

HPCI共用ストレージ: 更新後

東大またはAICSによる単独運用可能な高いサービス継続性

遠隔地間を含むメタデータフェイルオーバーによる高可用性

クラウドストレージの活用によるアーカイブサービス

Cloud Storage

1. メタデータを参照

2. データを書き込み

3. メタデータに登録

Meta Data Storage

Meta Data Storage

DataStorage
50 PB以上

理研計算科学研究機構

SINET5による広域・高帯域データ通信

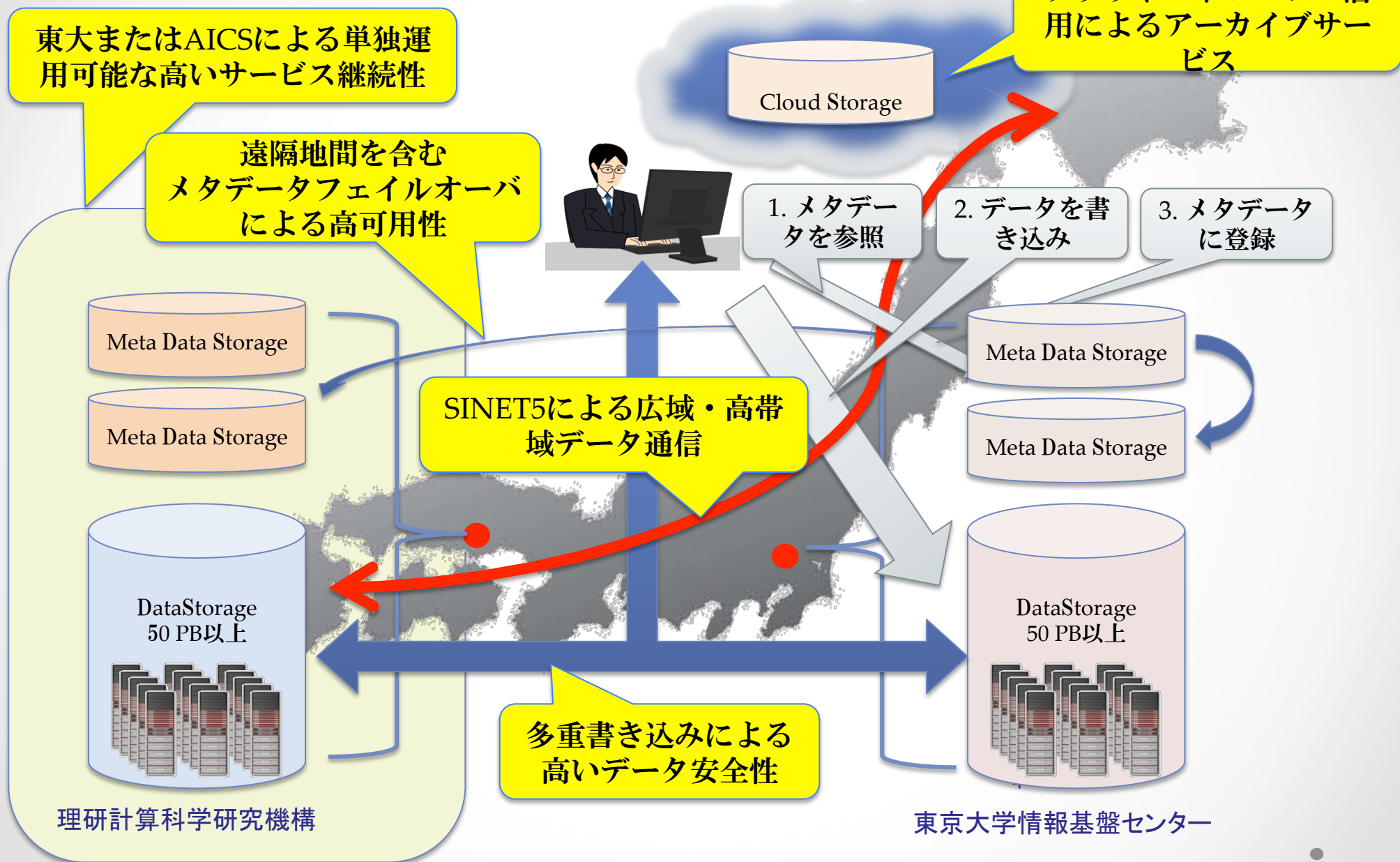
多重書き込みによる高いデータ安全性

Meta Data Storage

Meta Data Storage

DataStorage
50 PB以上

東京大学情報基盤センター



1. データ保護・安全性
2. サービス継続性
3. サービスメニューの多様化

過去の教訓・反省を踏まえつつ、、、
HPCI共用ストレージはどうあるべきか（目標
設定）
具体的にどうすべきか

更新後のサービスレベルと施策

データ保護・安全性

年間ファイル消失ゼロ

- データ消失事案の反省
 - 運用ガイドライン、障害対応指針の策定
 - 2015年4月末～、全データファイルの digest 照合
 - 更新ファイルのdigest 定期照合
 - digest不整合ファイルの検出、~2週間以内にユーザ通知体制
 - データ完全性チェックの自動化活用
 - データ破損・消失すると再計算資源の提供が大きな課題
 - 特にHPCIの場合は、他の資源提供機関の計算機が必要になる場合もある
- データ多重化・分散化
 - データ完全二重化導入
 - 現在は容量不足により、データ多重化はユーザ設定
 - AICS・東大でそれぞれ複製を一つ以上保持
 - Gfarmによるデータ完全性チェックの自動化（実現済み）
 - データ破損が発生しても迅速なユーザ通知を実現
 - 導入機材の単一障害点排除

