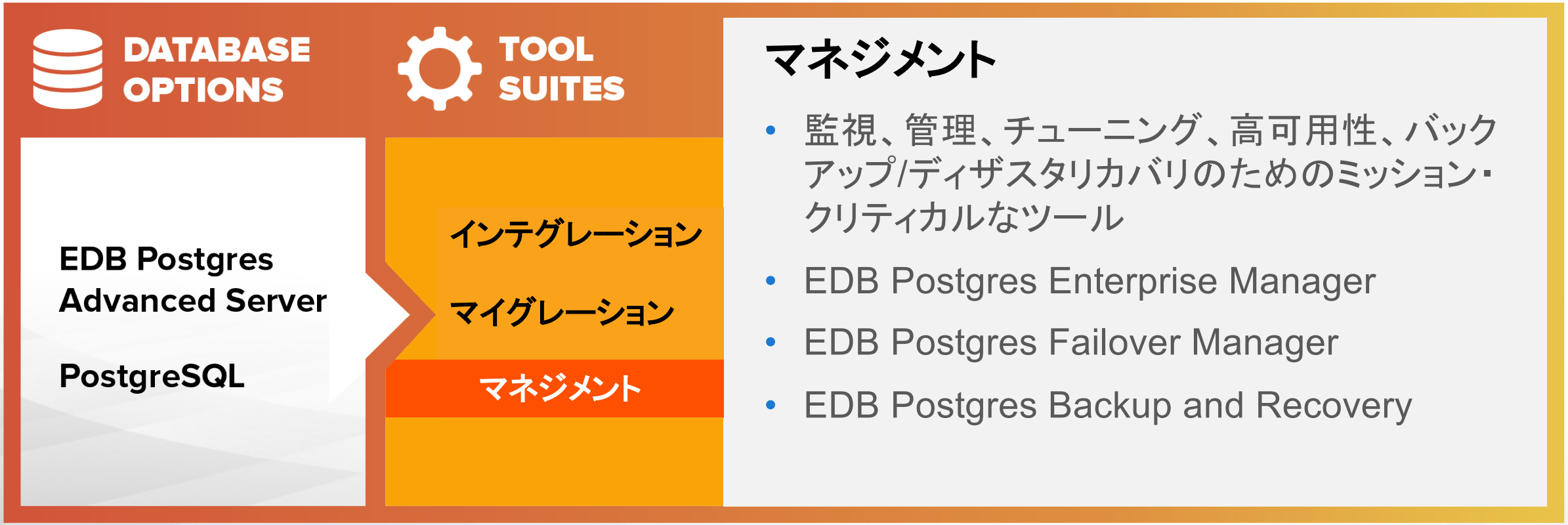
EDB POSTGRES PLATFORM MIGRATION SUITES



マイグレーション

- 高価なレガシデータベースからのデータの分析と移動
- サービス契約による EDB Postgres データベース移行評価
- EDB Postgres Migration Toolkit
- EDB Postgres Migration Portal

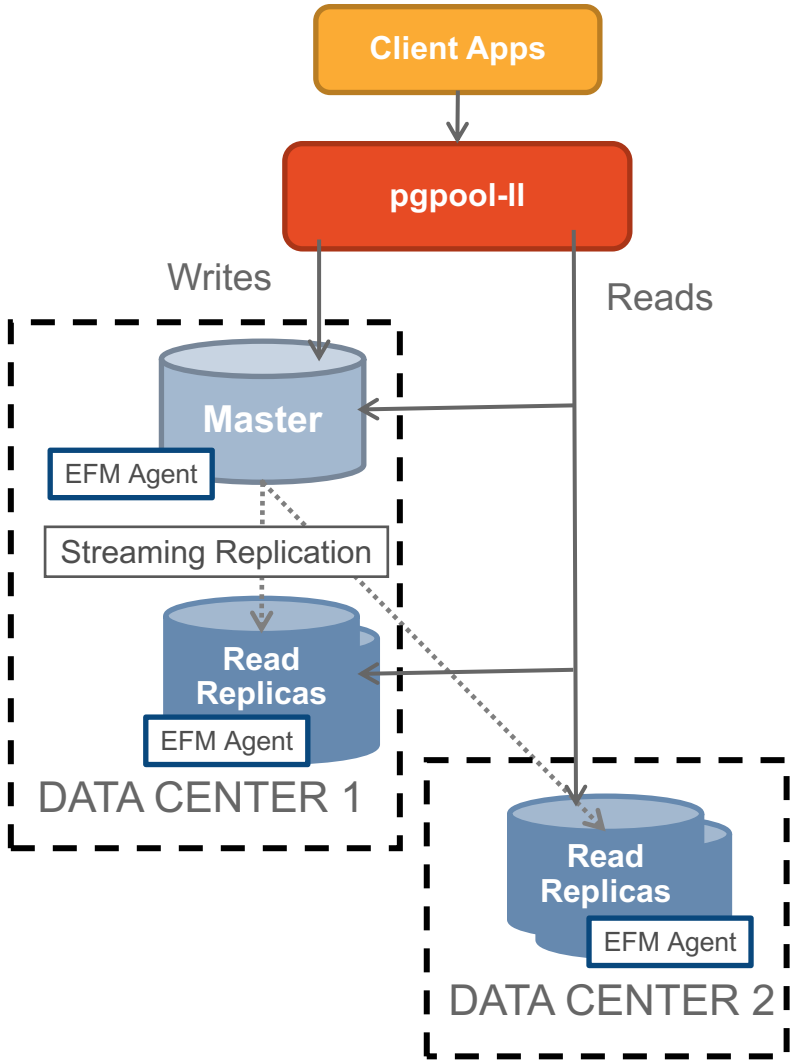
EDB POSTGRES PLATFORM MANAGEMENT SUITES



EDB Postgres Failover Manager (EFM)

EDB Postgres Failover Manager

- 簡単なセットアップ
- エージェント型 + Witnessにより、データベースのSplit-Brainの防止
- 仮想IP(VIP)を提供
- 複数のSlaveサーバの優先度を動的に指定可能
- pgpool-II との連携し、Read型のスケールアウトが可能
- データベースのヘルス・チェック & カスタム・モニタリング機能
- ストリーミング・レプリケーションの同期状態のチェック
- エージェント間のヘルス・チェック
- マスタ障害時自動フェイル・オーバとクラスタ再構成
- 容易なスイッチ・オーバ/スイッチ・バック機能
- Fencingスクリプトによる外部リソースの制御
- クラスタ構成変更時のメール通知とHook処理の実行
- クラスタ構成変更時のメール送信機能