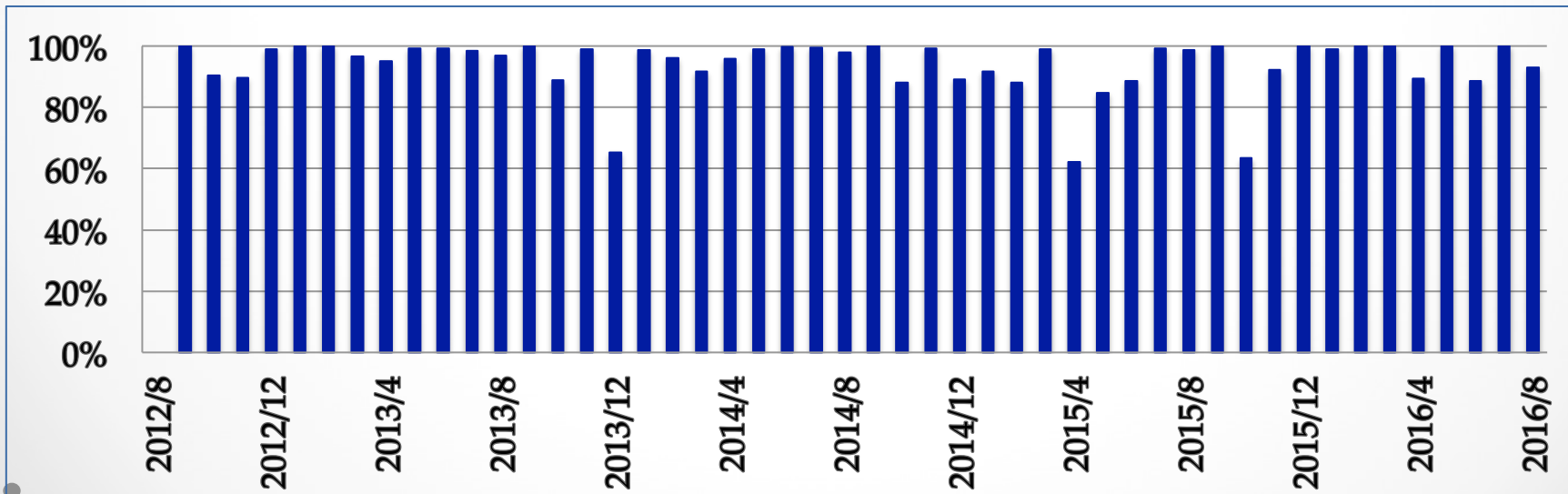
更新後のサービスレベルと施策

サービスの継続性・安定性

年間350日以上サービス提供

- データ消失事案の発生した2015年度は特に低稼働率
- 設備点検などの全館停電による長期サービス停止
- 東大・AICS片肺運転可能なシステム導入
 - 片肺運転中も全ファイルシステムにアクセス可能
 - 設備点検など、1機関全停止しても継続運転可能なシステム
- 可用性向上
 - メタデータフェイルオーバー（東大、AICS間フェイルオーバー）
 - データサーバフェイルオーバー
 - ストレージコントローラフェイルオーバー（導入済み）
 - データサーバ用ストレージはマルチパス構成
- 保守サービス時間の夜間・休日対応（東大が来年度試験導入）
- ネットワーク冗長化
 - 構成機関設置機材の基幹スイッチ・トポロジーの多重化

データサーバの稼働率



東大、AICSサイト間のデータ冗長化運用によるサービス継続性改善予測

- 2012年9月から2016年8月までの全サービス停止事案を調査
 - システム全停止だけでなく、部分停止障害も含む
 - 全メンテナンスも調査
 - 停止時間(何時から何時まで)、原因、停止台数、停止ノード時間積、検証のための記録、、、