EFM 3.x の修正履歴

リリース	Ver.	修正内容 (抜粋)
2018/2	3.0	• EPAS10/PostgreSQL10対応 • フェィル・オーバ時の、昇格ノード以外の処理の実行 [script.remote.{pre, post}.promotion] • マスタ・ノードが孤立した場合にマスタDBの停止を制御 [stop.isolated.master] • カスタム・モニタリング機能を追加 [script.custom.monitor, custom.monitor.{interval, timeout, mode}]
2018/5	3.1	• PROMOTOコマンドのパラメタ追加 [-sourcenode , -quiet] • 複数VIP対応 [virtualIp.single] • EFMエージェントがマスタDB接続不可の際のマスタDBの停止を制御 [stop.failed.master] • クラスタ構成変更時の efm.nodes の変更回避 [stable.nodes.file] • 通知レベルの制御 [notification.level]
2018/8	3.2	• ロードバランサーやpgpool-II 等との連携 [script.load.balancer.{attach, detach}] • ホスト名のロギング
2018/10	3.3	• EPAS11/PostgreSQL11対応 • フェィル・オーバ前のVIP存在チェックの制御 [check.vip.before.promotion] • Syslogロギングのサポート [syslog.{host, port, protocol, facility, enabled}, file.log.enabled] • ロギング・レベルの変更 (TRACE/ DEBUG/INFO/WARN/ERROR) • VIPプロパティでのホスト名の使用
2019/1	3.4	• マスタ・ノード上のEFMエージェント停止時のフェィル・オーバ [master.shutdown.as.failure] • VIPのNICがbind.address のNICと異なる場合、VIP有無チェックをリトライ

EFM 3.x の修正履歴

リリース	Ver.	修正内容
2019/4	3.5	• script.load.balancer.{attach, detach} での新パラメタ [%t] • サーバ再起動後[restart.connection.timeout] • 通知メール送信者の設定 [from.email] • Recovery.conf 内のaplication_name対応 [application.name] • PROMOTOコマンドのパラメタ追加 [-noscripts]
2019/8	3.6	• 緩やかなメモリ・リークに対するFIX • より多くのスタンバイDB情報を用いた確実な、フェィル・オーバ • レポートでの、Received/Replayed 情報出力 • Debianサポート [db.config.dir]
2019/10	3.7	• EPAS12/PostgreSQL12対応 [db.recovery.conf.dir -> db.recovery.dir]

EDB Postgres Backup and Recovery Tool (BART)

EDB POSTGRES BACKUP AND RECOVERY TOOL (BART)

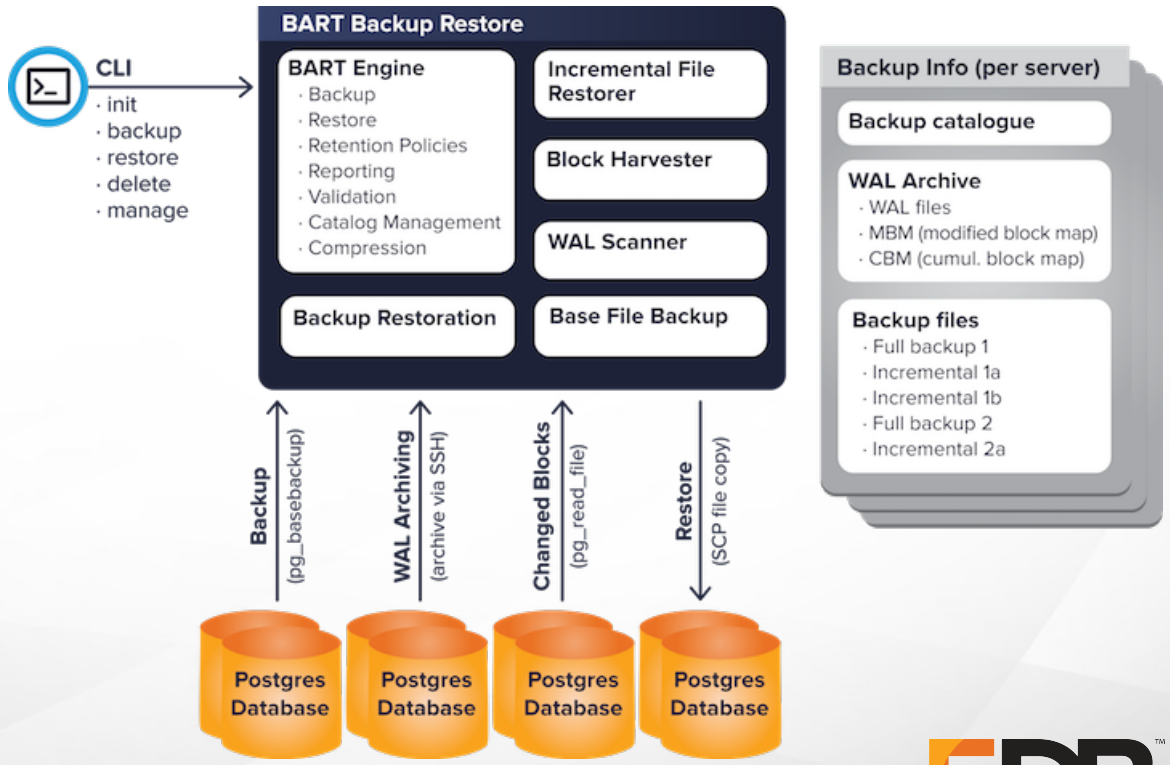
機能

- パラレルでのフル・バックアップ
- ブロックレベルの増分バックアップ
- ファイルの圧縮と検証
- 保存ポリシーに基づくバックアップ・ファイルの管理
- 容易なバックアップ管理とレポート作成
- PITR を含む容易なリストア処理

利点

- 簡単ですぐに使用できるバックアップ・ソリューション
- 安心できる信頼性
- 高速なバックアップ
- 必要ストレージの削減

BART アーキテクチャ



BART 2.x の修正履歴

リリース	Ver.	修正内容
2018/1	2.0	• ブロック・レベル・インクリメンタル・バックアップ • BART CHECK-CONFIG による構成ファイルとDB設定のチェック
2018/2	2.1	• EPAS10/PostgreSQL10対応
2018/11	2.2	• パラレル・フルバックアップ & 圧縮 • ブロック・レベル・インクリメンタル・バックアップのパラレル・リストア • pg_basebackup コマンドのオプション化
2019/2	2.3	• EPAS11/PostgreSQL11対応 • receovery.conf ファイルの常時作成 • 新規パラメタ “mbm-scan-timeout” の導入
2019/6	2.4	• 非アーカイブ 環境のサポート • Ubuntu/Debianサポート
2019/10	2.5	• EPAS12/PostgreSQL12対応