月	発生時刻	終了時刻	停止概要	エビデンス	分類	メンテナンス停止時間	障害停止時間	停止台数	定期メンテ停止時間(ノードh)	障害・緊急メンテ停
	2013/4/1 0:00	2014/4/1 0:00	2013年度運用期間			8760				
4	2013/4/10 18:00	2013/4/11 15:00	XFS障害	[hpci-support:1124] Re: [esci-storage-alert:0858] Re: oss29	障害		21	1		
4	2013/4/16 10:00	2013/4/16 12:00	証明書更新の作業ミス	Re: [hpci-support:1174] R	障害		2	ALL		
4	2013/4/25 9:00	2013/4/26 17:00	定期メンテナンス(空調工事含む)	運用会議#2 UT Report	定期メンテ	32		ALL	1728	
5	2013/5/27 9:00	2013/5/27 15:00	定期メンテナンス	運用会議#3 UT Report	定期メンテ	6		ALL	324	
6	2013/6/24 9:00	2013/6/24 15:00	定期メンテナンス	運用会議#4 UT Report	定期メンテ	6		ALL	324	
7	2013/7/24 14:00	2013/7/24 17:01	gfmd ダウン	[hpci-support:2250] gfmdがダウンしました	ソフト障害		3	ALL		
7	2013/7/29 9:00	2013/7/29 18:00	定期メンテナンス	運用会議#5 UT Report	定期メンテ	9		ALL	486	
8	2013/8/14 13:00	2013/8/15 16:00	東大 富士通のラッシュカレントによるUPS過負荷により1Fの全減		障害		27	28		
8	2013/8/19 9:00	2013/8/20 12:00	DOS攻撃	[hpci-support:2434] Re: ネットワーク疎通	障害		27	4		
8	2013/8/26 9:00	2013/8/26 15:00	定期メンテナンス	運用会議#5 UT Report	定期メンテ	6		ALL	324	
8	2013/8/30 5:45	2013/8/30 19:45	AICS wgfs008 DDNが再起動	[hpci-support:2588]	ハード障害		14	1		
9	2013/9/30 23:00	2013/10/1 13:00	oss27 down	[hpci-support:3018] Re: hpcieast-oss27 ダウン	ハード障害		14	1		
10	2013/10/4 10:00	2013/10/7 13:00	定期メンテナンス(柏停電)	運用会議#6 UT Report	定期メンテ	75		ALL	4050	
10	2013/10/28 9:00	2013/10/28 17:00	定期メンテナンス	運用会議#7 UT Report	定期メンテ	8		ALL	432	
11	2013/11/1 14:00	2013/11/1 15:01	oss21 ネットエラーでリポート	[hpci-support:4873] Re: oss	ハード障害		1	1		
11	2013/11/29 9:00	2013/11/29 17:00	定期メンテナンス	運用会議#8 UT Report	定期メンテ	8		ALL	432	
12	2013/12/16 8:00	2013/12/16 10:00	epgfd101 サーマルエラー	[hpci-support:5492] esci-epgfd101再起動	空調		2	1		
12	2013/12/23 9:00	2013/12/23 10:10	epgfd101 サーマルエラー	[hpci-support:5567] esci-epgfd101.gfsd再起動	空調		1	1		
12	2013/12/27 12:00	2014/1/6 12:00	定期メンテナンス	運用会議#8 UT Report	定期メンテ	240		ALL	12960	
1	2014/1/27 9:00	2014/1/27 14:00	定期メンテナンス	運用会議#9 UT Report	定期メンテ	5		ALL	270	
2	2014/2/9 17:00	2014/2/10 9:00	ソフト障害	[hpci-support:6111] 2014/2/9 17:10 東拠点メタデータサーバ再起動	ソフト障害		16	ALL		
2	2014/2/24 9:00	2014/2/24 18:00	定期メンテナンス	運用会議#10 UT Report	定期メンテ	9		ALL	486	
3	2014/3/25 9:00	2014/3/27 17:00	定期メンテナンス	運用会議#10 UT Report	定期メンテ	56		ALL	3024	
12	2013/12/15 22:11	2013/12/17 14:34	wgfs002 リポート障害	システムログから抽出※2 / [hpci-support:5491] esci-wgfs002:再起動しています	ソフト障害		40	1		
12	2013/12/16 22:16	2013/12/17 14:38	wgfs007 リポート障害	システムログから抽出※2 / [hpci-support:5491] esci-wgfs002:再起動しています	ソフト障害		16	1		
12	2013/12/22 8:00	2013/12/24 8:32	wgfs014 リポート障害	システムログから抽出※2	ソフト障害		49	1		
12	2013/12/23 20:40	2013/12/24 11:58	wgfs006 リポート障害	システムログから抽出※2	ソフト障害		15	1		
12	2013/12/23 20:40	2013/12/24 11:57	wgfs007 リポート障害	システムログから抽出※2	ソフト障害		15	1		
12	2013/12/25 14:28	2014/1/6 9:52	wgfs006 リポート障害	システムログから抽出※2	ソフト障害		283	1		
12	2013/12/25 14:28	2014/1/6 9:51	wgfs007 リポート障害	システムログから抽出※2	ソフト障害		283	1		
12	2013/12/26 14:34	2014/1/6 13:25	wgfs003 リポート障害	システムログから抽出※2	ソフト障害		263	1		
1	2014/1/6 10:27	2014/1/6 13:24	wgfs004リポート障害	システムログから抽出※2	ソフト障害		3	1		
1	2014/1/10 6:35	2014/1/10 9:40	wgfs013リポート障害	システムログから抽出※2	ソフト障害		3	1		
1	2014/1/15 13:41	2014/1/15 14:09	wgfs007リポート障害	システムログから抽出※2	ソフト障害		0	1		

共用ストレージ課題管理 / SS-185

AICSストレージ装置: sf11c102 のI/O Module故障(2016/10/11 - 10/13 17:08)

編集

コメント

割り当て

その他

課題の解決

課題のクローズ

詳細

タイプ:

インシデント

優先度:

中

ラベル:

なし

ステータス:

未解決

解決状況:

未解決

説明

概要: 2016/10/11(火) AICS全館停電後の復電時に発生。AICSストレージ装置 sf11c102 の SF12E02エンクロージャーのI/O Module(コントローラ#1と接続されている)障害レベル: レベル1

原因: I/O Moduleが故障。

影響があるホスト: esci-wgfs009..012 (修理完了までReadOnly運用)

破損ファイルの疑いのあるファイル: なし

修理日時: 2016/10/12(水) 14:00 ~ 16:00 (オンラインで交換のためサービス停止なし)

ReadOnly解除日時: 2016/10/13(木) 17:08

時系列: =====

■10/11(火):

08:30 - 08:57: 復電によりエンクロージャーの電源ON。SF12E02エンクロージャーのアラートLEDがオレンジ点灯

09:08 - 09:30: sf11c102コントローラ起動

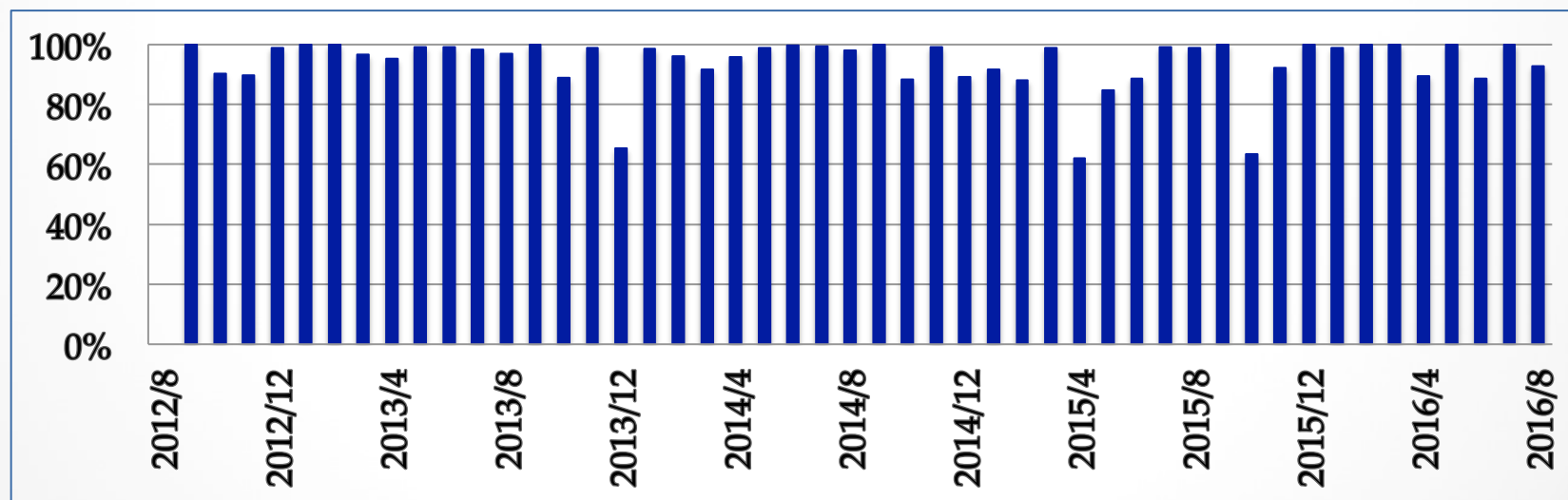
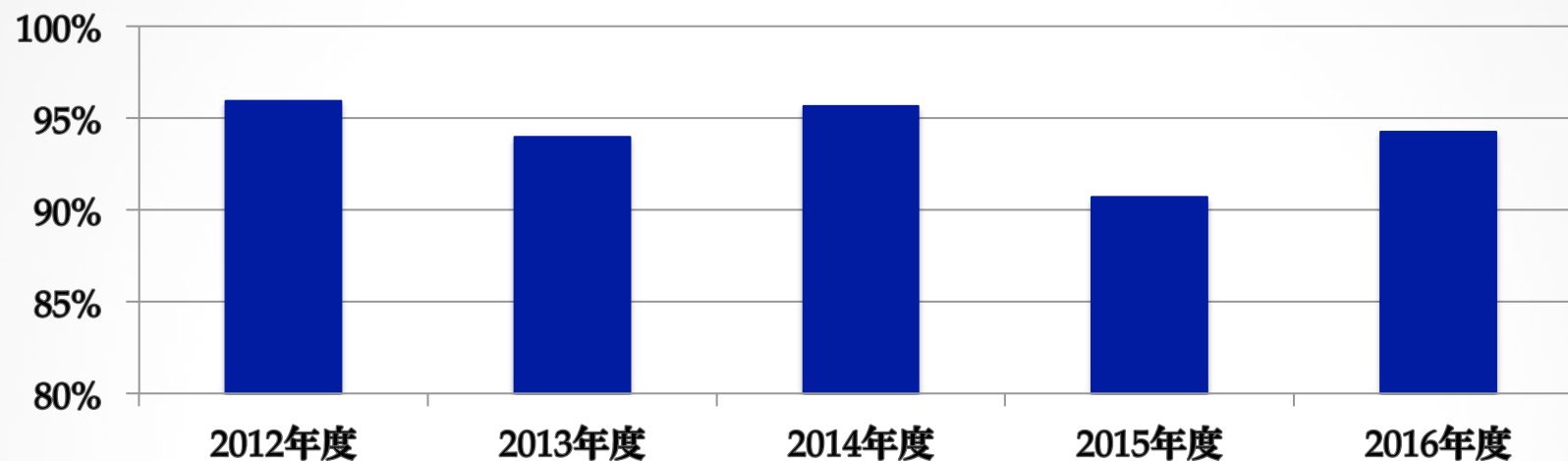
09:08: SF12E02エンクロージャーがsf11c102 コントローラ#1より認識できず、コピーバック開始