EDB Postgres High Availability Architecture

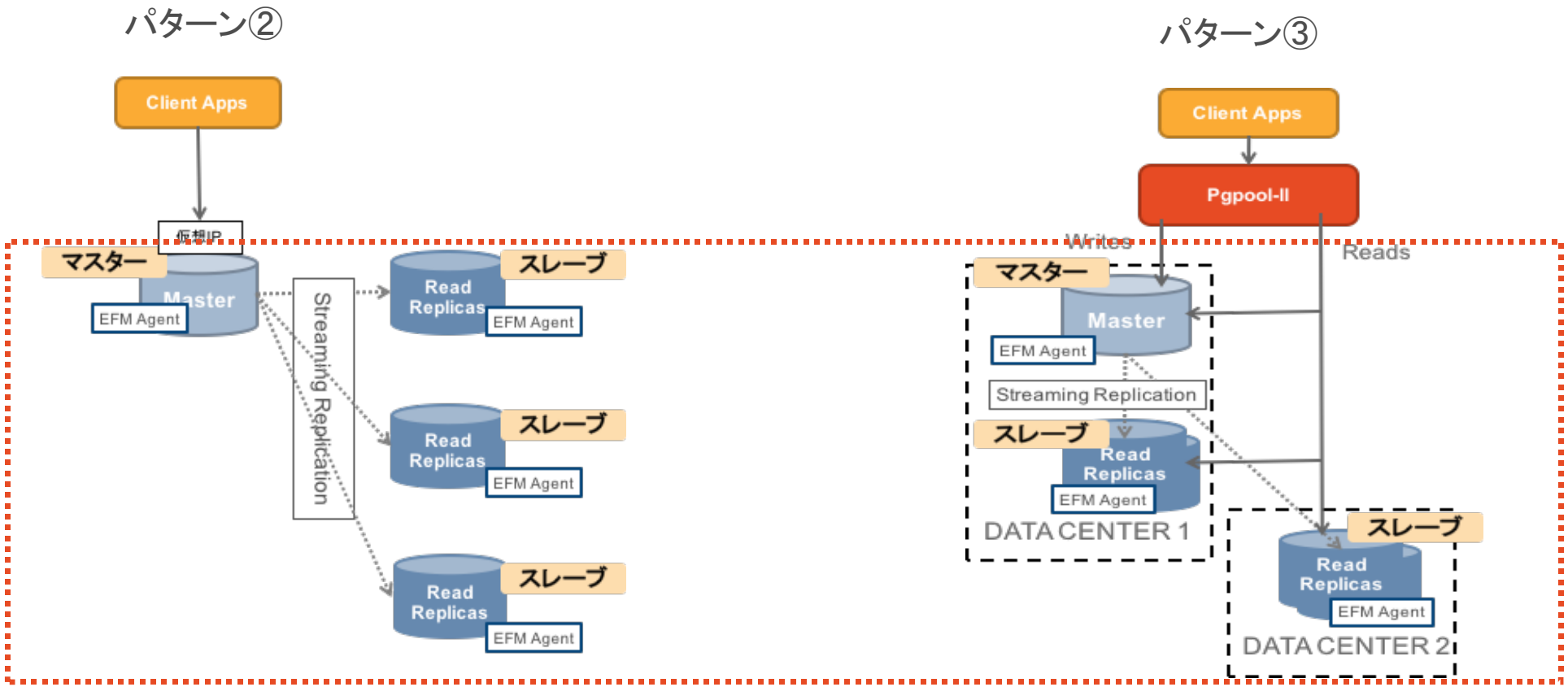
EDB Postgres におけるHA構成パターン

- EPASでは、下記構成が可能
  - ① Pgpool-II による Master – Slave 構成
  - ② EFM による Master – Slave 構成
  - ③ Pgpool-II & EFM による Master – Slave 構成
  - ④ EDB Replication Server による Active –Active 構成
  - ⑤ クラスタ・ソフトによる Active – Passive 構成
- EDBは、pgpool-IIもサポート(EPASインストール時、pgpool-IIもインストールされる)。但し、サポートされる機能は以下の通り。
  - ① ロード・バランシング
  - ② コネクション・プーリング
  - ③ Pgpool-II自体の可用性のための使用、WatchdogプロセスとMaster-Slaveモード