10:30: 障害原因判明、SF12E02エンクロージャーのコントローラ#1側のI/O Moduleの故障

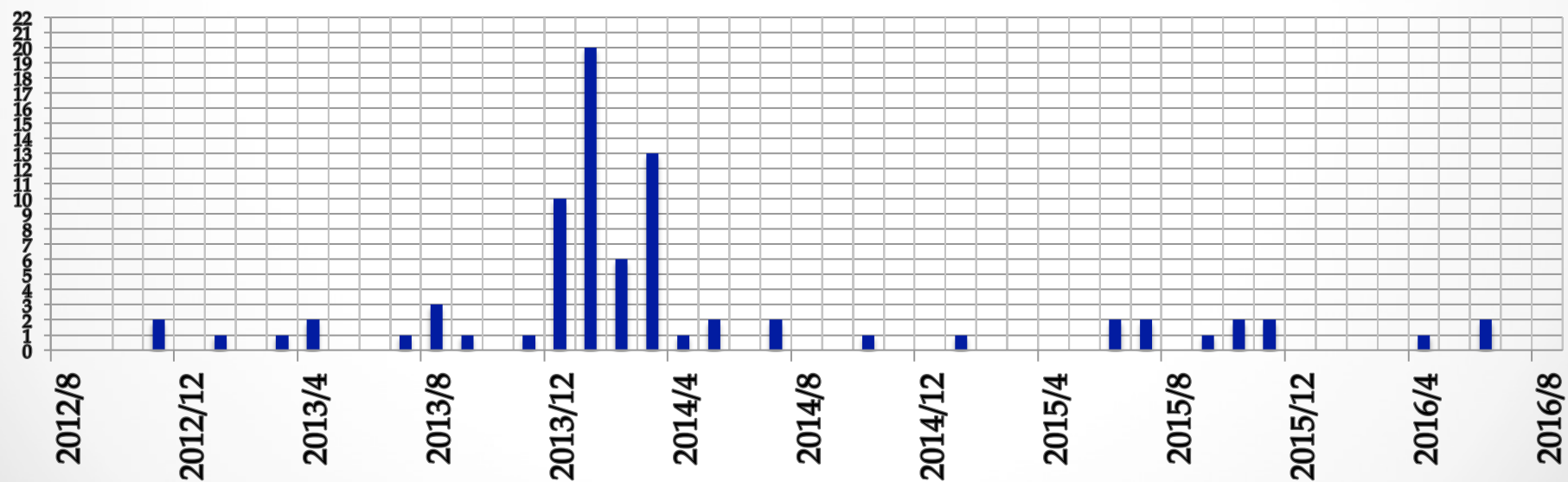
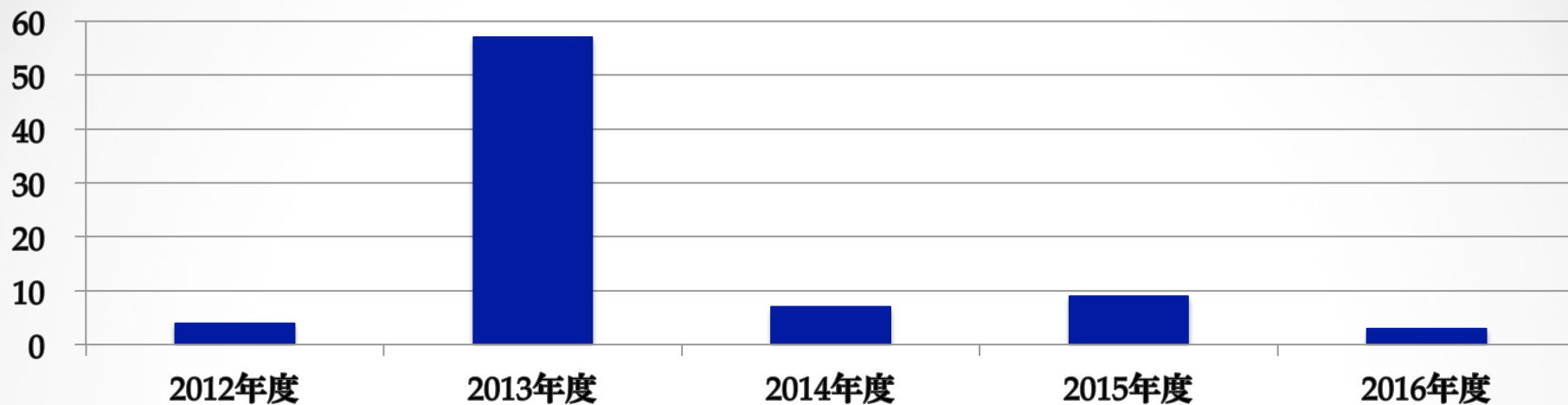
13:45 - 14:10: DDNがコントローラの片寄せ実行もコピーバックが発生している箇所での片寄せ失敗

=> 片寄せ失敗は正常な挙動

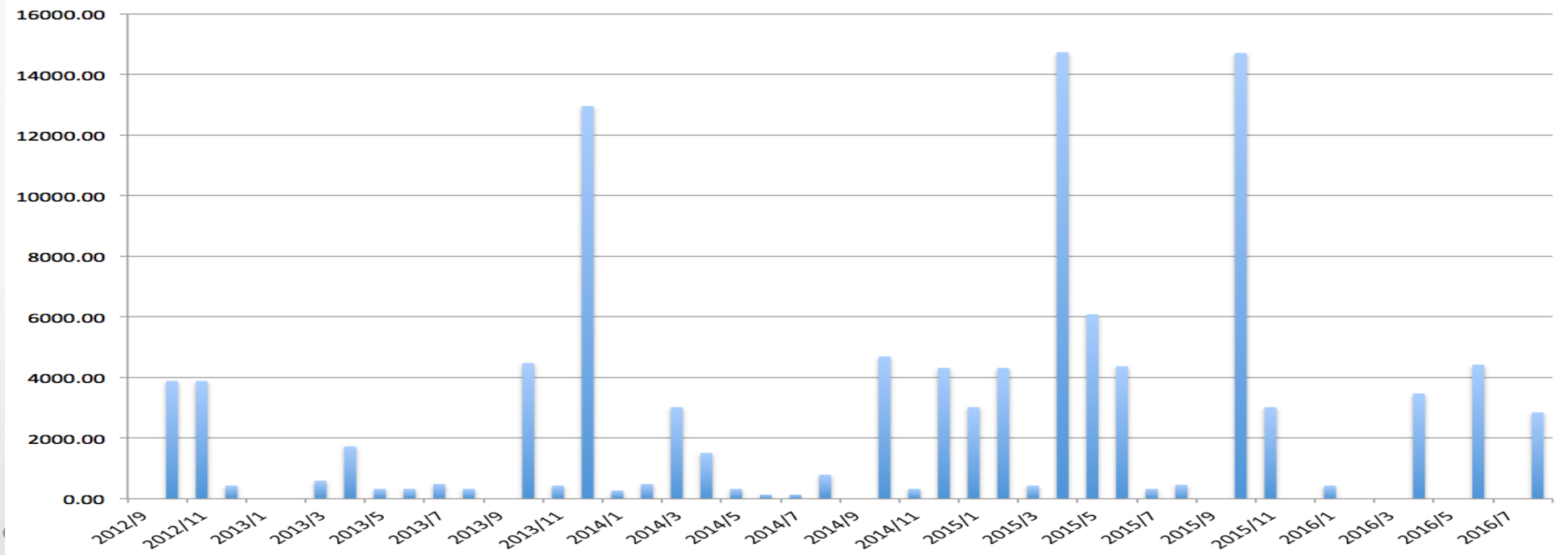
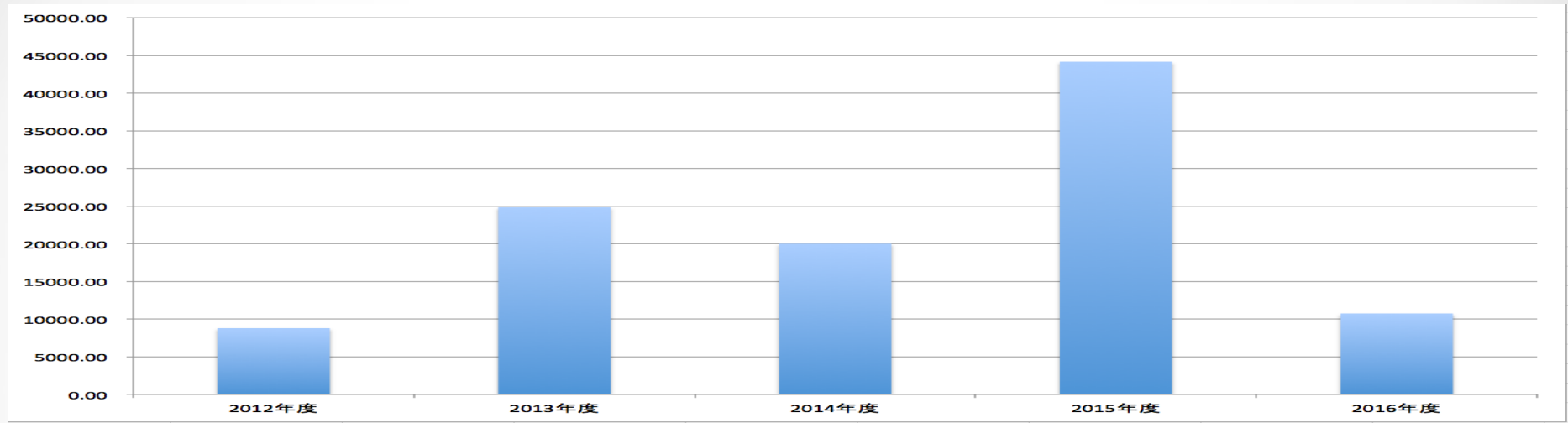
稼働率（年度、月）