  ＊pgpool-IIの「レプリケーション機能」と「パラレル・クエリ機能」はサポート対象外

Postgres HAクラスタ

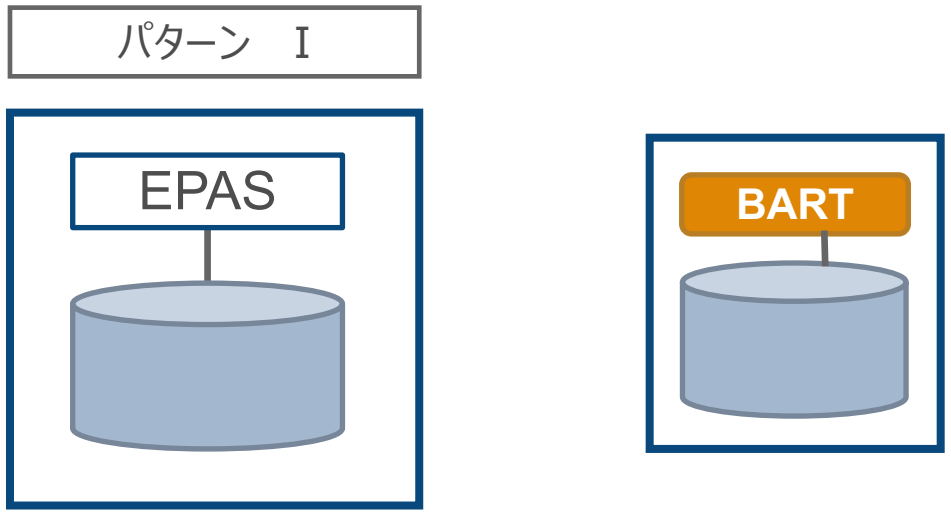
- PosgresHAクラスタのデータ同期方式⇒ マスターDBの複製の方式



High Availability Architecture

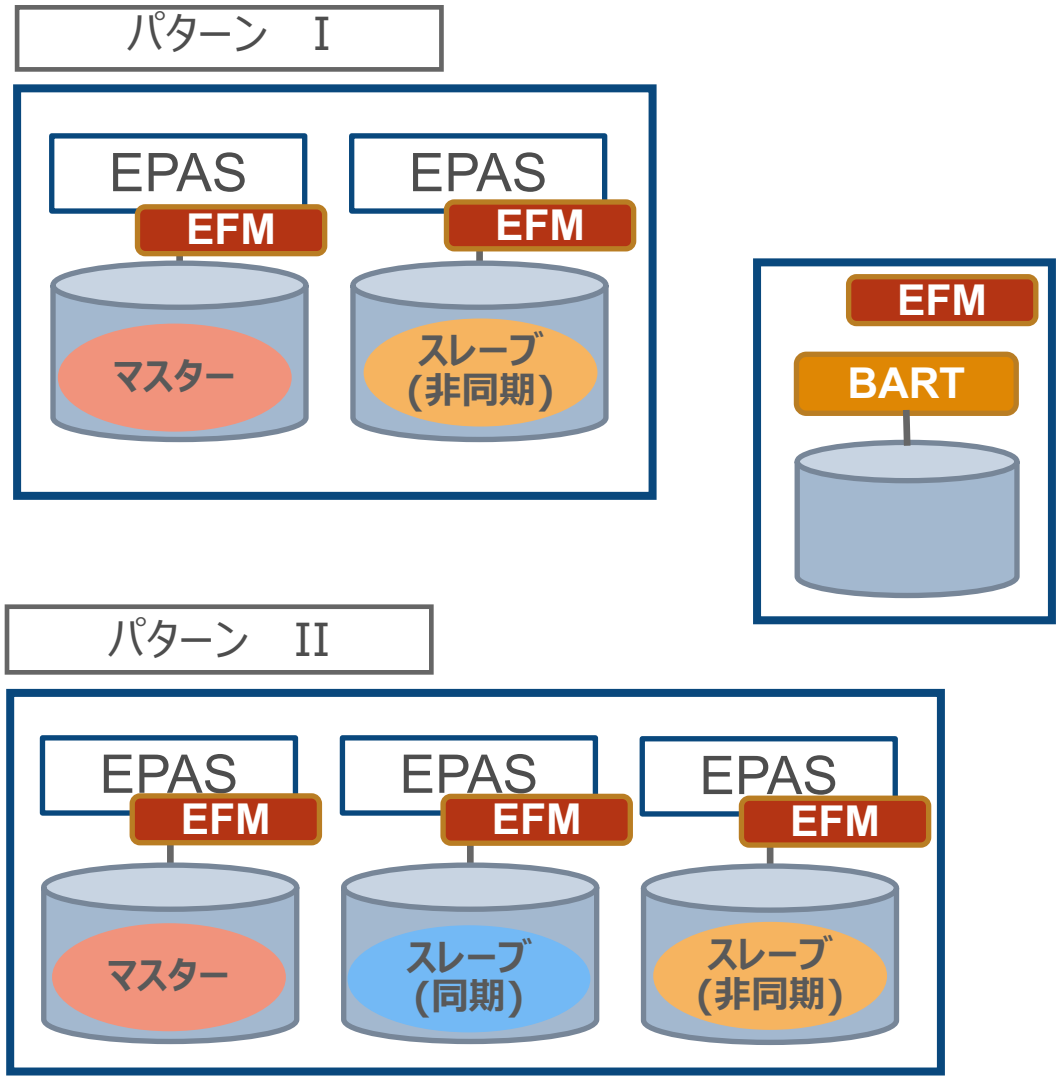
	バックアップ	ローカル・サイトのHA		DR
		Postgres-HAクラスタ	障害の許容台数	
ブロンズ	Yes	No	N/A	No
	BARTを用いたポリシー・ベースのバックアップ			
シルバー	Yes	Yes	1	No
	BARTを用いたポリシー・ベースのバックアップ	EFMを用いた、master-slave構成		
ゴールド	Yes	Yes	2	No
	BARTを用いたポリシー・ベースのバックアップ	EFMを用いた、master-slave構成		
プラチナ	Yes	Yes	>2	Yes
	BARTを用いたポリシー・ベースのバックアップ	EFMを用いた、master-slave構成		リモート・サイトを含めたmaster-slave構成

Bronze/ブロンズ



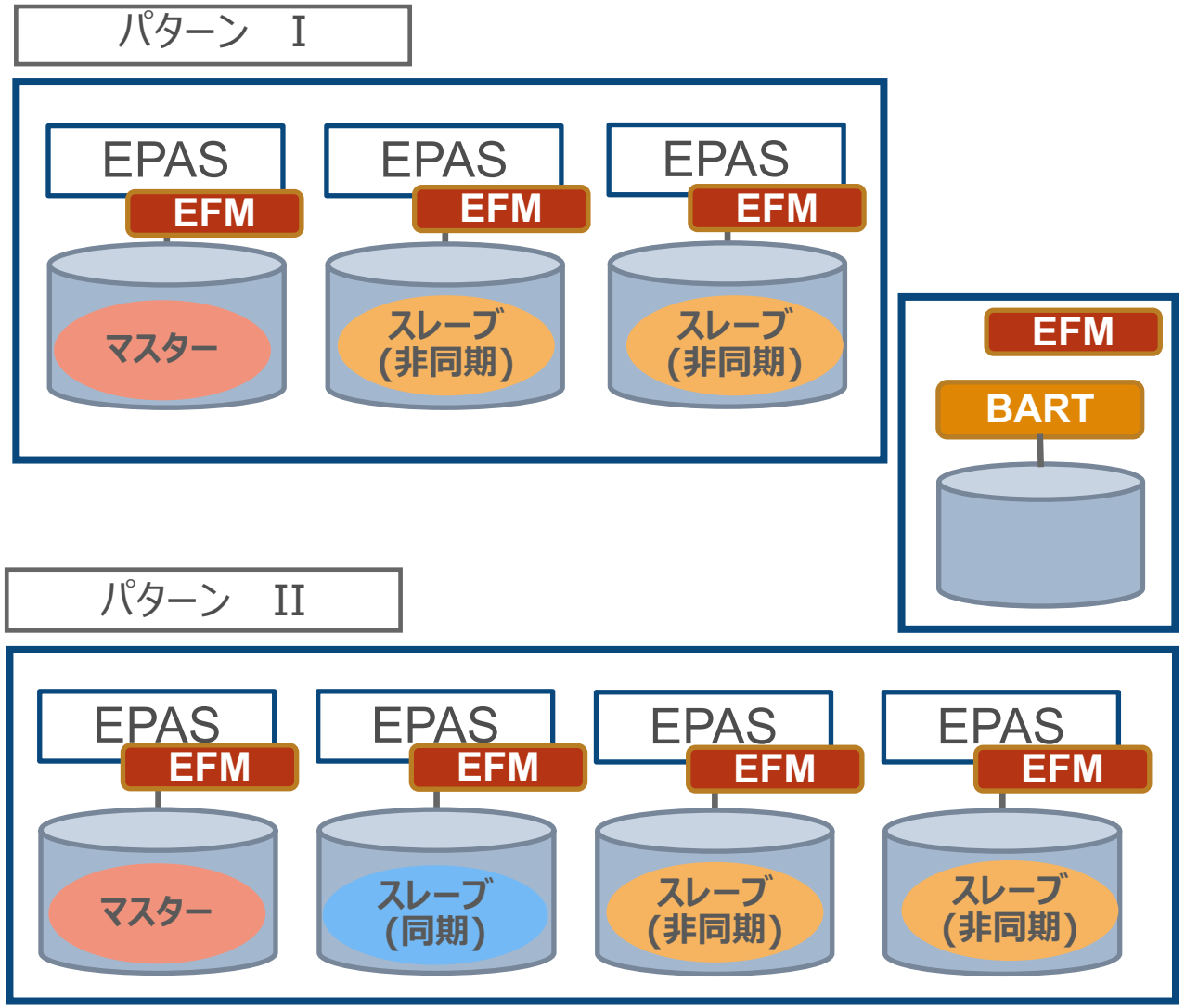
- データベースサーバのHigh Availabilityは、EDB製品等は使用せずにvSphereHAなどで対応
- 2世代以上の保持ポリシーで、BARTを用いたデータベースのバックアップを、定期的に取得
- データベース部分の障害が発生した場合は、バックアップを用いてリカバリ。リストア中は、サービスを停止する必要がある。

Silver/シルバー



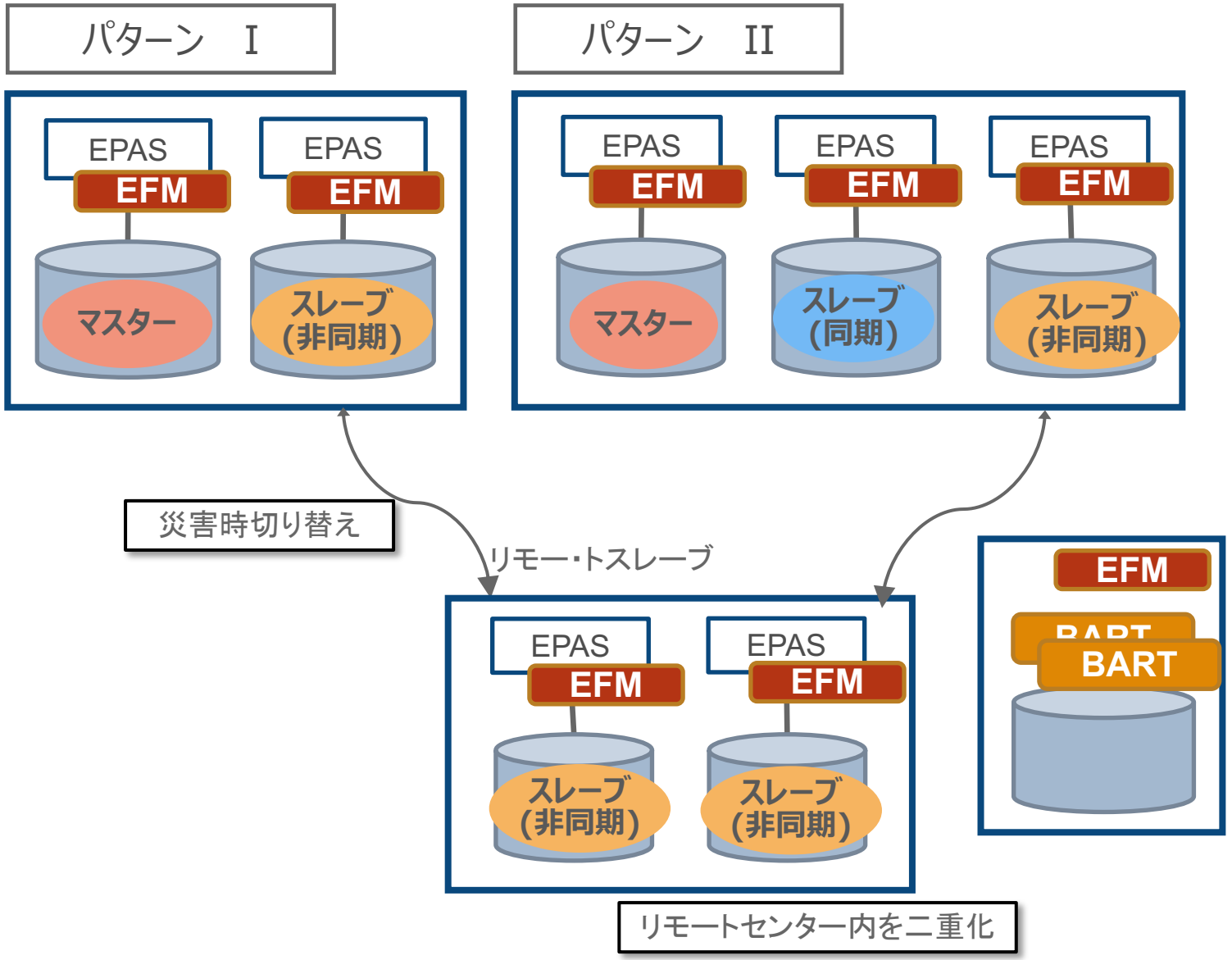
- マスター・スレーブ型アーキテクチャ
- 1台のDB障害に対して、サービス継続が可能
- Pgpool-II との併用により、以下が可能
  - ロード・バランシング機能によるREAD型のスケール・アウト対応
  - コネクション・プーリングによる大規模コネクション対応
- バック・アップをスレーブから取得することで、マスター側の負荷を軽減
- パターン1
  - 非同期レプリケーションによる2ノード構成
  - マスターが障害の場合、スレーブ(非同期)が昇格
  - データ・ロスの可能性はあるが、レプリケーションによるマスター側のオーバヘッドは軽微
- パターン2
  - 同期レプリケーションによる3ノード構成
  - マスターが障害の場合、スレーブ(同期)が昇格
  - データ・ロスはないが、レプリケーションによるマスタ側のオーバヘッドを考慮する必要がある