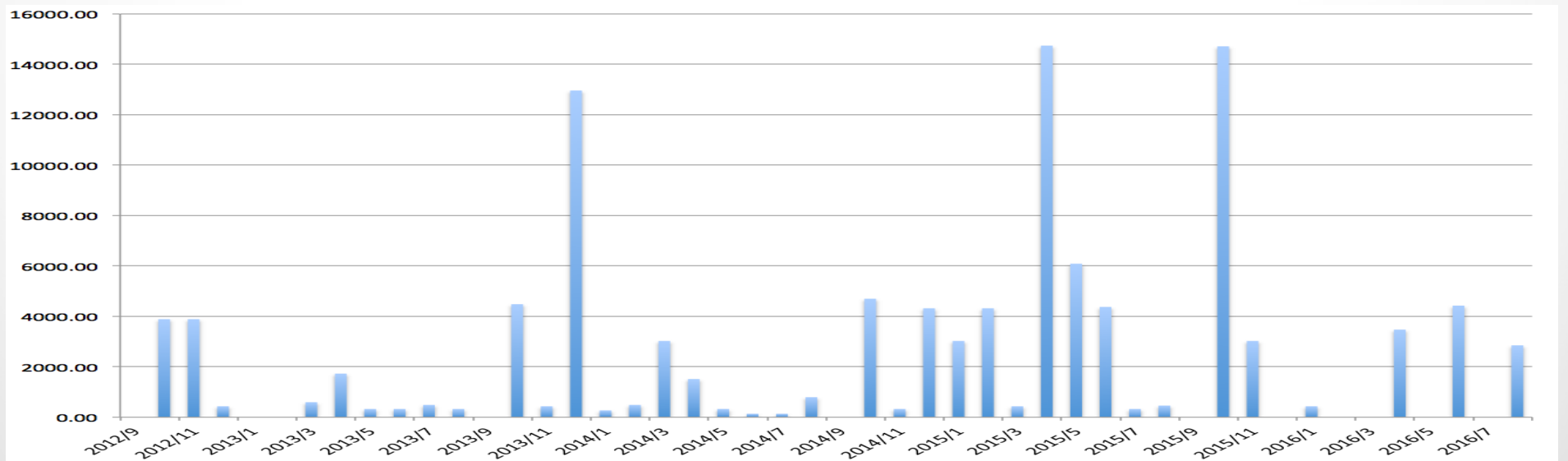
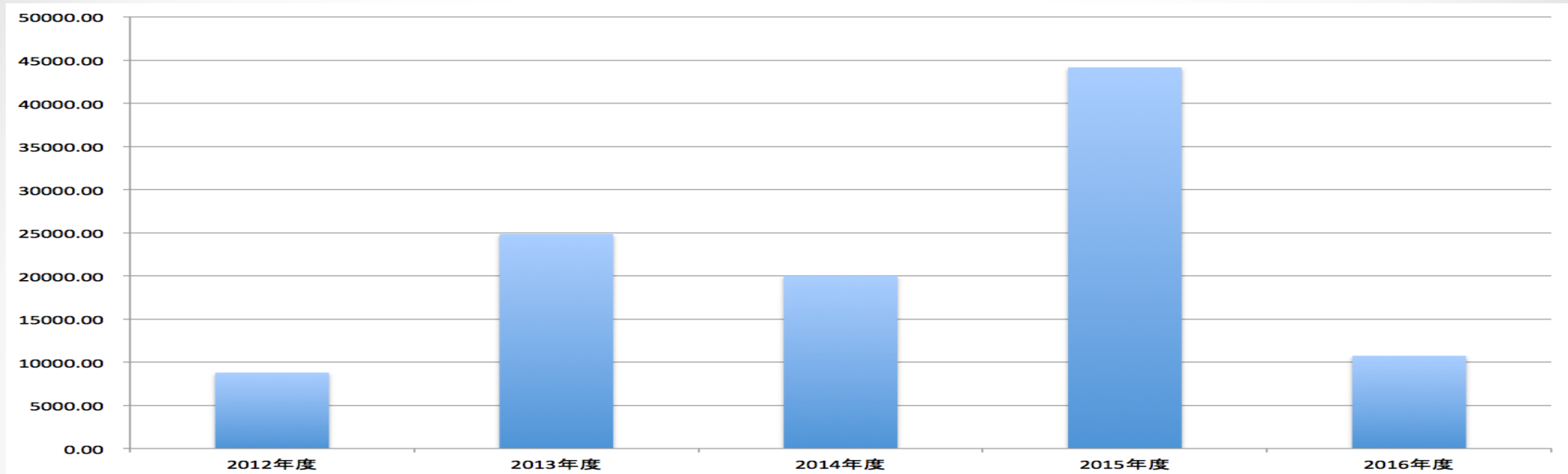
障害発生件数（年度、月）