Gold/ゴールド



- マスター・スレーブ型アーキテクチャ
- 2台までのDB障害に対して、サービス継続が可能
- Pgpool-II との併用により、以下が可能
  - ロード・バランシング機能によるREAD型のスケール・アウト対応
  - コネクション・プーリングによる大規模コネクション対応
- バック・アップをスレーブから取得することで、マスター側の負荷を軽減
- パターン1
  - 非同期レプリケーションによる3ノード構成
  - マスターが障害の場合、スレーブ(非同期)が昇格
  - データ・ロスの可能性はあるが、レプリケーションによるマスター側のオーバヘッドは軽微
- パターン2
  - 同期レプリケーションによる4ノード構成
  - マスターが障害の場合、スレーブ(同期)が昇格
  - データ・ロスはないが、レプリケーションによるマスタ側のオーバヘッドを考慮する必要がある

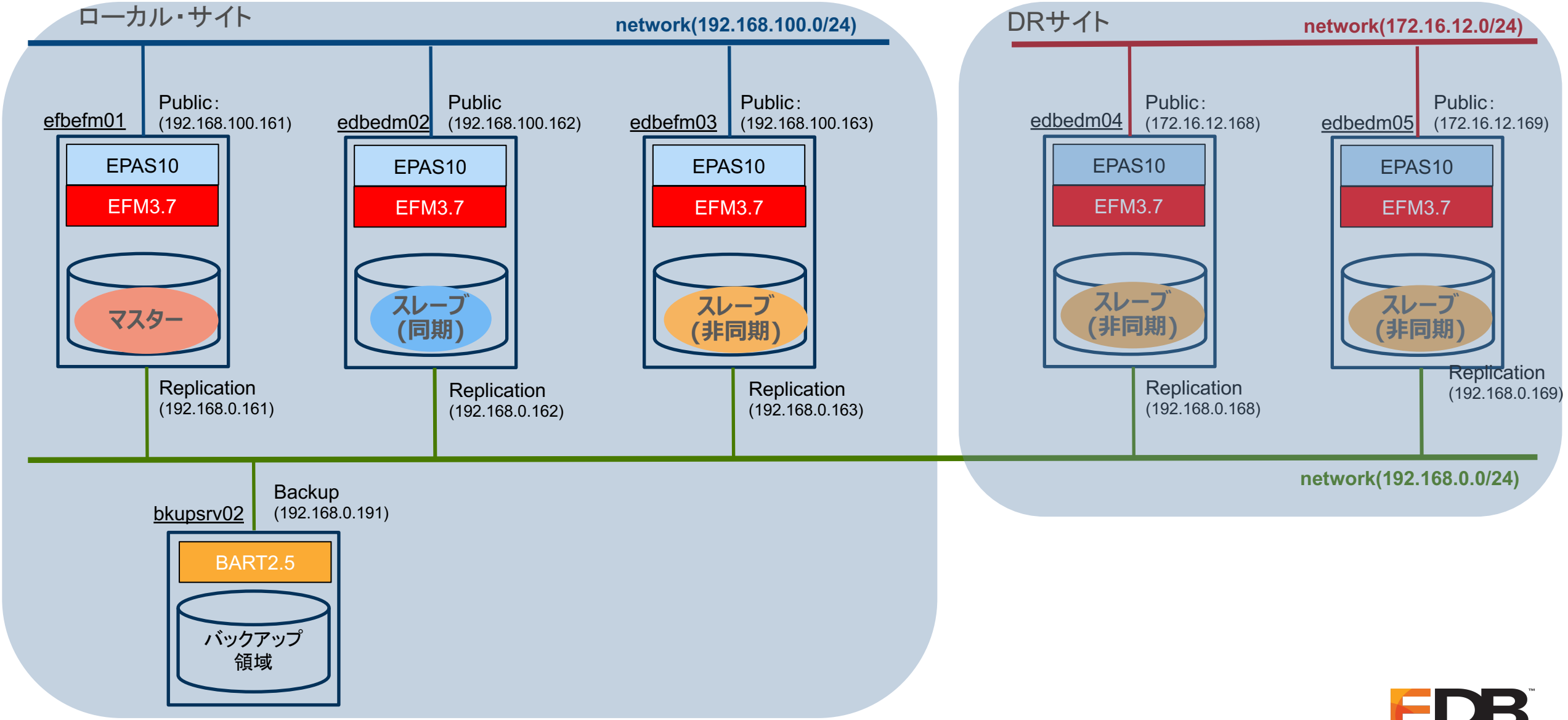
Platinum/プラチナ



- DR対応として、リモート・サイトにスレーブDBを配置
- ローカル・サイトのDBサーバが全て停止した場合に、リモート・サイトのDBが昇格し、新マスターDBとしてサービスを継続
- ローカル・サイトは、要件により、シルバー、ゴールドを選択可能

プラチナ構成(例)

システム構成



EFM3.7パラメタ

カテゴリ	パラメタ	efm01	efm02	efm03	efm04	efm05
データベース関連	db.user	enterprisedb	←	←	←	←
	db.password.encrypted	暗号化されたパスワード	←	←	←	←
	db.port	5444	←	←	←	←
	db.database	edb	←	←	←	←
	db.service.owner	enterprisedb	←	←	←	←
	db.service.name	edb-as-10	←	←	←	←
	db.bin	/opt/edb/as10/bin	←	←	←	←
	db.recovery.dir	/opt/edb/as10/data	←	←	←	←
	db.config.dir		←	←	←	←
	application.name	efm01	efm02	efm03	efm04	efm05
エージェント基本設定	bind.address	192.168.0.161:7801	192.168.0.162:7801	192.168.0.163:7801	192.168.0.168:7801	192.168.0.169:7801
	admin.port	7802	←	←	←	←
	is.witness	FALSE	←	←	←	←
ノード管理	auto.allow.hosts	FALSE	←	←	←	←
	stable.nodes.file	FALSE	←	←	←	←
モニタリング関連	local.period	10	←	←	←	←
	local.timeout	60	←	←	←	←
	local.timeout.final	10	←	←	←	←
	remote.timeout	10	←	←	←	←
	node.timeout	50	←	←	←	←
	pingServerIp	192.168.0.1	←	←	←	←
	pingServerCommand	/bin/ping -q -c3 -w5	←	←	←	←
	db.reuse.connection.count	0	←	←	←	←
	auto.resume.period	10	←	←	←	←
	recovery.check.period	2	←	←	←	←
	restart.connection.timeout	60	←	←	←	←
	check.vip.before.promotion	TRUE	←	←	←	←