障害停止ノード時間（年度、月）



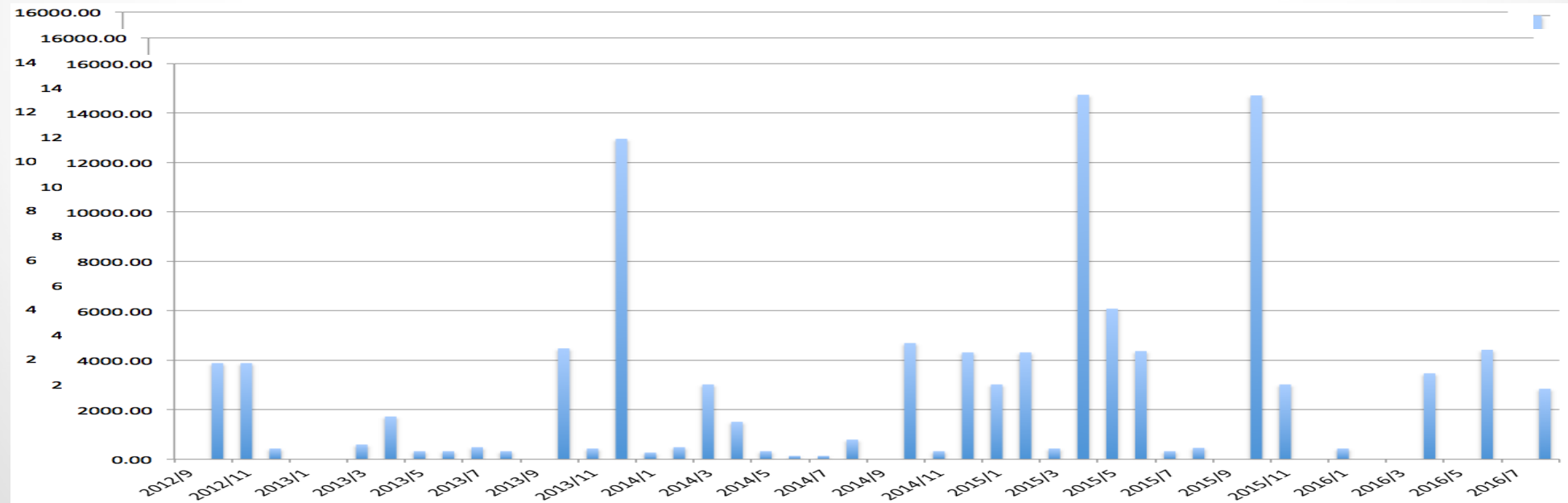
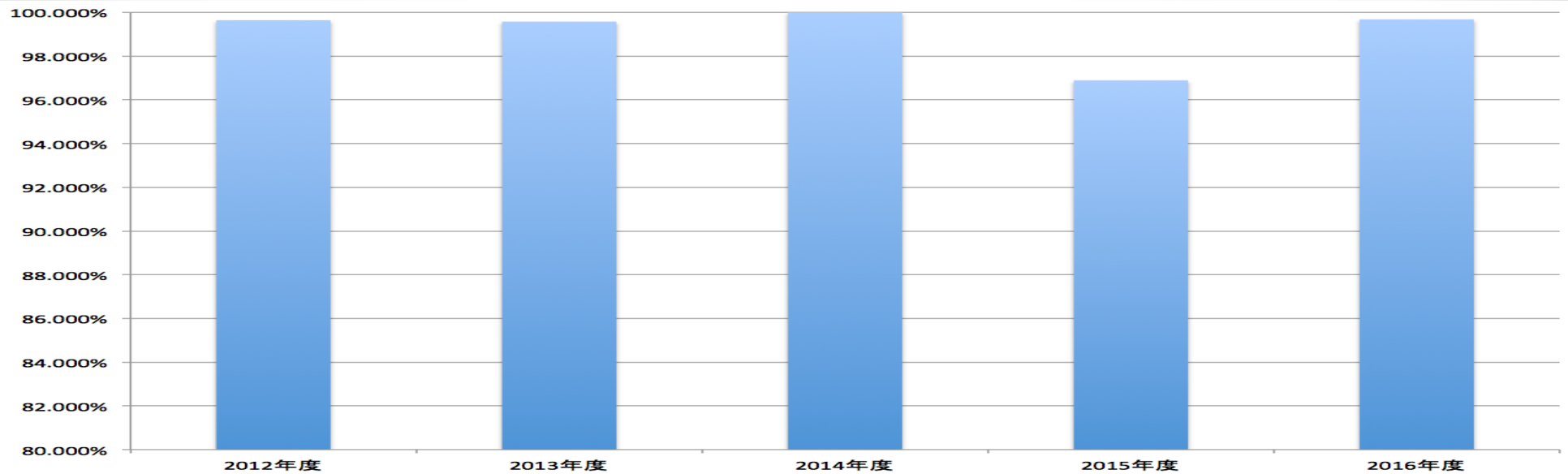
定期メンテナンスによる停止ノード時間（年度、月）