EFM3.7パラメタ (続き)

カテゴリ	パラメタ	efm01	efm02	efm03	efm04	efm05
カスタム・モニタリング	script.custom.monitor		←	←	←	←
	custom.monitor.interval		←	←	←	←
	custom.monitor.timeout		←	←	←	←
	custom.monitor.safe.mode	FALSE	←	←	←	←
クラスタ制御	auto.failover	TRUE	←	←	←	←
	auto.reconfigure	TRUE	←	←	←	←
	promotable	TRUE	←	←	←	←
	minimum.standbys		←	←	←	←
	stop.isolated.master	TRUE	←	←	←	←
	stop.failed.master	TRUE	←	←	←	←
	master.shutdown.as.failure	TRUE	←	←	←	←
仮想IP	virtualIp	192.168.100.160	←	←	172.16.12.160	←
	virtualIp.interface	ens34	←	←	ens33	←
	virtualIp.prefix	24	←	←	24	←
	virtualIp.single	FALSE	←	←	←	←
pgpool/LB連携	script.load.balancer.attach		←	←	←	←
	script.load.balancer.detach		←	←	←	←
Hook処理	script.fence		←	←	←	←
	script.post.promotion		←	←	←	←
	script.resumed	/var/efm/efmResumedScript.bash	←	←	←	←
	script.db.failure		←	←	←	←
	script.master.isolated		←	←	←	←
	script.remote.pre.promotion		←	←	←	←
	script.remote.post.promotion		←	←	←	←
通知	user.email	katsuji.takatsuru@enterprisedb.com	←	←	←	←
	from.email		←	←	←	←
	notification.level	INFO	←	←	←	←
	script.notification		←	←	←	←

EFM3.7パラメタ (続き)

カテゴリ	パラメタ	efm01	efm02	efm03	efm04	efm05
ログ関連	log.dir		←	←	←	←
	syslog.host	localhost	←	←	←	←
	syslog.port	514	←	←	←	←
	syslog.protocol	UDP	←	←	←	←
	syslog.facility	LOCAL1	←	←	←	←
	file.log.enabled	TRUE	←	←	←	←
	syslog.enabled	FALSE	←	←	←	←
	jgroups.loglevel	DEBUG	←	←	←	←
	efm.loglevel	DEBUG	←	←	←	←
その他	jdbc.sslmode		←	←	←	←
	sudo.command	sudo	←	←	←	←
	sudo.user.command	sudo -u %u	←	←	←	←
	lock.dir		←	←	←	←
	jvm.options	-Xmx128m	←	←	←	←

EFM & 同期ストリーミング・レプリケーション

- /var/efm/efmResumedScript.bash(例)

```
#!/bin/bash
#################################################################################################################
##              Resumed Script on EFM
#################################################################################################################
#################################################################################################################
# 変数設定
#################################################################################################################
EDBHOME=/opt/edb/as10
PGDATA=/opt/edb/as10/data
WAL_ARCHIVE=/pgsql_bkup/archives
PSQL=$EDBHOME/bin/psql
DBHOST=192.168.0.161                    ## Local Private IP Address
DBPORT=5444
DBNAME=edb
DBUSER=enterprisedb
DBPASSWD=******
LOGGERTAG=efm-3.7                       ## depend on the version of EFM
today=`date +%Y%m%d-%H%M%S`
OLDMASTER=$2
NEWMASTER=$1

#################################################################################################################
# 初期処理
#################################################################################################################
logger -i -s -t $LOGGERTAG 'EFM Resumed Start!'
MSG="OLD Master: ${OLDMASTER} => NEW Master: ${NEWMASTER}"
logger -i -s -t $LOGGERTAG $MSG
#SYNCSLAVE=`/usr/edb/efm-3.7/bin/efm cluster-status efm|grep -A 1 'Standby priority host list' | grep 192.168.0 | awk '{print $2}'`
MSG=`/usr/edb/efm-3.7/bin/efm cluster-status efm|grep -A 1 'Standby priority host list' | grep 192.168.0`
logger -i -s -t $LOGGERTAG $MSG
/usr/edb/efm-3.7/bin/efm set-priority efm 192.168.0.162 1
/usr/edb/efm-3.7/bin/efm set-priority efm 192.168.0.161 2
/usr/edb/efm-3.7/bin/efm set-priority efm 192.168.0.163 3
/usr/edb/efm-3.7/bin/efm set-priority efm 192.168.0.168 4
/usr/edb/efm-3.7/bin/efm set-priority efm 192.168.0.169 5
MSG=`/usr/edb/efm-3.7/bin/efm cluster-status efm|grep -A 1 'Standby priority host list' | grep 192.168.0`
logger -i -s -t $LOGGERTAG $MSG
logger -i -s -t $LOGGERTAG 'EFM Resumed End!'
exit 0;
```

データベース・パラメタ

パラメタ	efm01	efm02	efm03	efm04	efm05
synchronous_standby_names	1(efm01,efm02,efm03)	1(efm01,efm02,efm03)	1(efm01,efm02,efm03)	(NULL)	(NULL)
synchronous_commit	remote_write	remote_write	remote_write	(NULL)	(NULL)

- DRサイトのデータベースのpostgresql.confでは、同期レプリケーション関連の設定は行わない。復旧後、サイト間で同期レプリケーションが行われないようにするため。

- recovery.conf for edbefm02

```
primary_conninfo = 'host=192.168.0.161 port=5444 user=replicator password=mewtec application_name="efm02"
trigger_file='/tmp/postgresql.trigger'
recovery_target_timeline = 'latest'
```