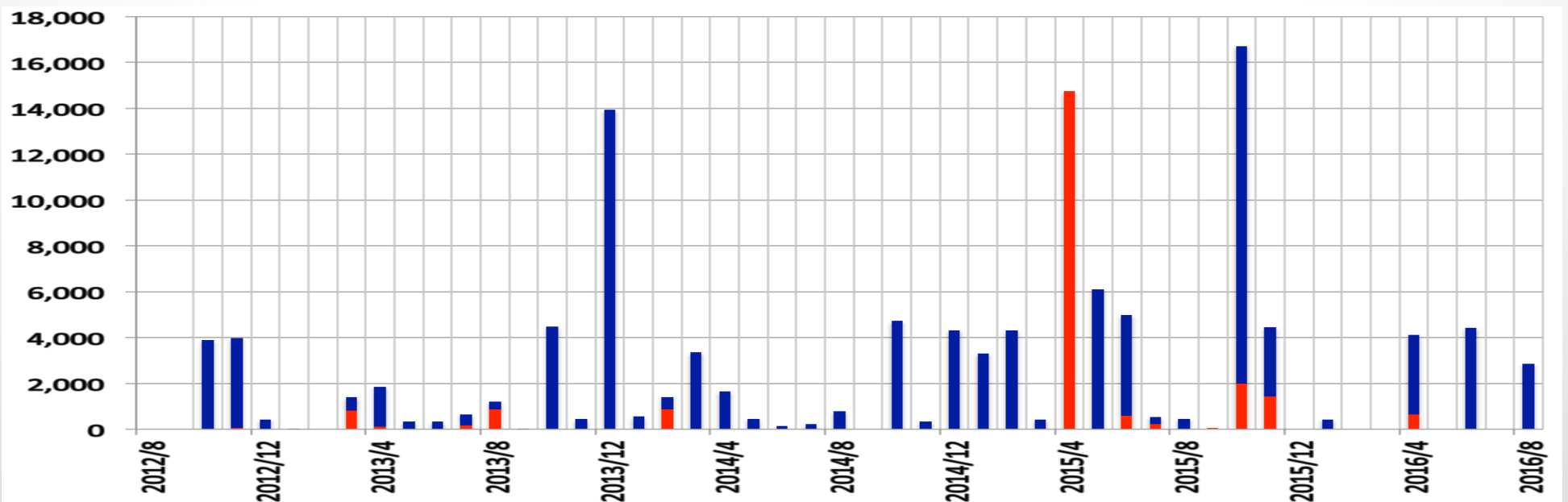
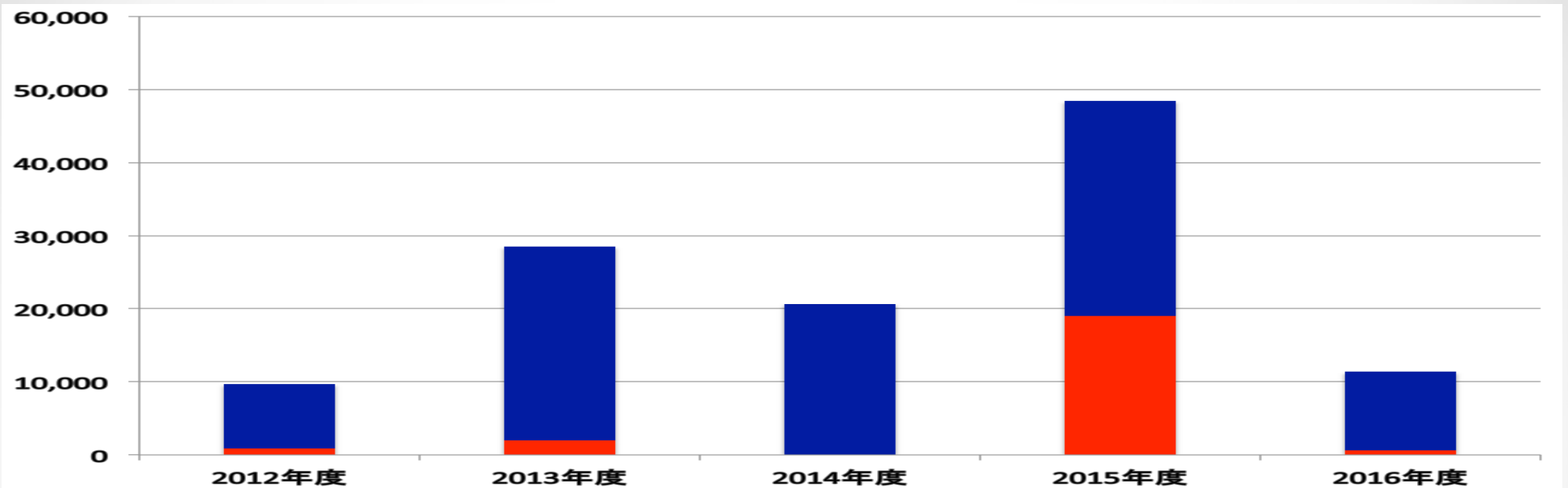
東大、AICSサイト間のデータ冗長化運用 によるサービス継続性改善予測

- データがAICS、東大で多重化されているなら、どちらか一方のサイトだけでも完全な運用が可能（全てのデータファイルをユーザに提供できる）
- AICS、東大で同時に障害が発生しない限り運用は継続可能
 - 計画停電も時期がずれていれば運用可能
 - 定期メンテナンスも時期をずらして個別に実施可能
- 過去の全障害、メンテナンスについてサービスの継続が可能だったか検証した
- サイト間データ多重化でも防げない障害
 - データ消失障害による長期メンテナンス
 - 東大のネームサーバ障害
 - etc

稼働率の改善予測（年度、月）



運用可能となるサービス停止ノード時間（年度、月）



東大、AICSサイト間のデータ冗長化運用 によるサービス継続性改善予測

- サービス停止ノード時間の81%がサービス継続可能に
 - サービス停止回数の87%がサービス継続可能に
 - MTBFは404時間から2154時間に改善
 - 実運用においても可能な限り提供サービスを改善したい
-
- データ移行に関してもなるべくサービス継続できるように検討している真っ最中
 - クラウドストレージを活用した長期保存サービスなどの新たなサービスの導入
 - 大容量、高帯域、高安全性、高可用なサービスを目指します！