クラスタ状況 (正常状態)

```
[root@edbefm01 ~]# /usr/edb/efm-3.7/bin/efm cluster-status efm
Cluster Status: efm

        Agent Type  Address              Agent  DB       VIP
        ------------------------------------------------------------------------
        Master      192.168.0.161        UP     UP       192.168.100.160*
        Standby     192.168.0.162        UP     UP       192.168.100.160
        Standby     192.168.0.163        UP     UP       192.168.100.160
        Standby     192.168.0.168        UP     UP       172.16.12.160
        Standby     192.168.0.169        UP     UP       172.16.12.160

Allowed node host list:
        192.168.0.161 192.168.0.191 192.168.0.162 192.168.0.163 192.168.0.168 192.168.0.169

Membership coordinator: 192.168.0.168

Standby priority host list:
        192.168.0.162 192.168.0.163 192.168.0.168 192.168.0.169

Promote Status:

        DB Type     Address              WAL Received LSN   WAL Replayed LSN   Info
        ------------------------------------------------------------------------
        Master      192.168.0.161                          6/1E0
        Standby     192.168.0.169        6/1E0             6/1E0
        Standby     192.168.0.163        6/1E0             6/1E0
        Standby     192.168.0.162        6/1E0             6/1E0
        Standby     192.168.0.168        6/1E0             6/1E0

 Standby database(s) in sync with master. It is safe to promote.
```

```
edb=# ¥! hostname
edbefm01.enterprisedb.com
edb=# select client_addr , application_name , state, sync_priority , sync_state from pg_stat_replication
order by client_addr;
  client_addr  | application_name |   state   | sync_priority | sync_state
---------------+------------------+-----------+---------------+------------
 192.168.0.162 | efm02            | streaming |             2 | sync
 192.168.0.163 | efm03            | streaming |             3 | potential
 192.168.0.168 | efm04            | streaming |             0 | async
 192.168.0.169 | efm05            | streaming |             0 | async
```

クラスタ状況 (efm01のマスタDBダウン時)

```
[root@edbefm01 ~]# /usr/edb/efm-3.7/bin/efm cluster-status efm
Cluster Status: efm

        Agent Type  Address              Agent  DB       VIP
        ------------------------------------------------------------------------
        Idle        192.168.0.161        UP     UNKNOWN  192.168.100.160
        Master      192.168.0.162        UP     UP       192.168.100.160*
        Standby     192.168.0.163        UP     UP       192.168.100.160
        Standby     192.168.0.168        UP     UP       172.16.12.160
        Standby     192.168.0.169        UP     UP       172.16.12.160

Allowed node host list:
        192.168.0.161 192.168.0.191 192.168.0.162 192.168.0.163 192.168.0.168 192.168.0.169

Membership coordinator: 192.168.0.168

Standby priority host list:
        192.168.0.163 192.168.0.168 192.168.0.169

Promote Status:

        DB Type     Address              WAL Received LSN   WAL Replayed LSN   Info
        ------------------------------------------------------------------------
        Master      192.168.0.162                          6/10001A8
        Standby     192.168.0.169        6/10001A8         6/10001A8
        Standby     192.168.0.168        6/10001A8         6/10001A8
        Standby     192.168.0.163        6/1000000         6/10001A8

        Standby database(s) in sync with master. It is safe to promote.

Idle Node Status (idle nodes ignored in WAL LSN comparisons):

        Address              WAL Received LSN   WAL Replayed LSN   Info
        ------------------------------------------------------------------------
        192.168.0.161        UNKNOWN            UNKNOWN            192.168.0.161:5444 への接続が拒絶されました。ホスト名とポート番号が正しいことと、postmaster がTCP/IP接続を受け付けていることを確認してください


edb=# ¥! hostname
edbefm02.enterprisedb.com
edb=# select client_addr , application_name , state, sync_priority , sync_state from pg_stat_replication order by client_addr;
  client_addr  | application_name |   state   | sync_priority | sync_state
---------------+------------------+-----------+---------------+------------
 192.168.0.163 | efm03            | streaming |             3 | sync
 192.168.0.168 | efm04            | streaming |             0 | async
 192.168.0.169 | efm05            | streaming |             0 | async
(3 行)
```

クラスタ状況 (ローカル・サイトの全DBダウン時)

```
[root@edbefm05 ~]# /usr/edb/efm-3.7/bin/efm cluster-status efm
Cluster Status: efm

        Agent Type  Address              Agent  DB       VIP
        -----------------------------------------------------------------------
        Idle        192.168.0.161        UP     UNKNOWN  192.168.100.160
        Idle        192.168.0.162        UP     UNKNOWN  192.168.100.160
        Idle        192.168.0.163        UP     UNKNOWN  192.168.100.160
        Master      192.168.0.168        UP     UP       172.16.12.160*
        Standby     192.168.0.169        UP     UP       172.16.12.160

Allowed node host list:
        192.168.0.161 192.168.0.191 192.168.0.162 192.168.0.163 192.168.0.168 192.168.0.169

Membership coordinator: 192.168.0.168

Standby priority host list:
        192.168.0.169

Promote Status:

        DB Type     Address              WAL Received LSN   WAL Replayed LSN   Info
        ---------------------------------------------------------------------------
        Master      192.168.0.168                          6/40001A8
        Standby     192.168.0.169        6/40001A8         6/40001A8

        Standby database(s) in sync with master. It is safe to promote.

Idle Node Status (idle nodes ignored in WAL LSN comparisons):

        Address              WAL Received LSN   WAL Replayed LSN   Info
        ---------------------------------------------------------------
        192.168.0.163        UNKNOWN            UNKNOWN            192.168.0.163:5444 への接続が拒絶されました。ホスト名とポート番号が正しいことと、postmaster がTCP/IP接続を受け付けていることを確認してください。
        192.168.0.161        UNKNOWN            UNKNOWN            192.168.0.161:5444 への接続が拒絶されました。ホスト名とポート番号が正しいことと、postmaster がTCP/IP接続を受け付けていることを確認してください。
        192.168.0.162        UNKNOWN            UNKNOWN            192.168.0.162:5444 への接続が拒絶されました。ホスト名とポート番号が正しいことと、postmaster がTCP/IP接続を受け付けていることを確認してください

edb=# ¥! hostname
edbefm04.enterprisedb.com
edb=# select client_addr , application_name , state, sync_priority , sync_state from pg_stat_replication order by client_addr;
  client_addr  | application_name |   state   | sync_priority | sync_state
---------------+------------------+-----------+---------------+------------
 192.168.0.169 | efm05            | streaming |             0 | async
(1 行)
```